# Contents

1 Some Basic Concepts  
1.1 The Lemma On \( \pi \) Systems ................................................................ 9  
1.2 Metric Space .................................................................................. 10  
1.3 Compactness In Metric Space .......................................................... 12  
1.4 Simple Functions And Measurable Functions .................................... 14  
1.5 One Dimensional Lebesgue Stieltjes Measure ................................... 16  
1.6 The Distribution Function ................................................................. 17  
1.7 Good Lambda Inequality ................................................................... 19  
1.8 Regularity Of Measures In Polish Space ........................................... 20  
1.9 The Ergodic Theorem ....................................................................... 22  

2 Some Extension Theorems  
2.1 Measures From Outer Measures ....................................................... 27  
2.2 Algebras ............................................................................................ 30  
2.3 Caratheodory Extension Theorem ..................................................... 31  
2.4 The Tychonoff Theorem .................................................................... 33  
2.5 Kolmogorov Extension Theorem For Polish Spaces ............................ 35  

3 Fourier Transforms  
3.1 An Algebra Of Special Functions ..................................................... 39  
3.2 Fourier Transforms Of Functions In \( G \) .......................................... 40  
3.3 Fourier Transforms Of Just About Anything ..................................... 41  
3.3.1 Fourier Transforms Of \( G \) ......................................................... 41  
3.3.2 Fourier Transforms Of Functions In \( L^1(\mathbb{R}^n) \) .................... 44  
3.3.3 Fourier Transforms Of Functions In \( L^2(\mathbb{R}^n) \) .................... 46  
3.3.4 The Schwartz Class .................................................................... 49  
3.3.5 Convolution ............................................................................... 51  

4 Banach Spaces  
4.1 Theorems Based On Baire Category ................................................ 53  
4.1.1 Baire Category Theorem ............................................................... 53  
4.1.2 Uniform Boundedness Theorem .................................................. 56  
4.1.3 Open Mapping Theorem ............................................................... 56  
4.1.4 Closed Graph Theorem ............................................................... 58  
4.2 Hahn Banach Theorem .................................................................... 59  
4.3 Weak And Weak * Topologies ......................................................... 64  
4.3.1 Basic Definitions ......................................................................... 64  
4.3.2 Banach Alaoglu Theorem ............................................................. 65  
4.3.3 Eberlein Smulian Theorem ......................................................... 67  

5 Locally Convex Topological Vector Spaces ........................................ 69  
5.1 Fundamental Considerations ............................................................ 69  
5.2 Separation Theorems ...................................................................... 72  
5.2.1 Convex Functionals .................................................................... 75  
5.2.2 More Separation Theorems ......................................................... 76  
5.3 The Weak And Weak* Topologies ................................................... 78
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Discrete Martingales</td>
<td>223</td>
</tr>
<tr>
<td>12.1</td>
<td>Conditional Expectation</td>
<td>223</td>
</tr>
<tr>
<td>12.2</td>
<td>Discrete Martingales</td>
<td>226</td>
</tr>
<tr>
<td>12.2.1</td>
<td>Upcrossings</td>
<td>227</td>
</tr>
<tr>
<td>12.2.2</td>
<td>The Submartingale Convergence Theorem</td>
<td>228</td>
</tr>
<tr>
<td>12.2.3</td>
<td>Doob Submartingale Estimate</td>
<td>229</td>
</tr>
<tr>
<td>12.3</td>
<td>Optional Sampling And Stopping Times</td>
<td>229</td>
</tr>
<tr>
<td>12.3.1</td>
<td>Stopping Times And Their Properties</td>
<td>229</td>
</tr>
<tr>
<td>12.3.2</td>
<td>Optional Stopping Times And Martingales</td>
<td>233</td>
</tr>
<tr>
<td>12.4</td>
<td>Stopping Times And Their Properties</td>
<td>233</td>
</tr>
<tr>
<td>12.5</td>
<td>Submartingale Convergence Theorem</td>
<td>238</td>
</tr>
<tr>
<td>12.5.1</td>
<td>Upcrossings</td>
<td>238</td>
</tr>
<tr>
<td>12.5.2</td>
<td>Maximal Inequalities</td>
<td>239</td>
</tr>
<tr>
<td>12.5.3</td>
<td>The Upcrossing Estimate</td>
<td>241</td>
</tr>
<tr>
<td>12.6</td>
<td>The Submartingale Convergence Theorem</td>
<td>243</td>
</tr>
<tr>
<td>12.7</td>
<td>A Reverse Submartingale Convergence Theorem</td>
<td>246</td>
</tr>
<tr>
<td>12.8</td>
<td>Strong Law Of Large Numbers</td>
<td>248</td>
</tr>
<tr>
<td>13</td>
<td>Stochastic Processes In Banach Space</td>
<td>251</td>
</tr>
<tr>
<td>13.1</td>
<td>Conditional Expectation In Banach Spaces</td>
<td>251</td>
</tr>
<tr>
<td>13.2</td>
<td>Properties Of Stochastic Processes</td>
<td>253</td>
</tr>
<tr>
<td>13.3</td>
<td>Kolmogorov Cenlsof Continuity Theorem</td>
<td>255</td>
</tr>
<tr>
<td>13.4</td>
<td>Filtrations</td>
<td>262</td>
</tr>
<tr>
<td>13.5</td>
<td>Martingales</td>
<td>268</td>
</tr>
<tr>
<td>13.6</td>
<td>Some Maximal Estimates</td>
<td>269</td>
</tr>
<tr>
<td>13.7</td>
<td>Optional Sampling Theorems</td>
<td>271</td>
</tr>
<tr>
<td>13.7.1</td>
<td>Stopping Times And Their Properties</td>
<td>271</td>
</tr>
<tr>
<td>13.7.2</td>
<td>Doob Optional Sampling Theorem</td>
<td>274</td>
</tr>
<tr>
<td>13.8</td>
<td>Doob Optional Sampling Continuous Case</td>
<td>276</td>
</tr>
<tr>
<td>13.8.1</td>
<td>Stopping Times</td>
<td>276</td>
</tr>
<tr>
<td>13.8.2</td>
<td>The Optional Sampling Theorem Continuous Case</td>
<td>279</td>
</tr>
<tr>
<td>13.9</td>
<td>Some Maximal Inequalities</td>
<td>284</td>
</tr>
<tr>
<td>13.10</td>
<td>Continuous Submartingale Convergence Theorem</td>
<td>287</td>
</tr>
<tr>
<td>13.11</td>
<td>Hitting This Before That</td>
<td>290</td>
</tr>
<tr>
<td>13.12</td>
<td>The Space $\mathcal{M}_r^p(E)$</td>
<td>293</td>
</tr>
<tr>
<td>14</td>
<td>The Quadratic Variation Of A Martingale</td>
<td>297</td>
</tr>
<tr>
<td>14.1</td>
<td>How To Recognize A Martingale</td>
<td>297</td>
</tr>
<tr>
<td>14.2</td>
<td>The Quadratic Variation</td>
<td>301</td>
</tr>
<tr>
<td>14.3</td>
<td>The Covariation</td>
<td>306</td>
</tr>
<tr>
<td>14.4</td>
<td>The Burkholder Davis Gundy Inequality</td>
<td>309</td>
</tr>
<tr>
<td>14.5</td>
<td>$\mathcal{M}_r^p(H)$ Is A Hilbert Space</td>
<td>312</td>
</tr>
<tr>
<td>14.6</td>
<td>The Quadratic Variation And Stochastic Integration</td>
<td>313</td>
</tr>
<tr>
<td>14.7</td>
<td>Another Limit For Quadratic Variation</td>
<td>319</td>
</tr>
<tr>
<td>15</td>
<td>Gaussian Measures</td>
<td>325</td>
</tr>
<tr>
<td>15.1</td>
<td>Definitions And Basic Properties</td>
<td>325</td>
</tr>
<tr>
<td>15.2</td>
<td>Fernique's Theorem</td>
<td>327</td>
</tr>
<tr>
<td>15.3</td>
<td>Gaussian Measures For A Separable Hilbert Space</td>
<td>332</td>
</tr>
<tr>
<td>16</td>
<td>Wiener Processes</td>
<td>339</td>
</tr>
<tr>
<td>16.1</td>
<td>Real Wiener Processes</td>
<td>339</td>
</tr>
<tr>
<td>16.2</td>
<td>Nowhere Differentiability Of Wiener Processes</td>
<td>343</td>
</tr>
<tr>
<td>16.3</td>
<td>An Example Of Martingales, Independent Increments</td>
<td>344</td>
</tr>
<tr>
<td>16.4</td>
<td>Hilbert Space Valued Wiener Processes</td>
<td>346</td>
</tr>
<tr>
<td>16.5</td>
<td>Wiener Processes, Another Approach</td>
<td>358</td>
</tr>
<tr>
<td>16.5.1</td>
<td>Lots Of Independent Normally Distributed Random Variables</td>
<td>358</td>
</tr>
<tr>
<td>16.5.2</td>
<td>The Wiener Processes</td>
<td>364</td>
</tr>
</tbody>
</table>
17 Stochastic Integration 371
| 17.1 Integrals Of Elementary Processes | 371 |
| 17.2 Different Definition Of Elementary Functions | 376 |
| 17.3 Approximating With Elementary Functions | 376 |
| 17.4 Some Hilbert Space Theory | 379 |
| 17.5 The General Integral | 383 |
| 17.6 The Case That $Q$ Is Trace Class | 388 |
| 17.7 A Short Comment On Measurability | 389 |
| 17.8 Localization For Elementary Functions | 390 |
| 17.9 Localization In General | 391 |
| 17.10 The Stochastic Integral As A Local Martingale | 392 |
| 17.11 The Quadratic Variation Of The Stochastic Integral | 394 |
| 17.12 The Holder Continuity Of The Integral | 397 |
| 17.13 Taking Out A Linear Transformation | 397 |
| 17.14A Technical Integration By Parts Result | 399 |

18 The Integral $\int_0^t (Y, dM)_t$ 405

19 A Different Kind Of Stochastic Integration 417
| 19.1 Hermite Polynomials | 417 |
| 19.2 A Multiple Integral | 422 |
| 19.3 The Skorokhod Integral |
| 19.3.1 The Derivative | 431 |
| 19.3.2 The Integral | 432 |
| 19.3.3 The Ito And Skorokhod Integrals | 433 |

20 The Easy Ito Formula 441
| 20.1 The Situation | 441 |
| 20.2 Assumptions And A Lemma | 441 |
| 20.3 A Special Case | 442 |
| 20.4 The Case Of Elementary Functions | 445 |
| 20.5 The Integrable Case | 446 |
| 20.6 The General Stochastically Integrable Case | 447 |
| 20.7 Remembering The Formula | 448 |
| 20.8 An Interesting Formula | 449 |
| 20.9 Some Representation Theorems | 449 |

21 Gelfand Triples 461
| 21.1 An Unnatural Example | 462 |
| 21.2 Some Standard Techniques In Evolution Equations | 466 |
| 21.3 An Important Formula | 474 |
| 21.4 The Implicit Case | 479 |
| 21.5 The Implicit Case, $B = B(t)$ | 491 |
| 21.6 Some Imbedding Theorems | 499 |

22 Measurability Without Uniqueness 505
| 22.1 Multifunctions And Their Measurability | 505 |
| 22.2 A Measurable Selection | 506 |
| 22.3 Measurability In Finite Dimensional Problems | 510 |
| 22.4 The Navier–Stokes Equations | 514 |
| 22.5 A Friction contact problem |
| 22.5.1 The Abstract Problem | 520 |
| 22.5.2 An Approximate Problem | 521 |
| 22.5.3 Discontinuous coefficient of friction | 526 |
## 23 Stochastic O.D.E. One Space

- 23.1 Adapted Solutions With Uniqueness ........................................ 531
- 23.2 Including Stochastic Integrals .............................................. 532
- 23.3 Stochastic Differential Equations In A Hilbert Space ................. 535
  - 23.3.1 The Lipschitz Case ...................................................... 535
  - 23.3.2 The Locally Lipschitz Case ........................................... 537

## 24 The Hard Ito Formula

- 24.1 Predictable And Stochastic Continuity ..................................... 541
- 24.2 Approximating With Step Functions ....................................... 542
- 24.3 The Situation ........................................................................ 544
- 24.4 A Simplification Of The Formula ........................................... 546
- 24.5 Converging In Probability .................................................... 551
- 24.6 The Ito Formula .................................................................. 551

## 25 The Hard Ito Formula, Implicit Case

- 25.1 Approximating With Step Functions ....................................... 557
- 25.2 The Situation ........................................................................ 558
- 25.3 Preliminary Results .................................................................. 559
- 25.4 The Main Estimate .................................................................. 563
- 25.5 A Simplification Of The Formula ........................................... 570
- 25.6 Convergence ........................................................................ 572
- 25.7 The Ito Formula .................................................................. 576

## 26 A More Attractive Version

- 26.1 The Situation ........................................................................ 586
- 26.2 Preliminary Results .................................................................. 587
- 26.3 The Main Estimate .................................................................. 589
- 26.4 A Simplification Of The Formula ........................................... 595
- 26.5 Converged ............................................................................ 596
- 26.6 The Ito Formula .................................................................. 598

## 27 Nonlinear Operators

- 27.1 An Assortment Of Nonlinear Operators .................................. 607
- 27.2 Duality Maps ......................................................................... 612
- 27.3 Penalization And Projection Operators .................................... 617
- 27.4 Set-Valued Maps, Pseudomonotone Operators ......................... 619
- 27.5 Generalized Gradients ............................................................ 626
- 27.6 Maximal Monotone Operators ............................................... 628
  - 27.6.1 Maximal Monotone, Equivalent Conditions ....................... 629
  - 27.6.2 Surjectivity Theorems ...................................................... 637
  - 27.6.3 Approximation Theorems ............................................... 644
  - 27.6.4 Sum Of Maximal Monotone Operators ............................... 650
- 27.6.5 Convex Functions, An Example ......................................... 653
- 27.7 Perturbation Theorems ............................................................ 663

## 28 Implicit Stochastic Equations

- 28.1 Introduction .......................................................................... 673
- 28.2 Preliminary Results .................................................................. 673
- 28.3 The Existence Of Approximate Solutions .................................. 677
- 28.4 The General Case .................................................................. 686
- 28.5 Replacing $\Phi$ With $\sigma(u)$ .................................................. 699
- 28.6 Examples ............................................................................. 704
- 28.7 Other Examples, Inclusions .................................................... 709
29 The Yankov von Neumann Aumann theorem ........................................ 717
  29.1 Multifunctions And Their Measurability ........................................ 724
      29.1.1 The General Case .......................................................... 724
      29.1.2 A Special Case Which Is Easier ........................................ 726

30 Stochastic Inclusions ........................................................................ 727
  30.1 The General Context ................................................................. 727
  30.2 Some Fundamental Theorems ....................................................... 727
  30.3 Preliminary Results ....................................................................... 739
  30.4 Measurable Approximate Solutions .............................................. 741
  30.5 The Main Result ......................................................................... 746
  30.6 Variational Inequalities ............................................................... 755
  30.7 Progressively Measurable Solutions ............................................. 758
  30.8 Adding Another Operator ............................................................ 759
  30.9 Including Stochastic Integrals ..................................................... 768
Chapter 1

Some Basic Concepts

1.1 The Lemma On \( \pi \) Systems

**Definition 1.1.1** \( \mathcal{F} \subseteq \mathcal{P}(\Omega) \), the set of all subsets of \( \Omega \), is called a \( \sigma \) algebra if it contains \( \emptyset, \Omega \), and is closed with respect to countable unions and complements. That is, if \( \{ A_n \}_{n=1}^{\infty} \) is countable and each \( A_n \in \mathcal{F} \), then \( \cup_{n=1}^{\infty} A_n \in \mathcal{F} \) also and if \( A \in \mathcal{F} \), then \( \Omega \setminus A \in \mathcal{F} \). It is clear that any intersection of \( \sigma \) algebras is a \( \sigma \) algebra. If \( \mathcal{K} \subseteq \mathcal{P}(\Omega) \), \( \sigma(\mathcal{K}) \) is the smallest \( \sigma \) algebra which contains \( \mathcal{K} \).

**Definition 1.1.2** Let \( \Omega \) be a set and let \( \mathcal{K} \) be a collection of subsets of \( \Omega \). Then \( \mathcal{K} \) is called a \( \pi \) system if \( \emptyset, \Omega \in \mathcal{K} \) and whenever \( A, B \in \mathcal{K} \), it follows \( A \cap B \in \mathcal{K} \).

Obviously an example of a \( \pi \) system is the set of measurable rectangles because

\[
A \times B \cap A' \times B' = (A \cap A') \times (B \cap B').
\]

The following is the fundamental lemma which shows these \( \pi \) systems are useful. This is due to Dynkin.

**Lemma 1.1.3** Let \( \mathcal{K} \) be a \( \pi \) system of subsets of \( \Omega \), a set. Also let \( \mathcal{G} \) be a collection of subsets of \( \Omega \) which satisfies the following three properties.

1. \( \mathcal{K} \subseteq \mathcal{G} \)
2. If \( A \in \mathcal{G} \), then \( A^C \in \mathcal{G} \)
3. If \( \{ A_i \}_{i=1}^{\infty} \) is a sequence of disjoint sets from \( \mathcal{G} \) then \( \cup_{i=1}^{\infty} A_i \in \mathcal{G} \).

Then \( \mathcal{G} \supseteq \sigma(\mathcal{K}) \), where \( \sigma(\mathcal{K}) \) is the smallest \( \sigma \) algebra which contains \( \mathcal{K} \).

**Proof:** First note that if

\[
\mathcal{H} = \{ \mathcal{G} : \text{ all hold} \}
\]

then \( \cap \mathcal{H} \) yields a collection of sets which also satisfies \( \text{all} \). Therefore, I will assume in the argument that \( \mathcal{G} \) is the smallest collection satisfying \( \text{all} \). Let \( A \in \mathcal{K} \) and define

\[
\mathcal{G}_A = \{ B \in \mathcal{G} : A \cap B \in \mathcal{G} \}.
\]

I want to show \( \mathcal{G}_A \) satisfies \( \text{all} \) because then it must equal \( \mathcal{G} \) since \( \mathcal{G} \) is the smallest collection of subsets of \( \Omega \) which satisfies \( \text{all} \). This will give the conclusion that for \( A \in \mathcal{K} \) and \( B \in \mathcal{G} \), \( A \cap B \in \mathcal{G} \). This information will then be used to show that if \( A, B \in \mathcal{G} \) then \( A \cap B \in \mathcal{G} \). From this it will follow very easily that \( \mathcal{G} \) is a \( \sigma \) algebra which will imply it contains \( \sigma(\mathcal{K}) \). Now here are the details of the argument.

Since \( \mathcal{K} \) is given to be a \( \pi \) system, \( \mathcal{K} \subseteq \mathcal{G}_A \). Property \( \text{all} \) is obvious because if \( \{ B_i \} \) is a sequence of disjoint sets in \( \mathcal{G}_A \), then

\[
A \cap \bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A \cap B_i \in \mathcal{G}
\]

because \( A \cap B_i \in \mathcal{G} \) and the property \( \text{all} \) of \( \mathcal{G} \).

It remains to verify Property \( \text{all} \) so let \( B \in \mathcal{G}_A \). I need to verify that \( B^C \in \mathcal{G}_A \). In other words, I need to show that \( A \cap B^C \in \mathcal{G} \). However,

\[
A \cap B^C = (A^C \cup (A \cap B))^C \in \mathcal{G}
\]
Here is why. Since $B \in \mathcal{G}_A$, $A \cap B \in \mathcal{G}$ and since $A \in K \subseteq \mathcal{G}$ it follows $A^C \in \mathcal{G}$ by assumption \( \Box \). It follows from assumption \( \Box \) the union of the disjoint sets, $A^C$ and $(A \cap B)$ is in $\mathcal{G}$ and then from \( \Box \) the complement of their union is in $\mathcal{G}$. Thus $\mathcal{G}_A$ satisfies \( \Box - \Box \) and this implies since $\mathcal{G}$ is the smallest such, that $\mathcal{G}_A \supseteq \mathcal{G}$. However, $\mathcal{G}_A$ is constructed as a subset of $\mathcal{G}$. This proves that for every $B \in \mathcal{G}$ and $A \in K$, $A \cap B \in \mathcal{G}$. Now pick $B \in \mathcal{G}$ and consider

$$
\mathcal{G}_B \equiv \left\{ A \in \mathcal{G} : A \cap B \in \mathcal{G} \right\}.
$$

I just proved $K \subseteq \mathcal{G}_B$. The other arguments are identical to show $\mathcal{G}_B$ satisfies \( \Box - \Box \) and is therefore equal to $\mathcal{G}$. This shows that whenever $A, B \in \mathcal{G}$ it follows $A \cap B \in \mathcal{G}$.

This implies $\mathcal{G}$ is a $\sigma$ algebra. To show this, all that is left is to verify $\mathcal{G}$ is closed under countable unions because then it follows $\mathcal{G}$ is a $\sigma$ algebra. Let $\{A_i\} \subseteq \mathcal{G}$. Then let $A'_i = A_1$ and

$$
A'_{n+1} \equiv A_{n+1} \setminus (\bigcup_{i=1}^n A_i)
= A_{n+1} \cap (\bigcap_{i=1}^n A_i^C)
= \bigcap_{i=1}^n (A_{n+1} \cap A_i^C) \in \mathcal{G}
$$

because finite intersections of sets of $\mathcal{G}$ are in $\mathcal{G}$. Since the $A'_i$ are disjoint, it follows

$$
\bigcup_{i=1}^\infty A_i = \bigcup_{i=1}^\infty A'_i \in \mathcal{G}
$$

Therefore, $\mathcal{G} \supseteq \sigma (\mathcal{K})$ and this proves the Lemma.

### 1.2 Metric Space

**Definition 1.2.1** A metric space is a set, $X$ and a function $d : X \times X \to [0, \infty)$ which satisfies the following properties.

- $d(x, y) = d(y, x)$
- $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$
- $d(x, y) \leq d(x, z) + d(z, y)$.

You can check that $\mathbb{R}^n$ and $\mathbb{C}^n$ are metric spaces with $d(x, y) = |x - y|$. However, there are many others. The definitions of open and closed sets are the same for a metric space as they are for $\mathbb{R}^n$.

**Definition 1.2.2** A set, $U$ in a metric space is open if whenever $x \in U$, there exists $r > 0$ such that $B(x, r) \subseteq U$. As before, $B(x, r) \equiv \{y : d(x, y) < r\}$. Closed sets are those whose complements are open. A point $p$ is a limit point of a set, $S$ if for every $r > 0$, $B(p, r)$ contains infinitely many points of $S$. A sequence, $\{x_n\}$ converges to a point $x$ if for every $\varepsilon > 0$ there exists $N$ such that if $n \geq N$, then $d(x_n, x) < \varepsilon$.\{x_n\} is a Cauchy sequence if for every $\varepsilon > 0$ there exists $N$ such that if $m, n \geq N$, then $d(x_m, x_n) < \varepsilon$.

**Lemma 1.2.3** In a metric space, $X$ every ball, $B(x, r)$ is open. A set is closed if and only if it contains all its limit points. If $p$ is a limit point of $S$, then $S$ contains a sequence of distinct points of $S, \{x_n\}$ such that $\lim_{n \to \infty} x_n = p$.

**Proof:** Let $z \in B(x, r)$. Let $\delta = r - d(x, z)$. Then if $w \in B(z, \delta)$,

$$
d(w, x) \leq d(w, z) + d(z, x) < d(x, z) + r - d(x, z) = r.
$$

Therefore, $B(z, \delta) \subseteq B(x, r)$ and this shows $B(x, r)$ is open.

The properties of balls are presented in the following theorem.

**Theorem 1.2.4** Suppose $(X, d)$ is a metric space. Then the sets $\{B(x, r) : r > 0, x \in X\}$ satisfy

$$
\bigcup \{B(x, r) : r > 0, x \in X\} = X \quad (1.2.1)
$$

If $p \in B(x, r_1) \cap B(z, r_2)$, there exists $r > 0$ such that

$$
B(p, r) \subseteq B(x, r_1) \cap B(z, r_2) \quad (1.2.2)
$$
Proof: Observe that the union of these balls includes the whole space, $X$ so \[\text{Lemma 1.2.7}\] is obvious. Consider \[\text{Lemma 1.2.5}\].

Let $p \in B(x,r_1) \cap B(z,r_2)$. Consider
\[r \equiv \min (r_1 - d(x,p), r_2 - d(z,p))\]
and suppose $y \in B(p,r)$. Then
\[d(y,x) \leq d(y,p) + d(p,x) < r_1 - d(x,p) + d(x,p) = r_1\]
and so $B(p,r) \subseteq B(x,r_1)$. By similar reasoning, $B(p,r) \subseteq B(z,r_2)$. This proves the theorem.

Let $K$ be a closed set. This means $K^C \equiv X \setminus K$ is an open set. Let $p$ be a limit point of $K$. If $p \in K^C$, then since $K^C$ is open, there exists $B(p,r) \subseteq K^C$. But this contradicts $p$ being a limit point because there are no points of $K$ in this ball. Hence all limit points of $K$ must be in $K$.

Suppose next that $K$ contains its limit points. Is $K^C$ open? Let $p \in K^C$. Then $p$ is not a limit point of $K$. Therefore, there exists $B(p,r)$ which contains at most finitely many points of $K$. Since $p \notin K$, it follows that by making $r$ smaller if necessary, $B(p,r)$ contains no points of $K$. That is $B(p,r) \subseteq K^C$ showing $K^C$ is open. Therefore, $K$ is closed.

Suppose now that $p$ is a limit point of $S$. Let $x_1 \in (S \setminus \{p\}) \cap B(p,1)$. If $x_1, \ldots, x_k$ have been chosen, let
\[r_{k+1} \equiv \min \left\{ d(p,x_i), i = 1, \ldots, k, \frac{1}{k+1} \right\}, \]
Let $x_{k+1} \in (S \setminus \{p\}) \cap B(p,r_{k+1})$. This proves the lemma.

Lemma 1.2.5 If $\{x_n\}$ is a Cauchy sequence in a metric space, $X$ and if some subsequence, $\{x_{n_k}\}$ converges to $x$, then $\{x_n\}$ converges to $x$. Also if a sequence converges, then it is a Cauchy sequence.

Proof: Note first that $n_k \geq k$ because in a subsequence, the indices, $n_1, n_2, \ldots$ are strictly increasing. Let $\varepsilon > 0$ be given and let $N$ be such that for $k > N, d(x,x_{n_k}) < \varepsilon/2$ and for $m,n \geq N, d(x_m,x_n) < \varepsilon/2$. Pick $k > n$. Then if $n > N$,
\[d(x_n,x) \leq d(x_n,x_{n_k}) + d(x_{n_k},x) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\]
Finally, suppose $\lim_{n \to \infty} x_n = x$. Then there exists $N$ such that if $n > N$, then $d(x_n,x) < \varepsilon/2$. It follows that for $m,n > N$,
\[d(x_n,x_m) \leq d(x_n,x) + d(x,x_m) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\]
This proves the lemma.

A useful idea is the idea of distance from a point to a set.

Definition 1.2.6 Let $(X,d)$ be a metric space and let $S$ be a nonempty set in $X$. Then
\[\text{dist} (x,S) \equiv \inf \{ d(x,y) : y \in S \}.\]

The following lemma is the fundamental result.

Lemma 1.2.7 The function, $x \rightarrow \text{dist} (x,S)$ is continuous and in fact satisfies
\[| \text{dist} (x,S) - \text{dist} (y,S) | \leq d(x,y).\]

Proof: Suppose dist $(x,y)$ is as least as large as dist $(y,S)$. Then pick $z \in S$ such that $d(y,z) \leq \text{dist} (y,S) + \varepsilon$. Then
\[| \text{dist} (x,S) - \text{dist} (y,S) | = \text{dist} (x,S) - \text{dist} (y,S) \leq d(x,z) - d(y,z) - \varepsilon = d(x,z) - d(y,z) + \varepsilon \leq d(x,y) + d(y,z) - d(y,z) + \varepsilon = d(x,y) + \varepsilon.\]
Since $\varepsilon > 0$ is arbitrary, this proves the lemma.
1.3 Compactness In Metric Space

Many existence theorems in analysis depend on some set being compact. Therefore, it is important to be able to identify compact sets. The purpose of this section is to describe compact sets in a metric space.

**Definition 1.3.1** Let $A$ be a subset of $X$. $A$ is compact if whenever $A$ is contained in the union of a set of open sets, there exists finitely many of these open sets whose union contains $A$. (Every open cover admits a finite subcover.)

$A$ is “sequentially compact” means every sequence has a convergent subsequence converging to an element of $A$.

In a metric space compact is not the same as closed and bounded!

**Example 1.3.2** Let $X$ be any infinite set and define $d (x, y) = 1$ if $x \neq y$ while $d (x, y) = 0$ if $x = y$.

You should verify the details that this is a metric space because it satisfies the axioms of a metric. The set $X$ is closed and bounded because its complement is $\emptyset$ which is clearly open because every point of $\emptyset$ is an interior point. (There are none.) Also $X$ is bounded because $X = B (x, 2)$. However, $X$ is clearly not compact because $\{B \left( x, \frac{1}{n} \right) : x \in X \}$ is a collection of open sets whose union contains $X$ but since they are all disjoint and nonempty, there is no finite subset of these whose union contains $X$. In fact $B \left( x, \frac{1}{n} \right) = \{x\}$.

From this example it is clear something more than closed and bounded is needed. If you are not familiar with the issues just discussed, ignore them and continue.

**Definition 1.3.3** In any metric space, a set $E$ is totally bounded if for every $\varepsilon > 0$ there exists a finite set of points $\{x_1, \ldots, x_n\}$ such that $E \subseteq \bigcup_{i=1}^{n} B (x_i, \varepsilon)$.

This finite set of points is called an $\varepsilon$ net.

The following proposition tells which sets in a metric space are compact. First here is an interesting lemma.

**Lemma 1.3.4** Let $X$ be a metric space and suppose $D$ is a countable dense subset of $X$. In other words, it is being assumed $X$ is a separable metric space. Consider the open sets of the form $B (d, r)$ where $r$ is a positive rational number and $d \in D$. Denote this countable collection of open sets by $B$. Then every open set is the union of sets of $B$. Furthermore, if $\mathcal{C}$ is any collection of open sets, there exists a countable subset, $\{U_n\} \subseteq \mathcal{C}$ such that $\bigcup_{n=1}^{\infty} U_n = \bigcup \mathcal{C}$.

**Proof:** Let $U$ be an open set and let $x \in U$. Let $B (x, \delta) \subseteq U$. Then by density of $D$, there exists $d \in D \cap B (x, \delta/4)$.

Now pick $r \in \mathbb{Q} \cap (\delta/4, 3\delta/4)$ and consider $B (d, r)$. Clearly, $B (d, r)$ contains the point $x$ because $r > \delta/4$. Is $B (d, r) \subseteq B (x, \delta)$? if so, this proves the lemma because $x$ was an arbitrary point of $U$. Suppose $z \in B (d, r)$. Then

$$d (z, x) \leq d (z, d) + d (d, x) < r + \frac{\delta}{4} < \frac{3\delta}{4} + \frac{\delta}{4} = \delta$$

Now let $\mathcal{C}$ be any collection of open sets. Each set in this collection is the union of countably many sets of $B$. Let $\mathcal{B}'$ denote the sets of $B$ which are contained in some set of $\mathcal{C}$. Thus $\bigcup \mathcal{B}' = \bigcup \mathcal{C}$. Then for each $B \in \mathcal{B}'$, pick $U_B \in \mathcal{C}$ such that $B \subseteq U_B$. Then $\{U_B : B \in \mathcal{B}'\}$ is a countable collection of sets of $\mathcal{C}$ whose union equals $\bigcup \mathcal{C}$. Therefore, this proves the lemma.

**Proposition 1.3.5** Let $(X, d)$ be a metric space. Then the following are equivalent.

(1.3.3) $(X, d)$ is compact,

(1.3.4) $(X, d)$ is sequentially compact,

(1.3.5) $(X, d)$ is complete and totally bounded.

**Proof:** Suppose $\operatorname{Lim}$ and let $\{x_k\}$ be a sequence. Suppose $\{x_k\}$ has no convergent subsequence. If this is so, then no value of the sequence is repeated more than finitely many times. Also $\{x_k\}$ has no limit point because if it did, there would exist a subsequence which converges. To see this, suppose $p$ is a limit point of $\{x_k\}$. Then in $B (p, 1)$ there are infinitely many points of $\{x_k\}$. Pick one called $x_k_1$. Now if $x_{k_1}, x_{k_2}, \ldots, x_{k_n}$ have been picked with $x_{k_n} \in B (p, 1/n)$, consider $B (p, 1/(n + 1))$. There are infinitely many points of $\{x_k\}$ in this ball also. Pick $x_{k_{n+1}}$ such that $k_{n+1} > k_n$. Then $\{x_{k_{n+1}}\}_{n=1}^{\infty}$ is a subsequence which converges to $p$ and it is assumed this does not happen.

Thus $\{x_k\}$ has no limit points. It follows the set

$$C_n = \bigcup \{x_k : k \geq n\}$$
1.3. COMPACTNESS IN METRIC SPACE

is a closed set because it has no limit points and if

\[ U_n = C_n, \]

then

\[ X = \bigcup_{n=1}^{\infty} U_n \]

but there is no finite subcovering, because no value of the sequence is repeated more than finitely many times. This contradicts compactness of \((X,d)\). Note \(x_k\) is not in \(U_n\) whenever \(k > n\). Thus implies .

Now suppose and let \(\{x_n\}\) be a Cauchy sequence. Is \(\{x_n\}\) convergent? By sequential compactness \(x_{n_k} \to x\) for some subsequence. By Lemma it follows that \(\{x_n\}\) also converges to \(x\) showing that \((X,d)\) is complete. If \((X,d)\) is not totally bounded, then there exists \(\varepsilon > 0\) for which there is no \(\varepsilon\) net. Hence there exists a sequence \(\{x_k\}\) with \(d(x_k, x_l) \geq \varepsilon\) for all \(l \neq k\). By Lemma again, this contradicts because no subsequence can be a Cauchy sequence and so no subsequence can converge. This shows implies .

Now suppose . What about ? Let \(\{p_n\}\) be a sequence and let \(\{x^n_i\}_{i=1}^m\) be a \(2^{-n}\) net for \(n = 1, 2, \cdots\). Let

\[ B_n = B \left( x^n_i, 2^{-n} \right) \]

be such that \(B_n\) contains \(p_k\) for infinitely many values of \(k\) and \(B_n \cap B_{n+1} \neq \emptyset\). To do this, suppose \(B_n\) contains \(p_k\) for infinitely many values of \(k\). Then one of the sets which intersect \(B_n, B \left( x_i^{n+1}, 2^{-(n+1)} \right)\) must contain \(p_k\) for infinitely many values of \(k\) because all these indices of points from \(\{p_n\}\) contained in \(B_n\) must be accounted for in one of finitely many sets, \(B \left( x_i^{n+1}, 2^{-(n+1)} \right)\). Thus there exists a strictly increasing sequence of integers, \(n_k\) such that \(p_{n_k} \in B_k\).

Then if \(k \geq l\),

\[
d(p_{n_k}, p_{n_l}) \leq \sum_{i=l}^{k-1} d(p_{n_{i+1}}, p_{n_i}) < \sum_{i=l}^{k-1} 2^{-(i-1)} < 2^{-(l-2)},
\]

Consequently \(\{p_{n_k}\}\) is a Cauchy sequence. Hence it converges because the metric space is complete. This proves .

Now suppose and which have now been shown to be equivalent. Let \(D_n\) be a \(n^{-1}\) net for \(n = 1, 2, \cdots\) and let

\[ D = \bigcup_{n=1}^\infty D_n. \]

Thus \(D\) is a countable dense subset of \((X,d)\).

Now let \(C\) be any set of open sets such that \(\bigcup C \supseteq X\). By Lemma , there exists a countable subset of \(C\),

\[ \tilde{C} = \{U_n\}_{n=1}^\infty \]

such that \(\bigcup \tilde{C} = \bigcup C\). If \(C\) admits no finite subcover, then neither does \(\tilde{C}\) and there exists \(p_n \in X \setminus \bigcup_{k=1}^n U_k\). Then since \(X\) is sequentially compact, there is a subsequence \(\{p_{n_k}\}\) such that \(\{p_{n_k}\}\) converges. Say

\[ p = \lim_{k \to \infty} p_{n_k}. \]

All but finitely many points of \(\{p_{n_k}\}\) are in \(X \setminus \bigcup_{k=1}^n U_k\). Therefore \(p \in X \setminus \bigcup_{k=1}^n U_k\) for each \(n\). Hence

\[ p \notin \bigcup_{k=1}^\infty U_k \]

contradicting the construction of \(\{U_n\}_{n=1}^\infty\) which required that \(\bigcup_{n=1}^\infty U_n \supseteq X\). Hence \(X\) is compact. This proves the proposition.

Consider \(\mathbb{R}^n\). In this setting totally bounded and bounded are the same. This will yield a proof of the Heine Borel theorem from advanced calculus.

**Lemma 1.3.6** A subset of \(\mathbb{R}^n\) is totally bounded if and only if it is bounded.
\[ |z - 0| \leq |z - x_j| + |x_j| < 1 + r. \]

Thus \( A \subseteq B(0, r + 1) \) and so \( A \) is bounded.

Now suppose \( A \) is bounded and suppose \( A \) is not totally bounded. Then there exists \( \varepsilon > 0 \) such that there is no \( \varepsilon \) net for \( A \). Therefore, there exists a sequence of points \( \{a_i\} \) with \( |a_i - a_j| \geq \varepsilon \) if \( i \neq j \). Since \( A \) is bounded, there exists \( r > 0 \) such that

\[ A \subseteq [-r, r]^n. \]

\((x \in [-r, r]^n \) means \( x_i \in [-r, r) \) for each \( i \).

Now define \( S \) to be all cubes of the form

\[ \prod_{k=1}^{n} (a_k, b_k) \]

where

\[ a_k = -r + i2^{-p}r, \quad b_k = -r + (i + 1)2^{-p}r, \]

for \( i \in \{0, 1, \cdots, 2^{p+1} - 1\} \). Thus \( S \) is a collection of \( (2^{p+1})^n \) non overlapping cubes whose union equals \([-r, r]^n\) and whose diameters are all equal to \( 2^{-p}r \sqrt{n} \). Now choose \( p \) large enough that the diameter of these cubes is less than \( \varepsilon \). This yields a contradiction because one of the cubes must contain infinitely many points of \( \{a_i\} \). This proves the lemma.

The next theorem is called the Heine Borel theorem and it characterizes the compact sets in \( \mathbb{R}^n \).

**Theorem 1.3.7** A subset of \( \mathbb{R}^n \) is compact if and only if it is closed and bounded.

**Proof:** Since a set in \( \mathbb{R}^n \) is totally bounded if and only if it is bounded, this theorem follows from Proposition \ref{prop:compact} and the observation that a subset of \( \mathbb{R}^n \) is closed if and only if it is complete. This proves the theorem.

### 1.4 Simple Functions And Measurable Functions

Recall that a \( \sigma \) algebra is a collection of subsets of a set \( \Omega \) which includes \( \emptyset, \Omega \), and is closed with respect to countable unions and complements.

**Definition 1.4.1** Let \((\Omega, F)\) be a measurable space, one for which \( F \) is a \( \sigma \) algebra contained in \( P(\Omega) \). Let \( f : \Omega \to X \) where \( X \) is a topological space. Then \( f \) is measurable means that \( f^{-1}(U) \in F \) whenever \( U \) is open.

It is important to have a theorem about pointwise limits of measurable functions, those with the property that inverse images of open sets are measurable. The following is a fairly general such theorem which holds in the situations to be considered in these notes.

**Theorem 1.4.2** Let \( \{f_n\} \) be a sequence of measurable functions mapping \( \Omega \) to \((X, d)\) where \((X, d)\) is a metric space and \((\Omega, F)\) is a measure space. Suppose also that \( f(\omega) = \lim_{n \to \infty} f_n(\omega) \) for all \( \omega \). Then \( f \) is also a measurable function.

**Proof:** It is required to show \( f^{-1}(U) \) is measurable for all \( U \) open. Let

\[ V_m = \left\{ x \in U : \text{dist} \left( x, U^C \right) > \frac{1}{m} \right\}. \]

Thus

\[ V_m \subseteq \left\{ x \in U : \text{dist} \left( x, U^C \right) \geq \frac{1}{m} \right\} \]

and \( V_m \subseteq \overline{V_m} \subseteq V_{m+1} \) and \( \bigcup_m V_m = U \). Then since \( V_m \) is open,

\[ f^{-1}(V_m) = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} f_k^{-1}(V_m) \]
and so
\[
    f^{-1}(U) = \bigcup_{m=1}^{\infty} f^{-1}(V_m) = \bigcup_{m=1}^{\infty} \bigcap_{n=1}^{\infty} f_k^{-1}(V_m) \subseteq \bigcup_{m=1}^{\infty} f^{-1}(V_m) = f^{-1}(U)
\]
which shows \(f^{-1}(U)\) is measurable. This proves the theorem.

There is a fundamental theorem about the relationship of simple functions to measurable functions given in the next theorem.

**Theorem 1.4.3** Let \(f \geq 0\) be measurable. Then there exists a sequence of nonnegative simple functions \(\{s_n\}\) satisfying
\[
    0 \leq s_n(\omega) \leq \cdots \leq s_{n+1}(\omega) \cdots
\]
\[
\lim_{n \to \infty} s_n(\omega) = f(\omega)
\]
for all \(\omega \in \Omega\). If \(f\) is bounded the convergence is actually uniform.

**Proof:** Letting \(I \equiv \{\omega : f(\omega) = \infty\}\), define
\[
t_n(\omega) = \sum_{k=0}^{2^n} \mathcal{X}_{\left\{k/n \leq f < (k+1)/n\right\}}(\omega) + n \mathcal{X}_I(\omega).
\]
Then \(t_n(\omega) \leq f(\omega)\) for all \(\omega\) and \(\lim_{n \to \infty} t_n(\omega) = f(\omega)\) for all \(\omega\). This is because \(t_n(\omega) = n\) for \(\omega \in I\) and if \(f(\omega) \in [0, 2^{n+1}/n]\), then
\[
0 \leq f(\omega) - t_n(\omega) \leq \frac{1}{n}.
\]
Thus whenever \(\omega \notin I\), the above inequality will hold for all \(n\) large enough. Let
\[
s_1 = t_1, \quad s_2 = \max(t_1, t_2), \quad s_3 = \max(t_1, t_2, t_3), \ldots
\]
Then the sequence \(\{s_n\}\) satisfies \(\text{[I]}\).

To verify the last claim, note that in this case the term \(n \mathcal{X}_I(\omega)\) is not present. Therefore, for all \(n\) large enough, \(\text{[I]}\) holds for all \(\omega\). Thus the convergence is uniform. This proves the theorem.

Although it is not needed here, there is a similar theorem which applies to measurable functions which have values in a separable metric space. In this context, a simple function is one which is of the form
\[
\sum_{k=1}^{m} x_k \mathcal{X}_{E_k}(\omega)
\]
where the \(E_k\) are disjoint measurable sets and the \(x_k\) are in \(X\). I am abusing notation somewhat by using a sum. You can’t add in a general metric space. The symbol means the function has value \(x_k\) on the set \(E_k\).

**Theorem 1.4.4** Let \((\Omega, \mathcal{F})\) be a measure space and let \(f : \Omega \to X\) where \((X, d)\) is a separable metric space. Then \(f\) is a measurable function if and only if there exists a sequence of simple functions \(\{f_n\}\) such that for each \(\omega \in \Omega\) and \(n \in \mathbb{N}\),
\[
d(f_n(\omega), f(\omega)) \geq d(f_{n+1}(\omega), f(\omega))
\]
and
\[
\lim_{n \to \infty} d(f_n(\omega), f(\omega)) = 0.
\]

**Proof:** Let \(D = \{x_k\}_{k=1}^{\infty}\) be a countable dense subset of \(X\). First suppose \(f\) is measurable. Then since in a metric space every open set is the countable intersection of closed sets, it follows \(f^{-1}(\text{closed set}) \in \mathcal{F}\). Now let \(D_n = \{x_k\}_{k=1}^{n}\). Let
\[
A_1 \equiv \left\{ \omega : d(x_1, f(\omega)) = \min_{k \leq n} d(x_k, f(\omega)) \right\}
\]
That is, \( A_1 \) are those \( \omega \) such that \( f(\omega) \) is approximated best out of \( D_n \) by \( x_1 \). Why is this a measurable set? It is because \( \omega \rightarrow d(x, f(\omega)) \) is a real valued measurable function, being the composition of a continuous function, \( y \rightarrow d(x, y) \) and a measurable function, \( \omega \rightarrow f(\omega) \). Next let

\[
A_2 = \left\{ \omega \in A_1 : d(x_2, f(\omega)) = \min_{k \leq n} d(x_k, f(\omega)) \right\}
\]

and continue in this manner obtaining disjoint measurable sets, \( \{A_k\}_{k=1}^n \) such that for \( \omega \in A_k \) the best approximation to \( f(\omega) \) from \( D_n \) is \( x_k \). Then

\[
f_n(\omega) = \sum_{k=1}^n x_k A_k(\omega).
\]

Note

\[
\min_{k \leq n+1} d(x_k, f(\omega)) \leq \min_{k \leq n} d(x_k, f(\omega))
\]

and so this verifies 1.4.9. It remains to verify 1.4.10.

Let \( \varepsilon > 0 \) be given and pick \( \omega \in \Omega \). Then there exists \( x_n \in D \) such that \( d(x_n, f(\omega)) < \varepsilon \). It follows from the construction that \( d(f_n(\omega), f(\omega)) \leq d(x_n, f(\omega)) < \varepsilon \). This proves the first half.

Now suppose the existence of the sequence of simple functions as described above. Each \( f_n \) is a measurable function because \( f_n^{-1}(U) = \bigcup \{A_k : x_k \in U\} \). Therefore, the conclusion that \( f \) is measurable follows from Theorem 1.4.2 on Page 14.

### 1.5 One Dimensional Lebesgue Stieltjes Measure

First here is a definition of a term.

**Definition 1.5.1** A measure \( \mu \) defined on a \( \sigma \) algebra of sets \( F \) of \( X \) a topological set will be called a Radon measure if it is complete and outer and inner regular. This means that for all \( E \in F \),

\[
\mu(E) = \sup \{ \mu(K) : K \subseteq E, K \text{ compact} \}
\]

\[
\mu(E) = \inf \{ \mu(V) : V \supseteq E, V \text{ open} \}
\]

One dimensional Lebesgue Stieltjes measure is just a generalization of Lebesgue measure. Instead of the functional,

\[
L_f = \int f(x) \, dx, \quad f \in C_c(\mathbb{R}),
\]

you use the functional

\[
L_f = \int f(x) \, dF(x), \quad f \in C_c(\mathbb{R}),
\]

where \( F \) is an increasing function defined on \( \mathbb{R} \) and this is just the Riemann Stieltjes integral. Therefore, by the Riesz representation theorem, there exists a unique Radon measure \( \mu \) representing the functional. Thus

\[
\int_{\mathbb{R}} fd\mu = \int_{\mathbb{R}} fdF
\]

for all \( f \in C_c(\mathbb{R}) \). Now consider what this measure does to intervals. To begin with, consider what it does to the closed interval, \([a, b]\). The following picture may help.

![Diagram](image)

In this picture \( \{a_n\} \) increases to \( a \) and \( b_n \) decreases to \( b \). Also suppose \( a, b \) are points of continuity of \( F \). Therefore,

\[
F(b) - F(a) \leq Lf_n = \int_{\mathbb{R}} f_n d\mu \leq F(b_n) - F(a_n)
\]
Passing to the limit and using the dominated convergence theorem, this shows
\[ \mu([a, b]) = F(b) - F(a) = F(b^+) - F(a^-). \]

Next suppose \(a, b\) are arbitrary, maybe not points of continuity of \(F\). Then letting \(a_n\) and \(b_n\) be as in the above picture which are points of continuity of \(F\),
\[ \mu([a, b]) = \lim_{n \to \infty} \mu([a_n, b_n]) = \lim_{n \to \infty} F(b_n) - F(a_n) = F(b^+) - F(a^-). \]

In particular \(\mu(a) = F(a^+) - F(a^-)\) and so
\[ \mu((a, b)) = F(b^+) - F(a^-) - (F(a^+) - F(a^-)) - (F(b^+) - F(b^-)) = F(b^-) - F(a^+) \]

This shows what \(\mu\) does to intervals. This is stated as the following proposition.

**Proposition 1.5.2** Let \(\mu\) be the measure representing the functional
\[ Lf \equiv \int f dF, \quad f \in C_c(\mathbb{R}) \]
for \(F\) an increasing function defined on \(\mathbb{R}\). Then
\[ \mu([a, b]) = F(b^+) - F(a^-) \]
\[ \mu((a, b)) = F(b^-) - F(a^+) \]
\[ \mu(a) = F(a^+) - F(a^-). \]

**Observation 1.5.3** Note that all the above would work as well if
\[ Lf \equiv \int f dF, \quad f \in C_c([0, \infty)) \]
where \(F\) is continuous at 0 and \(\nu\) is the measure representing this functional. This is because you could just extend \(F(x)\) to equal \(F(0)\) for \(x \leq 0\) and apply the above to the extended \(F\). In this case, \(\nu([0, b]) = F(b^+) - F(0)\).

## 1.6 The Distribution Function

There is an interesting connection between the Lebesgue integral of a nonnegative function with something called the distribution function.

**Definition 1.6.1** Let \(f \geq 0\) and suppose \(f\) is measurable. The distribution function is the function defined by
\[ t \rightarrow \mu([t < f]). \]

**Lemma 1.6.2** If \(\{f_n\}\) is an increasing sequence of functions converging pointwise to \(f\) then
\[ \mu([f > t]) = \lim_{n \to \infty} \mu([f_n > t]) \]

**Proof:** The sets, \([f_n > t]\) are increasing and their union is \([f > t]\) because if \(f(\omega) > t\), then for all \(n\) large enough, \(f_n(\omega) > t\) also. Therefore, the desired conclusion follows from properties of measures.

**Lemma 1.6.3** Suppose \(s \geq 0\) is a measurable simple function,
\[ s(\omega) = \sum_{k=1}^{n} a_k \chi_{E_k}(\omega) \]
where the \(a_k\) are the distinct nonzero values of \(s, 0 < a_1 < a_2 < \cdots < a_n\). Suppose \(\phi\) is a \(C^1\) function defined on \([0, \infty)\) which has the property that \(\phi(0) = 0, \phi'(t) > 0\) for all \(t\). Then
\[ \int_0^\infty \phi'(t) \mu([s > t]) \, dm = \int \phi(s) \, d\mu. \]
Proof: First note that if \( \mu (E_k) = \infty \) for any \( k \) then both sides equal \( \infty \) and so without loss of generality, assume \( \mu (E_k) < \infty \) for all \( k \). Letting \( a_0 \equiv 0 \), the left side equals

\[
\sum_{k=1}^{n} \int_{a_{k-1}}^{a_k} \phi' (t) \mu ([s > t]) \, dm(t) = \sum_{k=1}^{n} \int_{a_{k-1}}^{a_k} \phi' (t) \sum_{i=k}^{n} \mu (E_i) \, dm \\
= \sum_{k=1}^{n} \sum_{i=k}^{n} \mu (E_i) \int_{a_{k-1}}^{a_k} \phi' (t) \, dm \\
= \sum_{k=1}^{n} \sum_{i=k}^{n} \mu (E_i) (\phi (a_k) - \phi (a_{k-1})) \\
= \sum_{i=1}^{n} \mu (E_i) \sum_{k=1}^{i} (\phi (a_k) - \phi (a_{k-1})) \\
= \sum_{i=1}^{n} \mu (E_i) \phi (a_i) = \int \phi (s) \, d\mu. \]

With this lemma the next theorem which is the main result follows easily.

**Theorem 1.6.4** Let \( f \geq 0 \) be measurable and let \( \phi \) be a \( C^1 \) function defined on \([0, \infty)\) which satisfies \( \phi' (t) > 0 \) for all \( t > 0 \) and \( \phi (0) = 0 \). Then

\[
\int \phi (f) \, d\mu = \int_{0}^{\infty} \phi' (t) \mu ([f > t]) \, dm.
\]

Proof: By Theorem 1.6.3 on Page 17 there exists an increasing sequence of nonnegative simple functions, \( \{s_n\} \) which converges pointwise to \( f \). By the monotone convergence theorem and Lemma 1.6.4,

\[
\int \phi (f) \, d\mu = \lim_{n \to \infty} \int \phi (s_n) \, d\mu = \lim_{n \to \infty} \int_{0}^{\infty} \phi' (t) \mu ([s_n > t]) \, dm \\
= \int_{0}^{\infty} \phi' (t) \mu ([f > t]) \, dm \]

This theorem can be generalized to a situation in which \( \phi \) is only increasing and continuous. In the generalization I will replace the symbol \( \phi \) with \( F \) to coincide with earlier notation.

**Lemma 1.6.5** Suppose \( s \geq 0 \) is a measurable simple function,

\[
s (\omega) \equiv \sum_{k=1}^{n} a_k \chi_{E_k} (\omega)
\]

where the \( a_k \) are the distinct nonzero values of \( s, a_1 < a_2 < \cdots < a_n \). Suppose \( F \) is an increasing function defined on \([0, \infty), F (0) = 0, F' \) being continuous at 0 from the right and continuous at every \( a_k \). Then letting \( \mu \) be a measure and \((\Omega, F, \mu)\) a measure space,

\[
\int_{(0, \infty]} \mu ([s > t]) \, d\nu = \int_{\Omega} F (s) \, d\mu.
\]

where the integral on the left is the Lebesgue integral for the measure \( \nu \) given as the Radon measure representing the functional

\[
\int_{0}^{\infty} g \, dF
\]

for \( g \in C_c ([0, \infty)) \).

Proof: This follows from the following computation and Proposition 1.6.4. Since \( F \) is continuous at 0 and the values \( a_k \),

\[
\int_{0}^{\infty} \mu ([s > t]) \, d\nu (t) = \sum_{k=1}^{n} \int_{(a_{k-1}, a_k]} \mu ([s > t]) \, d\nu (t)
\]
\[= \sum_{k=1}^{n} \int_{(a_{k-1}, a_k]} \sum_{j=k}^{n} \mu(E_j) \, dF(t) = \sum_{j=1}^{n} \mu(E_j) \sum_{k=1}^{j} \nu((a_{k-1}, a_k]) \]

\[
= \sum_{j=1}^{n} \mu(E_j) \sum_{k=1}^{j} (F(a_k) - F(a_{k-1})) = \sum_{j=1}^{n} \mu(E_j) F(a_j) = \int_{\Omega} F(s) \, d\mu
\]

Now here is the generalization to nonnegative measurable functions.

**Theorem 1.6.6** Let \( f \geq 0 \) be measurable with respect to \( F \) where \((\Omega, F, \mu)\) a measure space, and let \( F \) be an increasing continuous function defined on \([0, \infty)\) and \( F(0) = 0 \). Then

\[
\int_{\Omega} F(f) \, d\mu = \int_{(0, \infty]} \mu([f > t]) \, d\nu(t)
\]

where \( \nu \) is the Radon measure representing

\[
Lg = \int_{0}^{\infty} gdF
\]

for \( g \in C_{c}([0, \infty)) \).

**Proof:** By Theorem 1.6.5 on Page 16 there exists an increasing sequence of nonnegative simple functions, \( \{s_{n}\} \) which converges pointwise to \( f \). By the monotone convergence theorem and Lemma 1.6.3,

\[
\int_{\Omega} F(f) \, d\mu = \lim_{n \to \infty} \int_{\Omega} F(s_{n}) \, d\mu = \lim_{n \to \infty} \int_{(0, \infty]} \mu([s_{n} > t]) \, d\nu
\]

\[
= \int_{(0, \infty]} \mu([f > t]) \, d\nu
\]

Note that the function \( t \to \mu([f > t]) \) is a decreasing function. Therefore, one can make sense of an improper Riemann Stieltjes integral

\[
\int_{0}^{\infty} \mu([f > t]) \, dF(t).
\]

With more work, one can have this equal to the corresponding Lebesgue integral above.

### 1.7 Good Lambda Inequality

There is a very interesting and important inequality called the good lambda inequality (I am not sure if there is a bad lambda inequality.) which follows from the above theory of distribution functions. It involves the inequality

\[\mu([f > \beta \lambda] \cap [g \leq \delta \lambda]) \leq \phi(\delta) \mu([f > \lambda])\]

for \( \beta > 1 \), nonnegative functions \( f, g \) and is supposed to hold for all small positive \( \delta \) and \( \phi(\delta) \to 0 \) as \( \delta \to 0 \). Note the left side is small when \( g \) is large and \( f \) is small. The inequality involves dominating an integral involving \( f \) with one involving \( g \) as described below. As above, \( \nu \) is the measure which comes from the functional \( \int_{\mathbb{R}} gdF \) for \( g \in C_{c}(\mathbb{R}) \).

**Theorem 1.7.1** Let \((\Omega, F, \mu)\) be a finite measure space and let \( F \) be a continuous increasing function defined on \([0, \infty)\) such that \( F(0) = 0 \). Suppose also that for all \( \alpha > 1 \), there exists a constant \( C_{\alpha} \) such that for all \( x \in [0, \infty) \),

\[F(\alpha x) \leq C_{\alpha} F(x)\]

Also suppose \( f, g \) are nonnegative measurable functions and there exists \( \beta > 1, 0 < r \leq 1 \), such that for all \( \lambda > 0 \) and \( 1 > \delta > 0 \),

\[\mu([f > \beta \lambda] \cap [g \leq r \delta \lambda]) \leq \phi(\delta) \mu([f > \lambda])\]

where \( \lim_{\delta \to 0^{+}} \phi(\delta) = 0 \) and \( \phi \) is increasing. Under these conditions, there exists a constant \( C \) depending only on \( \beta, \phi, r \) such that

\[
\int_{\Omega} F(f(\omega)) \, d\mu(\omega) \leq C \int_{\Omega} F(g(\omega)) \, d\mu(\omega).
\]
Proof: Let $\beta > 1$ be as given above. First suppose $f$ is bounded.

\[
\int_{\Omega} F(f) \, d\mu = \int_{\Omega} F\left(\frac{f}{\beta}\right) \, d\mu \leq C_{\beta} \int_{\Omega} F\left(\frac{f}{\beta}\right) \, d\mu
\]

\[
= C_{\beta} \int_{0}^{\infty} \mu (\{f > \beta \lambda\}) \, d\nu
\]

Now using the given inequality,

\[
= C_{\beta} \int_{0}^{\infty} \mu (\{f > \beta \lambda\} \cap [g \leq r\delta \lambda]) \, d\nu
\]

\[+ C_{\beta} \int_{0}^{\infty} \mu (\{f > \beta \lambda\} \cap [g > r\delta \lambda]) \, d\nu \leq C_{\beta}\phi(\delta) \int_{0}^{\infty} \mu (\{f > \lambda\}) \, d\nu + C_{\beta} \int_{0}^{\infty} \mu (\{g > r\delta \lambda\}) \, d\nu
\]

\[
\leq C_{\beta}\phi(\delta) \int_{\Omega} F(f) \, d\mu + C_{\beta} \int_{\Omega} F\left(\frac{g}{r\delta}\right) \, d\mu
\]

Now choose $\delta$ small enough that $C_{\beta}\phi(\delta) < \frac{1}{2}$ and then subtract the first term on the right in the above from both sides. It follows from the properties of $F$ again that

\[
\frac{1}{2} \int_{\Omega} F(f) \, d\mu \leq C_{\beta} C_{(r\delta)^{-1}} \int_{\Omega} F(g) \, d\mu.
\]

This establishes the inequality in the case where $f$ is bounded.

In general, let $f_n = \min (f, n)$. Then for $n \leq \lambda$, the inequality

\[
\mu (\{f > \beta \lambda\} \cap [g \leq r\delta \lambda]) \leq \phi(\delta) \mu (\{f > \lambda\})
\]

holds with $f$ replaced with $f_n$ because both sides equal 0 thanks to $\beta > 1$. If $n > \lambda$, then $\{f > \lambda\} = \{f_n > \lambda\}$ and so the inequality still holds because in this case,

\[
\mu (\{f_n > \beta \lambda\} \cap [g \leq r\delta \lambda]) \leq \phi(\delta) \mu (\{f_n > \lambda\})
\]

Therefore, (1.7.11) is valid with $f$ replaced with $f_n$. Now pass to the limit as $n \to \infty$ and use the monotone convergence theorem. 

1.8 Regularity Of Measures In Polish Space

Definition 1.8.1 A Polish space is a complete separable metric space.

For example, $\mathbb{R}$ is a Polish space as is any separable Banach space. Amazing things can be said about finite measures on Polish spaces.

Definition 1.8.2 A measure, $\mu$ defined on $\mathcal{B}(E)$ will be called inner regular if for all $F \in \mathcal{B}(E)$,

\[
\mu (F) = \sup \{ \mu (K) : K \subseteq F \text{ and } K \text{ is closed} \}
\]

A measure, $\mu$ defined on $\mathcal{B}(E)$ will be called outer regular if for all $F \in \mathcal{B}(E)$,

\[
\mu (F) = \inf \{ \mu (V) : V \supseteq F \text{ and } V \text{ is open} \}
\]

When a measure is both inner and outer regular, it is called regular.

For probability measures, the above definition of regularity tends to come free. Note it is a little weaker than the usual definition of regularity because $K$ is only assumed to be closed, not compact.

Lemma 1.8.3 Let $\mu$ be a finite measure defined on $\mathcal{B}(E)$ where $E$ is a metric space. Then $\mu$ is regular.
Proof: First note every open set is the countable union of closed sets and every closed set is the countable intersection of open sets. Here is why. Let $V$ be an open set and let

$$K_k \equiv \{ x \in V : \text{dist}(x, V^c) \geq 1/k \}.$$  

Then clearly the union of the $K_k$ equals $V$. Next, for $K$ closed let

$$V_k \equiv \{ x \in E : \text{dist}(x, K) < 1/k \}.$$  

Clearly the intersection of the $V_k$ equals $K$. Therefore, letting $V$ denote an open set and $K$ a closed set,

$$\mu(V) = \sup \{ \mu(K) : K \subseteq V \text{ and } K \text{ is closed} \}$$  

$$\mu(K) = \inf \{ \mu(V) : V \supseteq K \text{ and } V \text{ is open} \}.$$  

Also since $V$ is open and $K$ is closed,

$$\mu(V) = \inf \{ \mu(U) : U \supseteq V \text{ and } V \text{ is open} \}$$  

$$\mu(K) = \sup \{ \mu(L) : L \subseteq K \text{ and } L \text{ is closed} \}.$$  

In words, $\mu$ is regular on open and closed sets. Let

$$\mathcal{F} \equiv \{ F \in \mathcal{B}(E) \text{ such that } \mu \text{ is regular on } F \}.$$  

Then $\mathcal{F}$ contains the open sets and the closed sets.

Suppose $F \in \mathcal{F}$. Then there exists $V \supseteq F$ with $\mu(V \setminus F) < \varepsilon$. It follows $V^c \subseteq F^c$ and

$$\mu(F^c \setminus V^c) = \mu(V \setminus F) < \varepsilon.$$  

Thus $F^c$ is inner regular. Since $F \in \mathcal{F}$, there exists $K \subseteq F$ where $K$ is closed and $\mu(F \setminus K) < \varepsilon$. Then also $K^c \supseteq F^c$ and

$$\mu(K^c \setminus F^c) = \mu(F \setminus K) < \varepsilon.$$  

Thus if $F \in \mathcal{F}$ so is $F^c$.

Suppose now that $\{F_i\} \subseteq \mathcal{F}$, the $F_i$ being disjoint. Is $\bigcup F_i \in \mathcal{F}$? There exists $K_i \subseteq F_i$ such that $\mu(K_i) + \varepsilon/2^i > \mu(F_i)$. Then

$$\mu(\bigcup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} \mu(F_i) \leq \varepsilon + \sum_{i=1}^{\infty} \mu(K_i)$$

$$< 2\varepsilon + \sum_{i=1}^{N} \mu(K_i) = 2\varepsilon + \mu(\bigcup_{i=1}^{N} K_i)$$

provided $N$ is large enough. Thus it follows $\bigcup_{i=1}^{\infty} F_i$ is inner regular. Why is it outer regular? Let $V_i \supseteq F_i$ such that $\mu(F_i) + \varepsilon/2^i > \mu(V_i)$ and

$$\mu(\bigcup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} \mu(F_i) \geq -\varepsilon + \sum_{i=1}^{\infty} \mu(V_i) \geq -\varepsilon + \mu(\bigcup_{i=1}^{\infty} V_i)$$

which shows $\bigcup_{i=1}^{\infty} F_i$ is outer regular. It follows $\mathcal{F}$ contains the $\pi$ system consisting of open sets and so by the Lemma on $\pi$ systems, Lemma 1.1.3, $\mathcal{F}$ contains $\sigma(\tau)$ where $\tau$ is the set of open sets. Hence $\mathcal{F}$ contains the Borel sets and is itself a subset of the Borel sets by definition. Therefore, $\mathcal{F} = \mathcal{B}(E)$. 

One can say more if the metric space is complete and separable. In fact in this case the above definition of inner regularity can be shown to imply the usual one.

Lemma 1.8.4 Let $\mu$ be a finite measure on a $\sigma$ algebra containing $\mathcal{B}(E)$, the Borel sets of $E$, a separable complete metric space. Then if $C$ is a closed set,

$$\mu(C) = \sup \{ \mu(K) : K \subseteq C \text{ and } K \text{ is compact} \}.$$  

Proof: Let \( \{a_k\} \) be a countable dense subset of \( C \). Thus \( \bigcup_{k=1}^{\infty} B\left(a_k, \frac{1}{n}\right) \supseteq C \). Therefore, there exists \( m_n \) such that
\[
\mu\left(C \setminus \bigcup_{k=1}^{m_n} B\left(a_k, \frac{1}{n}\right)\right) = \mu(C \setminus C_n) < \frac{\varepsilon}{2n}.
\]

Now let \( K = C \cap \left(\cap_{n=1}^{\infty} C_n\right) \). Then \( K \) is a subset of \( C_n \) for each \( n \) and so for each \( \varepsilon > 0 \) there exists an \( \varepsilon \) net for \( K \) since \( C_n \) has a \( 1/n \) net, namely \( a_1, \ldots, a_{m_n} \). Since \( K \) is closed, it is complete and so it is also compact since it is complete and totally bounded. Now
\[
\mu(C \setminus K) = \mu\left(\bigcup_{n=1}^{\infty} (C \setminus C_n)\right) < \sum_{n=1}^{\infty} \frac{\varepsilon}{2n} = \varepsilon.
\]

Thus \( \mu(C) \) can be approximated by \( \mu(K) \) for \( K \) a compact subset of \( C \). \( \blacksquare \)

**Definition 1.8.5** A measurable function \( X : (\Omega, \mathcal{F}, \mu) \to Z \) a topological space is called a random variable when \( \mu(\Omega) = 1 \). For such a random variable, one can define a distribution measure \( \lambda_X \) on the Borel sets of \( Z \) as follows.
\[
\lambda_X(G) = \mu(\{X^{-1}(G) \in \mathcal{F}\}).
\]

This is a well defined measure on the Borel sets of \( Z \) because it makes sense for every \( G \) open and \( G \equiv \{G \subseteq Z : X^{-1}(G) \in \mathcal{F}\} \) is a \( \sigma \) algebra which contains the open sets, hence the Borel sets. Such a random variable is also called a random vector when \( Z \) is a vector space.

**Corollary 1.8.6** Let \( X \) be a random variable with values in a separable complete metric space, \( Z \). Then \( \lambda_X \) is an inner and outer regular measure defined on \( B(Z) \).

### 1.9 The Ergodic Theorem

I am putting this theorem here because it seems to fit in well with the material of this chapter.

In this section \( (\Omega, \mathcal{F}, \mu) \) will be a finite measure space. This means that \( \mu(\Omega) < \infty \). The mapping, \( T : \Omega \to \Omega \) will satisfy the following condition.
\[
T(A), T^{-1}(A) \in \mathcal{F} \text{ whenever } A \in \mathcal{F}, \text{ } T \text{ is one to one.} \tag{1.9.12}
\]

For example, you could have \( T \) a homeomorphism on some topological space \( X \) and the \( \sigma \) algebra could be the Borel sets.

**Lemma 1.9.1** If \( T \) satisfies \( \text{(1.9.12)} \), then \( f \circ T \) is measurable whenever \( f \) is measurable.

Proof: Let \( U \) be an open set. Then
\[
(f \circ T)^{-1}(U) = T^{-1}(f^{-1}(U)) \in \mathcal{F}
\]
by \( \text{(1.9.12)} \). \( \blacksquare \)

Now suppose that in addition to \( \text{(1.9.12)} \), \( T \) also satisfies
\[
\mu(T^{-1}A) = \mu(A), \tag{1.9.13}
\]
for all \( A \in \mathcal{F} \). In words, \( T^{-1} \) is measure preserving. Note that also
\[
\mu(TA) = \mu(T^{-1}TA) = \mu(A)
\]
so also \( T \) is measure preserving. Then for \( T \) satisfying \( \text{(1.9.12)} \) and \( \text{(1.9.13)} \), we have the following simple lemma.

**Lemma 1.9.2** If \( T \) satisfies \( \text{(1.9.12)} \) and \( \text{(1.9.13)} \) then whenever \( f \) is nonnegative and measurable,
\[
\int_{\Omega} f(\omega) \, d\mu = \int_{\Omega} f(T\omega) \, d\mu. \tag{1.9.14}
\]

Also \( \text{(1.9.14)} \) holds whenever \( f \in L^1(\Omega) \).
Proof: Let \( f \geq 0 \) and \( f \) is measurable. Let \( A \in \mathcal{F} \). Then from \( \text{Lemma 1.9.4} \),

\[
\int_{\Omega} \mathcal{X}_A (\omega) \, d\mu = \mu (A) = \mu (T^{-1} (A)) = \int_{\Omega} \mathcal{X}_{T^{-1} (A)} (\omega) \, d\mu = \int_{\Omega} \mathcal{X}_A (T (\omega)) \, d\mu.
\]

It follows that whenever \( s \) is a simple function,

\[
\int_{\Omega} s (\omega) \, d\mu = \int_{\Omega} s (T \omega) \, d\mu.
\]

If \( f \geq 0 \) and measurable, Theorem [3.13] on Page 13 implies there exists an increasing sequence of simple functions, \( \{s_n\} \) converging pointwise to \( f \). Then the result follows from monotone convergence theorem. Splitting \( f \in L^1 \) into real and imaginary parts we apply this to the positive and negative parts of these and obtain [3.14] in this case also.

Definition 1.9.3 A measurable function, \( f \), is said to be invariant if

\[
f (T \omega) = f (\omega).
\]

A set, \( A \in \mathcal{F} \) is said to be invariant if \( \mathcal{X}_A \) is an invariant function. Thus a set is invariant if and only if \( T^{-1} A = A \). (\( \mathcal{X}_A (T \omega) = \mathcal{X}_{T^{-1} (A)} (\omega) \) so to say that \( \mathcal{X}_A \) is invariant is to say that \( T^{-1} A = A \).)

The following theorem, the individual ergodic theorem, is the main result. Define \( T^0 (\omega) = \omega \). Let

\[
S_n f (\omega) \equiv \sum_{k=1}^{n} f (T^{k-1} \omega), \quad S_0 f (\omega) \equiv 0.
\]

Also define the following maximal type function \( M_\infty f (\omega) \)

\[
M_\infty f (\omega) \equiv \sup \{ S_k f (\omega) : 0 \leq k \} \quad (1.9.15)
\]

and let

\[
M_n f (\omega) \equiv \sup \{ S_k f (\omega) : 0 \leq k \leq n \} \quad (1.9.16)
\]

Then one can prove the following interesting lemma.

Lemma 1.9.4 Let \( f \in L^1 (\mu) \) where \( f \) has real values. Then \( \int_{\{M_\infty f > 0\}} f \, d\mu \geq 0 \).

Proof: First note that \( M_n f (\omega) \geq 0 \) for all \( n \) and \( \omega \). This follows easily from the observation that by definition, \( S_0 f (\omega) = 0 \) and so \( M_n f (\omega) \) is at least as large. There is certainly something to show here because the integrand is not known to be nonnegative.

Let \( T^* h = h \circ T \). Thus \( T^* \) is linear and maps measurable functions to measurable functions by Lemma [3.18]. It is also clear that if \( h \geq 0 \), then \( T^* h \geq 0 \) also. Therefore, for large \( k \),

\[
S_k f (\omega) = \sum_{j=1}^{k} f (T^{j-1} \omega) = f (\omega) + \sum_{j=2}^{k} f (T^{j-1} \omega) = f (\omega) + T^* \sum_{j=1}^{k-1} f (T^{j-1} \omega)
\]

and so, taking the supremum for \( k \leq n \),

\[
M_n f (\omega) \leq f (\omega) + T^* M_n f (\omega).
\]

Now since \( M_n f \geq 0 \),

\[
\int_{\Omega} M_n f (\omega) \, d\mu = \int_{\{M_n f > 0\}} M_n f (\omega) \, d\mu
\]

\[
\leq \int_{\{M_n f > 0\}} f (\omega) \, d\mu + \int_{\Omega} T^* M_n f (\omega) \, d\mu
\]

\[
= \int_{\{M_n f > 0\}} f (\omega) \, d\mu + \int_{\Omega} M_n f (\omega) \, d\mu
\]
by Lemma [lemma-number]. It follows that
\[ \int_{[M_n f > 0]} f(\omega) \, d\mu \geq 0 \]
for each \( n \). Also, since \( M_n f(\omega) \to M_\infty f(\omega) \), the following pointwise convergence holds.
\[ \mathcal{X}_{[M_n f > 0]}(\omega) f(\omega) \to \mathcal{X}_{[M_\infty f > 0]}(\omega) f(\omega) \]
Since \( f \) is in \( L^1 \), the dominated convergence theorem implies
\[ \int_{[M_\infty f > 0]} f(\omega) \, d\mu = \lim_{n \to \infty} \int_{[M_n f > 0]} f(\omega) \, d\mu \geq 0. \]

**Theorem 1.9.5** Let \((\Omega, \mathcal{F}, \mu)\) be a probability space and let \( T : \Omega \to \Omega \) satisfy \([lemma-number]\) and \([lemma-number]\). \( T^{-1} \) is measure preserving and \( T^{-1} \) maps \( \mathcal{F} \) to \( \mathcal{F} \) and \( T \) is one to one. Then if \( f \in L^1(\Omega) \) having real or complex values and
\[ S_n f(\omega) \equiv \sum_{k=1}^{n} f(T^{k-1}\omega), \quad S_0 f(\omega) \equiv 0, \quad (1.9.17) \]
it follows there exists a set of measure zero \( N \), and an invariant function \( g \) such that for all \( \omega \notin N \),
\[ \lim_{n \to \infty} \frac{1}{n} S_n f(\omega) = g(\omega). \quad (1.9.18) \]

**Proof:** To begin with, we assume \( f \) has real values. Now if \( A \) is an invariant set, \( \mathcal{X}_A(T^n \omega) = \mathcal{X}_A(\omega) \) and so
\[ S_n(\mathcal{X}_A f)(\omega) \equiv \sum_{k=1}^{n} f(T^{k-1}\omega) \mathcal{X}_A(T^{k-1}\omega) = \sum_{k=1}^{n} f(T^{k-1}\omega) \mathcal{X}_A(\omega) \]
\[ = \mathcal{X}_A(\omega) \sum_{k=1}^{n} f(T^{k-1}\omega) = \mathcal{X}_A(\omega) S_n f(\omega). \]
Therefore, for such an invariant set,
\[ M_n(\mathcal{X}_A f)(\omega) = \mathcal{X}_A(\omega) M_n f(\omega), \quad M_\infty(\mathcal{X}_A f)(\omega) = \mathcal{X}_A(\omega) M_\infty f(\omega). \quad (1.9.19) \]
Let \(-\infty < a < b < \infty\) and define
\[ N_{ab} \equiv \left[ -\infty < \lim \inf_{n \to \infty} \frac{1}{n} S_n f(\omega) < a < b < \lim \sup_{n \to \infty} \frac{1}{n} S_n f(\omega) < \infty \right] \quad (1.9.20) \]
Observe that from the definition,
\[ \lim \inf_{n \to \infty} \frac{1}{n} S_n f(\omega) = \lim \inf_{n \to \infty} \frac{1}{n} S_n f(T \omega) \]
and
\[ \lim \sup_{n \to \infty} \frac{1}{n} S_n f(\omega) = \lim \sup_{n \to \infty} \frac{1}{n} S_n f(T \omega). \]
Thus if \( \omega \in N_{ab} \), it follows that \( T \omega \in N_{ab} \) and if \( T \omega \in N_{ab} \), then so is \( \omega \). Thus \( N_{ab} \) is an invariant set. Also, if \( \omega \in N_{ab} \), then
\[ a - \lim \inf_{n \to \infty} \frac{1}{n} S_n f(\omega) = \lim \sup_{n \to \infty} \left( a - \frac{1}{n} S_n f(\omega) \right) > 0 \]
and
\[ \lim \sup_{n \to \infty} \left( \frac{1}{n} S_n f(\omega) - b \right) > 0 \]
It follows that
\[ N_{ab} \subseteq [M_\infty (f - b) > 0] \cap [M_\infty (a - f) > 0]. \]
Consequently, since $N_{ab}$ is invariant, argued above, $\mathcal{X}_{N_{ab}} M_\infty (f - b) = M_\infty (\mathcal{X}_{N_{ab}} (f - b))$ and so from Lemma 1.9.22,

$$\int_{N_{ab}} (f (\omega) - b) \, d\mu = \int_{[\mathcal{X}_{N_{ab}} M_\infty (f - b) > 0]} \mathcal{X}_{N_{ab}} (f (\omega) - b) \, d\mu$$

$$= \int_{[M_\infty (\mathcal{X}_{N_{ab}} (f - b)) > 0]} \mathcal{X}_{N_{ab}} (f (\omega) - b) \, d\mu \geq 0 \quad (1.9.21)$$

and

$$\int_{N_{ab}} (a - f (\omega)) \, d\mu = \int_{[\mathcal{X}_{N_{ab}} M_\infty (a - f) > 0]} \mathcal{X}_{N_{ab}} (a - f (\omega)) \, d\mu$$

$$= \int_{[M_\infty (\mathcal{X}_{N_{ab}} (a - f)) > 0]} \mathcal{X}_{N_{ab}} (a - f (\omega)) \, d\mu \geq 0 \quad (1.9.22)$$

It follows that

$$a\mu (N_{ab}) \geq \int_{N_{ab}} f \, d\mu \geq b\mu (N_{ab}). \quad (1.9.23)$$

Since $a < b$, it follows that $\mu (N_{ab}) = 0$.

Now let

$$N \equiv \cup \{ N_{ab} : a < b, \ a, b \in \mathbb{Q} \}.$$

It follows that $\mu (N) = 0$. Now $TN_{a,b} = N_{a,b}$ and so

$$T (N) = \cup_{a,b} T (N_{a,b}) = \cup_{a,b} N_{a,b} = N.$$

Thus, $T^n N = N$ for all $n \in \mathbb{N}$. For $\omega \notin N$, $\lim_{n \to \infty} \frac{1}{n} S_n f (\omega)$ exists. Now let

$$g (\omega) \equiv \begin{cases} 0 & \text{if } \omega \in N \\ \lim_{n \to \infty} \frac{1}{n} S_n f (\omega) & \text{if } \omega \notin N \end{cases}.$$

Then it is clear $g$ satisfies the conditions of the theorem because if $\omega \in N$, then $T \omega \in N$ also and so in this case, $g (T \omega) = g (\omega) = 0$. On the other hand, if $\omega \notin N$, then

$$g (T \omega) = \lim_{n \to \infty} \frac{1}{n} S_n f (T \omega) = \lim_{n \to \infty} \frac{1}{n} S_n f (\omega) = g (\omega).$$

Which shows that $g$ is invariant. Also, from Lemma 1.9.24,

$$\int \frac{1}{n} S_n f (\omega) \, d\mu \leq \liminf_{n \to \infty} \int \frac{1}{n} S_n f \bigg| \, d\mu \leq \liminf_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \int |f (T^{k-1} \omega)| \, d\mu$$

$$= \liminf_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \int |f (\omega)| \, d\mu = \|f\|_L^1,$$

so $g \in L^1 (\Omega, \mu)$.

The last claim about convergence in $L^1$ follows from the Vitali convergence theorem if we verify the sequence, \(\{\frac{1}{n} S_n f\}_{n=1}^{\infty}\) is uniformly integrable. To see this is the case, we know $f \in L^1 (\Omega)$ and so if $\varepsilon > 0$ is given, there exists $\delta > 0$ such that whenever $B \in \mathcal{F}$ and $\mu (B) \leq \delta$, then $|\int_B f (\omega) \, d\mu| < \varepsilon$. Taking $\mu (A) < \delta$, it follows

$$\left| \int_A \frac{1}{n} S_n f (\omega) \, d\mu \right| = \left| \frac{1}{n} \sum_{k=1}^{n} \int_A f (T^{k-1} \omega) \, d\mu \right| = \frac{1}{n} \sum_{k=1}^{n} \int_{\Omega} \mathcal{X}_A (\omega) f (T^{k-1} \omega) \, d\mu$$

$$= \frac{1}{n} \sum_{k=1}^{n} \int_{\Omega} \mathcal{X}_A (T^{k-1} T^{-(k-1)} \omega) f (T^{k-1} \omega) \, d\mu = \frac{1}{n} \sum_{k=1}^{n} \int_{\Omega} \mathcal{X}_A (T^{-(k-1)} \omega) f (\omega) \, d\mu$$

$$\leq \frac{1}{n} \sum_{k=1}^{n} \int_{T^{k-1} (A)} f (\omega) \, d\mu \leq \frac{1}{n} \sum_{k=1}^{n} \int_{T^{k-1} (A)} f (\omega) \, d\mu < \frac{1}{n} \sum_{k=1}^{n} \varepsilon = \varepsilon.$$
because $\mu (T^{(k-1)}A) = \mu (A)$ by assumption. This proves the above sequence is uniformly integrable and so, by the Vitali convergence theorem,

$$\lim_{n \to \infty} \int_{\Omega} \frac{1}{n} S_n f - g \, d\mu = 0.$$  

This proves the theorem in the case the function has real values. In the case where $f$ has complex values, apply the above result to the real and imaginary parts of $f$. ■

**Definition 1.9.6** The above mapping $T$ is ergodic if the only invariant sets have measure 0 or 1.

If the map, $T$ is ergodic, the following corollary holds.

**Corollary 1.9.7** In the situation of Theorem 1.9.5, if $T$ is ergodic, then

$$g(\omega) = \int f(\omega) \, d\mu$$

for a.e. $\omega$.

**Proof:** Let $g$ be the function of Theorem 1.9.5 and let $R_1$ be a rectangle in $\mathbb{R}^2 = \mathbb{C}$ of the form $[-a,a] \times [-a,a]$ such that $g^{-1}(R_1)$ has measure greater than 0. This set is invariant because the function, $g$ is invariant and so it must have measure 1. Divide $R_1$ into four equal rectangles, $R_1', R_2', R_3', R_4'$. Then one of these, renamed $R_2$ has the property that $g^{-1}(R_2)$ has positive measure. Therefore, since the set is invariant, it must have measure 1. Continue in this way obtaining a sequence of closed rectangles, $\{R_i\}$ such that the diameter of $R_i$ converges to zero and $g^{-1}(R_i)$ has measure 1. Then let $c = \cap_{i=1}^{\infty} R_i$. We know $\mu(g^{-1}(c)) = \lim_{n \to \infty} \mu(g^{-1}(R_i)) = 1$. It follows that $g(\omega) = c$ for a.e. $\omega$. Now from Theorem 1.9.7,

$$c = \int c \, d\mu = \lim_{n \to \infty} \frac{1}{n} \int S_n f \, d\mu = \int f \, d\mu.$$ ■
Chapter 2

Some Extension Theorems

2.1 Measures From Outer Measures

Definition 2.1.1 Let $\Omega$ be a nonempty set. A function mapping $\mathcal{P}(\Omega) \to [0, \infty]$ is called an outer measure if it satisfies: If $A \subseteq B$, then $0 \leq \mu(A) \leq \mu(B)$, $\mu(\emptyset) = 0$. Also $\mu(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$.

There is a general approach to obtaining measures from outer measures which is due to Caratheodory in about 1918.

Definition 2.1.2 Let $\Omega$ be a nonempty set and let $\mu : \mathcal{P}(\Omega) \to [0, \infty]$ be an outer measure. For $E \subseteq \Omega$, $E$ is $\mu$ measurable if for all $S \subseteq \Omega$,

$$\mu(S) = \mu(S \setminus E) + \mu(S \cap E).$$

To help in remembering 2.1.1, think of a measurable set $E$, as a process which divides a given set into two pieces, the part in $E$ and the part not in $E$ as in 2.1.1. In the Bible, there are several incidents recorded in which a process of division resulted in more stuff than was originally present. Measurable sets are exactly those which are incapable of such a miracle. You might think of the measurable sets as the nonmiraculous sets. The idea is to show that they form a $\sigma$ algebra on which the outer measure $\mu$ is a measure.

First here is a definition and a lemma.

Definition 2.1.3 $(\mu|_S)(A) \equiv \mu(S \cap A)$ for all $A \subseteq \Omega$. Thus $\mu|_S$ is the name of a new outer measure, called $\mu$ restricted to $S$.

The next lemma indicates that the property of measurability is not lost by considering this restricted measure.

Lemma 2.1.4 If $A$ is $\mu$ measurable, then $A$ is $\mu|_S$ measurable.

Proof: Suppose $A$ is $\mu$ measurable. It is desired to show that for all $T \subseteq \Omega$,

$$(\mu|_S)(T) = (\mu|_S)(T \cap A) + (\mu|_S)(T \setminus A).$$

Thus it is desired to show

$$\mu(S \cap T) = \mu(T \cap A \cap S) + \mu(T \cap S \cap A^C).$$

But 2.1.2 holds because $A$ is $\mu$ measurable. Apply Definition 2.1.1 to $S \cap T$ instead of $S$.

If $A$ is $\mu|_S$ measurable, it does not follow that $A$ is $\mu$ measurable. Indeed, if you believe in the existence of nonmeasurable sets, you could let $A = S$ for such a $\mu$ nonmeasurable set and verify that $S$ is $\mu|_S$ measurable. In fact there do exist nonmeasurable sets but this is another topic.

The next theorem is the main result on outer measures which shows that starting with an outer measure you can obtain a measure.

---

1 Kings 17, 2 Kings 4, Mathew 14, and Mathew 15 all contain such descriptions. The stuff involved was either oil, bread, flour or fish. In mathematics such things have also been done with sets. In the book by Bruckner Bruckner and Thompson there is an interesting discussion of the Banach Tarski paradox which says it is possible to divide a ball in $\mathbb{R}^3$ into five disjoint pieces and assemble the pieces to form two disjoint balls of the same size as the first. The details can be found in: The Banach Tarski Paradox by Wagon, Cambridge University press. 1985. It is known that all such examples must involve the axiom of choice.
**Theorem 2.1.5** Let \( \Omega \) be a set and let \( \mu \) be an outer measure on \( \mathcal{P}(\Omega) \). The collection of \( \mu \)-measurable sets \( \mathcal{S} \), forms a \( \sigma \)-algebra and

\[
\text{If } F_i \in \mathcal{S}, \ F_i \cap F_j = \emptyset, \text{ then } \mu(\bigcup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} \mu(F_i). \tag{2.1.3}
\]

If \( \cdots F_n \subseteq F_{n+1} \subseteq \cdots \), then if \( F = \bigcup_{n=1}^{\infty} F_n \) and \( F_n \in \mathcal{S} \), it follows that

\[
\mu(F) = \lim_{n \to \infty} \mu(F_n). \tag{2.1.4}
\]

If \( \cdots F_n \supseteq F_{n+1} \supseteq \cdots \), and if \( F = \bigcap_{n=1}^{\infty} F_n \) for \( F_n \in \mathcal{S} \) then if \( \mu(F_1) < \infty \),

\[
\mu(F) = \lim_{n \to \infty} \mu(F_n). \tag{2.1.5}
\]

This measure space is also complete which means that if \( \mu(F) = 0 \) for some \( F \in \mathcal{S} \) then if \( G \subseteq F \), it follows \( G \in \mathcal{S} \) also.

**Proof:** First note that \( \emptyset \) and \( \Omega \) are obviously in \( \mathcal{S} \). Now suppose \( A, B \in \mathcal{S} \). I will show \( A \setminus B \equiv A \cap B^C \) is in \( \mathcal{S} \). To do so, consider the following picture.

![Diagram](image_url)

First note that

\[
S \setminus (A \setminus B) = (S \setminus A) \cup (S \cap B) = (S \setminus A) \cup (A \cap B \cap S).
\]

Since \( \mu \) is subadditive and \( A, B \in \mathcal{S} \),

\[
\mu(S) \leq \mu(S \setminus (A \setminus B)) + \mu(S \cap (A \setminus B))
\]

\[
\leq \mu(S \setminus A) + \mu(S \cap B \setminus A) + \mu(S \cap A \setminus B^C)
\]

\[
= \mu(S \setminus A) + \mu(S \cap A) = \mu(S)
\]

and so all the inequalities are equal signs. Therefore, since \( S \) is arbitrary, this shows \( A \setminus B \in \mathcal{S} \).

Since \( \Omega \in \mathcal{S} \), this shows that \( A \in \mathcal{S} \) if and only if \( A^C \in \mathcal{S} \). Now if \( A, B \in \mathcal{S} \), \( A \cup B = (A^C \cap B^C)^C = (A^C \setminus B)^C \in \mathcal{S} \).

By induction, if \( A_1, \cdots, A_n \in \mathcal{S} \), then so is \( \bigcup_{i=1}^{n} A_i \). If \( A, B \in \mathcal{S} \) with \( A \cap B = \emptyset \),

\[
\mu(A \cup B) = \mu((A \cup B) \cap A) + \mu((A \cup B) \setminus A) = \mu(A) + \mu(B).
\]

By induction, if \( A_i \cap A_j = \emptyset \) and \( A_i \in \mathcal{S} \),

\[
\mu(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i). \tag{2.1.6}
\]

Now let \( A = \bigcup_{i=1}^{\infty} A_i \), where \( A_i \cap A_j = \emptyset \) for \( i \neq j \).

\[
\sum_{i=1}^{\infty} \mu(A_i) \geq \mu(A) \geq \mu(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i).
\]

Since this holds for all \( n \), you can take the limit as \( n \to \infty \) and conclude,

\[
\sum_{i=1}^{\infty} \mu(A_i) = \mu(A)
\]
which establishes \( \mu \). Without loss of generality \( \mu (F_k) < \infty \) for all \( k \) since otherwise there is nothing to show. Suppose \( \{ F_k \} \) is an increasing sequence of sets of \( S \). Then letting \( F_0 = 0, \{ F_{k+1} \setminus F_k \}_{k=0}^{\infty} \) is a sequence of disjoint sets of \( S \) since it was shown above that the difference of two sets of \( S \) is in \( S \). Also note that from (2.1.0)

\[
\mu (F_{k+1} \setminus F_k) + \mu (F_k) = \mu (F_{k+1})
\]

and so if \( \mu (F_k) < \infty \), then

\[
\mu (F_{k+1} \setminus F_k) = \mu (F_{k+1}) - \mu (F_k).
\]

Therefore, letting

\[
F = \bigcup_{k=1}^{\infty} F_k
\]

which also equals

\[
\bigcup_{k=1}^{\infty} (F_{k+1} \setminus F_k),
\]

it follows from part (2.1.6) just shown that

\[
\mu (F) = \sum_{k=1}^{\infty} \mu (F_{k+1} \setminus F_k) = \lim_{n \to \infty} \sum_{k=1}^{n} \mu (F_{k+1} \setminus F_k)
\]

\[
= \lim_{n \to \infty} \sum_{k=1}^{n} \mu (F_{k+1}) - \mu (F_k) = \lim_{n \to \infty} \mu (F_{n+1}).
\]

In order to establish (2.1.7) let the \( F_n \) be as given there. Then, since \( (F_1 \setminus F_n) \) increases to \( (F_1 \setminus F) \), (2.1.3) implies

\[
\lim_{n \to \infty} (\mu (F_1) - \mu (F_n)) = \mu (F_1 \setminus F).
\]

Now \( \mu (F_1 \setminus F) + \mu (F) \geq \mu (F_1) \) and so \( \mu (F_1 \setminus F) \geq \mu (F_1) - \mu (F) \). Hence

\[
\lim_{n \to \infty} (\mu (F_1) - \mu (F_n)) = \mu (F_1 \setminus F) \geq \mu (F_1) - \mu (F)
\]

which implies

\[
\lim_{n \to \infty} \mu (F_n) \leq \mu (F).
\]

But since \( F \subseteq F_n \),

\[
\mu (F) \leq \lim_{n \to \infty} \mu (F_n)
\]

and this establishes (2.1.7). Note that it was assumed \( \mu (F_1) < \infty \) because \( \mu (F_1) \) was subtracted from both sides.

It remains to show \( S \) is closed under countable unions. Recall that if \( A \in S \), then \( A^c \in S \) and \( S \) is closed under finite unions. Let \( A_i \in S, A = \bigcup_{i=1}^{\infty} A_i, B_n = \bigcup_{i=1}^{n} A_i. \) Then

\[
\mu (S) = \mu (S \cap B_n) + \mu (S \setminus B_n)
\]

\[
= (\mu (S) (B_n) + (\mu (S) (B_n^c)).
\]

By Lemma (2.1.8) \( B_n \) is \( (\mu (S) \) measurable and so is \( B_n^c \). I want to show \( \mu (S) \geq \mu (S \setminus A) + \mu (S \cap A) \). If \( \mu (S) = \infty \), there is nothing to prove. Assume \( \mu (S) < \infty \). Then apply Parts (2.1.7) and (2.1.4) to the outer measure \( \mu (S) \) in (2.1.7) and let \( n \to \infty \). Thus

\[
B_n \uparrow A, B_n^c \downarrow A^c
\]

and this yields \( \mu (S) = (\mu (S) (A) + (\mu (S) (A^c) = \mu (S \cap A) + \mu (S \setminus A) \).

Therefore \( A \in S \) and this proves Parts (2.1.8), (2.1.4), and (2.1.6).

It only remains to verify the assertion about completeness. Letting \( G \) and \( F \) be as described above, let \( S \subseteq \Omega \). I need to verify

\[
\mu (S) \geq \mu (S \cap G) + \mu (S \setminus G)
\]

However,

\[
\mu (S \cap G) + \mu (S \setminus G) \leq \mu (S \cap F) + \mu (S \setminus F) + \mu (F \setminus G)
\]

\[
= \mu (S \cap F) + \mu (S \setminus F) = \mu (S)
\]

because by assumption, \( \mu (F \setminus G) \leq \mu (F) = 0. \)
Corollary 2.1.6 Completeness is the same as saying that if \((E \setminus E') \cup (E' \setminus E) \subseteq N \in \mathcal{F}\) and \(\mu(N) = 0\), then if \(E \in \mathcal{F}\), it follows that \(E' \in \mathcal{F}\) also.

Proof: If the new condition holds, then suppose \(G \subseteq F\) where \(\mu(F) = 0\), \(F \in \mathcal{F}\). Then \((G \setminus F) \cup (F \setminus G) \subseteq F\) and \(\mu(F)\) is given to equal 0. Therefore, \(G \in \mathcal{F}\).

Now suppose the earlier version of completeness and let
\[
(E \setminus E') \cup (E' \setminus E) \subseteq N \in \mathcal{F}
\]
where \(\mu(N) = 0\) and \(E \in \mathcal{F}\). Then we know
\[
(E \setminus E'), (E' \setminus E) \in \mathcal{F}
\]
and all have measure zero. It follows \(E \setminus (E \setminus E') = E \cap E' \in \mathcal{F}\). Hence
\[
E' = (E \cap E') \cup (E' \setminus E) \in \mathcal{F}
\]

2.2 Algebras

First of all, here is the definition of an algebra and theorems which tell how to recognize one when you see it. An algebra is like a \(\sigma\) algebra except it is only closed with respect to finite unions.

Definition 2.2.1 \(A\) is said to be an algebra of subsets of a set, \(Z\) if \(Z \in A\), \(\emptyset \in A\), and when \(E,F \in A\), \(E \cup F\) and \(E \setminus F\) are both in \(A\).

It is important to note that if \(A\) is an algebra, then it is also closed under finite intersections. This is because \(E \cap F = (E^C \cup F^C)^C \in A\) since \(E^C = Z \setminus E \in A\) and \(F^C = Z \setminus F \in A\). Note that every \(\sigma\) algebra is an algebra but not the other way around.

Something satisfying the above definition is called an algebra because union is like addition, the set difference is like subtraction and intersection is like multiplication. Furthermore, only finitely many operations are done at a time and so there is nothing like a limit involved.

How can you recognize an algebra when you see one? The answer to this question is the purpose of the following lemma.

Lemma 2.2.2 Suppose \(\mathcal{R}\) and \(\mathcal{E}\) are subsets of \(\mathcal{P}(Z)\) such that \(\mathcal{E}\) is defined as the set of all finite disjoint unions of sets of \(\mathcal{R}\). Suppose also
\[
\emptyset, Z \in \mathcal{R}
\]
\[
A \cap B \in \mathcal{R} \text{ whenever } A, B \in \mathcal{R},
\]
\[
A \setminus B \in \mathcal{E} \text{ whenever } A, B \in \mathcal{R}.
\]

Then \(\mathcal{E}\) is an algebra of sets of \(Z\).

Proof: Note first that if \(A \in \mathcal{R}\), then \(A^C \in \mathcal{E}\) because \(A^C = Z \setminus A\).

Now suppose that \(E_1\) and \(E_2\) are in \(\mathcal{E}\),
\[
E_1 = \bigcup_{i=1}^{m} R_i, \quad E_2 = \bigcup_{j=1}^{n} R_j
\]
where the \(R_i\) are disjoint sets in \(\mathcal{R}\) and the \(R_j\) are disjoint sets in \(\mathcal{R}\). Then
\[
E_1 \cap E_2 = \bigcup_{i=1}^{m} \bigcup_{j=1}^{n} R_i \cap R_j
\]
which is clearly an element of \(\mathcal{E}\) because no two of the sets in the union can intersect and by assumption they are all in \(\mathcal{R}\). Thus by induction, finite intersections of sets of \(\mathcal{E}\) are in \(\mathcal{E}\). Consider the difference of two elements of \(\mathcal{E}\) next.

If \(E = \bigcup_{i=1}^{m} R_i \in \mathcal{E}\),
\[
E^C = \bigcap_{i=1}^{m} R_i^C = \text{ finite intersection of sets of } \mathcal{E}
\]
which was just shown to be in \( \mathcal{E} \). Now, if \( E_1, E_2 \in \mathcal{E} \),
\[
E_1 \setminus E_2 = E_1 \cap E_2^C \in \mathcal{E}
\]
from what was just shown about finite intersections.

Finally consider finite unions of sets of \( \mathcal{E} \). Let \( E_1 \) and \( E_2 \) be sets of \( \mathcal{E} \). Then
\[
E_1 \cup E_2 = (E_1 \setminus E_2) \cup E_2 \in \mathcal{E}
\]
because \( E_1 \setminus E_2 \) consists of a finite disjoint union of sets of \( \mathcal{R} \) and these sets must be disjoint from the sets of \( \mathcal{R} \) whose union yields \( E_2 \) because \( (E_1 \setminus E_2) \cap E_2 = \emptyset \). This proves the lemma.

The following corollary is particularly helpful in verifying the conditions of the above lemma.

**Corollary 2.2.3** Let \((Z_1, \mathcal{R}_1, \mathcal{E}_1)\) and \((Z_2, \mathcal{R}_2, \mathcal{E}_2)\) be as described in Lemma 2.2.2. Then \((Z_1 \times Z_2, \mathcal{R}, \mathcal{E})\) also satisfies the conditions of Lemma 2.2.2 if \( \mathcal{R} \) is defined as
\[
\mathcal{R} \equiv \{ R_1 \times R_2 : R_i \in \mathcal{R}_i \}
\]
and
\[
\mathcal{E} \equiv \{ \text{finite disjoint unions of sets of } \mathcal{R} \}.
\]
Consequently, \( \mathcal{E} \) is an algebra of sets.

**Proof:** It is clear \( \emptyset, Z_1 \times Z_2 \in \mathcal{R} \). Let \( A \times B \) and \( C \times D \) be two elements of \( \mathcal{R} \).
\[
A \times B \cap C \times D = A \cap C \times B \cap D \in \mathcal{R}
\]
by assumption.
\[
A \times B \setminus (C \times D) =
\]
\[
\underbrace{A \times (B \setminus D)}_{\in \mathcal{E}_2} \cup \underbrace{(A \setminus C) \times (D \cap B)}_{\in \mathcal{E}_1} \cup \underbrace{(A \times Q) \cup (P \times R)}_{\in \mathcal{R}_2}
\]
where \( Q \in \mathcal{E}_2, P \in \mathcal{E}_1, \) and \( R \in \mathcal{R}_2 \).

Since \( A \times Q \) and \( P \times R \) do not intersect, it follows the above expression is in \( \mathcal{E} \) because each of these terms are. This proves the corollary.

### 2.3 Caratheodory Extension Theorem

The Carathéodory extension theorem is a fundamental result which makes possible the consideration of measures on infinite products among other things. The idea is that if a finite measure defined only on an algebra is trying to be a measure, then in fact it can be extended to a measure.

**Definition 2.3.1** Let \( \mathcal{E} \) be an algebra of sets of \( \Omega \) and let \( \mu_0 \) be a finite measure on \( \mathcal{E} \). This means \( \mu_0 \) is finitely additive and if \( E_1, E \) are sets of \( \mathcal{E} \) with the \( E_1 \) disjoint and
\[
E = \bigcup_{i=1}^{\infty} E_i,
\]
then
\[
\mu_0(E) = \sum_{i=1}^{\infty} \mu_0(E_i)
\]
while \( \mu_0(\Omega) < \infty \).
In this definition, \( \mu_0 \) is trying to be a measure and acts like one whenever possible. Under these conditions, \( \mu_0 \) can be extended uniquely to a complete measure, \( \mu \), defined on a \( \sigma \) algebra of sets containing \( \mathcal{E} \) such that \( \mu \) agrees with \( \mu_0 \) on \( \mathcal{E} \). The following is the main result.

**Theorem 2.3.2** Let \( \mu_0 \) be a measure on an algebra of sets, \( \mathcal{E} \), which satisfies \( \mu_0(\Omega) < \infty \). Then there exists a complete measure space \( (\Omega, \mathcal{S}, \mu) \) such that

\[
\mu(E) = \mu_0(E)
\]

for all \( E \in \mathcal{E} \). Also if \( \nu \) is any such measure which agrees with \( \mu_0 \) on \( \mathcal{E} \), then \( \nu = \mu \) on \( \sigma(\mathcal{E}) \), the \( \sigma \) algebra generated by \( \mathcal{E} \).

**Proof:** Define an outer measure as follows.

\[
\mu(S) \equiv \inf \left\{ \sum_{i=1}^{\infty} \mu_0(E_i) : S \subseteq \bigcup_{i=1}^{\infty} E_i, E_i \in \mathcal{E} \right\}
\]

Claim 1: \( \mu \) is an outer measure.

**Proof of Claim 1:** Let \( S \subseteq \bigcup_{i=1}^{\infty} S_i \) and let \( S_i \subseteq \bigcup_{j=1}^{\infty} E_{ij} \), where

\[
\mu(S_i) + \frac{\varepsilon}{2^i} \geq \sum_{j=1}^{\infty} \mu(E_{ij}).
\]

Then

\[
\mu(S) \leq \sum_i \sum_j \mu(E_{ij}) = \sum_i \left( \mu(S_i) + \frac{\varepsilon}{2^i} \right) = \sum_i \mu(S_i) + \varepsilon.
\]

Since \( \varepsilon \) is arbitrary, this shows \( \mu \) is an outer measure as claimed.

By the Caratheodory procedure, there exists a unique \( \sigma \) algebra, \( \mathcal{S} \), consisting of the \( \mu \) measurable sets such that

\[
(\Omega, \mathcal{S}, \mu)
\]

is a complete measure space. It remains to show \( \mu \) extends \( \mu_0 \).

Claim 2: If \( \mathcal{S} \) is the \( \sigma \) algebra of \( \mu \) measurable sets, \( \mathcal{S} \supseteq \mathcal{E} \) and \( \mu = \mu_0 \) on \( \mathcal{E} \).

**Proof of Claim 2:** First observe that if \( A \in \mathcal{E} \), then \( \mu(A) \leq \mu_0(A) \) by definition. Letting

\[
\mu(A) + \varepsilon > \sum_{i=1}^{\infty} \mu_0(E_i), \quad \bigcup_{i=1}^{\infty} E_i \supseteq A, \quad E_i \in \mathcal{E},
\]

it follows

\[
\mu(A) + \varepsilon > \sum_{i=1}^{\infty} \mu_0(E_i \cap A) \geq \mu_0(A)
\]

since \( A = \bigcup_{i=1}^{\infty} E_i \cap A \). Therefore, \( \mu = \mu_0 \) on \( \mathcal{E} \).

Consider the assertion that \( \mathcal{E} \subseteq \mathcal{S} \). Let \( A \in \mathcal{E} \) and let \( S \subseteq \Omega \) be any set. There exist sets \( \{E_i\} \subseteq \mathcal{E} \) such that \( \bigcup_{i=1}^{\infty} E_i \supseteq S \) but

\[
\mu(S) + \varepsilon > \sum_{i=1}^{\infty} \mu(E_i).
\]

Then

\[
\mu(S) \leq \mu(S \cap A) + \mu(S \setminus A) \\
\leq \mu(\bigcup_{i=1}^{\infty} E_i \cap A) + \mu(\bigcup_{i=1}^{\infty} (E_i \setminus A))
\]

\[
\leq \sum_{i=1}^{\infty} \mu(E_i \cap A) + \sum_{i=1}^{\infty} \mu(E_i \setminus A) = \sum_{i=1}^{\infty} \mu(E_i) < \mu(S) + \varepsilon.
\]

Since \( \varepsilon \) is arbitrary, this shows \( A \in \mathcal{S} \).

This has proved the existence part of the theorem. To verify uniqueness, Let

\[
\mathcal{G} \equiv \{E \in \sigma(\mathcal{E}) : \mu(E) = \nu(E)\}.
\]

Then \( \mathcal{G} \) is given to contain \( \mathcal{E} \) and is obviously closed with respect to countable disjoint unions and complements. Therefore by Lemma \( \Box \), \( \mathcal{G} \supseteq \sigma(\mathcal{E}) \) and this proves the lemma.

The following lemma is also very significant.
Lemma 2.3.3 Let $M$ be a metric space with the closed balls compact and suppose $\mu$ is a measure defined on the Borel sets of $M$ which is finite on compact sets. Then there exists a unique Radon measure, $\overline{\mu}$ which equals $\mu$ on the Borel sets. In particular $\mu$ must be both inner and outer regular on all Borel sets.

Proof: Define a positive linear functional, $\Lambda (f) = \int fd\mu$. Let $\overline{\mu}$ be the Radon measure which comes from the Riesz representation theorem for positive linear functionals. Thus for all $f \in C_0 (M)$,

$$\int fd\mu = \int fd\overline{\mu}.$$ 

If $V$ is an open set, let $\{ f_n \}$ be a sequence of continuous functions in $C_0 (M)$ which is increasing and converges to $X_V$ pointwise. Then applying the monotone convergence theorem,

$$\int X_V d\mu = \mu (V) = \int X_V d\overline{\mu} = \overline{\mu} (V)$$

and so the two measures coincide on all open sets. Every compact set is a countable intersection of open sets and so the two measures coincide on all compact sets. Now let $B (a, n)$ be a ball of radius $n$ and let $E$ be a Borel set contained in this ball. Then by regularity of $\overline{\mu}$ there exist sets $F, G$ such that $G$ is a countable intersection of open sets and $F$ is a countable union of compact sets such that $F \subseteq E \subseteq G$ and $\overline{\mu} (G \setminus F) = 0$. Now $\mu (G) = \overline{\mu} (G)$ and $\mu (F) = \overline{\mu} (F)$. Thus

$$\overline{\mu} (G \setminus F) + \mu (F) = \overline{\mu} (G) = \mu (G) = \mu (G \setminus F) + \mu (F)$$

and so $\mu (G \setminus F) = \overline{\mu} (G \setminus F)$. It follows

$$\mu (E) = \mu (F) = \overline{\mu} (F) = \overline{\mu} (G) = \overline{\mu} (E).$$

If $E$ is an arbitrary Borel set, then

$$\mu (E \cap B (a, n)) = \overline{\mu} (E \cap B (a, n))$$

and letting $n \to \infty$, this yields $\mu (E) = \overline{\mu} (E)$.

2.4 The Tychonoff Theorem

Sometimes it is necessary to consider infinite Cartesian products of topological spaces. When you have finitely many topological spaces in the product and each is compact, it can be shown that the Cartesian product is compact with the product topology. It turns out that the same thing holds for infinite products but you have to be careful how you define the topology. The first thing likely to come to mind by analogy with finite products is not the right way to do it.

First recall the Hausdorff maximal principle.

Theorem 2.4.1 (Hausdorff maximal principle) Let $F$ be a nonempty partially ordered set. Then there exists a maximal chain.

The main tool in the study of products of compact topological spaces is the Alexander subbasis theorem which is presented next. Recall a set is compact if every basic open cover admits a finite subcover. This was pretty easy to prove. However, there is a much smaller set of open sets called a subbasis which has this property. The proof of this result is much harder.

Definition 2.4.2 $S \subseteq \tau$ is called a subbasis for the topology $\tau$ if the set $B$ of finite intersections of sets of $S$ is a basis for the topology, $\tau$.

Theorem 2.4.3 Let $(X, \tau)$ be a topological space and let $S \subseteq \tau$ be a subbasis for $\tau$. Then if $H \subseteq X$, $H$ is compact if and only if every open cover of $H$ consisting entirely of sets of $S$ admits a finite subcover.

Proof: The only if part is obvious because the subbasic sets are themselves open.

If every basic open cover admits a finite subcover then the set in question is compact. Suppose then that $H$ is a subset of $X$ having the property that subbasic open covers admit finite subcovers. Is $H$ compact? Assume this
is not so. Then what was just observed about basic covers implies there exists a basic open cover of \( H, \mathcal{O} \) which admits no finite subcover. Let \( \mathcal{F} \) be defined as

\[
\{ \mathcal{O} : \mathcal{O} \text{ is a basic open cover of } H \text{ which admits no finite subcover} \}.
\]

The assumption is that \( \mathcal{F} \) is nonempty. Partially order \( \mathcal{F} \) by set inclusion and use the Hausdorff maximal principle to obtain a maximal chain, \( \mathcal{C} \), of such open covers and let

\[
\mathcal{D} = \cup \mathcal{C}.
\]

If \( \mathcal{D} \) admits a finite subcover, then since \( \mathcal{C} \) is a chain and the finite subcover has only finitely many sets, some element of \( \mathcal{C} \) would also admit a finite subcover, contrary to the definition of \( \mathcal{F} \). Therefore, \( \mathcal{D} \) admits no finite subcover. If \( \mathcal{D}' \) properly contains \( \mathcal{D} \) and \( \mathcal{D}' \) is a basic open cover of \( H \), then \( \mathcal{D}' \) has a finite subcover of \( H \) since otherwise, \( \mathcal{C} \) would fail to be a maximal chain, being properly contained in \( \mathcal{C} \cup \{ \mathcal{D}' \} \). Every set of \( \mathcal{D} \) is of the form

\[
U = \bigcap_{i=1}^{m} B_i, \quad B_i \in \mathcal{S}
\]

because they are all basic open sets. If it is the case that for all \( U \in \mathcal{D} \) one of the \( B_i \) is found in \( \mathcal{D} \), then replace each such \( U \) with the subbasic set from \( \mathcal{D} \) containing it. But then this would be a subbasic open cover of \( H \) which by assumption would admit a finite subcover contrary to the properties of \( \mathcal{D} \). Therefore, one of the sets of \( \mathcal{D} \), denoted by \( U \), has the property that

\[
U = \bigcap_{i=1}^{m} B_i, \quad B_i \in \mathcal{S}
\]

and no \( B_i \) is in \( \mathcal{D} \). Thus \( \mathcal{D} \cup \{ B_i \} \) admits a finite subcover, for each of the above \( B_i \) because it is strictly larger than \( \mathcal{D} \). Let this finite subcover corresponding to \( B_i \) be denoted by

\[
V_{i}^{1}, \cdots, V_{i}^{m}, B_i
\]

Consider

\[
\{U, V_{j}^{j}, \quad j = 1, \cdots, m, \quad i = 1, \cdots, m\}.
\]

If \( p \in H \setminus \cup \{V_{j}^{j}\} \), then \( p \in B_i \) for each \( i \) and so \( p \in U \). This is therefore a finite subcover of \( \mathcal{D} \) contradicting the properties of \( \mathcal{D} \). Therefore, \( \mathcal{F} \) must be empty and this proves the theorem.

**Definition 2.4.4** Let \( I \) be a set and suppose for each \( i \in I \), \( (X_i, \tau_i) \) is a nonempty topological space. The Cartesian product of the \( X_i \), denoted by \( \prod_{i \in I} X_i \), consists of the set of all choice functions defined on \( I \) which select a single element of each \( X_i \). Thus \( f \in \prod_{i \in I} X_i \) means for every \( i \in I \), \( f(i) \in X_i \). The axiom of choice says \( \prod_{i \in I} X_i \) is nonempty. Let

\[
P_j(A) \equiv \prod_{i \in I} B_i
\]

where \( B_i \equiv X_i \) if \( i \neq j \) and \( B_j = A \). A subbasis for a topology on the product space consists of all sets \( P_j(A) \) where \( A \in \tau_j \). (These sets have an open set from the topology of \( X_j \) in the \( j \)th slot and the whole space in the other slots.) Thus a basis consists of finite intersections of these sets. Note that the intersection of two of these basic sets is another basic set and their union yields \( \prod_{i \in I} X_i \). Therefore, they satisfy the condition needed for a collection of sets to serve as a basis for a topology. This topology is called the product topology and is denoted by \( \prod \tau_i \).

**Proposition 2.4.5** The product topology is the smallest topology \( \tau \) for \( X \equiv \prod_{i \in I} X_i \) such that each \( \pi_i \) is continuous. Here \( \pi_i \) is defined in the following manner. For \( \mathbf{x} \in X \), \( \pi_i(\mathbf{x}) \equiv x_i \). Thus \( \pi_i \) delivers the \( i \)th entry of \( \mathbf{x} \).

**Proof:** If each \( \pi_i \) is continuous, then for \( A \in \tau_i, \pi_i^{-1}(A) \) must be in \( \tau \). However, \( \pi_i^{-1}(A) = P_j(A) \) having \( A \) in the \( i \)th slot and \( X_j \) in every other. Therefore, \( \tau \) must contain the sets \( P_j(A) \). Since it must be a topology, it must also contain all finite intersections of these sets. Thus the topology \( \tau \) must contain the product topology described in the above definition. Is it any larger? No, because if it were, it would not be the smallest topology making the coordinate maps continuous, due to the observation that these coordinate maps are indeed continuous with respect to the product topology. ■

It is tempting to define a basis for a topology to be sets of the form \( \prod_{i \in I} A_i \) where \( A_i \) is open in \( X_i \). This is not the same thing at all. Note that the basis just described has at most finitely many slots filled with an open set which is not the whole space. The thing just mentioned in which every slot may be filled by a proper open set is called the box topology and there exist people who are interested in it.

The Alexander subbasis theorem is used to prove the Tychonoff theorem which says that if each \( X_i \) is a compact topological space, then in the product topology, \( \prod_{i \in I} X_i \) is also compact.
2.5. KOLMOGOROV EXTENSION THEOREM FOR POLISH SPACES

Theorem 2.4.6 If \((X_i, \tau_i)\) is compact, then so is \((\prod_{i \in I} X_i, \tau)\) where \(\tau\) is the product topology.

Proof: By the Alexander subbasis theorem, the theorem will be proved if every subbasic open cover admits a finite subcover. Therefore, let \(\mathcal{O}\) be a subbasic open cover of \(X = \prod_{i \in I} X_i\). Let
\[
\mathcal{O}_j = \{Q \in \mathcal{O} : \tau_i Q = X_i \text{ for } i \neq j\}
\]
\[
\pi_j \mathcal{O}_j = \{\pi_j Q : Q \in \mathcal{O}_j\}
\]
Thus \(\mathcal{O}_j\) are those sets of \(\mathcal{O}\) which might have a proper open subset of \(X_j\) in the \(j^{th}\) position. If each \(\pi_j \mathcal{O}_j\) fails to cover \(X_j\) then there exists
\[
f \in \prod_{j \in I} X_j \setminus \cup \pi_j \mathcal{O}_j
\]
Now \(f\) is contained in some open set from \(\mathcal{O}\) which must be in some \(\mathcal{O}_j\). Hence \(\pi_j f = f(j) \in \cup \pi_j \mathcal{O}_j\) but this does not happen. Hence for some \(j, \pi_j \mathcal{O}_j\) must cover \(X_j\).
\[
X_j = \cup \pi_j \mathcal{O}_j
\]
and so by compactness of \(X_j\), there exist \(A_1, \cdots, A_m\) sets in \(\tau_j\) such that \(X_j \subseteq \bigcup_{k=1}^m A_k\) and letting \(\pi_j U_k = A_k\) for \(U_k \in \mathcal{O}_j\), \(\{U_k\}_{k=1}^m\) covers \(\prod_{i \in I} X_i\). By the Alexander subbasis theorem this proves \(\prod_{i \in I} X_i\) is compact. ■

2.5 Kolmogorov Extension Theorem For Polish Spaces

Let \(M_t\) be a complete separable metric space. This is called a Polish space. \(I\) will denote a totally ordered index set, (Like \(\mathbb{R}\)) and the interest will be in building a measure on the product space, \(\prod_{t \in I} M_t\). By the well ordering principle, you can always put an order on any index set so this order is no restriction, but we do not insist on a well order and in fact, index sets of great interest are \(\mathbb{R}\) or \([0, \infty)\). Also for \(X\) a topological space, \(B(X, X)\) will denote the Borel sets.

Notation 2.5.1 The symbol \(J\) will denote a finite subset of \(I, J = (t_1, \cdots, t_n)\), the \(i^{th}\) taken in order. \(E_J\) will denote a set which has a set \(E_t\) of \(B(M_t)\) in the \(t^{th}\) position for \(t \in J\) and for \(t \notin J\), the set in the \(t^{th}\) position will be \(M_t\). \(K_J\) will denote a set which has a compact set in the \(t^{th}\) position for \(t \in J\) and for \(t \notin J\), the set in the \(t^{th}\) position will be \(M_t\). Also denote by \(R_J\) the sets \(E_J\) and \(R\) the union of all such \(R_J\). Let \(E_J\) denote finite disjoint unions of sets of \(R_J\) and \(E\) denote finite disjoint unions of sets of \(R\). Thus if \(F \in \Omega\), there exists \(J\) such that \(F\) is a finite disjoint union of sets of \(R_J\). For \(F \in \Omega\), denote by \(\pi_J(F)\) the set \(\prod_{t \in I} F_t\) where \(F = \prod_{t \in I} F_t\).

Lemma 2.5.2 The sets, \(E, E_J\) defined above form an algebra of sets of \(\prod_{t \in I} M_t\).

Proof: First consider \(R_J\). If \(A, B \in R_J\), then \(A \cap B \in R_J\) also. Is \(A \setminus B\) a finite disjoint union of sets of \(R_J\)? It suffices to verify that \(\pi_J(A \setminus B)\) is a finite disjoint union of \(\pi_J(R_J)\). Let \(|J|\) denote the number of indices in \(J\). If \(|J| = 1\), then it is obvious that \(\pi_J(A \setminus B)\) is a finite disjoint union of sets of \(\pi_J(R_J)\). In fact, letting \(J = (t)\) and the \(t^{th}\) entry of \(A\) is \(A_t\) and the \(t^{th}\) entry of \(B\) is \(B_t\), then the \(t^{th}\) entry of \(A \setminus B\) is \(A_t \setminus B_t\), a Borel set of \(M_t\), a finite disjoint union of Borel sets of \(M_t\).

Suppose then that for \(A, B\) sets of \(R_J\), \(\pi_J(A \setminus B)\) is a finite disjoint union of sets of \(\pi_J(R_J)\) for \(|J| \leq n\), and consider \(J = (t_1, \cdots, t_n, t_{n+1})\). Let the \(t_{n+1}^{th}\) entry of \(A\) and \(B\) be respectively \(A_t\) and \(B_t\). It follows that \(\pi_J(A \setminus B)\) has the following in the entries for \(J\)
\[
(A_1 \times A_2 \times \cdots \times A_n \times A_{n+1}) \setminus (B_1 \times B_2 \times \cdots \times B_n \times B_{n+1})
\]
Letting \(A\) represent \(A_1 \times A_2 \times \cdots \times A_n\) and \(B\) represent \(B_1 \times B_2 \times \cdots \times B_n\), this is of the form
\[
A \setminus (A_{n+1} \setminus B_{n+1}) \cup (A \setminus B) \times (A_{n+1} \cap B_{n+1})
\]
By induction, \((A \setminus B)\) is the finite disjoint union of sets of \(R(t_1, \cdots, t_n)\). Therefore, the above is the finite disjoint union of sets of \(R_J\). It follows that \(E_J\) is an algebra.

Now suppose \(A, B \in R\). Then for some finite set \(J\), both are in \(R_J\). Then from what was just shown,
\[
A \setminus B \in E_J \subseteq E, A \cap B \in R.
\]
By Lemma [this page] on Page [this page] this shows \(E\) is an algebra. ■

With this preparation, here is the Kolmogorov extension theorem. In the statement and proof of the theorem, \(F_t, G_t, E_t\) will denote Borel sets. Any list of indices from \(I\) will always be assumed to be taken in order. Thus, if \(J \subseteq I\) and \(J = (t_1, \cdots, t_n)\), it will always be assumed \(t_1 < t_2 < \cdots < t_n\).
Theorem 2.5.3  For each finite set
\[ J = (t_1, \cdots, t_n) \subseteq I, \]
suppose there exists a Borel probability measure, \( \nu_J = \nu_{t_1, \cdots, t_n} \) defined on the Borel sets of \( \prod_{t \in J} M_t \) such that the following consistency condition holds. If
\[ (t_1, \cdots, t_n) \subseteq (s_1, \cdots, s_p), \]
then
\[ \nu_{t_1, \cdots, t_n} (F_{t_1} \times \cdots \times F_{t_n}) = \nu_{s_1, \cdots, s_p} (G_{s_1} \times \cdots \times G_{s_p}) \]
(2.5.8)
where if \( s_i = t_j \), then \( G_{s_i} = F_{t_j} \) and if \( s_i \) is not equal to any of the indices, \( t_k \), then \( G_{s_i} = M_{s_i} \). Then for \( E \) defined in Notation 2.5.2, there exists a probability measure, \( P \) and a \( \sigma \) algebra \( \mathcal{F} = \sigma (E) \) such that
\[ \left( \prod_{t \in I} M_t, P, \mathcal{F} \right) \]
is a probability space. Also there exist measurable functions, \( X_s : \prod_{t \in I} M_t \rightarrow M_s \) defined as
\[ X_s x = x_s \]
for each \( s \in I \) such that for each \( (t_1, \cdots, t_n) \subseteq I, \)
\[ \nu_{t_1, \cdots, t_n} (F_{t_1} \times \cdots \times F_{t_n}) = P ([X_{t_1} \in F_{t_1}] \cap \cdots \cap [X_{t_n} \in F_{t_n}]) \]
\[ = P \left( (X_{t_1}, \cdots, X_{t_n}) \in \prod_{j=1}^n F_{t_j} \right) = P \left( \prod_{t \in I} F_t \right) \]
(2.5.9)
where \( F_t = M_t \) for every \( t \notin \{t_1, \cdots, t_n\} \) and \( F_{t_i} \) is a Borel set. Also if \( f \) is a nonnegative function of finitely many variables, \( x_{t_1}, \cdots, x_{t_n} \), measurable with respect to \( B \left( \prod_{j=1}^n M_{t_j} \right) \), then \( f \) is also measurable with respect to \( \mathcal{F} \) and
\[ \int_{M_{t_1} \times \cdots \times M_{t_n}} f(x_{t_1}, \cdots, x_{t_n}) \, d\nu_{t_1, \cdots, t_n} \]
\[ = \int_{\prod_{t \in I} M_t} f(x_{t_1}, \cdots, x_{t_n}) \, dP \]
(2.5.10)

**Proof:** Let \( E \) be the algebra of sets defined in the above notation. I want to define a measure on \( E \). For \( F \in E \), there exists \( J \) such that \( F \) is the finite disjoint union of sets of \( \mathcal{R}_J \). Define
\[ P_0 (F) \equiv \nu_J (\pi_J (F)) \]
Then \( P_0 \) is well defined because of the consistency condition on the measures \( \nu_J \). \( P_0 \) is clearly finitely additive because the \( \nu_J \) are measures and one can pick \( J \) as large as desired to include all \( t \) where there may be something other than \( M_t \). Also, from the definition,
\[ P_0 (\Omega) = P_0 \left( \prod_{t \in I} M_t \right) = \nu_{t_1} (M_{t_1}) = 1. \]
Next I will show \( P_0 \) is a finite measure on \( E \). After this it is only a matter of using the Caratheodory extension theorem to get the existence of the desired probability measure \( P \).

**Claim:** Suppose \( E^n \) is in \( E \) and suppose \( E^n \uparrow 0 \). Then \( P_0 \left( E^n \right) \uparrow 0 \).

**Proof of the claim:** If not, there exists a sequence such that although \( E^n \uparrow 0, P_0 \left( E^n \right) \uparrow \varepsilon > 0 \). Let \( E^n \in \mathcal{E}_{J^n} \). Thus it is a finite disjoint union of sets of \( \mathcal{R}_{J^n} \). By regularity of the measures \( \nu_J \), which follows from Lemmas 1.5.3 and 1.5.4, there exists \( K_{J^n} \subseteq E^n \) such that
\[ \nu_{J^n} \left( \pi_{J^n} (K_{J^n}) \right) + \frac{\varepsilon}{2^{n+2}} > \nu_{J^n} \left( \pi_{J^n} (E^n) \right) \]
Thus
\[ P_0 \left( K_{J^n} \right) + \frac{\varepsilon}{2^{n+2}} = \nu_{J^n} \left( \pi_{J^n} (K_{J^n}) \right) + \frac{\varepsilon}{2^{n+2}} > \nu_{J^n} \left( \pi_{J^n} (E^n) \right) \equiv P_0 \left( E^n \right) \]
The interesting thing about these $K_{j_n}$ is: they have the finite intersection property. Here is why.

$$\varepsilon \leq P_0 (\cap_{k=1}^m K_{j_k}) + P_0 (E^m \setminus \cap_{k=1}^m K_{j_k})$$

$$\leq P_0 (\cap_{k=1}^m K_{j_k}) + P_0 (\cup_{k=1}^m E^k \setminus K_{j_k})$$

$$< P_0 (\cap_{k=1}^m K_{j_k}) + \sum_{k=1}^\infty \frac{\varepsilon}{2^{k+2}} < P_0 (\cap_{k=1}^m K_{j_k}) + \varepsilon/2,$$

and so $P_0 (\cap_{k=1}^m K_{j_k}) > \varepsilon/2$. In considering all the $E^n$, there are countably many entries in the product space which have something other than $M_i$ in them. Say these are $\{t_1, t_2, \cdots\}$. Let $p_i$ be a point which is in the intersection of the $t_i$ components of the sets $K_{j_n}$. The compact sets in the $t_i$ position must have the finite intersection property also because if not, the sets $K_{j_n}$ can’t have it. Thus there is such a point. As to the other positions, use the axiom of choice to pick something in each of these. Thus the intersection of these $K_{j_n}$ contains a point which is contrary to $E^n \downarrow \emptyset$ because these sets are contained in the $E^n$.

With the claim, it follows $P_0$ is a measure on $\mathcal{E}$. Here is why: If $E = \cup_{k=1}^\infty E^k$ where $E, E^k \in \mathcal{E}$, then $(E \setminus \cup_{k=1}^m E^k) \downarrow \emptyset$ and so

$$P_0 (\cup_{k=1}^\infty E^k) \rightarrow P_0 (E).$$

Hence if the $E_k$ are disjoint, $P_0 (\cup_{k=1}^n E^k) = \sum_{k=1}^n P_0 (E_k) \rightarrow P_0 (E)$. Thus for disjoint $E_k$ having $\cup_k E_k = E \in \mathcal{E}$,

$$P_0 (\cup_{k=1}^\infty E^k) = \sum_{k=1}^\infty P_0 (E_k).$$

Now to conclude the proof, apply the Caratheodory extension theorem to obtain $P$ a probability measure which extends $P_0$ to a $\sigma$ algebra which contains $\sigma (\mathcal{E})$ the sigma algebra generated by $\mathcal{E}$ with $P = P_0$ on $\mathcal{E}$. Thus for $E_J \in \mathcal{E}$, $P (E_J) = P_0 (E_J) = \nu_J (P_J)$.

Next, let $(\prod_{t \in I} M_t, F, P)$ be the probability space and for $x \in \prod_{t \in I} M_t$ let $X_t (x) = x_t$, the $t^{th}$ entry of $x$. It follows $X_t$ is measurable (also continuous) because if $U$ is open in $M_t$, then $X_t^{-1} (U)$ has a $U$ in the $t^{th}$ slot and $M_s$ everywhere else for $s \neq t$. Thus inverse images of open sets are measurable. Also, letting $J$ be a finite subset of $I$ and for $J = (t_1, \cdots, t_n)$, and $F_{t_1}, \cdots, F_{t_n}$ Borel sets in $M_{t_1}, \cdots, M_{t_n}$ respectively, it follows $F_J$, where $F_J$ has $F_{t_i}$ in the $t_i^{th}$ entry, is in $\mathcal{E}$ and therefore,

$$P ([X_{t_1} \in F_{t_1}] \cap [X_{t_2} \in F_{t_2}] \cap \cdots \cap [X_{t_n} \in F_{t_n}]) =$$

$$P (\{X_{t_1}, X_{t_2}, \cdots, X_{t_n}\} \in F_{t_1} \times \cdots \times F_{t_n}) = P (F_J) = P_0 (F_J)$$

$$= \nu_{t_1, \cdots, t_n} (F_{t_1} \times \cdots \times F_{t_n})$$

Finally consider the claim about the integrals. Suppose $f (x_{t_1}, \cdots, x_{t_n}) = x_F$ where $F$ is a Borel set of $\prod_{t \in I} M_t$ where $J = (t_1, \cdots, t_n)$. To begin with suppose

$$F = F_{t_1} \times \cdots \times F_{t_n} \quad (2.5.11)$$

where each $F_{t_j}$ is in $B (M_{t_j})$. Then

$$\int_{M_{t_1} \times \cdots \times M_{t_n}} x_F (x_{t_1}, \cdots, x_{t_n}) d\nu_{t_1 \cdots t_n} = \nu_{t_1 \cdots t_n} (F_{t_1} \times \cdots \times F_{t_n})$$

$$= P \left( \prod_{t \in J} F_t \right) = \int_\Omega x_F (x) F_t dP$$

$$= \int_\Omega x_F (x_{t_1}, \cdots, x_{t_n}) dP \quad (2.5.12)$$

where $F_t = M_t$ if $t \notin J$. Let $K$ denote sets, $F$ of the sort in (2.5.11). It is clearly a $\pi$ system. Now let $G$ denote those sets $F$ in $B (\prod_{t \in I} M_t)$ such that (2.5.12) holds. Thus $G \supseteq K$. It is clear that $G$ is closed with respect to countable disjoint unions and complements. Hence $G \supseteq \sigma (K)$ but $\sigma (K) = B (\prod_{t \in I} M_t)$ because every open set in $\prod_{t \in I} M_t$ is the countable union of rectangles like $M_{t_1} \times \cdots \times M_{t_n}$ in which each $F_{t_i}$ is open. Therefore, (2.5.11) holds for every $F \in B (\prod_{t \in I} M_t)$. 

2.5. KOLMOGOROV EXTENSION THEOREM FOR POLISH SPACES 

37
Passing to simple functions and then using the monotone convergence theorem yields the final claim of the theorem.

As a special case, you can obtain a version of product measure for possibly infinitely many factors. Suppose in the context of the above theorem that \( \nu_t \) is a probability measure defined on the Borel sets of \( M_t \equiv \mathbb{R}^{n_t} \) for \( n_t \) a positive integer, and let the measures, \( \nu_{t_1 \cdots t_n} \) be defined on the Borel sets of \( \prod_{t=1}^{n} M_{t_i} \) by

\[
\nu_{t_1 \cdots t_n} (E) \equiv (\nu_{t_1} \times \cdots \times \nu_{t_n})(E).
\]

Then these measures satisfy the necessary consistency condition and so the Kolmogorov extension theorem given above can be applied to obtain a measure \( P \) defined on a \( (\prod_{t \in I} M_t, \mathcal{F}) \) and measurable functions \( X_s : \prod_{t \in I} M_t \to M_s \) such that for \( F_t \) a Borel set in \( M_{t_i} \),

\[
P \left( (X_{t_1}, \cdots, X_{t_n}) \in \prod_{i=1}^{n} F_{t_i} \right) = \nu_{t_1 \cdots t_n} (F_{t_1} \times \cdots \times F_{t_n})
\]

\[
= \nu_{t_1} (F_{t_1}) \cdots \nu_{t_n} (F_{t_n}).
\]

(2.5.13)

In particular, \( P (X_t \in F_t) = \nu_t (F_t) \). Then \( P \) in the resulting probability space,

\[
\left( \prod_{t \in I} M_t, \mathcal{F}, P \right)
\]

will be denoted as \( \prod_{t \in I} \nu_t \). This proves the following theorem which describes an infinite product measure.

**Theorem 2.5.4** Let \( M_t \) for \( t \in I \) be given as in Theorem 2.5.3 and let \( \nu_t \) be a Borel probability measure defined on the Borel sets of \( M_t \). Then there exists a measure \( P \) and a \( \sigma \) algebra \( \mathcal{F} = \sigma(\mathcal{E}) \) where \( \mathcal{E} \) is given in the Notation 2.5.1 such that \( (\prod_{t \in I} M_t, \mathcal{F}, P) \) is a probability space satisfying 2.5.13 whenever each \( F_t \) is a Borel set of \( M_t \). This probability measure is sometimes denoted as \( \prod_{t \in I} \nu_t \).
Chapter 3

Fourier Transforms

3.1 An Algebra Of Special Functions

First recall the following definition of a polynomial.

**Definition 3.1.1** \( \alpha = (\alpha_1, \ldots, \alpha_n) \) for \( \alpha_1 \cdots \alpha_n \) positive integers is called a multi-index. For \( \alpha \) a multi-index, \(|\alpha| \equiv \alpha_1 + \cdots + \alpha_n \) and if \( x \in \mathbb{R}^n \),

\[
x = (x_1, \ldots, x_n),
\]

and \( f \) a function, define

\[
x^\alpha \equiv x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}.
\]

A polynomial in \( n \) variables of degree \( m \) is a function of the form

\[
p(x) = \sum_{|\alpha| \leq m} a_\alpha x^\alpha.
\]

Here \( \alpha \) is a multi-index as just described and \( a_\alpha \in \mathbb{C} \). Also define for \( \alpha = (\alpha_1, \ldots, \alpha_n) \) a multi-index

\[
D^\alpha f(x) \equiv \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}.
\]

**Definition 3.1.2** Define \( G_1 \) to be the functions of the form \( p(x) e^{-a|x|^2} \) where \( a > 0 \) and \( p(x) \) is a polynomial. Let \( G \) be all finite sums of functions in \( G_1 \). Thus \( G \) is an algebra of functions which has the property that if \( f \in G \) then \( f \in G \).

It is always assumed, unless stated otherwise that the measure will be Lebesgue measure.

**Lemma 3.1.3** \( G \) is dense in \( C_0(\mathbb{R}^n) \) with respect to the norm,

\[
||f||_\infty \equiv \sup \{|f(x)| : x \in \mathbb{R}^n\}
\]

**Proof:** By the Weierstrass approximation theorem, it suffices to show \( G \) separates the points and annihilates no point. It was already observed in the above definition that \( f \in G \) whenever \( f \in G \). If \( y_1 \neq y_2 \) suppose first that \( |y_1| \neq |y_2| \). Then in this case, you can let \( f(x) \equiv e^{-a|x|^2} \) and \( f \in G \) and \( f(y_1) \neq f(y_2) \). If \( |y_1| = |y_2| \), then suppose \( y_{1k} \neq y_{2k} \). This must happen for some \( k \) because \( y_1 \neq y_2 \). Then let \( f(x) \equiv x_k e^{-|x|^2} \). Thus \( G \) separates points. Now \( e^{-|x|^2} \) is never equal to zero and so \( G \) annihilates no point of \( \mathbb{R}^n \). This proves the lemma.

These functions are clearly quite specialized. Therefore, the following theorem is somewhat surprising.

**Theorem 3.1.4** For each \( p \geq 1, p < \infty \), \( G \) is dense in \( L^p(\mathbb{R}^n) \).

**Proof:** Let \( f \in L^p(\mathbb{R}^n) \). Then there exists \( g \in C_c(\mathbb{R}^n) \) such that \( ||f - g||_p < \varepsilon \). Now let \( b > 0 \) be large enough that

\[
\int_{\mathbb{R}^n} (e^{-b|x|^2})^p \, dx < \varepsilon^p.
\]
Then \( x \to g(x) e^{b|x|^2} \) is in \( C_c(\mathbb{R}^n) \subseteq C_0(\mathbb{R}^n) \). Therefore, from Lemma 3.1.4 there exists \( \psi \in G \) such that
\[
\left\| g e^{b|x|^2} - \psi \right\|_\infty < 1
\]
Therefore, letting \( \phi(x) \equiv e^{-b|x|^2} \psi(x) \) it follows that \( \phi \in G \) and for all \( x \in \mathbb{R}^n \),
\[
|g(x) - \phi(x)| < e^{-b|x|^2}
\]
Therefore,
\[
\left( \int_{\mathbb{R}^n} |g(x) - \phi(x)|^p \, dx \right)^{1/p} \leq \left( \int_{\mathbb{R}^n} (e^{-b|x|^2})^p \, dx \right)^{1/p} < \varepsilon.
\]
It follows
\[
\|f - \phi\|_p \leq \|f - g\|_p + \|g - \phi\|_p < 2\varepsilon.
\]
Since \( \varepsilon > 0 \) is arbitrary, this proves the theorem.

The following lemma is also interesting even if it is obvious.

**Lemma 3.1.5** For \( \psi \in G \), \( p \) a polynomial, and \( \alpha, \beta \) multiindices, \( D^\alpha \psi \in G \) and \( p\psi \in G \).

\[
\sup\{|x^\beta D^\alpha \psi(x)| : x \in \mathbb{R}^n\} < \infty
\]

### 3.2 Fourier Transforms Of Functions In \( G \)

**Definition 3.2.1** For \( \psi \in G \) Define the Fourier transform, \( F \) and the inverse Fourier transform, \( F^{-1} \) by
\[
F\psi(t) \equiv (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-it\cdot x} \psi(x) \, dx,
\]
\[
F^{-1}\psi(t) \equiv (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{it\cdot x} \psi(x) \, dx.
\]
where \( t \cdot x \equiv \sum_{i=1}^n t_i x_i \). Note there is no problem with this definition because \( \psi \) is in \( L^1(\mathbb{R}^n) \) and therefore,
\[
\left| e^{it\cdot x} \psi(x) \right| \leq |\psi(x)|,
\]
an integrable function.

One reason for using the functions, \( G \) is that it is very easy to compute the Fourier transform of these functions. The first thing to do is to verify \( F \) and \( F^{-1} \) map \( G \) to \( G \) and that \( F^{-1} \circ F(\psi) = \psi \).

**Lemma 3.2.2** The following formulas are true. \((c > 0)\)
\[
\int_{\mathbb{R}} e^{-ct^2} e^{-ist} \, dt = \int_{\mathbb{R}} e^{-ct^2} e^{ist} \, dt = e^{-\frac{s^2}{4c}} \sqrt{\frac{\pi}{c}}, \quad (3.2.1)
\]
\[
\int_{\mathbb{R}^n} e^{-c|t|^2} e^{-ist} \, dt = \int_{\mathbb{R}^n} e^{-c|t|^2} e^{ist} \, dt = e^{-\frac{|s|^2}{4c}} \left( \frac{\sqrt{\pi}}{\sqrt{c}} \right)^n. \quad (3.2.2)
\]

**Proof:** Consider the first one. Let \( h(s) \) be given by the left side. Then
\[
H(s) \equiv \int_{\mathbb{R}} e^{-ct^2} e^{-ist} \, dt = \int_{\mathbb{R}} e^{-ct^2} \cos(st) \, dt
\]
Then using the dominated convergence theorem to differentiate,
\[
H'(s) = \int_{\mathbb{R}} -e^{-ct^2} t \sin(st) \, dt = \frac{e^{-ct^2}}{2c} \sin(st) \bigg|_\infty^\infty - \frac{s}{2c} \int_{\mathbb{R}} e^{-ct^2} \cos(st) \, dt = -\frac{s}{2c} H(s).
\]
Also \( H(0) = \int_{\mathbb{R}} e^{-ct^2} \, dt \). Thus \( H(0) = \int_{\mathbb{R}} e^{-ct^2} \, dx \equiv I \) and so
\[
I^2 = \int_{\mathbb{R}^2} e^{-c(x^2+y^2)} \, dx \, dy = \int_0^\infty \int_0^{2\pi} e^{-cr^2} r \, dr \, d\theta = \frac{\pi}{c}.
\]
Hence

$$H'(s) + \frac{s}{2c} H(s) = 0, \ H(0) = \sqrt{\frac{\pi}{c}}.$$  

It follows that $H(s) = e^{-\frac{s^2}{2}} \sqrt{\pi}$. The second formula follows right away from Fubini’s theorem. ■

With these formulas, it is easy to verify $F, F^{-1}$ map $\mathcal{G}$ to $\mathcal{G}$ and $F \circ F^{-1} = F^{-1} \circ F = id.$

**Theorem 3.2.3** Each of $F$ and $F^{-1}$ map $\mathcal{G}$ to $\mathcal{G}$. Also $F^{-1} \circ F(\psi) = \psi$ and $F \circ F^{-1}(\psi) = \psi$.

**Proof:** The first claim will be shown if it is shown that $F\psi \in \mathcal{G}$ for $\psi(x) = x^\alpha e^{-b|x|^2}$ because an arbitrary function of $\mathcal{G}$ is a finite sum of scalar multiples of functions such as $\psi$. Using Lemma 6.2.4,

$$F\psi(t) = \left(\frac{1}{2\pi}\right)^{n/2} \int_{\mathbb{R}^n} e^{-it \cdot x} x^\alpha e^{-b|x|^2} dx$$

$$= \left(\frac{1}{2\pi}\right)^{n/2} (i)^{-[\alpha]} D_t^\alpha \left(\int_{\mathbb{R}^n} e^{-it \cdot x} e^{-b|x|^2} dx\right)$$

$$= \left(\frac{1}{2\pi}\right)^{n/2} (i)^{-[\alpha]} D_t^\alpha \left(\sqrt{\frac{\pi}{b}}\right)^n$$

and this is clearly in $\mathcal{G}$ because it equals a polynomial times $e^{-\frac{|t|^2}{2\pi}}$.

It remains to verify the other assertion. As in the first case, it suffices to consider $\psi(x) = x^\alpha e^{-b|x|^2}$.

$$F^{-1} \circ F(\psi)(s) = \left(\frac{1}{2\pi}\right)^{-n/2} \int_{\mathbb{R}^n} e^{is \cdot t} F(\psi)(t) dt$$

$$= \left(\frac{1}{2\pi}\right)^{n/2} \int_{\mathbb{R}^n} e^{is \cdot t} \left(\frac{1}{2\pi}\right)^{n/2} (i)^{-[\alpha]} D_t^\alpha \left(e^{-\frac{|t|^2}{2\pi}}\right) dt$$

$$= \left(\frac{1}{2\pi}\right)^n (i)^{-[\alpha]} \int_{\mathbb{R}^n} e^{is \cdot t} D_t^\alpha \left(e^{-\frac{|t|^2}{2\pi}}\right) dt$$

$$= \left(\frac{1}{2\pi}\right)^n (i)^{-[\alpha]} \int_{\mathbb{R}^n} (i)^{[\alpha]} s^\alpha e^{is \cdot t} \left(e^{-\frac{|t|^2}{2\pi}}\right) dt$$

and by Lemma 6.2.4,

$$= \left(\frac{1}{2\pi}\right)^n \left(\frac{\sqrt{\pi}}{\sqrt{b}}\right)^n \int_{\mathbb{R}^n} s^\alpha e^{-\frac{|s|^2}{2\pi}} dt = 1$$

$$= \left(\frac{1}{2\pi}\right)^n \left(\frac{\sqrt{\pi}}{\sqrt{b}}\right)^n \left(\frac{\sqrt{\pi}}{\sqrt{1/4b}}\right)^n s^\alpha e^{-b|s|^2} = \psi(s).$$

### 3.3 Fourier Transforms Of Just About Anything

#### 3.3.1 Fourier Transforms Of $\mathcal{G}^*$

**Definition 3.3.1** Let $\mathcal{G}^*$ denote the vector space of linear functions defined on $\mathcal{G}$ which have values in $\mathbb{C}$. Thus $T \in \mathcal{G}^*$ means $T : \mathcal{G} \rightarrow \mathbb{C}$ and $T$ is linear,

$$T(a\psi + b\phi) = aT(\psi) + bT(\phi) \text{ for all } a, b \in \mathbb{C}, \ \psi, \phi \in \mathcal{G}.$$  

Let $\psi \in \mathcal{G}$. Then we can regard $\psi$ as an element of $\mathcal{G}^*$ by defining

$$\psi(\phi) \equiv \int_{\mathbb{R}^n} \psi(x) \phi(x) dx.$$  

Then we have the following important lemma.
Lemma 3.3.2 The following is obtained for all $\phi, \psi \in \mathcal{G}$.

$$F\psi(\phi) = \psi(F\phi), F^{-1}\psi(\phi) = \psi(F^{-1}\phi)$$

Also if $\psi \in \mathcal{G}$ and $\psi = 0$ in $\mathcal{G}^*$ so that $\psi(\phi) = 0$ for all $\phi \in \mathcal{G}$, then $\psi = 0$ as a function.

Proof:

$$F\psi(\phi) \equiv \int_{\mathbb{R}^n} F\psi(t) \phi(t) \, dt = \int_{\mathbb{R}^n} \left(\frac{1}{2\pi}\right)^{n/2} \int_{\mathbb{R}^n} e^{-it \cdot x} \psi(x) \phi(t) \, dx \, dt = \int_{\mathbb{R}^n} \psi(x) F\phi(x) \, dx = \psi(F\phi)$$

The other claim is similar.

Suppose now $\psi(\phi) = 0$ for all $\phi \in \mathcal{G}$. Then

$$\int_{\mathbb{R}^n} \psi\phi \, dx = 0$$

for all $\phi \in \mathcal{G}$. Therefore, this is true for $\phi = \psi$ and so $\psi = 0$. ■

This lemma suggests a way to define the Fourier transform of something in $\mathcal{G}^*$.

Definition 3.3.3 For $T \in \mathcal{G}^*$, define $FT, F^{-1}T \in \mathcal{G}^*$ by

$$FT(\phi) \equiv T(F\phi), F^{-1}T(\phi) \equiv T(F^{-1}\phi)$$

Lemma 3.3.4 $F$ and $F^{-1}$ are both one to one, onto, and are inverses of each other.

Proof: First note $F$ and $F^{-1}$ are both linear. This follows directly from the definition. Suppose now $FT = 0$. Then $FT(\phi) = T(F\phi) = 0$ for all $\phi \in \mathcal{G}$. But $F$ and $F^{-1}$ map $\mathcal{G}$ onto $\mathcal{G}$ because if $\psi \in \mathcal{G}$, then as shown above, $\psi = F(F^{-1}(\psi))$. Therefore, $T = 0$ and so $F$ is one to one. Similarly $F^{-1}$ is one to one. Now

$$F^{-1}(FT)(\phi) \equiv (FT)(F^{-1}\phi) \equiv T(F(F^{-1}(\phi))) = T\phi.$$  

Therefore, $F^{-1} \circ F(T) = T$. Similarly, $F \circ F^{-1}(T) = T$. Thus both $F$ and $F^{-1}$ are one to one and onto and are inverses of each other as suggested by the notation. ■

Probably the most interesting things in $\mathcal{G}^*$ are functions of various kinds. The following lemma will be useful in considering this situation.

Lemma 3.3.5 If $f \in L^1_{loc}(\mathbb{R}^n)$ and $\int_{\mathbb{R}^n} f\phi \, dx = 0$ for all $\phi \in C_c(\mathbb{R}^n)$, then $f = 0$ a.e.

Proof: For $r > 0$, let

$$E \equiv \{x : f(x) \geq r\}, E_R \equiv E \cap B(0,R).$$

Let $K_m$ be an increasing sequence of compact sets, and let $V_m$ be a decreasing sequence of open sets satisfying

$$K_m \subseteq E_R \subseteq V_m, m_n(V_m) \leq m_n(K_m) + 2^{-m}, V_1 \subseteq B(0,R).$$

Therefore,

$$m_n(V_m \setminus K_m) \leq 2^{-m}.$$  

Let

$$\phi_m \in C_c(V_m), K_m \prec \phi_m \prec V_m.$$  

The statement $K_m \prec \phi_m \prec V_m$ means that $\phi_m$ equals 1 on $K_m$, has compact support in $V_m$, maps into $[0,1]$, and is continuous. Then $\phi_m(x) \to \chi_{E_R}(x)$ a.e. because the set where $\phi_m(x)$ fails to converge to this set is contained in the set of all $x$ which are in infinitely many of the sets $V_m \setminus K_m$. This set has measure zero because

$$\sum_{m=1}^{\infty} m_n(V_m \setminus K_m) < \infty$$
Thus $\phi_m$ converges pointwise a.e to $X_{E_R}$ and so, by the dominated convergence theorem,

$$0 = \lim_{m \to \infty} \int_{E_R} f \phi_m \, dx = \lim_{m \to \infty} \int_{Y_1} f \phi_m \, dx = \int_{E_R} f \, dx \geq rm(E_R).$$

Thus, $m_n(E_R) = 0$ and therefore $m_n(E) = \lim_{R \to \infty} m_n(E_R) = 0$. Since $r > 0$ is arbitrary, it follows

$$m_n([f > 0]) = 0 \text{ and therefore } m_n([f > 0]) = \lim_{R \to \infty} m_n([f > 0]) = 0.$$ 

Hence $f^+ = 0$ a.e. It follows that $\int f^- \phi \, dx = 0$ for all $\phi \in C_c(\mathbb{R}^n)$ because

$$\int f^- \phi \, dx = \int f^+ \phi - \int f \phi = 0.$$ 

Thus from what was just shown, with $f^-$ taking the place of $f$, it follows $|f^-| + f^- = 0$ and so $f^- = 0$ a.e. also. ■

**Corollary 3.3.6** Let $f \in L^1(\mathbb{R}^n)$ and suppose

$$\int_{\mathbb{R}^n} f(x) \phi(x) \, dx = 0$$

for all $\phi \in \mathcal{G}$. Then $f = 0$ a.e.

**Proof:** Let $\psi \in C_c(\mathbb{R}^n)$. Then by the Stone Weierstrass approximation theorem, there exists a sequence of functions, $\{\phi_k\} \subseteq \mathcal{G}$ such that $\phi_k \to \psi$ uniformly. Then by the dominated convergence theorem,

$$\int f \psi \, dx = \lim_{k \to \infty} \int f \phi_k \, dx = 0.$$ 

By Lemma 3.3.5, $f = 0$. ■

The next theorem is the main result of this sort.

**Theorem 3.3.7** Let $f \in L^p(\mathbb{R}^n)$, $p \geq 1$, or suppose $f$ is measurable and has polynomial growth,

$$|f(x)| \leq K \left(1 + |x|^m\right)^m$$

for some $m \in \mathbb{N}$. Then if

$$\int f \psi \, dx = 0$$

for all $\psi \in \mathcal{G}$, then it follows $f = 0$.

**Proof:** First note that if $f \in L^p(\mathbb{R}^n)$ or has polynomial growth, then it makes sense to write the integral $\int f \psi \, dx$ described above. This is obvious in the case of polynomial growth. In the case where $f \in L^p(\mathbb{R}^n)$ it also makes sense because

$$\int |f| \psi \, dx \leq \left(\int |f|^p \, dx\right)^{1/p} \left(\int |\psi|^{p'} \, dx\right)^{1/p'} < \infty$$

due to the fact mentioned above that all these functions in $\mathcal{G}$ are in $L^p(\mathbb{R}^n)$ for every $p \geq 1$. Suppose now that $f \in L^p, p \geq 1$. The case where $f \in L^1(\mathbb{R}^n)$ was dealt with in Corollary 3.3.5. Suppose $f \in L^p(\mathbb{R}^n)$ for $p > 1$. Then

$$|f|^{p-2} f \in L^{p'}(\mathbb{R}^n), \quad p' = q, \quad \frac{1}{p} + \frac{1}{q} = 1$$

and by density of $\mathcal{G}$ in $L^{p'}(\mathbb{R}^n)$ (Theorem 3.3.5), there exists a sequence $\{g_k\} \subseteq \mathcal{G}$ such that

$$\left\|g_k - |f|^{p-2} f\right\|_{p'} \to 0.$$
Then
\[
\int_{\mathbb{R}^n} |f|^p \, dx = \int_{\mathbb{R}^n} f \left( |f|^{p-2} \bar{f} - g_k \right) \, dx + \int_{\mathbb{R}^n} fg_k \, dx
\]
\[
= \int_{\mathbb{R}^n} f \left( |f|^{p-2} \bar{f} - g_k \right) \, dx
\]
\[
\leq \|f\|_{L^p} \|g_k - |f|^{p-2} \bar{f}\|_{p'}
\]
which converges to 0. Hence \( f = 0 \).

It remains to consider the case where \( f \) has polynomial growth. Thus \( x \rightarrow f(x) e^{-|x|^2} \in L^1(\mathbb{R}^n) \). Therefore, for all \( \psi \in \mathcal{G} \),
\[
0 = \int f(x) e^{-|x|^2} \psi(x) \, dx
\]
because \( e^{-|x|^2} \psi(x) \in \mathcal{G} \). Therefore, by the first part, \( f(x) e^{-|x|^2} = 0 \) a.e. \( \square \)

The following theorem shows that you can consider most functions you are likely to encounter as elements of \( \mathcal{G}^* \).

**Theorem 3.3.8** Let \( f \) be a measurable function with polynomial growth,
\[|f(x)| \leq C \left( 1 + |x|^2 \right)^N\]
for some \( N \), or let \( f \in L^p(\mathbb{R}^n) \) for some \( p \in [1, \infty] \). Then \( f \in \mathcal{G}^* \) if
\[f(\phi) \equiv \int f \phi \, dx.
\]

**Proof:** Let \( f \) have polynomial growth first. Then the above integral is clearly well defined and so in this case, \( f \in \mathcal{G}^* \).

Next suppose \( f \in L^p(\mathbb{R}^n) \) with \( \infty > p \geq 1 \). Then it is clear again that the above integral is well defined because of the fact that \( \phi \) is a sum of polynomials times exponentials of the form \( e^{-c|x|^2} \) and these are in \( L^{p'}(\mathbb{R}^n) \). Also \( \phi \rightarrow f(\phi) \) is clearly linear in both cases. \( \square \)

This has shown that for nearly any reasonable function, you can define its Fourier transform as described above. You could also define the Fourier transform of a finite Borel measure \( \mu \) because for such a measure
\[\psi \rightarrow \int_{\mathbb{R}^n} \psi d\mu \]
is a linear functional on \( \mathcal{G} \). This includes the very important case of probability distribution measures. The theoretical basis for this assertion will be given a little later.

### 3.3.2 Fourier Transforms Of Functions In \( L^1(\mathbb{R}^n) \)

First suppose \( f \in L^1(\mathbb{R}^n) \).

**Theorem 3.3.9** Let \( f \in L^1(\mathbb{R}^n) \). Then \( Ff(\phi) = \int_{\mathbb{R}^n} g \phi dt \) where
\[g(t) = \left( \frac{1}{2\pi} \right)^{n/2} \int_{\mathbb{R}^n} e^{-it \cdot x} f(x) \, dx \]
and \( F^{-1}f(\phi) = \int_{\mathbb{R}^n} g \phi dt \) where \( g(t) = \left( \frac{1}{2\pi} \right)^{n/2} \int_{\mathbb{R}^n} e^{it \cdot x} f(x) \, dx \). In short,
\[Ff(t) \equiv (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-it \cdot x} f(x) \, dx, \]
\[F^{-1}f(t) \equiv (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{it \cdot x} f(x) \, dx. \]
3.3. FOURIER TRANSFORMS OF JUST ABOUT ANYTHING

Proof: From the definition and Fubini’s theorem,

$$Ff(\phi) = \int_{\mathbb{R}^n} f(t) F\phi(t) \, dt = \int_{\mathbb{R}^n} f(t) \left(\frac{1}{2\pi}\right)^{n/2} \int_{\mathbb{R}^n} e^{-it \cdot x} \phi(x) \, dx \, dt$$

$$= \int_{\mathbb{R}^n} \left(\frac{1}{2\pi}\right)^{n/2} \int_{\mathbb{R}^n} f(t) e^{-it \cdot x} \, dt \phi(x) \, dx.$$ 

Since $\phi \in \mathcal{G}$ is arbitrary, it follows from Theorem 3.3.4 that $Ff(x)$ is given by the claimed formula. The case of $F^{-1}$ is identical.

Here are interesting properties of these Fourier transforms of functions in $L^1$.

Theorem 3.3.10 If $f \in L^1(\mathbb{R}^n)$ and $||f_k - f||_1 \to 0$, then $Ff_k$ and $F^{-1}f_k$ converge uniformly to $Ff$ and $F^{-1}f$ respectively. If $f \in L^1(\mathbb{R}^n)$, then $F^{-1}f$ and $Ff$ are both continuous and bounded. Also,

$$\lim_{|x| \to \infty} F^{-1}f(x) = \lim_{|x| \to \infty} Ff(x) = 0. \quad (3.3.3)$$

Furthermore, for $f \in L^1(\mathbb{R}^n)$ both $Ff$ and $F^{-1}f$ are uniformly continuous.

Proof: The first claim follows from the following inequality.

$$|Ff_k(t) - Ff(t)| \leq (2\pi)^{-n/2} \int_{\mathbb{R}^n} |e^{-it \cdot x} f_k(x) - e^{-it \cdot x} f(x)| \, dx$$

$$= (2\pi)^{-n/2} \int_{\mathbb{R}^n} |f_k(x) - f(x)| \, dx$$

$$= (2\pi)^{-n/2} ||f - f_k||_1.$$ 

which a similar argument holding for $F^{-1}$.

Now consider the second claim of the theorem.

$$|Ff(t) - Ff(t')| \leq (2\pi)^{-n/2} \int_{\mathbb{R}^n} |e^{-it \cdot x} - e^{-it' \cdot x}| |f(x)| \, dx$$

The integrand is bounded by $2|f(x)|$, a function in $L^1(\mathbb{R}^n)$ and converges to 0 as $t' \to t$ and so the dominated convergence theorem implies $Ff$ is continuous. To see $Ff(t)$ is uniformly bounded,

$$|Ff(t)| \leq (2\pi)^{-n/2} \int_{\mathbb{R}^n} |f(x)| \, dx < \infty.$$ 

A similar argument gives the same conclusions for $F^{-1}$.

It remains to verify 3.3.2 and the claim that $Ff$ and $F^{-1}f$ are uniformly continuous.

$$|Ff(t)| \leq \left| (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-it \cdot x} f(x) dx \right|$$

Now let $\varepsilon > 0$ be given and let $g \in C_0^\infty(\mathbb{R}^n)$ such that $(2\pi)^{-n/2} ||g - f||_1 < \varepsilon/2$. Then

$$|Ff(t)| \leq (2\pi)^{-n/2} \int_{\mathbb{R}^n} |f(x) - g(x)| \, dx$$

$$+ |(2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-it \cdot x} g(x) dx|$$

$$\leq \varepsilon/2 + |(2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-it \cdot x} g(x) dx|.$$ 

Now integrating by parts, it follows that for $||t||_\infty = \max \{||t_j|| : j = 1, \cdots, n\} > 0$

$$|Ff(t)| \leq \varepsilon/2 + (2\pi)^{-n/2} \left| \frac{1}{||t||_\infty} \int_{\mathbb{R}^n} \sum_{j=1}^n \frac{\partial g(x)}{\partial x_j} \, dx \right| \quad (3.3.4)$$
and this last expression converges to zero as $||t||_\infty \to \infty$. The reason for this is that if $t_j \neq 0$, integration by parts with respect to $x_j$ gives

$$(2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-it \cdot x} g(x) dx = (2\pi)^{-n/2} \frac{1}{-it_j} \int_{\mathbb{R}^n} e^{-it \cdot x} \frac{\partial g(x)}{\partial x_j} dx.$$ 

Therefore, choose the $j$ for which $||t||_\infty = |t_j|$ and the result of Theorem 3.3.1 holds. Therefore, from Theorem 3.3.2 if $||t||_\infty$ is large enough, $|Ff(t)| < \epsilon$. Similarly, $\lim_{||t||_\infty \to \infty} F^{-1}(t) = 0$. Consider the claim about uniform continuity. Let $\epsilon > 0$ be given. Then there exists $R$ such that if $||t||_\infty > R$, then $|Ff(t)| < \frac{\epsilon}{2}$. Since $Ff$ is continuous, it is uniformly continuous on the compact set $[-R-1, R+1]^n$. Therefore, there exists $\delta_1$ such that if $||t-t'||_\infty < \delta_1$ for $t, t' \in [-R-1, R+1]^n$, then

$$|Ff(t) - Ff(t')| < \epsilon/2. \quad (3.3.5)$$

Now let $0 < \delta < \min(\delta_1, 1)$ and suppose $||t-t'||_\infty < \delta$. If both $t, t'$ are contained in $[-R, R]^n$, then Theorem 3.3.2 holds. If $t \in [-R, R]^n$ and $t' \notin [-R, R]^n$, then both are contained in $[-R-1, R+1]^n$ and so this verifies Theorem 3.3.2 in this case. The other case is that neither point is in $[-R, R]^n$ and in this case,

$$|Ff(t) - Ff(t')| \leq |Ff(t)| + |Ff(t')| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \blacksquare$$

There is a very interesting relation between the Fourier transform and convolutions.

**Theorem 3.3.11** Let $f, g \in L^1(\mathbb{R}^n)$. Then $f * g \in L^1$ and $F(f * g) = (2\pi)^{n/2} Ff Fg$.

**Proof:** Consider

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |f(x-y)g(y)| dy dx.$$

The function, $(x, y) \to |f(x-y)g(y)|$ is Lebesgue measurable and so by Fubini's theorem,

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |f(x-y)g(y)| dy dx = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |f(x-y)g(y)| dx dy = ||f||_1 ||g||_1 < \infty.$$

It follows that for a.e. $x$, $\int_{\mathbb{R}^n} |f(x-y)g(y)| dy < \infty$ and for each of these values of $x$, it follows that $\int_{\mathbb{R}^n} f(x-y)g(y) dy$ exists and equals a function of $x$ which is in $L^1(\mathbb{R}^n)$, $f * g (x)$. Now

$$F(f * g) (t) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-it \cdot x} f * g (x) dx$$

$$= (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-it \cdot x} \int_{\mathbb{R}^n} f(x-y)g(y) dy dx$$

$$= (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-it \cdot y} g(y) \int_{\mathbb{R}^n} e^{-it \cdot (x-y)} f(x-y) dx dy$$

$$= (2\pi)^{n/2} Ff(t) Fg(t). \blacksquare$$

There are many other considerations involving Fourier transforms of functions in $L^1(\mathbb{R}^n)$.

### 3.3.3 Fourier Transforms Of Functions In $L^2(\mathbb{R}^n)$

Consider $Ff$ and $F^{-1}f$ for $f \in L^2(\mathbb{R}^n)$. First note that the formula given for $Ff$ and $F^{-1}f$ when $f \in L^1(\mathbb{R}^n)$ will not work for $f \in L^2(\mathbb{R}^n)$ unless $f$ is also in $L^1(\mathbb{R}^n)$. Recall that $a + ib = a - ib$.

**Theorem 3.3.12** For $\phi \in \mathcal{G}$, $||F\phi||_2 = ||F^{-1}\phi||_2 = ||\phi||_2$.

**Proof:** First note that for $\psi \in \mathcal{G}$,

$$F(\overline{\psi}) = \overline{F(\psi)}, \quad F^{-1}(\overline{\psi}) = \overline{F(\psi)}. \quad (3.3.6)$$

This follows from the definition. For example,

$$F\overline{\psi}(t) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-it \cdot x} \overline{\psi}(x) dx$$

$$= (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{it \cdot x} \psi(x) dx.$$
3.3. FOURIER TRANSFORMS OF JUST ABOUT ANYTHING

Let \( \phi, \psi \in \mathcal{G} \). It was shown above that

\[
\int_{\mathbb{R}^n} (F\phi)\psi(t)dt = \int_{\mathbb{R}^n} \phi(F\psi)dx.
\]

Similarly,

\[
\int_{\mathbb{R}^n} \phi(F^{-1}\psi)dx = \int_{\mathbb{R}^n} (F^{-1}\phi)\psi dt.
\] (3.3.7)

Now, (3.3.6) and (3.3.7) imply

\[
\int_{\mathbb{R}^n} |\phi|^2 dx = \int_{\mathbb{R}^n} \phi F^{-1}(F\phi)dx = \int_{\mathbb{R}^n} \phi F\overline{\phi}dx = \int_{\mathbb{R}^n} F\phi(\overline{\phi})dx = \int_{\mathbb{R}^n} |F\phi|^2dx.
\]

Similarly

\[
||\phi||_2 = ||F^{-1}\phi||_2. \quad \blacksquare
\]

Lemma 3.3.13 Let \( f \in L^2(\mathbb{R}^n) \) and let \( \phi_k \to f \) in \( L^2(\mathbb{R}^n) \) where \( \phi_k \in \mathcal{G} \). (Such a sequence exists because of density of \( \mathcal{G} \) in \( L^2(\mathbb{R}^n) \).) Then \( Ff \) and \( F^{-1}f \) are both in \( L^2(\mathbb{R}^n) \) and the following limits take place in \( L^2 \):

\[
\lim_{k \to \infty} F(\phi_k) = F(f), \quad \lim_{k \to \infty} F^{-1}(\phi_k) = F^{-1}(f).
\]

**Proof:** Let \( \psi \in \mathcal{G} \) be given. Then

\[
Ff(\psi) = \int_{\mathbb{R}^n} f(\psi) = \int_{\mathbb{R}^n} f(\psi) = \int_{\mathbb{R}^n} Ff(x)\psi(x) dx
\]

\[
= \lim_{k \to \infty} \int_{\mathbb{R}^n} \phi_k(x)F\psi(x) dx = \lim_{k \to \infty} \int_{\mathbb{R}^n} F\phi_k(x)\psi(x) dx.
\]

Also by Theorem (3.3.10) \( \{F\phi_k\}_{k=1}^{\infty} \) is Cauchy in \( L^2(\mathbb{R}^n) \) and so it converges to some \( h \in L^2(\mathbb{R}^n) \). Therefore, from the above,

\[
Ff(\psi) = \int_{\mathbb{R}^n} h(x)\psi(x)
\]

which shows that \( F(f) \in L^2(\mathbb{R}^n) \) and \( h = F(f) \). The case of \( F^{-1} \) is entirely similar. \( \blacksquare \)

Since \( Ff \) and \( F^{-1}f \) are in \( L^2(\mathbb{R}^n) \), this also proves the following theorem.

Theorem 3.3.14 If \( f \in L^2(\mathbb{R}^n) \), \( Ff \) and \( F^{-1}f \) are the unique elements of \( L^2(\mathbb{R}^n) \) such that for all \( \phi \in \mathcal{G} \),

\[
\int_{\mathbb{R}^n} Ff(x)\phi(x)dx = \int_{\mathbb{R}^n} f(x)F\phi(x)dx, \quad (3.3.8)
\]

\[
\int_{\mathbb{R}^n} F^{-1}f(x)\phi(x)dx = \int_{\mathbb{R}^n} f(x)F^{-1}\phi(x)dx. \quad (3.3.9)
\]

Theorem 3.3.15 (Plancherel)

\[
||f||_2 = ||Ff||_2 = ||F^{-1}f||_2. \quad (3.3.10)
\]

**Proof:** Use the density of \( \mathcal{G} \) in \( L^2(\mathbb{R}^n) \) to obtain a sequence, \( \{\phi_k\} \) converging to \( f \) in \( L^2(\mathbb{R}^n) \). Then by Lemma 3.3.13

\[
||Ff||_2 = \lim_{k \to \infty} ||F\phi_k||_2 = \lim_{k \to \infty} ||\phi_k||_2 = ||f||_2.
\]

Similarly,

\[
||f||_2 = ||F^{-1}f||_2. \quad \blacksquare
\]

The following corollary is a simple generalization of this. To prove this corollary, use the following simple lemma which comes as a consequence of the Cauchy Schwarz inequality.

Lemma 3.3.16 Suppose \( f_k \to f \) in \( L^2(\mathbb{R}^n) \) and \( g_k \to g \) in \( L^2(\mathbb{R}^n) \). Then

\[
\lim_{k \to \infty} \int_{\mathbb{R}^n} f_k g_k dx = \int_{\mathbb{R}^n} f g dx
\]
Proof:
\[
\left| \int_{\mathbb{R}^n} f_k g_k dx - \int_{\mathbb{R}^n} f g dx \right| \leq \left| \int_{\mathbb{R}^n} f_k g_k dx - \int_{\mathbb{R}^n} f_k g dx \right| + \left| \int_{\mathbb{R}^n} f_k g dx - \int_{\mathbb{R}^n} f g dx \right|
\]
\[
\leq \|f_k\|_2 \|g - g_k\|_2 + \|g\|_2 \|f_k - f\|_2.
\]
Now \(\|f_k\|_2\) is a Cauchy sequence and so it is bounded independent of \(k\). Therefore, the above expression is smaller than \(\varepsilon\) whenever \(k\) is large enough. ■

**Corollary 3.3.17** For \(f, g \in L^2(\mathbb{R}^n)\),
\[
\int_{\mathbb{R}^n} f \overline{g} dx = \int_{\mathbb{R}^n} Ff \overline{Fg} dx = \int_{\mathbb{R}^n} F^{-1} f \overline{F^{-1} g} dx.
\]

**Proof:** First note the above formula is obvious if \(f, g \in \mathcal{S}\). To see this, note
\[
\int_{\mathbb{R}^n} Ff \overline{Fg} dx = \int_{\mathbb{R}^n} Ff(x) \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-ix \cdot t} g(t) dt dx
\]
\[
= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{ix \cdot t} Ff(x) dx \overline{g(t)} dt
\]
\[
= \int_{\mathbb{R}^n} f(t) \overline{g(t)} dt.
\]

The formula with \(F^{-1}\) is exactly similar.

Now to verify the corollary, let \(\phi_k \to f\) in \(L^2(\mathbb{R}^n)\) and let \(\psi_k \to g\) in \(L^2(\mathbb{R}^n)\). Then by Lemma 3.3.16,
\[
\int_{\mathbb{R}^n} Ff \overline{Fg} dx = \lim_{k \to \infty} \int_{\mathbb{R}^n} F\phi_k \overline{F\psi_k} dx = \lim_{k \to \infty} \int_{\mathbb{R}^n} \phi_k \psi_k dx = \int_{\mathbb{R}^n} f \overline{g} dx
\]
A similar argument holds for \(F^{-1}\).

**Theorem 3.3.18** For \(f \in L^2(\mathbb{R}^n)\), let \(f_r = f \chi_{E_r}\) where \(E_r\) is a bounded measurable set with \(E_r \uparrow \mathbb{R}^n\). Then the following limits hold in \(L^2(\mathbb{R}^n)\).
\[
Ff = \lim_{r \to \infty} Ff_r, \quad F^{-1} f = \lim_{r \to \infty} F^{-1} f_r.
\]

**Proof:** \(\|f - f_r\|_2 \to 0\) and so \(\|F f - F f_r\|_2 \to 0\) and \(\|F^{-1} f - F^{-1} f_r\|_2 \to 0\) by Plancherel’s Theorem. ■

What are \(F f_r\) and \(F^{-1} f_r\)? Let \(\phi \in \mathcal{S}\)
\[
\int_{\mathbb{R}^n} F f_r \phi dx = \int_{\mathbb{R}^n} f_r \phi dx
\]
\[
= (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f_r(x) e^{-i x \cdot y} \phi(y) dy dx
\]
\[
= \int_{\mathbb{R}^n} \left( (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} f_r(x) e^{-i x \cdot y} dx \right) \phi(y) dy.
\]
Since this holds for all \(\phi \in \mathcal{S}\), a dense subset of \(L^2(\mathbb{R}^n)\), it follows that
\[
F f_r(y) = (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} f_r(x) e^{-i x \cdot y} dx.
\]
Similarly
\[
F^{-1} f_r(y) = (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} f_r(x) e^{i x \cdot y} dx.
\]
This shows that to take the Fourier transform of a function in \(L^2(\mathbb{R}^n)\), it suffices to take the limit as \(r \to \infty\) in \(L^2(\mathbb{R}^n)\) of \((2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} f_r(x) e^{-i x \cdot y} dx\). A similar procedure works for the inverse Fourier transform.

Note this reduces to the earlier definition in case \(f \in L^1(\mathbb{R}^n)\). Now consider the convolution of a function in \(L^2\) with one in \(L^1\).
3.3. FOURIER TRANSFORMS OF JUST ABOUT ANYTHING

Theorem 3.3.19 Let \( h \in L^2(\mathbb{R}^n) \) and let \( f \in L^1(\mathbb{R}^n) \). Then \( h * f \in L^2(\mathbb{R}^n) \),

\[
F^{-1}(h * f) = (2\pi)^{n/2} F^{-1}h F^{-1}f, \\
F(h * f) = (2\pi)^{n/2} Fh Ff,
\]

and

\[
||h * f||_2 \leq ||h||_2 ||f||_1. \tag{3.3.11}
\]

Proof: An application of Minkowski’s inequality yields

\[
\left( \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^n} |h(x - y)| |f(y)| dy \right)^2 dx \right)^{1/2} \leq ||f||_1 ||h||_2. \tag{3.3.12}
\]

Hence \( \int |h(x - y)||f(y)|dy < \infty \) a.e. \( x \) and

\[
x \to \int h(x - y)f(y)dy
\]
is in \( L^2(\mathbb{R}^n) \). Let \( E_r \uparrow \mathbb{R}^n, m(E_r) < \infty \). Thus,

\[
h_r \equiv \chi_{E_r} h \in L^2(\mathbb{R}^n) \cap L^1(\mathbb{R}^n),
\]

and letting \( \phi \in \mathcal{G} \),

\[
\int F(h_r * f)(\phi)dx
\]

\[
= \int (h_r * f)(F\phi)dx
\]

\[
= (2\pi)^{-n/2} \int \int h_r(x - y)f(y)e^{-ix \cdot t}\phi(t)dtdydx
\]

\[
= (2\pi)^{-n/2} \int \int \left( \int h_r(x - y)e^{i(x - y) \cdot t}dy \right) f(y)e^{-iy \cdot t}d\phi(t)dt
\]

\[
= \int (2\pi)^{n/2} Fh_r(t) Ff(t) \phi(t)dt.
\]

Since \( \phi \) is arbitrary and \( \mathcal{G} \) is dense in \( L^2(\mathbb{R}^n) \),

\[
F(h_r * f) = (2\pi)^{n/2} Fh_r Ff.
\]

Now by Minkowski’s Inequality, \( h_r * f \to h * f \) in \( L^2(\mathbb{R}^n) \) and also it is clear that \( h_r \to h \) in \( L^2(\mathbb{R}^n) \); so, by Plancherel’s theorem, you may take the limit in the above and conclude

\[
F(h * f) = (2\pi)^{n/2} Fh Ff.
\]

The assertion for \( F^{-1} \) is similar and (3.3.11) follows from (3.3.12).

3.3.4 The Schwartz Class

The problem with \( \mathcal{G} \) is that it does not contain \( C_c^\infty(\mathbb{R}^n) \). I have used it in presenting the Fourier transform because the functions in \( \mathcal{G} \) have a very specific form which made some technical details work out easier than in any other approach I have seen. The Schwartz class is a larger class of functions which does contain \( C_c^\infty(\mathbb{R}^n) \) and also has the same nice properties as \( \mathcal{G} \). The functions in the Schwartz class are infinitely differentiable and they vanish very rapidly as \( |x| \to \infty \) along with all their partial derivatives. This is the description of these functions, not a specific form involving polynomials times \( e^{-\alpha|x|^2} \). To describe this precisely requires some notation.
**Definition 3.3.20** Let $f \in \mathcal{S}$, the Schwartz class, if $f \in C^\infty(\mathbb{R}^n)$ and for all positive integers $N$,

$$\rho_N(f) < \infty$$

where

$$\rho_N(f) = \sup\{(1 + |x|^2)^N |D^\alpha f(x)| : x \in \mathbb{R}^n, |\alpha| \leq N\}.$$ 

Thus $f \in \mathcal{S}$ if and only if $f \in C^\infty(\mathbb{R}^n)$ and

$$\sup\{|x^\beta D^\alpha f(x)| : x \in \mathbb{R}^n\} < \infty$$

for all multi indices $\alpha$ and $\beta$.

Also note that if $f \in \mathcal{S}$, then $p(f) \in \mathcal{S}$ for any polynomial, $p$ with $p(0) = 0$ and that

$$\mathcal{S} \subseteq L^p(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$$

for any $p \geq 1$. To see this assertion about the $p(f)$, it suffices to consider the case of the product of two elements of the Schwartz class. If $f, g \in \mathcal{S}$, then $D^\alpha (fg)$ is a finite sum of derivatives of $f$ times derivatives of $g$. Therefore, $\rho_N(fg) < \infty$ for all $N$. You may wonder about examples of things in $\mathcal{S}$. Clearly any function in $C^\infty_c(\mathbb{R}^n)$ is in $\mathcal{S}$. However there are other functions in $\mathcal{S}$. For example $e^{-|x|^2}$ is in $\mathcal{S}$ as you can verify for yourself and so is any function from $\mathcal{G}$. Note also that the density of $C^\infty_c(\mathbb{R}^n)$ in $L^p(\mathbb{R}^n)$ shows that $\mathcal{S}$ is dense in $L^p(\mathbb{R}^n)$ for every $p$.

Recall the Fourier transform of a function in $L^1(\mathbb{R}^n)$ is given by

$$Ff(t) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-it \cdot x} f(x) dx.$$ 

Therefore, this gives the Fourier transform for $f \in \mathcal{S}$. The nice property which $\mathcal{S}$ has in common with $\mathcal{G}$ is that the Fourier transform and its inverse map $\mathcal{S}$ one to one onto $\mathcal{S}$. This means I could have presented the whole of the above theory in terms of $\mathcal{S}$ rather than in terms of $\mathcal{G}$. However, it is more technical.

**Theorem 3.3.21** If $f \in \mathcal{S}$, then $Ff$ and $F^{-1}f$ are also in $\mathcal{S}$.

**Proof:** To begin with, let $\alpha = e_j = (0, 0, \cdots, 1, 0, \cdots, 0)$, the 1 in the $j^{th}$ slot.

$$F^{-1}f(t + he_j) - F^{-1}f(t) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{it \cdot x} f(x) \left( e^{ihx_j} - 1 \right) dx.$$ 

(3.3.14)

Consider the integrand in $\mathbb{R}^n$:

$$\left| e^{it \cdot x} f(x) \left( e^{ihx_j} - 1 \right) \right| = |f(x)| \left| \left( e^{i(h/2)x_j} - e^{-i(h/2)x_j} \right) \right|$$

$$= |f(x)| \left| \frac{i \sin \left( \frac{h}{2} x_j \right)}{\frac{h}{2}} \right| \leq |f(x)| |x_j|$$

and this is a function in $L^1(\mathbb{R}^n)$ because $f \in \mathcal{S}$. Therefore by the Dominated Convergence Theorem,

$$\frac{\partial F^{-1}f(t)}{\partial t_j} = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{it \cdot x} i x_j f(x) dx$$

$$= i(2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{it \cdot x \cdot e_j} f(x) dx.$$ 

Now $x^\alpha f(x) \in \mathcal{S}$ and so one can continue in this way and take derivatives indefinitely. Thus $F^{-1}f \in C^\infty(\mathbb{R}^n)$ and from the above argument,

$$D^\alpha F^{-1}f(t) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{it \cdot (ix)} f(x) dx.$$ 

To complete showing $F^{-1}f \in \mathcal{S}$,

$$t^\beta D^\alpha F^{-1}f(t) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{it \cdot x \cdot t^\beta} f(x) dx.$$
Integrate this integral by parts to get
\[ \int e^{it_jx_j}t_j^{\beta_j}f(x)dx_j = \frac{e^{it_jx_j}t_j^{\beta_j}f(x)}{it_j} \bigg|_0^\infty + \int e^{it_jx_j}t_j^{\beta_j-1}D^{\beta_j}((ix)^nf(x))dx_j \]
where the boundary term vanishes because \( f \in \mathcal{S} \). Returning to Theorem 3.3.22 use the fact that \( |e^{ia}| = 1 \) to conclude
\[ |t^{\beta}D^{\alpha}F^{-1}f(t)| \leq C \int R^n |D^{\beta}((ix)^nf(x))|dx < \infty. \]

It follows \( F^{-1}f \in \mathcal{S} \). Similarly \( Ff \in \mathcal{S} \) whenever \( f \in \mathcal{S} \).

Of course \( \mathcal{S} \) can be considered a subset of \( \mathcal{G}^* \) as follows. For \( \psi \in \mathcal{S} \),
\[ \psi(\phi) \equiv \int R^n \psi\phi dx \]

**Theorem 3.3.22** Let \( \psi \in \mathcal{S} \). Then \( (F \circ F^{-1})(\psi) = \psi \) and \( (F^{-1} \circ F)(\psi) = \psi \) whenever \( \psi \in \mathcal{S} \). Also \( F \) and \( F^{-1} \) map \( \mathcal{S} \) one to one and onto \( \mathcal{S} \).

**Proof:** The first claim follows from the fact that \( F \) and \( F^{-1} \) are inverses of each other on \( \mathcal{G}^* \) which was established above. For the second, let \( \psi \in \mathcal{S} \). Then \( \psi = F(F^{-1}\psi) \). Thus \( F \) maps \( \mathcal{S} \) onto \( \mathcal{S} \). If \( F\psi = 0 \), then do \( F^{-1} \) to both sides to conclude \( \psi = 0 \). Thus \( F \) is one to one and onto. Similarly, \( F^{-1} \) is one to one and onto.

### 3.3.5 Convolution

To begin with it is necessary to discuss the meaning of \( \phi f \) where \( f \in \mathcal{G}^* \) and \( \phi \in \mathcal{G} \). What should it mean? First suppose \( f \in L^p(\mathbb{R}^n) \) or measurable with polynomial growth. Then \( \phi f \) also has these properties. Hence, it should be the case that \( \phi f \equiv \int R^n \phi f dx = \int R^n f(\phi x) dx \). This motivates the following definition.

**Definition 3.3.23** Let \( T \in \mathcal{G}^* \) and let \( \phi \in \mathcal{G} \). Then \( \phi T \equiv T\phi \in \mathcal{G}^* \) will be defined by
\[ \phi T(\psi) \equiv T(\phi \psi). \]

The next topic is that of convolution. It was just shown that
\[ F(f \ast \phi) = (2\pi)^{n/2} F\phi Ff, \quad F^{-1}(f \ast \phi) = (2\pi)^{n/2} F^{-1}\phi F^{-1}f \]
whenever \( f \in L^2(\mathbb{R}^n) \) and \( \phi \in \mathcal{G} \) so the same definition is retained in the general case because it makes perfect sense and agrees with the earlier definition.

**Definition 3.3.24** Let \( f \in \mathcal{G}^* \) and let \( \phi \in \mathcal{G} \). Then define the convolution of \( f \) with an element of \( \mathcal{G} \) as follows.
\[ f \ast \phi \equiv (2\pi)^{n/2} F^{-1}(F\phi Ff) \in \mathcal{G}^* \]

There is an obvious question. With this definition, is it true that \( F^{-1}(f \ast \phi) = (2\pi)^{n/2} F^{-1}\phi F^{-1}f \) as it was earlier?

**Theorem 3.3.25** Let \( f \in \mathcal{G}^* \) and let \( \phi \in \mathcal{G} \).
\[ F(f \ast \phi) = (2\pi)^{n/2} F\phi Ff, \]
\[ F^{-1}(f \ast \phi) = (2\pi)^{n/2} F^{-1}\phi F^{-1}f. \]
Proof: Note that \(3.3.16\) follows from Definition \(3.3.24\) and both assertions hold for \(f \in \mathcal{G}\). Consider \(3.3.17\). Here is a simple formula involving a pair of functions in \(\mathcal{G}\).

\[
(\psi * F^{-1}F^{-1}\phi)(x) = \left( \int \int \int \psi(x-y)e^{iy \cdot \lambda} e^{iy \cdot z} \phi(z) dz dy \lambda \right) (2\pi)^n
\]

= \( (\psi * FF\phi)(x) \).

Now for \(\psi \in \mathcal{G}\),

\[
(2\pi)^{n/2} F \left( F^{-1} \phi F^{-1} f \right)(\psi) \equiv (2\pi)^{n/2} \left( F^{-1} \phi F^{-1} f \right) \left( F \psi \right) \equiv (2\pi)^n F^{-1} f \left( F^{-1} \phi F \psi \right) \equiv f \left( (2\pi)^{n/2} F^{-1} \left( F \left( F^{-1} \phi F \psi \right) \right) \right) \equiv f \left( (2\pi)^{n/2} F^{-1} \left( F \left( F^{-1} \phi \right) \right) \right) \equiv f \left( (2\pi)^{n/2} F^{-1} \left( F \left( F^{-1} \phi \right) \right) \right) \equiv f \left( (2\pi)^{n/2} F^{-1} \left( F \left( F^{-1} \phi \right) \right) \right) \equiv f \left( (2\pi)^{n/2} F^{-1} \left( F \left( F^{-1} \phi \right) \right) \right) \equiv f \left( (2\pi)^{n/2} F^{-1} \left( F \left( F^{-1} \phi \right) \right) \right) \equiv f(\psi) \equiv \phi \equiv f(\psi) \equiv \phi = f(\psi) \equiv \phi.
\]

Also

\[
(2\pi)^{n/2} F F \left( F^{-1} \phi F \psi \right) \equiv (2\pi)^{n/2} \left( F \phi F \psi \right) \equiv \left( F \phi F \psi \right) \equiv f \left( F \left( (2\pi)^{n/2} \left( F \phi F \psi \right) \right) \right) \equiv f \left( F \left( (2\pi)^{n/2} \left( F \phi F \psi \right) \right) \right) \equiv f \left( F \left( (2\pi)^{n/2} \left( F \phi F \psi \right) \right) \right) \equiv f \left( F \left( (2\pi)^{n/2} \left( F \phi F \psi \right) \right) \right) \equiv f \left( F \left( (2\pi)^{n/2} \left( F \phi F \psi \right) \right) \right) \equiv f \left( F \left( (2\pi)^{n/2} \left( F \phi F \psi \right) \right) \right) \equiv f \left( F \left( (2\pi)^{n/2} \left( F \phi F \psi \right) \right) \right) \equiv f(\psi) \equiv \phi \equiv f(\psi) \equiv \phi.
\]

The last line follows from the following.

\[
\int F F \phi (x-y) \psi (y) dy = \int F \phi (x-y) F \psi (y) dy = \int F \psi (x-y) F \phi (y) dy = \int \psi (x-y) F F \phi (y) dy.
\]

From \(3.3.16\) and \(3.3.18\), since \(\psi\) was arbitrary,

\[
(2\pi)^{n/2} F \left( F^{-1} \phi F^{-1} f \right) = (2\pi)^{n/2} F^{-1} \left( F \phi F f \right) = f \phi
\]

which shows \(3.3.17\).
Chapter 4

Banach Spaces

4.1 Theorems Based On Baire Category

4.1.1 Baire Category Theorem

Some examples of Banach spaces that have been discussed up to now are \( \mathbb{R}^n \), \( \mathbb{C}^n \), and \( L^p(\Omega) \). Theorems about general Banach spaces are proved in this chapter. The main theorems to be presented here are the uniform boundedness theorem, the open mapping theorem, the closed graph theorem, and the Hahn Banach Theorem. The first three of these theorems come from the Baire category theorem which is about to be presented. They are topological in nature. The Hahn Banach theorem has nothing to do with topology. Banach spaces are all normed linear spaces and as such, they are all metric spaces because a normed linear space may be considered as a metric space with \( d(x, y) \equiv ||x - y|| \). You can check that this satisfies all the axioms of a metric. As usual, if every Cauchy sequence converges, the metric space is called complete.

Definition 4.1.1 A complete normed linear space is called a Banach space.

The following remarkable result is called the Baire category theorem. To get an idea of its meaning, imagine you draw a line in the plane. The complement of this line is an open set and is dense because every point, even those on the line, are limit points of this open set. Now draw another line. The complement of the two lines is still open and dense. Keep drawing lines and looking at the complements of the union of these lines. You always have an open set which is dense. Now what if there were countably many lines? The Baire category theorem implies the complement of the union of these lines is dense. In particular it is nonempty. Thus you cannot write the plane as a countable union of lines. This is a rather rough description of this very important theorem. The precise statement and proof follow.

Theorem 4.1.2 Let \((X, d)\) be a complete metric space and let \( \{U_n\}_{n=1}^{\infty} \) be a sequence of open subsets of \( X \) satisfying \( \bigcap_{n=1}^{\infty} U_n = X \) (\( U_n \) is dense). Then \( D \equiv \bigcap_{n=1}^{\infty} U_n \) is a dense subset of \( X \).

Proof: Let \( p \in X \) and let \( r_0 > 0 \). I need to show \( D \cap B(p, r_0) \neq \emptyset \). Since \( U_1 \) is dense, there exists \( p_1 \in U_1 \cap B(p, r_0) \), an open set. Let \( p_1 \in B(p_1, r_1) \subseteq \overline{B(p_1, r_1)} \subseteq U_1 \cap B(p, r_0) \) and \( r_1 < 2^{-1} \). This is possible because \( U_1 \cap B(p, r_0) \) is an open set and so there exists \( r_1 \) such that \( B(p_1, 2r_1) \subseteq U_1 \cap B(p, r_0) \). But

\[
B(p_1, r_1) \subseteq \overline{B(p_1, r_1)} \subseteq B(p_1, 2r_1)
\]

because \( \overline{B(p_1, r_1)} = \{ x \in X : d(x, p) \leq r_1 \} \). (Why?)

There exists \( p_2 \in U_2 \cap B(p_1, r_1) \) because \( U_2 \) is dense. Let

\[
p_2 \in B(p_2, r_2) \subseteq \overline{B(p_2, r_2)} \subseteq U_2 \cap B(p_1, r_1) \subseteq U_1 \cap U_2 \cap B(p, r_0).
\]
and let \( r_2 < 2^{-2} \). Continue in this way. Thus
\[
\lim_{n \to \infty} p_n = p_\infty.
\]
Since all but finitely many terms of \( \{ p_n \} \) are in \( \overline{B(p_m, r_m)} \), it follows that \( p_\infty \in \overline{B(p_m, r_m)} \) for each \( m \). Therefore,
\[
p_\infty \in \bigcap_{m=1}^{\infty} \overline{B(p_m, r_m)} \subseteq \bigcap_{i=1}^{\infty} U_i \cap B(p, r).\]
This proves the theorem.

The following corollary is also called the Baire category theorem.

**Corollary 4.1.3** Let \( X \) be a complete metric space and suppose \( X = \cup_{i=1}^{\infty} F_i \) where each \( F_i \) is a closed set. Then for some \( i \), interior \( F_i \neq \emptyset \).

**Proof:** If all \( F_i \) has empty interior, then \( F_i^C \) would be a dense open set. Therefore, from Theorem 4.1.2 it would follow that
\[
\emptyset = (\cup_{i=1}^{\infty} F_i)^C = \cap_{i=1}^{\infty} F_i^C \neq \emptyset.
\]

The set \( D \) of Theorem 4.1.2 is called a \( G_\delta \) set because it is the countable intersection of open sets. Thus \( D \) is a dense \( G_\delta \) set.

Recall that a norm satisfies:

a.) \( \|x\| \geq 0, \|x\| = 0 \) if and only if \( x = 0 \).

b.) \( \|x + y\| \leq \|x\| + \|y\| \).

c.) \( \|cx\| = |c| \|x\| \) if \( c \) is a scalar and \( x \in X \).

From the definition of continuity, it follows easily that a function is continuous if
\[
\lim_{n \to \infty} x_n = x
\]
implies
\[
\lim_{n \to \infty} f(x_n) = f(x).
\]

**Theorem 4.1.4** Let \( X \) and \( Y \) be two normed linear spaces and let \( L : X \to Y \) be linear \((L(ax + by) = aL(x) + bL(y) \) for \( a, b \) scalars and \( x, y \in X \)). The following are equivalent

a.) \( L \) is continuous at 0

b.) \( L \) is continuous

c.) There exists \( K > 0 \) such that \( \|Lx\|_Y \leq K \|x\|_X \) for all \( x \in X \) (\( L \) is bounded).

**Proof:** a.)\( \Rightarrow \)b.) Let \( x_n \to x \). It is necessary to show that \( Lx_n \to Lx \). But \( (x_n - x) \to 0 \) and so from continuity at 0, it follows
\[
L(x_n - x) = Lx_n - Lx \to 0
\]
so \( Lx_n \to Lx \). This shows a.) implies b.).

b.)\( \Rightarrow \)c.) Since \( L \) is continuous, \( L \) is continuous at 0. Hence \( \|Lx\|_Y < 1 \) whenever \( \|x\|_X \leq \delta \) for some \( \delta \). Therefore, suppressing the subscript on the \( \| \mid \mid \),
\[
\|L\left(\frac{\delta x}{\|x\|}\right)\| \leq 1.
\]
Hence
\[
\|Lx\| \leq \frac{1}{\delta} \|x\|.
\]
c.)\( \Rightarrow \)a.) follows from the inequality given in c.)
4.1. THEOREMS BASED ON BAIRE CATEGORY

Definition 4.1.5 Let \( L : X \to Y \) be linear and continuous where \( X \) and \( Y \) are normed linear spaces. Denote the set of all such continuous linear maps by \( \mathcal{L}(X,Y) \) and define

\[
|L| = \sup\{|Lx| : ||x|| \leq 1\}.
\]  

(4.1.1)

This is called the operator norm.

Note that from Theorem 4.1.3 \(|L|\) is well defined because of part c.) of that Theorem.

The next lemma follows immediately from the definition of the norm and the assumption that \( L \) is linear.

Lemma 4.1.6 With \(|L|\) defined in (4.1.1), \( \mathcal{L}(X,Y) \) is a normed linear space. Also \(|L||x| \leq |L||x|\).

Proof: Let \( x \neq 0 \) then \( x/|x| \) has norm equal to 1 and so

\[
\left| L \left( \frac{x}{|x|} \right) \right| \leq |L|.
\]

Therefore, multiplying both sides by \(|x|\), \(|Lx| \leq |L||x|\). This is obviously a linear space. It remains to verify the operator norm really is a norm. First of all, if \(|L| = 0\), then \( Lx = 0 \) for all \(|x| \leq 1\). It follows that for any \( x \neq 0, 0 = L \left( \frac{x}{|x|} \right) \) and so \( Lx = 0 \). Therefore, \( L = 0 \). Also, if \( c \) is a scalar,

\[
|cL| = \sup_{||x|| \leq 1} ||cL(x)|| = |c| \sup_{||x|| \leq 1} ||Lx|| = |c| ||L||.
\]

It remains to verify the triangle inequality. Let \( L, M \in \mathcal{L}(X,Y) \).

\[
||L + M|| = \sup_{||x|| \leq 1} ||(L + M)(x)|| \leq \sup_{||x|| \leq 1} (||Lx|| + ||Mx||)
\]

\[
\leq \sup_{||x|| \leq 1} ||Lx|| + \sup_{||x|| \leq 1} ||Mx|| = ||L|| + ||M||.
\]

This shows the operator norm is really a norm as hoped. This proves the lemma.

For example, consider the space of linear transformations defined on \( \mathbb{R}^n \) having values in \( \mathbb{R}^m \). The fact the transformation is linear automatically imparts continuity to it. You should give a proof of this fact. Recall that every such linear transformation can be realized in terms of matrix multiplication.

Thus, in finite dimensions the algebraic condition that an operator is linear is sufficient to imply the topological condition that the operator is continuous. The situation is not so simple in infinite dimensional spaces such as \( C(X;\mathbb{R}^n) \). This explains the imposition of the topological condition of continuity as a criterion for membership in \( \mathcal{L}(X,Y) \) in addition to the algebraic condition of linearity.

Theorem 4.1.7 If \( Y \) is a Banach space, then \( \mathcal{L}(X,Y) \) is also a Banach space.

Proof: Let \( \{L_n\} \) be a Cauchy sequence in \( \mathcal{L}(X,Y) \) and let \( x \in X \).

\[
||L_n x - L_m x|| \leq ||x|| ||L_n - L_m||.
\]

Thus \( \{L_n x\} \) is a Cauchy sequence. Let

\[
Lx = \lim_{n \to \infty} L_n x.
\]

Then, clearly, \( L \) is linear because if \( x_1, x_2 \) are in \( X \), and \( a, b \) are scalars, then

\[
L(ax_1 + bx_2) = \lim_{n \to \infty} L_n (ax_1 + bx_2)
\]

\[
= \lim_{n \to \infty} (aL_n x_1 + bL_n x_2)
\]

\[
= aLx_1 + bLx_2.
\]

Also \( L \) is continuous. To see this, note that \( \{||L_n||\} \) is a Cauchy sequence of real numbers because

\[
|||L_n|| - ||L_m||| \leq ||L_n - L_m||.
\]

Hence there exists \( K > \sup\{||L_n|| : n \in \mathbb{N}\} \). Thus, if \( x \in X \),

\[
||Lx|| = \lim_{n \to \infty} ||L_n x|| \leq K||x||.
\]

This proves the theorem.
4.1.2 Uniform Boundedness Theorem

The next big result is sometimes called the Uniform Boundedness theorem, or the Banach-Steinhaus theorem. This is a very surprising theorem which implies that for a collection of bounded linear operators, if they are bounded pointwise, then they are also bounded uniformly. As an example of a situation in which pointwise bounded does not imply uniformly bounded, consider the functions \( f_\alpha(x) \equiv X(\alpha,1)(x)x^{-1} \) for \( \alpha \in (0,1) \). Clearly each function is bounded and the collection of functions is bounded at each point of \((0,1)\), but there is no bound for all these functions taken together. One problem is that \((0,1)\) is not a Banach space. Therefore, the functions cannot be linear.

**Theorem 4.1.8** Let \( X \) be a Banach space and let \( Y \) be a normed linear space. Let \( \{L_\alpha\}_{\alpha \in \Lambda} \) be a collection of elements of \( \mathcal{L}(X,Y) \). Then one of the following happens.

a.) \( \sup\{|\|L_\alpha\|| : \alpha \in \Lambda\} < \infty \)

b.) There exists a dense \( G_\delta \) set, \( D \), such that for all \( x \in D \),

\[
\sup\{|\|L_\alpha x\|| : \alpha \in \Lambda\} = \infty.
\]

**Proof:** For each \( n \in \mathbb{N} \), define

\[
U_n = \{x \in X : \sup\{|\|L_\alpha x\|| : \alpha \in \Lambda\} > n\}.
\]

Then \( U_n \) is an open set because if \( x \in U_n \), then there exists \( \alpha \in \Lambda \) such that

\[
|\|L_\alpha x\|| > n
\]

But then, since \( L_\alpha \) is continuous, this situation persists for all \( y \) sufficiently close to \( x \), say for all \( y \in B(x,\delta) \). Then \( B(x,\delta) \subseteq U_n \) which shows \( U_n \) is open.

Case b.) is obtained from Theorem 4.1.3 if each \( U_n \) is dense.

The other case is that for some \( n \), \( U_n \) is not dense. If this occurs, there exists \( x_0 \) and \( r > 0 \) such that for all \( x \in B(x_0,r) \), \( |\|L_\alpha x\|| \leq n \) for all \( \alpha \). Now if \( y \in B(0,r) \), \( x_0 + y \in B(x_0,r) \). Consequently, for all such \( y \), \( |\|L_\alpha(x_0 + y)\|| \leq n \). This implies that for all \( \alpha \in \Lambda \) and \( |\|y\|| < r \),

\[
|\|L_\alpha y\|| \leq n + |\|L_\alpha(x_0)\|| \leq 2n.
\]

Therefore, if \( |\|y\|| \leq 1 \), \( |\|\frac{r}{2}y\|| < r \) and so for all \( \alpha \),

\[
|\|L_\alpha \left( \frac{r}{2}y \right)\|| \leq 2n.
\]

Now multiplying by \( r/2 \) it follows that whenever \( |\|y\|| \leq 1 \), \( |\|L_\alpha(y)\|| \leq 4n/r \). Hence case a.) holds.

4.1.3 Open Mapping Theorem

Another remarkable theorem which depends on the Baire category theorem is the open mapping theorem. Unlike Theorem 4.1.3, it requires both \( X \) and \( Y \) to be Banach spaces.

**Theorem 4.1.9** Let \( X \) and \( Y \) be Banach spaces, let \( L \in \mathcal{L}(X,Y) \), and suppose \( L \) is onto. Then \( L \) maps open sets onto open sets.

To aid in the proof, here is a lemma.

**Lemma 4.1.10** Let \( a \) and \( b \) be positive constants and suppose

\[
B(0,a) \subseteq \overline{L(B(0,b))}.
\]

Then

\[
\overline{L(B(0,b))} \subseteq L(B(0,2b)).
\]

**Proof of Lemma:** Let \( y \in \overline{L(B(0,b))} \). There exists \( x_1 \in B(0,b) \) such that \( |\|y - Lx_1\|| < \frac{a}{2} \). Now this implies

\[
2y - 2Lx_1 \in B(0,a) \subseteq \overline{L(B(0,b))}.
\]
Thus $2y - 2Lx_1 \in L(B(0,b))$ just like $y$ was. Therefore, there exists $x_2 \in B(0,b)$ such that $\|2y - 2Lx_1 - Lx_2\| < a/2$. Hence $\|4y - 4Lx_1 - 2Lx_2\| < a$, and there exists $x_3 \in B(0,b)$ such that $\|4y - 4Lx_1 - 2Lx_2 - Lx_3\| < a/2$. Continuing in this way, there exist $x_1, x_2, x_3, x_4, \ldots$ in $B(0,b)$ such that

$$\|2^ny - \sum_{i=1}^{n} 2^{n-(i-1)}L(x_i)\| < a$$

which implies

$$\|y - \sum_{i=1}^{n} 2^{-(i-1)}L(x_i)\| = \|y - L\left(\sum_{i=1}^{n} 2^{-(i-1)}(x_i)\right)\| < 2^{-n}a \tag{4.1.2}$$

Now consider the partial sums of the series, $\sum_{i=1}^{\infty} 2^{-(i-1)}x_i$.

$$\|\sum_{i=m}^{n} 2^{-(i-1)}x_i\| \leq b \sum_{i=m}^{\infty} 2^{-(i-1)} = b 2^{-m+2}.$$ 

Therefore, these partial sums form a Cauchy sequence and so since $X$ is complete, there exists $x = \sum_{i=1}^{\infty} 2^{-(i-1)}x_i$. Letting $n \to \infty$ in (4.1.2) yields $\|y - Lx\| = 0$. Now

$$\|x\| = \lim_{n \to \infty} \|\sum_{i=1}^{n} 2^{-(i-1)}x_i\|$$

$$\leq \lim_{n \to \infty} \sum_{i=1}^{n} 2^{-(i-1)}\|x_i\| < \lim_{n \to \infty} \sum_{i=1}^{n} 2^{-(i-1)}b = 2b.$$ 

This proves the lemma.

**Proof of Theorem 4.1.3.** Let $Y = \cup_{n=1}^{\infty} L(B(0,n))$. By Corollary 4.1.2, the set, $L(B(0,n))$ has nonempty interior for some $n_0$. Thus $B(y,r) \subseteq L(B(0,n))$ for some $y$ and some $r > 0$. Since $L$ is linear $B(-y,r) \subseteq L(B(0,n))$ also. Here is why. If $z \in B(-y,r)$, then $-z \in B(y,r)$ and so there exists $x_n \in B(0,n)$ such that $Lx_n \to z$. Therefore, $L(-x_n) \to z$ and $-x_n \in B(0,n)$ also. Therefore $z \in L(B(0,n))$. Then it follows that

$$B(0,r) \subseteq B(y,r) + B(-y,r)$$

$$\equiv \{ y_1 + y_2 : y_1 \in B(y,r) \text{ and } y_2 \in B(-y,r) \}$$

$$\subseteq L(B(0,2n_0))$$

The reason for the last inclusion is that from the above, if $y_1 \in B(y,r)$ and $y_2 \in B(-y,r)$, there exists $x_n, z_n \in B(0,n_0)$ such that

$$Lx_n \to y_1, \quad Lz_n \to y_2.$$ 

Therefore,

$$\|x_n + z_n\| \leq 2n_0$$

and so $(y_1 + y_2) \in L(B(0,2n_0))$.

By Lemma 4.1.2 \(L(B(0,2n_0)) \subseteq L(B(0,4n_0))\) which shows

$$B(0,r) \subseteq L(B(0,4n_0)).$$

Letting $a = r(4n_0)^{-1}$, it follows, since $L$ is linear, that $B(0,a) \subseteq L(B(0,1))$. It follows since $L$ is linear,

$$L(B(0,r)) \supseteq B(0,ar). \tag{4.1.3}$$

Now let $U$ be open in $X$ and let $x + B(0,r) = B(x,r) \subseteq U$. Using 4.1.2 \(L(U) \supseteq L(x + B(0,r))\)

$$= Lx + L(B(0,r)) \supseteq Lx + B(0,ar) = B(Lx,ar).$$

Hence

$$Lx \in B(Lx,ar) \subseteq L(U).$$

which shows that every point, $Lx \in LU$, is an interior point of $LU$ and so $LU$ is open. This proves the theorem.

This theorem is surprising because it implies that if $|\cdot|$ and $||\cdot||$ are two norms with respect to which a vector space $X$ is a Banach space such that $|\cdot| \leq K ||\cdot||$, then there exists a constant $k$, such that $||\cdot|| \leq k |\cdot|$. This can be useful because sometimes it is not clear how to compute $k$ when all that is needed is its existence. To see the open mapping theorem implies this, consider the identity map $id \colon x \to x$. Then $id : (X,||\cdot||) \to (X,|\cdot|)$ is continuous and onto. Hence $id$ is an open map which implies $id^{-1}$ is continuous. Theorem 4.1.2 gives the existence of the constant $k$. 

---

4.1. THEOREMS BASED ON BAIRE CATEGORY

57
4.1.4 Closed Graph Theorem

Definition 4.1.11 Let \( f : D \to E \). The set of all ordered pairs of the form \( \{(x, f(x)) : x \in D\} \) is called the graph of \( f \).

Definition 4.1.12 If \( X \) and \( Y \) are normed linear spaces, make \( X \times Y \) into a normed linear space by using the norm \( ||(x, y)|| = \max(||x||, ||y||) \) along with component-wise addition and scalar multiplication. Thus \( a(x, y) + b(z, w) \equiv (ax+bz, ay+bw) \).

There are other ways to give a norm for \( X \times Y \). For example, you could define \( ||(x, y)|| = ||x|| + ||y|| \)

Lemma 4.1.13 The norm defined in Definition 4.1.12 on \( X \times Y \) along with the definition of addition and scalar multiplication given there make \( X \times Y \) into a normed linear space.

Proof: The only axiom for a norm which is not obvious is the triangle inequality. Therefore, consider

\[
|| (x_1, y_1) + (x_2, y_2) || = || (x_1 + x_2, y_1 + y_2) ||
\]

\[
= \max (||x_1 + x_2||, ||y_1 + y_2||)
\]

\[
\leq \max (||x_1|| + ||x_2||, ||y_1|| + ||y_2||)
\]

\[
\leq \max (||x_1||, ||y_1||) + \max (||x_2||, ||y_2||)
\]

It is obvious \( X \times Y \) is a vector space from the above definition. This proves the lemma.

Lemma 4.1.14 If \( X \) and \( Y \) are Banach spaces, then \( X \times Y \) with the norm and vector space operations defined in Definition 4.1.12 is also a Banach space.

Proof: The only thing left to check is that the space is complete. But this follows from the simple observation that \( \{(x_n, y_n)\} \) is a Cauchy sequence in \( X \times Y \) if and only if \( \{x_n\} \) and \( \{y_n\} \) are Cauchy sequences in \( X \) and \( Y \) respectively. Thus if \( \{(x_n, y_n)\} \) is a Cauchy sequence in \( X \times Y \), it follows there exist \( x \) and \( y \) such that \( x_n \to x \) and \( y_n \to y \). But then from the definition of the norm, \( (x_n, y_n) \to (x, y) \).

Lemma 4.1.15 Every closed subspace of a Banach space is a Banach space.

Proof: If \( F \subseteq X \) where \( X \) is a Banach space and \( \{x_n\} \) is a Cauchy sequence in \( F \), then since \( X \) is complete, there exists a unique \( x \in X \) such that \( x_n \to x \). However this means \( x \in F = F \) since \( F \) is closed.

Definition 4.1.16 Let \( X \) and \( Y \) be Banach spaces and let \( D \subseteq X \) be a subspace. A linear map \( L : D \to Y \) is said to be closed if its graph is a closed subspace of \( X \times Y \). Equivalently, \( L \) is closed if \( x_n \to x \) and \( Lx_n \to y \) implies \( x \in D \) and \( y = Lx \).

Note the distinction between closed and continuous. If the operator is closed the assertion that \( y = Lx \) only follows if it is known that the sequence \( \{Lx_n\} \) converges. In the case of a continuous operator, the convergence of \( \{Lx_n\} \) follows from the assumption that \( x_n \to x \). It is not always the case that a mapping which is closed is necessarily continuous. Consider the function \( f(x) = \tan(x) \) if \( x \) is not an odd multiple of \( \frac{\pi}{2} \) and \( f(x) \equiv 0 \) at every odd multiple of \( \frac{\pi}{2} \). Then the graph is closed and the function is defined on \( \mathbb{R} \) but it clearly fails to be continuous. Of course this function is not linear. You could also consider the map,

\[
\frac{d}{dx} : \{y \in C^1([0,1]) : y(0) = 0\} \equiv D \to C([0,1]).
\]

where the norm is the uniform norm on \( C([0,1]) \), \( ||y||_\infty \). If \( y \in D \), then

\[
y(x) = \int_0^x y'(t) \, dt.
\]

Therefore, if \( \frac{dy_n}{dx} \to f \in C([0,1]) \) and if \( y_n \to y \) in \( C([0,1]) \) it follows that

\[
y_n(x) = \int_0^x \frac{dy_n(t)}{dx} \, dt \quad \downarrow \quad y(x) = \int_0^x f(t) \, dt
\]
and so by the fundamental theorem of calculus \( f(x) = y'(x) \) and so the mapping is closed. It is obviously not continuous because it takes \( y(x) \) and \( y(x) + \frac{1}{n} \sin(nx) \) to two functions which are far from each other even though these two functions are very close in \( C([0,1]) \). Furthermore, it is not defined on the whole space, \( C([0,1]) \).

The next theorem, the closed graph theorem, gives conditions under which closed implies continuous.

**Theorem 4.1.17** Let \( X \) and \( Y \) be Banach spaces and suppose \( L : X \to Y \) is closed and linear. Then \( L \) is continuous.

**Proof:** Let \( G \) be the graph of \( L \). \( G = \{(x, Lx) : x \in X\} \). By Lemma 4.1.11 it follows that \( G \) is a Banach space. Define \( P : G \to X \) by \( P(x, Lx) = x \). \( P \) maps the Banach space \( G \) onto the Banach space \( X \) and is continuous and linear. By the open mapping theorem, \( P \) maps open sets onto open sets. Since \( P \) is also one to one, this says that \( P^{-1} \) is continuous. Thus \( ||P^{-1}x|| \leq K||x|| \). Hence

\[
||Lx|| \leq \max(||x||, ||Lx||) \leq K||x||
\]

By Theorem 4.1.14 on Page 51, this shows \( L \) is continuous and proves the theorem.

The following corollary is quite useful. It shows how to obtain a new norm on the domain of a closed operator such that the domain with this new norm becomes a Banach space.

**Corollary 4.1.18** Let \( L : D \subseteq X \to Y \) where \( X, Y \) are a Banach spaces, and \( L \) is a closed operator. Then define a new norm on \( D \) by

\[
||x||_D \equiv ||x||_X + ||Lx||_Y .
\]

Then \( D \) with this new norm is a Banach space.

**Proof:** If \( \{x_n\} \) is a Cauchy sequence in \( D \) with this new norm, it follows both \( \{x_n\} \) and \( \{Lx_n\} \) are Cauchy sequences and therefore, they converge. Since \( L \) is closed, \( x_n \to x \) and \( Lx_n \to Lx \) for some \( x \in D \). Thus \( ||x_n - x||_D \to 0 \).

### 4.2 Hahn Banach Theorem

The closed graph, open mapping, and uniform boundedness theorems are the three major topological theorems in functional analysis. The other major theorem is the Hahn-Banach theorem which has nothing to do with topology. Before presenting this theorem, here are some preliminaries about partially ordered sets.

**Definition 4.2.1** Let \( \mathcal{F} \) be a nonempty set. \( \mathcal{F} \) is called a partially ordered set if there is a relation, denoted here by \( \leq \), such that

- \( x \leq x \) for all \( x \in \mathcal{F} \).
- If \( x \leq y \) and \( y \leq z \) then \( x \leq z \).

\( \mathcal{C} \subseteq \mathcal{F} \) is said to be a chain if every two elements of \( \mathcal{C} \) are related. This means that if \( x, y \in \mathcal{C} \), then either \( x \leq y \) or \( y \leq x \). Sometimes a chain is called a totally ordered set. \( \mathcal{C} \) is said to be a maximal chain if whenever \( \mathcal{D} \) is a chain containing \( \mathcal{C} \), \( \mathcal{D} = \mathcal{C} \).

The most common example of a partially ordered set is the power set of a given set with \( \subseteq \) being the relation. It is also helpful to visualize partially ordered sets as trees. Two points on the tree are related if they are on the same branch of the tree and one is higher than the other. Thus two points on different branches would not be related although they might both be larger than some point on the trunk. You might think of many other things which are best considered as partially ordered sets. Think of food for example. You might find it difficult to determine which of two favorite pies you like better although you may be able to say very easily that you would prefer either pie to a dish of lard topped with whipped cream and mustard. The following theorem is equivalent to the axiom of choice. For a discussion of this, see the appendix on the subject.

**Theorem 4.2.2** (Hausdorff Maximal Principle) Let \( \mathcal{F} \) be a nonempty partially ordered set. Then there exists a maximal chain.

**Definition 4.2.3** Let \( X \) be a real vector space \( \rho : X \to \mathbb{R} \) is called a gauge function if

\[
\rho(x + y) \leq \rho(x) + \rho(y),
\]

\[
\rho(ax) = a\rho(x) \text{ if } a \geq 0.
\] (4.2.4)
Suppose \( M \) is a subspace of \( X \) and \( z \notin M \). Suppose also that \( f \) is a linear real-valued function having the property that \( f(x) \leq \rho(x) \) for all \( x \in M \). Consider the problem of extending \( f \) to \( M \oplus \mathbb{R}z \) such that if \( F \) is the extended function, \( F(y) \leq \rho(y) \) for all \( y \in M \oplus \mathbb{R}z \) and \( F \) is linear. Since \( F \) is to be linear, it suffices to determine how to define \( F(z) \). Letting \( a > 0 \), it is required to define \( F(z) \) such that the following hold for all \( x, y \in M \).

\[
\begin{align*}
\frac{f(x)}{F(x)} + aF(z) &= F(x + az) \leq \rho(x + az), \\
\frac{f(y)}{F(y)} - aF(z) &= F(y - az) \leq \rho(y - az).
\end{align*}
\]

(4.2.5)

Now if these inequalities hold for all \( y/a \), they hold for all \( y \) because \( M \) is given to be a subspace. Therefore, multiplying by \( a^{-1} \) implies that what is needed is to choose \( F(z) \) such that for all \( x, y \in M \),

\[
f(x) + F(z) \leq \rho(x + z), \quad f(y) - \rho(y - z) \leq F(z)
\]

and that if \( F(z) \) can be chosen in this way, this will satisfy \( f \) for all \( x, y \) and the problem of extending \( f \) will be solved. Hence it is necessary to choose \( F(z) \) such that for all \( x, y \in M \)

\[
f(y) - \rho(y - z) \leq F(z) \leq \rho(x + z) - f(x).
\]

(4.2.6)

Is there any such number between \( f(y) - \rho(y - z) \) and \( \rho(x + z) - f(x) \) for every pair \( x, y \in M \)? This is where \( f(x) \leq \rho(x) \) on \( M \) and that \( f \) is linear is used. For \( x, y \in M \),

\[
\begin{align*}
\rho(x + z) - f(x) - [f(y) - \rho(y - z)] & = \rho(x + z) + \rho(y - z) - (f(x) + f(y)) \\
& \geq \rho(x + y) - f(x + y) \geq 0.
\end{align*}
\]

Therefore there exists a number between

\[
\sup \{ f(y) - \rho(y - z) : y \in M \}
\]

and

\[
\inf \{ \rho(x + z) - f(x) : x \in M \}
\]

Choose \( F(z) \) to satisfy \( \text{4.2.6} \). This has proved the following lemma.

**Lemma 4.2.4** Let \( M \) be a subspace of \( X \), a real linear space, and let \( \rho \) be a gauge function on \( X \). Suppose \( f : M \to \mathbb{R} \) is linear, \( z \notin M \), and \( f(x) \leq \rho(x) \) for all \( x \in M \). Then \( f \) can be extended to \( M \oplus \mathbb{R}z \) such that, if \( F \) is the extended function, \( F \) is linear and \( F(x) \leq \rho(x) \) for all \( x \in M \oplus \mathbb{R}z \).

With this lemma, the Hahn Banach theorem can be proved.

**Theorem 4.2.5** (Hahn Banach theorem) Let \( X \) be a real vector space, let \( M \) be a subspace of \( X \), let \( f : M \to \mathbb{R} \) be linear, let \( \rho \) be a gauge function on \( X \), and suppose \( f(x) \leq \rho(x) \) for all \( x \in M \). Then there exists a linear function, \( F : X \to \mathbb{R} \), such that

a.) \( F(x) = f(x) \) for all \( x \in M \)

b.) \( F(x) \leq \rho(x) \) for all \( x \in X \).

**Proof:** Let \( \mathcal{F} = \{(V, g) : V \supseteq M, \, V \text{ is a subspace of } X, \, g : V \to \mathbb{R} \text{ is linear, } g(x) = f(x) \text{ for all } x \in M, \text{ and } g(x) \leq \rho(x) \text{ for } x \in V \} \). Then \( (M, f) \in \mathcal{F} \) so \( \mathcal{F} \neq \emptyset \). Define a partial order by the following rule.

\[
(V, g) \leq (W, h)
\]

means

\[
V \subseteq W \text{ and } h(x) = g(x) \text{ if } x \in V.
\]

By Theorem \( \text{4.2.2} \), there exists a maximal chain, \( \mathcal{C} \subseteq \mathcal{F} \). Let \( Y = \bigcup \{ V : (V, g) \in \mathcal{C} \} \) and let \( h : Y \to \mathbb{R} \) be defined by \( h(x) = g(x) \) where \( x \in V \) and \( (V, g) \in \mathcal{C} \). This is well defined because if \( x \in V_1 \) and \( V_2 \) where \( (V_1, g_1) \) and \( (V_2, g_2) \) are both in the chain, then since \( \mathcal{C} \) is a chain, the two element related. Therefore, \( g_1(x) = g_2(x) \). Also \( h \) is linear because if \( ax + by \in Y \), then \( x \in V_1 \) and \( y \in V_2 \) where \( (V_1, g_1) \) and \( (V_2, g_2) \) are elements of \( \mathcal{C} \). Therefore, letting \( V \)
4.2. HAHN BANACH THEOREM

denote the larger of the two \( V_i \), and \( g \) be the function that goes with \( V \), it follows \( ax + by \in V \) where \((V, g) \in C \). Therefore,

\[
    h (ax + by) = g (ax + by) = ag(x) + bg(y) = ah(x) + bh(y).
\]

Also, \( h(x) = g(x) \leq \rho(x) \) for any \( x \in Y \) because for such \( x \), \( x \in V \) where \((V, g) \in C \).

Is \( Y = X? \) If not, there exists \( z \in X \setminus Y \) and there exists an extension of \( h \) to \( Y \oplus \mathbb{R}z \) using Lemma \[\text{2.3.}\]

Letting \( \overline{h} \) denote this extended function, contradicts the maximality of \( C \). Indeed, \( C \cup \{(Y \oplus \mathbb{R}z, \overline{h})\} \) would be a longer chain. This proves the Hahn Banach theorem.

This is the original version of the theorem. There is also a version of this theorem for complex vector spaces which is based on a trick.

Corollary 4.2.6 (Hahn Banach) Let \( M \) be a subspace of a complex normed linear space, \( X \), and suppose \( f : M \to C \) is linear and satisfies \( |f(x)| \leq K||x|| \) for all \( x \in M \). Then there exists a linear function, \( F \), defined on all of \( X \) such that \( F(x) = f(x) \) for all \( x \in M \) and \( |F(x)| \leq K||x|| \) for all \( x \).

Proof: First note \( f(x) = \text{Re} f(x) + i \text{Im} f(x) \) and so

\[
    \text{Re} f(ix) + i \text{Im} f(ix) = f(ix) = if(x) = i \text{Re} f(x) - \text{Im} f(x).
\]

Therefore, \( \text{Im} f(x) = - \text{Re} f(ix) \), and

\[
    f(x) = \text{Re} f(x) - i \text{Re} f(ix).
\]

This is important because it shows it is only necessary to consider \( \text{Re} f \) in understanding \( f \). Now it happens that \( \text{Re} f \) is linear with respect to real scalars so the above version of the Hahn Banach theorem applies. This is shown next.

If \( c \) is a real scalar

\[
    \text{Re} f(cx) - i \text{Re} f(icx) = cf(x) = c \text{Re} f(x) - ic \text{Re} f(ix).
\]

Thus \( \text{Re} f(cx) = c \text{Re} f(x) \). Also,

\[
    \text{Re} f(x + y) - i \text{Re} f(i (x + y)) = f(x + y) = f(x) + f(y) = \text{Re} f(x) - i \text{Re} f(ix) + \text{Re} f(y) - i \text{Re} f(iy).
\]

Equating real parts, \( \text{Re} f(x + y) = \text{Re} f(x) + \text{Re} f(y) \). Thus \( \text{Re} f \) is linear with respect to real scalars as hoped.

Consider \( X \) as a real vector space and let \( \rho(x) \equiv K||x|| \). Then for all \( x \in M \),

\[
    |\text{Re} f(x)| \leq |f(x)| \leq K||x|| = \rho(x).
\]

From Theorem \[\text{1.2.2.}\], \( \text{Re} f \) may be extended to a function, \( h \) which satisfies

\[
    h(ax + by) = ah(x) + bh(y) \text{ if } a, b \in \mathbb{R} \quad h(x) \leq K||x|| \text{ for all } x \in X.
\]

Actually, \( |h(x)| \leq K||x|| \). The reason for this is that \( h(-x) = -h(x) \leq K||-x|| = K||x|| \) and therefore, \( h(x) \geq -K||x|| \). Let

\[
    F(x) \equiv h(x) - ih(ix).
\]

By arguments similar to the above, \( F \) is linear.

\[
    F(ix) = h(ix) - ih(-x) = ih(x) + h(ix) = i(h(x) - ih(ix)) = iF(x).
\]

If \( c \) is a real scalar,

\[
    F(cx) = h(cx) - ih(icx) = ch(x) - cih(ix) = cF(x).
\]
Now
\[ F(x + y) = h(x + y) - ih(i(x + y)) \]
\[ = h(x) + h(y) - ih(ix) - ih(iy) \]
\[ = F(x) + F(y). \]

Thus
\[ F((a + ib)x) = F(ax) + F(ibx) \]
\[ = aF(x) + ibF(x) \]
\[ = (a + ib)F(x). \]

This shows \( F \) is linear as claimed.

Now \( wF(x) = |F(x)| \) for some \( |w| = 1 \). Therefore
\[ |F(x)| = wF(x) = h(wx) - ih(ixw) = h(wx) \]
\[ = |h(wx)| ≤ K||wx|| = K ||x||. \]

This proves the corollary.

**Definition 4.2.7** Let \( X \) be a Banach space. Denote by \( X' \) the space of continuous linear functions which map \( X \) to the field of scalars. Thus \( X' = \mathcal{L}(X, \mathbb{F}) \). By Theorem 4.1.7 on Page 55, \( X' \) is a Banach space. Remember with the norm defined on \( \mathcal{L}(X, \mathbb{F}) \),
\[ ||f|| = \sup\{|f(x)| : ||x|| ≤ 1\} \]
\( X' \) is called the dual space.

**Definition 4.2.8** Let \( X \) and \( Y \) be Banach spaces and suppose \( L ∈ \mathcal{L}(X, Y) \). Then define the adjoint map in \( \mathcal{L}(Y', X') \), denoted by \( L^* \), by
\[ L^*y^*(x) = y^*(Lx) \]
for all \( y^* ∈ Y' \).

The following diagram is a good one to help remember this definition.

\[ \begin{array}{ccc}
X' & \xleftarrow{L^*} & Y' \\
X & \xrightarrow{L} & Y \\
\end{array} \]

This is a generalization of the adjoint of a linear transformation on an inner product space. Recall
\[ (Ax, y) = (x, A^*y) \]

What is being done here is to generalize this algebraic concept to arbitrary Banach spaces. There are some issues which need to be discussed relative to the above definition. First of all, it must be shown that \( L^*y^* ∈ X' \). Also, it will be useful to have the following lemma which is a useful application of the Hahn Banach theorem.

**Lemma 4.2.9** Let \( X \) be a normed linear space and let \( x ∈ X \). Then there exists \( x^* ∈ X' \) such that \( ||x^*|| = 1 \) and \( x^*(x) = ||x|| \).

**Proof:** Let \( f : \mathbb{F}x → \mathbb{F} \) be defined by \( f(αx) = α||x|| \). Then for \( y = αx ∈ \mathbb{F}x \),
\[ |f(y)| = |f(αx)| = |α||x|| = |y|. \]

By the Hahn Banach theorem, there exists \( x^* ∈ X' \) such that \( x^*(αx) = f(αx) \) and \( ||x^*|| ≤ 1 \). Since \( x^*(x) = ||x|| \) it follows that \( ||x^*|| = 1 \) because
\[ ||x^*|| ≥ \left| x^* \left( \frac{x}{||x||} \right) \right| = \frac{||x||}{||x||} = 1. \]

This proves the lemma.
4.2. HAHN BANACH THEOREM

Theorem 4.2.10 Let \( L \in \mathcal{L}(X,Y) \) where \( X \) and \( Y \) are Banach spaces. Then

a.) \( L^* \in \mathcal{L}(Y^*,X^*) \) as claimed and \( ||L^*|| = ||L|| \).

b.) If \( L \) maps one to one onto a closed subspace of \( Y \), then \( L^* \) is onto.

c.) If \( L \) maps onto a dense subset of \( Y \), then \( L^* \) is one to one.

Proof: It is routine to verify \( L^*y^* \) and \( L^* \) are both linear. This follows immediately from the definition. As usual, the interesting thing concerns continuity.

\[
||L^*y^*|| = \sup_{||x|| \leq 1} |L^*y^*(x)| = \sup_{||x|| \leq 1} |y^*(Lx)| \leq ||y^*|| ||L||.
\]

Thus \( L^* \) is continuous as claimed and \( ||L^*|| \leq ||L|| \).

By Lemma 4.2.9, there exists \( y^*_x \in Y^* \) such that \( ||y^*_x|| = 1 \) and \( y^*_x(Lx) = ||Lx|| \). Therefore,

\[
||L^*|| = \sup_{||y^*|| \leq 1} ||L^*y^*|| = \sup_{||y^*|| \leq 1} \sup_{||x|| \leq 1} |L^*y^*(x)|
\]

\[
= \sup_{||x|| \leq 1} \sup_{||x|| \leq 1} |y^*(Lx)| = \sup_{||x|| \leq 1} |y^*(Lx)| \geq \sup_{||x|| \leq 1} |y^*_x(Lx)| = \sup_{||x|| \leq 1} ||Lx|| = ||L||
\]

showing that \( ||L^*|| \geq ||L|| \) and this shows part a.).

If \( L \) is one to one and onto a closed subset of \( Y \), then \( L(X) \) being a closed subspace of a Banach space, is itself a Banach space and so the open mapping theorem implies \( L^{-1} : L(X) \to X \) is continuous. Hence

\[
||x|| = ||L^{-1}Lx|| \leq ||L^{-1}|| ||Lx||
\]

Now let \( x^* \in X^* \) be given. Define \( f \in \mathcal{L}(L(X),\mathbb{C}) \) by \( f(Lx) = x^*(x) \). The function, \( f \), is well defined because if \( Lx_1 = Lx_2 \), then since \( L \) is one to one, it follows \( x_1 = x_2 \) and so \( f(L(x_1)) = x^*(x_1) = x^*(x_2) = f(L(x_1)) \). Also, \( f \) is linear because

\[
f(aL(x_1) + bL(x_2)) = f(L(ax_1 + bx_2)) = x^*(ax_1 + bx_2) = ax^*(x_1) + bx^*(x_2) = a f(L(x_1)) + b f(L(x_2)).
\]

In addition to this,

\[
||f(Lx)|| = ||x^*(x)|| \leq ||x^*|| ||x|| \leq ||x^*|| ||L^{-1}|| ||Lx||
\]

and so the norm of \( f \) on \( L(X) \) is no larger than \( ||x^*|| ||L^{-1}|| \). By the Hahn Banach theorem, there exists an extension of \( f \) to an element \( y^* \in Y^* \) such that \( ||y^*|| \leq ||x^*|| ||L^{-1}|| \). Then

\[
L^*y^*(x) = y^*(Lx) = f(Lx) = x^*(x)
\]

so \( L^*y^* = x^* \) because this holds for all \( x \). Since \( x^* \) was arbitrary, this shows \( L^* \) is onto and proves b.).

Consider the last assertion. Suppose \( L^*y^* = 0 \). Is \( y^* = 0 \)? In other words is \( y^*(y) = 0 \) for all \( y \in Y \)? Pick \( y \in Y \). Since \( L(X) \) is dense in \( Y \), there exists a sequence, \( \{Lx_n\} \) such that \( Lx_n \to y \). But then by continuity of \( y^* \),

\[
y^*(y) = \lim_{n \to \infty} y^*(Lx_n) = \lim_{n \to \infty} L^*y^*(x_n) = 0.
\]

Since \( y^*(y) = 0 \) for all \( y \), this implies \( y^* = 0 \) and so \( L^* \) is one to one.

Corollary 4.2.11 Suppose \( X \) and \( Y \) are Banach spaces, \( L \in \mathcal{L}(X,Y) \), and \( L \) is one to one and onto. Then \( L^* \) is also one to one and onto.

There exists a natural mapping, called the James map from a normed linear space, \( X \), to the dual of the dual space which is described in the following definition.

Definition 4.2.12 Define \( J : X \to X'' \) by \( J(x)(x^*) = x^*(x) \).

Theorem 4.2.13 The map, \( J \), has the following properties.

a.) \( J \) is one to one and linear.

b.) \( ||Jx|| = ||x|| \) and \( ||J|| = 1 \).

c.) \( J(X) \) is a closed subspace of \( X'' \) if \( X \) is complete.

Also if \( x^* \in X' \),

\[
||x^*|| = \sup \{ ||x^{**} (x^*)|| : ||x^{**}|| \leq 1, \ x^{**} \in X'' \}.
\]
Proof:

\[ J(ax + by)(x^*) = x^*(ax + by) = ax^*(x) + bx^*(y) = (aJ(x) + bJ(y))(x^*). \]

Since this holds for all \( x^* \in X' \), it follows that

\[ J(ax + by) = aJ(x) + bJ(y) \]

and so \( J \) is linear. If \( Jx = 0 \), then by Lemma 4.2.9 there exists \( x^* \) such that \( x^*(x) = ||x|| \) and \( ||x^*|| = 1 \). Then \( 0 = J(x)(x^*) = x^*(x) = ||x|| \).

This shows a.\).

To show b.), let \( x \in X \) and use Lemma 4.2.9 to obtain \( x^* \in X' \) such that \( x^*(x) = ||x|| \) with \( ||x^*|| = 1 \). Then

\[ ||x|| \geq \sup \{||y^*(x)|| : ||y^*|| \leq 1\} = \sup \{||J(x)(y^*)|| : ||y^*|| \leq 1\} = ||Jx|| \]
\[ \geq |J(x)(x^*)| = |x^*(x)| = ||x|| \]

Therefore, \( ||Jx|| = ||x|| \) as claimed. Therefore,

\[ ||J|| = \sup \{||Jx|| : ||x|| \leq 1\} = \sup \{||x|| : ||x|| \leq 1\} = 1. \]

This shows b.\).

To verify c.), use b.). If \( Jx_n \to y^{**} \in X'' \) then by b.), \( x_n \) is a Cauchy sequence converging to some \( x \in X \) because

\[ ||x_n - x_m|| = ||Jx_n - Jx_m|| \]

and \( \{Jx_n\} \) is a Cauchy sequence. Then \( Jx = \lim_{n \to \infty} Jx_n = y^{**} \).

Finally, to show the assertion about the norm of \( x^* \), use what was just shown applied to the James map from \( X' \) to \( X'' \) still referred to as \( J \).

\[ ||x^*|| = \sup \{|x^*(x)| : ||x|| \leq 1\} = \sup \{|J(x)(x^*)| : ||x|| \leq 1\} \]
\[ \leq \sup \{|x^{**}(x^*)| : ||x^*|| \leq 1\} = \sup \{|J(x^*)(x^{**})| : ||x^{**}|| \leq 1\} \]
\[ = ||Jx^*|| = ||x^*||. \]

This proves the theorem.

**Definition 4.2.14** When \( J \) maps \( X \) onto \( X'' \), \( X \) is called reflexive.

It happens the \( L^p \) spaces are reflexive whenever \( p > 1 \).

### 4.3 Weak And Weak ∗ Topologies

#### 4.3.1 Basic Definitions

Let \( X \) be a Banach space and let \( X' \) be its dual space.\(^\text{1}\) For \( A' \) a finite subset of \( X' \), denote by \( \rho_{A'} \) the function defined on \( X \)

\[ \rho_{A'}(x) \equiv \max_{x^* \in A'} |x^*(x)| \quad (4.3.7) \]

and also let \( B_{A'}(x, r) \) be defined by

\[ B_{A'}(x, r) \equiv \{ y \in X : \rho_{A'}(y - x) < r \} \quad (4.3.8) \]

\(^1\)Actually, all this works in much more general settings than this.
Then certain things are obvious. First of all, if \(a \in F\) and \(x, y \in X\),
\[
\rho_{A'}(x + y) \leq \rho_{A'}(x) + \rho_{A'}(y), \\
\rho_{A'}(ax) = |a| \rho_{A'}(x).
\]

Similarly, letting \(A\) be a finite subset of \(X\), denote by \(\rho_A\) the function defined on \(X'\)
\[
\rho_A(x^*) = \max_{x \in A} |x^*(x)|
\]  
and let \(B_A(x^*, r)\) be defined by
\[
B_A(x^*, r) = \{y^* \in X' : \rho_A(y^* - x^*) < r\}.
\]  
It is also clear that
\[
\rho_A(x^* + y^*) \leq \rho(x^*) + \rho_A(y^*), \\
\rho_A(ax^*) = |a| \rho_A(x^*).
\]

**Lemma 4.3.1** The sets, \(B_A(x, r)\) where \(A'\) is a finite subset of \(X'\) and \(x \in X\) form a basis for a topology on \(X\) known as the weak topology. The sets \(B_A(x^*, r)\) where \(A\) is a finite subset of \(X\) and \(x^* \in X'\) form a basis for a topology on \(X'\) known as the weak * topology.

**Proof:** The two assertions are very similar. I will verify the one for the weak topology. The union of these sets, \(B_A(x, r)\) for \(x \in X\) and \(r > 0\) is all of \(X\). Now suppose \(z\) is contained in the intersection of two of these sets. Say
\[
z \in B_{A'}(x, r) \cap B_{A'_1}(x_1, r_1)
\]
Then let \(C' = A' \cup A'_1\) and let
\[
0 < \delta \leq \min\left(r - \rho_{A'}(z - x), r_1 - \rho_{A'_1}(z - x_1)\right).
\]
Consider \(y \in B_{C'}(z, \delta)\). Then
\[
r - \rho_{A'}(z - x) \geq \delta > \rho_{C'}(y - z) \geq \rho_{A'}(y - z)
\]
and so
\[
r > \rho_{A'}(y - z) + \rho_{A'}(z - x) \geq \rho_{A'}(y - x)
\]
which shows \(y \in B_{A'}(x, r)\). Similar reasoning shows \(y \in B_{A'_1}(x_1, r_1)\) and so
\[
B_{C'}(z, \delta) \subseteq B_{A'}(x, r) \cap B_{A'_1}(x_1, r_1).
\]
Therefore, the weak topology consists of the union of all sets of the form \(B_A(x, r)\).

### 4.3.2 Banach Alaoglu Theorem

Why does anyone care about these topologies? The short answer is that in the weak * topology, closed unit ball in \(X'\) is compact. This is not true in the normal topology. This wonderful result is the Banach Alaoglu theorem. First recall the notion of the product topology, and the Tychonoff theorem, Theorem 2.4.11 on Page 63 which are stated here for convenience.

**Definition 4.3.2** Let \(I\) be a set and suppose for each \(i \in I\), \((X_i, \tau_i)\) is a nonempty topological space. The Cartesian product of the \(X_i\), denoted by \(\prod_{i \in I} X_i\), consists of the set of all choice functions defined on \(I\) which select a single element of each \(X_i\). Thus \(f \in \prod_{i \in I} X_i\) means for every \(i \in I\), \(f(i) \in X_i\). The axiom of choice says \(\prod_{i \in I} X_i\) is nonempty. Let
\[
P_J(A) = \prod_{i \in J} B_i
\]
where \(B_i = X_i\) if \(i \neq j\) and \(B_j = A\). A subbasis for a topology on the product space consists of all sets \(P_J(A)\) where \(A \in \tau_j\). (These sets have an open set from the topology of \(X_j\) in the \(j\)th slot and the whole space in the other slots.) Thus a basis consists of finite intersections of these sets. Note that the intersection of two of these basic sets is another basic set and their union yields \(\prod_{i \in I} X_i\). Therefore, they satisfy the condition needed for a collection of sets to serve as a basis for a topology. This topology is called the product topology and is denoted by \(\prod \tau_i\).
Theorem 4.3.3  If \((X_i, \tau_i)\) is compact, then so is \((\prod_{i \in I} X_i, \prod \tau_i)\).

The Banach Alaoglu theorem is as follows.

Theorem 4.3.4  Let \(B'\) be the closed unit ball in \(X'\). Then \(B'\) is compact in the weak * topology.

Proof: By the Tychonoff theorem, Theorem \ref{thm:tychonoff}

\[
P = \prod_{x \in X} B(0, \|x\|)
\]

is compact in the product topology where the topology on \(B(0, \|x\|)\) is the usual topology of \(F\). Recall \(P\) is the set of functions which map a point, \(x \in X\) to a point in \(B(0, \|x\|)\). Therefore, \(B' \subseteq P\). Also the basic open sets in the weak * topology on \(B'\) are obtained as the intersection of basic open sets in the product topology of \(P\) to \(B'\) and so it suffices to show \(B'\) is a closed subset of \(P\). Suppose then that \(f \in P \setminus B'\). Since \(|f(x)| \leq \|x\|\) for each \(x\), it follows \(f\) cannot be linear. There are two ways this can happen. One way is that for some \(x, y\)

\[
f(x + y) \neq f(x) + f(y)
\]

for some \(x, y \in X\). However, if \(g\) is close enough to \(f\) at the three points, \(x + y, x, \) and \(y, \) the above inequality will hold for \(g\) in place of \(f\). In other words there is a basic open set containing \(f\), such that for all \(g\) in this basic open set, \(g \notin B'\). A similar consideration applies in case \(f(\lambda x) \neq \lambda f(x)\) for some scalar \(\lambda\) and \(x\). Since \(P \setminus B'\) is open, it follows \(B'\) is a closed subset of \(P\) and is therefore, compact. ■

Sometimes one can consider the weak * topology in terms of a metric space.

Theorem 4.3.5  If \(K \subseteq X'\) is compact in the weak * topology and \(X\) is separable in the weak topology then there exists a metric, \(d\), on \(K\) such that if \(\tau_d\) is the topology on \(K\) induced by \(d\) and if \(\tau\) is the topology on \(K\) induced by the weak * topology of \(X'\), then \(\tau = \tau_d\). Thus one can consider \(K\) with the weak * topology as a metric space.

Proof: Let \(D = \{x_n\}\) be the dense countable subset in \(X\). The metric is

\[
d(f, g) = \sum_{n=1}^{\infty} 2^{-n} \frac{\rho_{x_n}(f - g)}{1 + \rho_{x_n}(f - g)}
\]

where \(\rho_{x_n}(f) = |f(x_n)|\). Clearly \(d(f, g) = d(g, f) \geq 0\). If \(d(f, g) = 0\), then this requires \(f(x_n) = g(x_n)\) for all \(x_n \in D\). Is it the case that \(f = g\)? \(B\{f, g\}(x, r)\) contains some \(x_n \in D\). Hence

\[
\max\{|f(x_n) - f(x)|, |g(x_n) - g(x)|\} < r
\]

and \(f(x_n) = g(x_n)\). It follows that \(|f(x) - g(x)| < 2r\). Since \(r\) is arbitrary, this implies \(f(x) = g(x)\). It is routine to verify the triangle inequality from the easy to establish inequality,

\[
\frac{x}{1 + x} + \frac{y}{1 + y} \geq \frac{x + y}{1 + x + y},
\]

valid whenever \(x, y \geq 0\). Therefore this is a metric.

Thus there are two topological spaces, \((K, \tau)\) and \((K, d)\), the first being \(K\) with the weak * topology and the second being \(K\) with this metric. It is clear that if \(i\) is the identity map, \(i : (K, \tau) \to (K, d)\), then \(i\) is continuous. Therefore, sets which are open in \((K, \tau)\) are open in \((K, \tau_d)\). Letting \(\tau_d\) denote those sets which are open with respect to the metric, \(\tau_d \subseteq \tau\).

Now suppose \(U \in \tau\). Is \(U\) in \(\tau_d\)? Since \(K\) is compact with respect to \(\tau\), it follows from the above that \(K\) is compact with respect to \(\tau_d \subseteq \tau\). Hence \(K \setminus U\) is compact with respect to \(\tau_d\) and so it is closed with respect to \(\tau_d\). Thus \(U\) is open with respect to \(\tau_d\). ■

The fact that this set with the weak * topology can be considered a metric space is very significant because if a point is a limit point in a metric space, one can extract a convergent sequence.

Note that if a Banach space is separable, then it is weakly separable.

Corollary 4.3.6  If \(X\) is weakly separable and \(K \subseteq X'\) is compact in the weak * topology, then \(K\) is sequentially compact. That is, if \(\{f_n\}_{n=1}^{\infty} \subseteq K\), then there exists a subsequence \(f_{n_k}\) and \(f \in K\) such that for all \(x \in X\),

\[
\lim_{k \to \infty} f_{n_k}(x) = f(x).
\]

Proof: By Theorem \ref{thm:banachalaoglu}, \(K\) is a metric space for the metric described there and it is compact. Therefore by the characterization of compact metric spaces, Proposition \ref{prop:characterizationcompact} on Page \pageref{prop:characterizationcompact}, \(K\) is sequentially compact. This proves the corollary. ■
4.3.3 Eberlein Smulian Theorem

Next consider the weak topology. The most interesting results have to do with a reflexive Banach space. The following lemma ties together the weak and weak * topologies in the case of a reflexive Banach space.

**Lemma 4.3.7** Let \( J : X \to X'' \) be the James map

\[
Jx(f) = f(x)
\]

and let \( X \) be reflexive so that \( J \) is onto. Then \( J \) is a homeomorphism of \((X, \text{weak topology})\) and \((X'', \text{weak * topology})\). This means \( J \) is one to one, onto, and both \( J \) and \( J^{-1} \) are continuous.

**Proof:** Let \( f \in X' \) and let

\[
B_f(x,r) = \{ y : |f(x) - f(y)| < r \}.
\]

Thus \( B_f(x,r) \) is a subbasic set for the weak topology on \( X \). Now by the definition of \( J \),

\[
y \in B_f(x,r) \text{ if and only if } |Jy(f) - Jx(f)| < r
\]

if and only if \( Jy \in B_f(Jx,r) \equiv \{ y'' \in X'': |y''(f) - J(x)(f)| < r \},
\]

a subbasic set for the weak * topology on \( X'' \). Since \( J^{-1} \) and \( J \) are one to one and onto and map subbasic sets to subbasic sets, it follows that \( J \) is a homeomorphism. \( \blacksquare \)

The following is an easy corollary.

**Corollary 4.3.8** If \( X \) is a reflexive Banach space, then the closed unit ball is weakly compact.

**Proof:** Let \( B \) be the closed unit ball. Then \( B = J^{-1}(B^{**}) \) where \( B^{**} \) is the unit ball in \( X'' \) which is compact in the weak * topology. Therefore \( B \) is weakly compact because \( J^{-1} \) is continuous.

**Corollary 4.3.9** Let \( X \) be a reflexive Banach space. If \( K \subseteq X \) is compact in the weak topology and \( X' \) is separable in the weak * topology, then there exists a metric \( d \) on \( K \) such that if \( \tau_d \) is the topology on \( K \) induced by \( d \) and if \( \tau \) is the topology on \( K \) induced by the weak topology of \( X \), then \( \tau = \tau_d \). Thus one can consider \( K \) with the weak topology as a metric space.

**Proof:** This follows from Theorem \([\text{??}] \) and Lemma \([\text{??}] \). Lemma \([\text{??}] \) implies \( J(K) \) is compact in \( X'' \). Then since \( X' \) is separable in the weak * topology, there is a metric, \( d'' \) on \( J(K) \) which delivers the weak * topology on \( J(K) \). Let \( d(x,y) \equiv d''(Jx,Jy) \). Then

\[
(K, \tau_d) \xrightarrow{J} (J(K), \tau_{d''}) \xrightarrow{id} (J(K), \tau_{\text{weak *}}) \xrightarrow{J^{-1}} (K, \tau_{\text{weak}})
\]

and all the maps are homeomorphisms. \( \blacksquare \)

Here is a useful lemma.

**Lemma 4.3.10** Let \( Y \) be a closed subspace of a Banach space \( X \) and let \( y \in X \setminus Y \). Then there exists \( x^* \in X' \) such that \( x^* \langle y \rangle = 0 \) but \( x^* \langle y \rangle \neq 0 \).

**Proof:** Define \( f(x + \alpha y) = \|y\| \alpha \). Thus \( f \) is linear on \( Y \oplus \mathbb{F}y \). I claim that \( f \) is also continuous on this subspace of \( X \). If not, then there exists \( x_n + \alpha_n y \to 0 \) but \( |f(x_n + \alpha_n y)| \geq \varepsilon > 0 \) for all \( n \). First suppose \( |\alpha_n| \) is bounded. Then, taking a further subsequence, we can assume \( \alpha_n \to \alpha \). It follows then that \( \{x_n\} \) must also converge to some \( x \in Y \) since \( Y \) is closed. Therefore, in this case, \( x + \alpha y = 0 \) and so \( \alpha = 0 \) since otherwise, \( y \in Y \). In the other case when \( \alpha_n \) is unbounded, you have \((x_n/\alpha_n + y) \to 0 \) and so it would require that \( y \in Y \) which cannot happen because \( Y \) is closed. Hence \( f \) is continuous as claimed. It follows that for some \( k \),

\[
|f(x + \alpha y)| \leq k \|x + \alpha y\|
\]

Now apply the Hahn Banach theorem to extend \( f \) to \( x^* \in X' \). \( \blacksquare \)

Next is the Eberlein Smulian theorem which states that a Banach space is reflexive if and only if the closed unit ball is weakly sequentially compact. Actually, only half the theorem is proved here, the more useful only if part. The book by Yoshida \([\text{??}] \) has the complete theorem discussed. First here is an interesting lemma for its own sake.
Lemma 4.3.11: A closed subspace of a reflexive Banach space is reflexive.

Proof: Let $Y$ be the closed subspace of the reflexive space, $X$. Consider the following diagram

\[
\begin{align*}
Y'' & \xrightarrow{i'' \circ \frac{1}{2}} X'' \\
Y' & \xrightarrow{i' \text{ onto}} X' \\
Y & \xrightarrow{i} X
\end{align*}
\]

This diagram follows from Theorem 4.3.10 on Page 63, the theorem on adjoints. Now let $y'' \in Y''$. Then $i''y'' = J_X(y)$ because $X$ is reflexive. I want to show that $y \in Y$. If it is not in $Y$ then since $Y$ is closed, there exists $x^* \in X'$ such that $x^*(y) \neq 0$ but $x^*(Y) = 0$. Then $i^*x^* = 0$. Hence

\[
0 = y''(i^*x^*) = i''y''(x^*) = J_X(y)(x^*) = x^*(y) \neq 0,
\]

a contradiction. Hence $y \in Y$. Letting $J_Y$ denote the James map from $Y$ to $Y''$ and $x^* \in X'$,

\[
y''(i^*x^*) = i''y''(x^*) = J_X(y)(x^*) = x^*(y) = x^*(iy) = i^*x^*(y) = J_Y(y)(i^*x^*)
\]

Since $i^*$ is onto, this shows $y'' = J_Y(y)$. ■

Theorem 4.3.12 (Eberlein Smulian) The closed unit ball in a reflexive Banach space $X$, is weakly sequentially compact. By this is meant that if $\{x_n\}$ is contained in the closed unit ball, there exists a subsequence, $\{x_{n_k}\}$ and $x \in X$ such that for all $x^* \in X'$,

\[
x^*(x_{n_k}) \to x^*(x).
\]

Proof: Let $\{x_n\} \subseteq B = \overline{B}(0,1)$. Let $Y$ be the closure of the linear span of $\{x_n\}$. Thus $Y$ is a separable. It is reflexive because it is a closed subspace of a reflexive space so the above lemma applies. By the Banach Alaoglu theorem, the closed unit ball $B^*$ in $Y'$ is weak * compact. Also by Theorem 4.3.10, $B^*$ is a metric space with a suitable metric.

\[
B^* \xrightarrow{\text{separable}} \text{weakly separable } Y' \xrightarrow{i^* \text{ onto}} X' \xrightarrow{\text{onto}} X
\]

Thus $B^*$ is complete and totally bounded with respect to this metric and it follows that $B^*$ with the weak topology is separable. This implies $Y'$ is also separable in the weak * topology. To see this, let $\{y_n\} \equiv D$ be a weak * dense set in $B^*$ and let $y^* \in Y'$. Let $p$ be a large enough positive rational number that $y^*/p \in B^*$. Then if $A$ is any finite set from $Y$, there exists $y^*_n \in D$ such that $\rho_A(y^*/p - y^*_n) < \frac{\varepsilon}{2}$. It follows $py^*_n \in B_1(y^*, \varepsilon)$ showing that rational multiples of $A$ are weak * dense in $Y'$. Since $Y$ is reflexive, the weak and weak * topologies on $Y'$ coincide and so $Y'$ is weakly separable. Since $Y'$ is weakly separable, Corollary 4.3.7 implies $B^*$, the closed unit ball in $Y''$ is weak * sequentially compact. Then by Lemma 4.3.7 $B$, the unit ball in $Y$, is weakly sequentially compact. It follows there exists a subsequence $x_{n_k}$, of the sequence $\{x_n\}$ and a point $x \in Y$, such that for all $f \in Y'$,

\[
f(x_{n_k}) \to f(x).
\]

Now if $x^* \in X'$, and $i$ is the inclusion map of $Y$ into $X$,

\[
x^*(x_{n_k}) = i^*x^*(x_{n_k}) \to i^*x^*(x) = x^*(x),
\]

which shows $x_{n_k}$ converges weakly and this shows the unit ball in $X$ is weakly sequentially compact. ■

Corollary 4.3.13: Let $\{x_n\}$ be any bounded sequence in a reflexive Banach space $X$. Then there exists $x \in X$ and a subsequence, $\{x_{n_k}\}$ such that for all $x^* \in X'$,

\[
\lim_{k \to \infty} x^*(x_{n_k}) = x^*(x).
\]

Proof: If a subsequence, $x_{n_k}$ has $\|x_{n_k}\| \to 0$, then the conclusion follows. Simply let $x = 0$. Suppose then that $\|x_n\|$ is bounded away from 0. That is, $\|x_n\| \in [\delta, C]$. Take a subsequence such that $\|x_{n_k}\| \to a$. Then consider $x_{n_k}/\|x_{n_k}\|$. By the Eberlein Smulian theorem, this subsequence has a further subsequence, $x_{n_{k_j}}/\|x_{n_{k_j}}\|$ which converges weakly to $x \in B$ where $B$ is the closed unit ball. It follows from routine considerations that $x_{n_{k_j}} \to ax$ weakly. This proves the corollary.
Chapter 5

Locally Convex Topological Vector Spaces

5.1 Fundamental Considerations

The right context to consider certain topics like separation theorems is in locally convex topological vector spaces, a generalization of normed linear spaces. Let $X$ be a vector space and let $\Psi$ be a collection of functions defined on $X$ such that if $\rho \in \Psi$,

$$
\rho(x + y) \leq \rho(x) + \rho(y),
$$

$$
\rho(ax) = |a| \rho(x) \text{ if } a \in \mathbb{F},
$$

$$
\rho(x) \geq 0,
$$

where $\mathbb{F}$ denotes the field of scalars, either $\mathbb{R}$ or $\mathbb{C}$, assumed to be $\mathbb{C}$ unless otherwise specified. These functions are called seminorms because it is not necessarily true that $x = 0$ when $\rho(x) = 0$. A basis for a topology, $\mathcal{B}$, is defined as follows.

**Definition 5.1.1** For $A$ a finite subset of $\Psi$ and $r > 0$,

$$
B_A(x, r) \equiv \{ y \in X : \rho(x - y) < r \text{ for all } \rho \in A \}.
$$

Then

$$
\mathcal{B} \equiv \{ B_A(x, r) : x \in X, r > 0, \text{ and } A \subseteq \Psi, A \text{ finite} \}. 
$$

That this really is a basis is the content of the next theorem.

**Theorem 5.1.2** $\mathcal{B}$ is the basis for a topology.

**Proof:** I need to show that if $B_A(x, r_1)$ and $B_B(y, r_2)$ are two elements of $\mathcal{B}$ and if $z \in B_A(x, r_1) \cap B_B(y, r_2)$, then there exists $U \in \mathcal{B}$ such that $z \in U \subseteq B_A(x, r_1) \cap B_B(y, r_2)$.

Let

$$
r = \min \left( \min \{ (r_1 - \rho(z - x)) : \rho \in A \}, \min \{ (r_2 - \rho(z - y)) : \rho \in B \} \right)
$$

and consider $B_{A \cup B}(z, r)$. If $w$ belongs to this set, then for $\rho \in A$,

$$
\rho(w - z) < r_1 - \rho(z - x).
$$

Hence

$$
\rho(w - x) \leq \rho(w - z) + \rho(z - x) < r_1
$$

for each $\rho \in A$ and so $B_{A \cup B}(z, r) \subseteq B_A(x, r_1)$. Similarly, $B_{A \cup B}(z, r) \subseteq B_B(y, r_2)$. This proves the theorem.

Let $\tau$ be the topology consisting of unions of all subsets of $\mathcal{B}$. Then $(X, \tau)$ is a locally convex topological vector space.
CHAPTER 5. LOCALLY CONVEX TOPOLOGICAL VECTOR SPACES

Theorem 5.1.3 The vector space operations of addition and scalar multiplication are continuous. More precisely,

\[ + : X \times X \to X, \cdot : F \times X \to X \]

are continuous.

Proof: It suffices to show \( +^{-1}(B) \) is open in \( X \times X \) and \( \cdot^{-1}(B) \) is open in \( F \times X \) if \( B \) is of the form

\[ B = \{ y \in X : \rho(y - x) < r \} \]

because finite intersections of such sets form the basis \( \mathcal{B} \). (This collection of sets is a subbasis.) Suppose \( u + v \in B \) where \( B \) is described above. Then

\[ \rho(u + v - x) < \lambda r \]

for some \( \lambda < 1 \). Consider

\[ B_{\rho}(u, \delta) \times B_{\rho}(v, \delta). \]

If \((u_1, v_1)\) is in this set, then

\[ \rho(u_1 + v_1 - x) \leq \rho(u + v - x) + \rho(u_1 - u) + \rho(v_1 - v) < \lambda r + 2\delta. \]

Let \( \delta \) be positive but small enough that

\[ 2\delta + \lambda r < r. \]

Thus this choice of \( \delta \) shows that \( +^{-1}(B) \) is open and this shows + is continuous.

Now suppose \( \alpha z \in B \). Then

\[ \rho(\alpha z - x) < \lambda r < r \]

for some \( \lambda \in (0, 1) \). Let \( \delta > 0 \) be small enough that \( \delta < 1 \) and also

\[ \lambda r + \delta (\rho(z) + 1) + \delta |\alpha| < r. \]

Then consider \((\beta, w) \in B(\alpha, \delta) \times B_{\rho}(z, \delta)\).

\[ \rho(\beta w - x) - \rho(\alpha z - x) \leq \rho(\beta w - \alpha z) \leq |\beta - \alpha| \rho(w) + \rho(w - z) |\alpha| \leq |\beta - \alpha| (\rho(z) + 1) + \rho(w - z) |\alpha| < \delta (\rho(z) + 1) + \delta |\alpha|. \]

Hence

\[ \rho(\beta w - x) < \lambda r + \delta (\rho(z) + 1) + \delta |\alpha| < r \]

and so

\[ B(\alpha, \delta) \times B_{\rho}(z, \delta) \subseteq \cdot^{-1}(B). \]

This proves the theorem.

Theorem 5.1.4 Let \( x \) be given and let \( f_x(y) = x + y \). Then \( f_x \) is \( 1 - 1 \), onto, and continuous. If \( \alpha \neq 0 \) and \( g_{\alpha}(x) = \alpha x \), then \( g_{\alpha} \) is also \( 1 - 1 \) onto and continuous.

Proof: The assertions about \( 1 - 1 \) and onto are obvious. It remains to show \( f_x \) and \( g_{\alpha} \) are continuous. Let \( B = B_{\rho}(z, r) \) and consider \( f_x^{-1}(B) \). Then it is easy to see that

\[ f_x^{-1}(B) = B_{\rho}(z - x, r) \]

and so \( f_x \) is continuous. To see that \( g_{\alpha} \) is continuous, note that

\[ g_{\alpha}^{-1}(B) = B_{\rho}\left(\frac{z}{\alpha}, \frac{r}{|\alpha|}\right). \]

This proves the theorem.

As in the case of a normed linear space, the vector space of continuous linear functionals, is denoted by \( X' \).
**Definition 5.1.5** Define, for \( A \) a finite subset of \( \Psi \),
\[
\rho_A(x) = \max \{ \rho(x) : \rho \in A \}.
\]

The following theorem is the equivalent to the earlier theorems concerning continuous linear functionals on normed linear spaces.

**Theorem 5.1.6** The following are equivalent for \( f \), a linear function mapping \( X \) to \( F \).
\[
f \text{ is continuous at } 0. \tag{5.1.1}
\]
For some \( A \subseteq \Psi, A \text{ finite},
\[
|f(x)| \leq C \rho_A(x) \tag{5.1.2}
\]
for all \( x \in X \) where the constant may depend on \( A \) but is independent of \( x \).
\[
f \text{ is continuous at } x. \tag{5.1.3}
\]
for all \( x \).

**Proof:** Clearly (5.1.3) implies (5.1.1). Suppose (5.1.1). Then
\[
0 = f(0) \in B(0,1) \subseteq \mathbb{F}.
\]
Since \( f \) is continuous at 0, \( 0 \in f^{-1}(B(0,1)) \) and there exists an open set \( V \in \tau \) such that
\[
0 \in V \subseteq f^{-1}(B(0,1)).
\]
Then \( 0 \in B_A(0,r) \subseteq V \) for some \( r \) and some \( A \subseteq \Psi, A \text{ finite}. \)
Hence
\[
|f(y)| < 1 \text{ if } \rho_A(y) < r.
\]
Since \( f \) is linear
\[
|f(x)| \leq \frac{2}{r} \rho_A(x).
\]
To see this, note that if \( x \neq 0 \), then
\[
\frac{rx}{2\rho_A(x)} \in B_A(0,r)
\]
and so
\[
\frac{|f(rx)|}{2\rho_A(x)} \leq 1
\]
which shows that (5.1.1) implies (5.1.2).

Now suppose (5.1.2) and suppose \( f(x) \in V \), an open set in \( F \). Then
\[
f(x) \in B(f(x),r) \subseteq V
\]
for some \( r > 0 \). Suppose \( \rho_A(x-y) < r(C_A + 1)^{-1} \). Then
\[
|f(x) - f(y)| = |f(x-y)| \leq C_A \rho_A(y-x) < r.
\]
Hence
\[
f(B_A(x, r(C_A + 1)^{-1})) \subseteq B(f(x),r) \subseteq V.
\]
Thus \( f \) is continuous at \( x \). This proves the theorem.

What are some examples of locally convex topological vector spaces? It is obvious that any normed linear space is such an example. More generally, here is a theorem which shows how to make any vector space into a locally convex topological vector space.

**Theorem 5.1.7** Let \( X \) be a vector space and let \( Y \) be a vector space of linear functionals defined on \( X \). For each \( y \in Y \), define
\[
\rho_y(x) \equiv |y(x)|.
\]
Then the collection of seminorms \( \{\rho_y\}_{y \in Y} \) defined on \( X \) makes \( X \) into a locally convex topological vector space and \( Y = X' \).
**Proof:** Clearly \( \{ \rho_y \}_{y \in Y} \) is a collection of seminorms defined on \( X \); so, \( X \) supplied with the topology induced by this collection of seminorms is a locally convex topological vector space. Is \( Y = X' \)?

Let \( y \in Y \), let \( U \subseteq F \) be open and let \( x \in y^{-1}(U) \). Then \( B(y(x), r) \subseteq U \) for some \( r > 0 \). Letting \( A = \{ y \} \), it is easy to see from the definition that \( B_A(x, r) \subseteq y^{-1}(U) \) and so \( y^{-1}(U) \) is an open set as desired. Thus, \( Y \subseteq X' \).

Now suppose \( z \in X' \). Then by \ref{5.1.2}, there exists a finite subset of \( Y \), \( A = \{ y_1, \ldots, y_n \} \), such that

\[
|z(x)| \leq C\rho_A(x).
\]

Let

\[
\pi(x) \equiv (y_1(x), \ldots, y_n(x))
\]

and let \( f \) be a linear map from \( \pi(X) \) to \( F \) defined by

\[
f(\pi x) \equiv z(x).
\]

(This is well defined because if \( \pi(x) = \pi(x_1) \), then \( y_i(x) = y_i(x_1) \) for \( i = 1, \ldots, n \) and so

\[
\rho_A(x - x_1) = 0.
\]

Thus,

\[
|z(x_1) - z(x)| = |z(x_1 - x)| \leq C\rho_A(x - x_1) = 0.
\]

Extend \( f \) to all of \( F^n \) and denote the resulting linear map by \( F \). Then there exists a vector

\[
\alpha = (\alpha_1, \ldots, \alpha_n) \in F^n
\]

with \( \alpha_i = F(e_i) \) such that

\[
F(\beta) = \alpha \cdot \beta.
\]

Hence for each \( x \in X \),

\[
z(x) = f(\pi x) = F(\pi x) = \sum_{i=1}^{n} \alpha_i y_i(x)
\]

and so

\[
z = \sum_{i=1}^{n} \alpha_i y_i \in Y.
\]

This proves the theorem.

### 5.2 Separation Theorems

It will always be assumed that \( X \) is a locally convex topological vector space. A set, \( K \), is said to be convex if whenever \( x, y \in K \),

\[
\lambda x + (1 - \lambda) y \in K
\]

for all \( \lambda \in [0, 1] \).

**Definition 5.2.1** Let \( U \) be an open convex set containing 0 and define

\[
m(x) \equiv \inf\{t > 0 : x/t \in U\}.
\]

This is called a Minkowski functional.

**Proposition 5.2.2** Let \( X \) be a locally convex topological vector space. Then \( m \) is defined on \( X \) and satisfies

\[
m(x + y) \leq m(x) + m(y) \tag{5.2.4}
\]

\[
m(\lambda x) = \lambda m(x) \quad \text{if} \quad \lambda > 0. \tag{5.2.5}
\]

Thus, \( m \) is a gauge function on \( X \).
**Proof:** Let \( x \in X \) be arbitrary. There exists \( A \subseteq \Psi \) such that

\[
0 \in B_A (0, r) \subseteq U.
\]

Then

\[
\frac{rx}{2p_A (x)} \in B_A (0, r) \subseteq U
\]

which implies

\[
\frac{2p_A (x)}{r} \geq m (x).
\] (5.2.6)

Thus \( m (x) \) is defined on \( X \).

Let \( x/t \in U, y/s \in U \). Then since \( U \) is convex,

\[
\frac{x + y}{t + s} = \left( \frac{t}{t + s} \right) \left( \frac{x}{t} \right) + \left( \frac{s}{t + s} \right) \left( \frac{y}{s} \right) \in U.
\]

It follows that

\[
m (x + y) \leq t + s.
\]

Choosing \( s, t \) such that \( t - \varepsilon < m (x) \) and \( s - \varepsilon < m (y) \),

\[
m (x + y) \leq m (x) + m (y) + 2\varepsilon.
\]

Since \( \varepsilon \) is arbitrary, this shows \( \frac{\lambda x}{\lambda t} \in U \). It remains to show \( \frac{\lambda m (x)}{\lambda t} \leq m \). Let \( x/t \in U \). Then if \( \lambda > 0 \),

\[
\frac{\lambda x}{\lambda t} \in U
\]

and so \( m (\lambda x) \leq \lambda t \). Thus \( m (\lambda x) \leq \lambda m (x) \) for all \( \lambda > 0 \). Hence

\[
m (x) = m (\lambda^{-1} \lambda x) \leq \lambda^{-1} m (\lambda x) \leq \lambda^{-1} \lambda m (x) = m (x)
\]

and so

\[
\lambda m (x) = m (\lambda x).
\]

This proves the proposition.

**Lemma 5.2.3** Let \( U \) be an open convex set containing \( 0 \) and let \( q \notin U \). Then there exists \( f \in X' \) such that

\[
\text{Re } f (q) > \text{Re } f (x)
\]

for all \( x \in U \).

**Proof:** Let \( m \) be the Minkowski functional just defined and let

\[
F (cq) = cm (q)
\]

for \( c \in \mathbb{R} \). If \( c > 0 \) then

\[
F (cq) = m (cq)
\]

while if \( c \leq 0 \),

\[
F (cq) = cm (q) \leq 0 \leq m (cq).
\]

By the Hahn Banach theorem, \( F \) has an extension, \( g \), defined on all of \( X \) satisfying

\[
g (x + y) = g (x) + g (y), \; g (cx) = cg (x)
\]

for all \( c \in \mathbb{R} \), and

\[
g (x) \leq m (x).
\]

Thus, \( g (-x) \leq m (-x) \) and so

\[
-m (-x) \leq g (x) \leq m (x).
\]
It follows as in 5.2.6 that for some \( A \subseteq \Psi \), \( A \) finite, and \( r > 0 \),
\[
|g(x)| \leq m(x) + m(-x) \\
\leq \frac{2}{r} \rho_A(x) + \frac{2}{r} \rho_A(-x) = 4 \rho_A(x)
\]
because
\[
\rho_A(-x) = |-1| \rho_A(x) = \rho_A(x).
\]
Hence \( g \) is continuous by Theorem 5.1.6. Now define
\[
f(x) \equiv g(x) - ig(ix).
\]
Thus \( f \) is linear and continuous so \( f \in X' \) and 
\[
\text{Re } f(x) = g(x).
\]
But for \( x \in U \), Theorem 5.1.3 implies that \( x/t \in U \) for some \( t < 1 \) and so \( m(x) < 1 \). Since \( U \) is convex and \( 0 \in U \), it follows \( q/t \notin U \) if \( t < 1 \) because if it were,
\[
q = t \left( \frac{q}{t} \right) + (1-t) 0 \notin U.
\]
Therefore, \( m(q) > 1 \) and for \( x \in U \),
\[
\text{Re } f(x) = g(x) \leq m(x) < 1 \leq m(q) = g(q) = \text{Re } f(q)
\]
and this proves the lemma.

**Theorem 5.2.4** Let \( K \) be closed and convex in a locally convex topological vector space and let \( p \notin K \). Then there exists a real number, \( c \), and \( f \in X' \) such that
\[
\text{Re } f(p) > c > \text{Re } f(k)
\]
for all \( k \in K \).

**Proof:** Since \( K \) is closed, and \( p \notin K \), there exists a finite subset of \( \Psi, A \), and a positive \( r > 0 \) such that
\[
K \cap B_A(p, 2r) = \emptyset.
\]
Pick \( k_0 \in K \) and let
\[
U = K + B_A(0, r) - k_0, \ q = p - k_0.
\]
It follows that \( U \) is an open convex set containing \( 0 \) and \( q \notin U \). Therefore, by Lemma 5.2.3, there exists \( f \in X' \) such that
\[
\text{Re } f(p - k_0) = \text{Re } f(q) > \text{Re } f(k + e - k_0)
\]
for all \( k \in K \) and \( e \in B_A(0, r) \). If \( \text{Re } f(e) = 0 \) for all \( e \in B_A(0, r) \), then \( \text{Re } f = 0 \) and 5.2.7 could not hold. Therefore, \( \text{Re } f(e) > 0 \) for some \( e \in B_A(0, r) \) and so,
\[
\text{Re } f(p) > \text{Re } f(k) + \text{Re } f(e)
\]
for all \( k \in K \). Let \( c_1 \equiv \sup \{ \text{Re } f(k) : k \in K \} \). Then for all \( k \in K \),
\[
\text{Re } f(p) \geq c_1 + \text{Re } f(e) > c_1 + \frac{\text{Re } f(e)}{2} > \text{Re } f(k).
\]
Let \( c = c_1 + \frac{\text{Re } f(e)}{2} \). This proves the theorem.

Note that if the field of scalars comes from \( \mathbb{R} \) rather than \( \mathbb{C} \) there is no essential change to the above conclusions. Just eliminate all references to the real part.
5.2. SEPARATION THEOREMS

5.2.1 Convex Functionals

As an important application, this theorem gives the basis for proving something about lower semicontinuity of functionals.

**Definition 5.2.5** Let $X$ be a Banach space and let $\phi : X \to (0, \infty]$ be convex and lower semicontinuous. This means whenever $x \in X$ and $\lim_{n \to \infty} x_n = x$,

$$\phi(x) \leq \lim \inf_{n \to \infty} \phi(x_n).$$

Also assume $\phi$ is not identically equal to $\infty$.

**Lemma 5.2.6** Let $X, Y$ be two Banach spaces. Then letting

$$||(x, y)|| \equiv \max(||x||_X, ||y||_Y),$$

it follows $X \times Y$ is a Banach space and $\phi \in (X \times Y)'$ if and only if there exist $x^* \in X'$ and $y^* \in Y'$ such that

$$\phi((x, y)) = x^*(x) + y^*(y).$$

The topology coming from this norm is called the strong topology.

**Proof:** Most of these conclusions are obvious. In particular it is clear $X \times Y$ is a Banach space with the given norm. Let $\phi \in (X \times Y)'$. Also let $\pi_X (x, y) \equiv (x, 0)$ and $\pi_Y (x, y) \equiv (0, y)$. Then each of $\pi_X$ and $\pi_Y$ is continuous and

$$\phi((x, y)) = \phi(\pi_X + \pi_Y)((x, y))$$
$$= \phi((x, 0)) + \phi((0, y)).$$

Thus $\phi \circ \pi_X$ and $\phi \circ \pi_Y$ are both continuous and their sum equals $\phi$. Let $x^* (x) \equiv \phi \circ \pi_X (x, 0)$ and let $y^* \equiv \phi \circ \pi_Y (x, 0)$. Then it is clear both $x^*$ and $y^*$ are continuous and linear defined on $X$ and $Y$ respectively. Also, if $(x^*, y^*) \in X' \times Y'$, then if $\phi((x, y)) \equiv x^*(x) + y^*(y)$, it follows $\phi \in (X \times Y)'$. This proves the lemma.

**Lemma 5.2.7** Let $\phi$ be a functional as described in Definition 5.2.5. Then $\phi$ is lower semicontinuous if and only if the epigraph of $\phi$ is closed in $X \times \mathbb{R}$ with the strong topology. Here the epigraph is defined as

$$\text{epi}(\phi) \equiv \{(x, y) : y \geq \phi(x)\}.$$ 

In this case the functional is called strongly lower semicontinuous.

**Proof:** First suppose $\text{epi}(\phi)$ is closed and suppose $x_n \to x$. Let $l < \phi(x)$. Then $(x, l) \notin \text{epi}(\phi)$ and so there exists $\delta > 0$ such that if $|x - y| < \delta$ and $|\alpha - l| < \delta$, then $\alpha < \phi(y)$. This implies that if $|x - y| < \delta$ and $\alpha < l + \delta$, then the above holds. Therefore, $(x_n, \phi(x_n))$, being in $\text{epi}(\phi)$ cannot satisfy both conditions,

$$|x_n - x| < \delta, \phi(x_n) < l + \delta.$$

However, for all $n$ large enough, the first condition is satisfied. Consequently, for all $n$ large enough, $\phi(x_n) \geq l + \delta \geq l$.

Thus

$$\lim \inf_{n \to \infty} \phi(x_n) \geq l$$
and since $l < \phi(x)$ is arbitrary, it follows

$$\lim \inf_{n \to \infty} \phi(x_n) \geq \phi(x).$$

Next suppose the condition about the $\lim \inf$. If $\text{epi}(\phi)$ is not closed, then there exists $(x, l) \notin \text{epi}(\phi)$ which is a limit point of points of $\text{epi}(\phi)$. Thus there exists $(x_n, l_n) \in \text{epi}(\phi)$ such that $(x_n, l_n) \to (x, l)$ and so

$$l = \lim \inf_{n \to \infty} l_n \geq \lim \inf_{n \to \infty} \phi(x_n) \geq \phi(x),$$

contradicting $(x, l) \notin \text{epi}(\phi)$. This proves the lemma.

**Definition 5.2.8** Let $\phi$ be convex and defined on $X$, a Banach space. Then $\phi$ is said to be weakly lower semicontinuous if $\text{epi}(\phi)$ is closed in $X \times \mathbb{R}$ where a basis for the topology of $X \times \mathbb{R}$ consists of sets of the form $U \times (a, b)$ for $U$ a weakly open set in $X$. 
CHAPTER 5. LOCALLY CONVEX TOPOLOGICAL VECTOR SPACES

Theorem 5.2.9  Let \( \phi \) be a lower semicontinuous convex functional as described in Definition 5.2.5 and let \( X \) be a real Banach space. Then \( \phi \) is also weakly lower semicontinuous.

Proof: By Lemma 5.2.7 \( \text{epi}(\phi) \) is closed in \( X \times \mathbb{R} \) with the strong topology as well as being convex. Letting \((z,l) \notin \text{epi}(\phi)\), it follows from Theorem 5.2.4 and Lemma 5.2.6 there exists \((x^*,\alpha) \in X' \times \mathbb{R} \) such that for some \( c \)

\[
x^*(z) + \alpha l > c > x^*(x) + \alpha \beta
\]

whenever \( \beta \geq \phi(x) \).

Consider \( B\{((x^*,\alpha)) \} \) where \( r \) is chosen so small that if \((y,\gamma) \in B\{((x^*,\alpha)) \}\), then

\[
x^*(y) + \alpha \gamma > c.
\]

This shows that the complement of \( \text{epi}(\phi) \) is weakly open and this proves the theorem.

Corollary 5.2.10  Let \( \phi \) be a lower semicontinuous convex functional as described in Definition 5.2.5 and let \( X \) be a real Banach space. Then if \( x_n \) converges weakly to \( x \), it follows that

\[
\phi(x) \leq \liminf_{n \to \infty} \phi(x_n).
\]

Proof: Let \( l < \phi(x) \) so that \((x,l) \notin \text{epi}(\phi)\). Then by Theorem 5.2.9 there exists \( B \times (-\infty, l + \delta) \) such that \( B \) is a weakly open set in \( X \) containing \( x \) and

\[
B \times (-\infty, l + \delta) \subseteq \text{epi}(\phi)^C.
\]

Thus \((x_n,\phi(x_n)) \notin B \times (-\infty, l + \delta) \) for all \( n \). However, \( x_n \in B \) for all \( n \) large enough. Therefore, for those values of \( n \), it must be the case that \( \phi(x_n) \notin (-\infty, l + \delta) \) and so

\[
\liminf_{n \to \infty} \phi(x_n) \geq l + \delta \geq l
\]

which shows, since \( l < \phi(x) \) is arbitrary that

\[
\liminf_{n \to \infty} \phi(x_n) \geq \phi(x).
\]

This proves the corollary.

5.2.2  More Separation Theorems

There are other separation theorems which can be proved in a similar way. The next theorem considers the separation of an open convex set from a convex set.

Theorem 5.2.11  Let \( A \) and \( B \) be disjoint, convex and nonempty sets with \( B \) open. Then there exists \( f \in X' \) such that

\[
\text{Re} f(a) < \text{Re} f(b)
\]

for all \( a \in A \) and \( b \in B \).

Proof: Let \( b_0 \in B, a_0 \in A \). Then the set

\[
B - A + a_0 - b_0
\]

is open, convex, contains \( 0 \), and does not contain \( a_0 - b_0 \). By Lemma 5.2.12 there exists \( f \in X' \) such that

\[
\text{Re} f(a_0 - b_0) > \text{Re} f(b - a + a_0 - b_0)
\]

for all \( a \in A \) and \( b \in B \). Therefore, for all \( a \in A, b \in B \),

\[
\text{Re} f(b) > \text{Re} f(a).
\]

Before giving another separation theorem, here is a lemma.

Lemma 5.2.12  If \( B \) is convex, then \( \text{int}(B) \equiv \text{union of all open sets contained in } B \) is convex. Also, if \( \text{int}(B) \neq \emptyset \), then \( B \subseteq \text{int}(B) \).
Proof: Suppose $x, y \in \text{int}(B)$. Then there exists $r > 0$ and a finite set $A \subseteq \Psi$ such that

$$B_A(x, r), B_A(y, r) \subseteq B.$$ 

Let

$$V \equiv \cup_{\lambda \in [0,1]} \lambda B_A(x, r) + (1 - \lambda) B_A(y, r).$$

Then $V$ is open, $V \subseteq B$, and if $\lambda \in [0,1]$, then

$$\lambda x + (1 - \lambda) y \in V \subseteq B.$$ 

Therefore, int$(B)$ is convex as claimed.

Now let $y \in B$ and $x \in \text{int}(B)$. Let

$$x \in B_A(x, r) \subseteq \text{int}(B)$$

and let $x_\lambda \equiv (1 - \lambda) x + \lambda y$. Define the open cone,

$$C \equiv \cup_{\lambda \in [0,1]} B_A(x_\lambda, (1 - \lambda) r).$$

Thus $C$ is represented in the following picture.

I claim $C \subseteq B$ as suggested in the picture. To see this, let

$$z \in B_A(x_\lambda, (1 - \lambda) r), \lambda \in (0,1).$$

Then

$$\rho_A(z - x_\lambda) < (1 - \lambda) r$$

and so

$$\rho_A\left(\frac{z}{1 - \lambda} - x - \frac{\lambda y}{1 - \lambda}\right) < r.$$ 

Therefore,

$$\frac{z}{1 - \lambda} - \frac{\lambda y}{1 - \lambda} \in B_A(x, r) \subseteq B.$$ 

It follows

$$(1 - \lambda)\left(\frac{z}{1 - \lambda} - \frac{\lambda y}{1 - \lambda}\right) + \lambda y = z \in B$$

and so $C \subseteq B$ as claimed. Now this shows $x_\lambda \in \text{int}(B)$ and $\lim_{\lambda \to 1} x_\lambda = y$. Thus, $y \in \text{int}(B)$ and this proves the lemma.

**Corollary 5.2.13** Let $A, B$ be convex, nonempty sets. Suppose int$(B) \neq \emptyset$ and $A \cap \text{int}(B) = \emptyset$. Then there exists $f \in X'$, $f \neq 0$, such that for all $a \in A$ and $b \in B$,

$$\text{Re } f(b) \geq \text{Re } f(a).$$
Proof: By Theorem [5.2.11], there exists $f \in X'$ such that for all $b \in \text{int}(B)$, and $a \in A$,
\[ \text{Re } f(b) > \text{Re } f(a). \]
Thus, in particular, $f \neq 0$. By Lemma [5.2.12], if $b \in B$ and $a \in A$,
\[ \text{Re } f(b) \geq \text{Re } f(a). \]
This proves the theorem.

Lemma 5.2.14 If $X$ is a topological Hausdorff space then compact implies closed.

Proof: Let $K$ be compact and suppose $K^C$ is not open. Then there exists $p \in K^C$ such that $V_p \cap K \neq \emptyset$ for all open sets $V_p$ containing $p$. Let
\[ C = \{(V_p)^C : V_p \text{ is an open set containing } p\}. \]
Then $C$ is an open cover of $K$ because if $q \in K$, there exist disjoint open sets $V_p$ and $V_q$ containing $p$ and $q$ respectively. Thus $q \in (V_p)^C$. This is an example of an open cover of $K$ which has no finite subcover, contradicting the assumption that $K$ is compact. This proves the lemma.

Lemma 5.2.15 If $X$ is a locally convex topological vector space, and if every point is a closed set, then the seminorms and $X'$ separate the points. This means if $x \neq y$, then for some $\rho \in \Psi$,
\[ \rho(x - y) \neq 0 \]
and for some $f \in X'$,
\[ f(x) \neq f(y). \]
In this case, $X$ is a Hausdorff space.

Proof: Let $x \neq y$. Then by Theorem [5.2.10], there exists $f \in X'$ such that $f(x) \neq f(y)$. Thus $X'$ separates the points. Since $f \in X'$, Theorem [5.1.6] implies
\[ |f(z)| \leq C\rho_A(z) \]
for some $A$ a finite subset of $\Psi$. Thus
\[ 0 < |f(x - y)| \leq C\rho_A(x - y) \]
and so $\rho(x - y) \neq 0$ for some $\rho \in A \subseteq \Psi$. Now to show $X$ is Hausdorff, let
\[ 0 < r < \rho(x - y)2^{-1}. \]
Then the two disjoint open sets containing $x$ and $y$ respectively are
\[ B_\rho(x, r) \text{ and } B_\rho(y, r). \]
This proves the lemma.

5.3 The Weak And Weak* Topologies

The weak and weak * topologies are examples which make the underlying vector space into a topological vector space. This section gives a description of these topologies. Unless otherwise specified, $X$ is a locally convex topological vector space. For $G$ a finite subset of $X'$ define $\delta_G : X \to [0, \infty)$ by
\[ \delta_G(x) = \max\{|f(x)| : f \in G\}. \]

Lemma 5.3.1 The functions $\delta_G$ for $G$ a finite subset of $X'$ are seminorms and the sets
\[ B_G(x, r) = \{y \in X : \delta_G(x - y) < r\} \]
form a basis for a topology on $X$. Furthermore, $X$ with this topology is a locally convex topological vector space. If each point in $X$ is a closed set, then the same is true of $X$ with respect to this new topology.
5.4. MEAN ERGODIC THEOREM

The following theorem is called the mean ergodic theorem.

5.4 Mean Ergodic Theorem

Proof: It is obvious that the functions \( \delta_G \) are seminorms and therefore the proof that the sets \( B_G (x, r) \) form a basis for a topology is the same as in Theorem 5.3.4. To see every point is a closed set in this new topology, assuming this is true for \( X \) with the original topology, use Lemma 5.3.15 to assert \( X' \) separates the points. Let \( x \in X \) and let \( y \neq x \). There exists \( f \in X' \) such that \( \langle f, x \rangle \neq \langle f, y \rangle \). Let \( G = \{ f \} \) and consider

\[
B_G (y, |f (x - y)|/2).
\]

Then this open set does not contain \( x \). Thus \( \{ x \}^C \) is open and so \( \{ x \} \) is closed. This proves the Lemma.

This topology for \( X \) is called the weak topology for \( X \). For \( F \) a finite subset of \( X \), define \( \gamma_F : X' \to [0, \infty) \) by

\[
\gamma_F (f) = \max \{|f (x)| : x \in F\}.
\]

Lemma 5.3.2 The functions \( \gamma_F \) for \( F \) a finite subset of \( X \) are seminorms and the sets

\[
B_F (f, r) = \{ g \in X' : \gamma_F (f - g) < r \}
\]

form a basis for a topology on \( X' \). Furthermore, \( X' \) with this topology is a locally convex topological vector space having the property that every point is a closed set.

Proof: The proof is similar to that of Lemma 5.3.1 but there is a difference in the part where every point is shown to be a closed set. Let \( f \in X' \) and let \( g \neq f \). Thus there exists \( x \in X \) such that \( \langle f, x \rangle \neq \langle g, x \rangle \). Let \( F = \{ x \} \). Then

\[
B_F (g, |f - g| (x))/2
\]

contains \( g \) but not \( f \). Thus \( \{ f \}^C \) is open and so \( \{ f \} \) is closed. \( \blacksquare \)

Note that it was not necessary to assume points in \( X \) are closed sets to get this.

The topology for \( X' \) just described is called the weak * topology. In terms of Theorem 5.4.7 the weak topology is obtained by letting \( Y = X' \) in that theorem while the weak * topology is obtained by letting \( Y = X \) with the understanding that \( X \) is a vector space of linear functionals on \( X' \) defined by

\[
x (x^*) = x^* (x).
\]

By Theorem 5.3.3 there is a useful result which follows immediately.

Theorem 5.3.3 Let \( K \) be closed and convex in a Banach space \( X \). Then it is also weakly closed. Furthermore, if \( p \notin K \), there exists \( x \in X' \) such that

\[
\text{Re } f (p) > c > \text{Re } f (k)
\]

for all \( k \in K \). If \( K^* \) is closed and convex in the dual of a Banach space, \( X' \), then it is also weak * closed.

Proof: By Theorem 5.3.3 there exists \( f \in X' \) such that 5.3.3 holds. Therefore, letting \( A = \{ f \} \), it follows that for \( r \) small enough, \( B_A (p, r) \cap K = \emptyset \). Thus \( K \) is weakly closed. This establishes the first part.

For the second part, the seminorms for the weak * topology are determined from \( X \) and the continuous linear functionals are of the form \( x^* \to x^* (x) \) where \( x \in X \). Thus if \( p^* \notin K^* \), it follows from Theorem 5.3.3 there exists \( x \in X \) such that

\[
\text{Re } p^* (x) > c > \text{Re } k^* (x)
\]

for all \( k^* \in K^* \). Therefore, letting \( A = \{ x \} \), \( B_A (p^*, r) \cap K^* = \emptyset \) whenever \( r \) is small enough and this shows \( K^* \) is weak * closed. \( \blacksquare \)

5.4 Mean Ergodic Theorem

The following theorem is called the mean ergodic theorem.

Theorem 5.4.1 Let \((\Omega, \mathcal{S}, \mu)\) be a finite measure space and let \( T : \Omega \to \Omega \) satisfy \( T^{-1} (E) \in \mathcal{S}, T (E) \in \mathcal{S} \) for all \( E \in \mathcal{S} \). Also suppose for all positive integers, \( n \), that

\[
\mu (T^{-n} (E)) \leq K \mu (E).
\]

For \( f \in L^p (\Omega) \), and \( p > 1 \), let

\[
T^* f = f \circ T.
\]
Then \( T^* \in \mathcal{L}(L^p(\Omega), L^p(\Omega)) \), the continuous linear mappings form \( L^p(\Omega) \) to itself with
\[
||T^{*n}|| \leq K^{1/p}.
\] (5.4.10)

Defining \( A_n \in \mathcal{L}(L^p(\Omega), L^p(\Omega)) \) by
\[
A_n = \frac{1}{n} \sum_{k=0}^{n-1} T^k,
\]
there exists \( A \in \mathcal{L}(L^p(\Omega), L^p(\Omega)) \) such that for all \( f \in L^p(\Omega) \),
\[
A_n f \to Af \text{ weakly} \quad (5.4.11)
\]
and \( A \) is a projection, \( A^2 = A \), onto the space of all \( f \in L^p(\Omega) \) such that \( T^* f = f \). (The invariant functions.) The norm of \( A \) satisfies
\[
||A|| \leq K^{1/p}.
\] (5.4.12)

**Proof:** To begin with, it follows from simple considerations that
\[
\int |X_A(T^m(\omega))|^p d\mu = \int |X_{T^{-n}(A)}(\omega)|^p d\mu = \mu(T^{-n}(A)) \leq K\mu(A)
\]
Hence
\[
||T^{*n}(X_A)|| \leq K^{1/p}\mu(A)^{1/p} = K^{1/p}||X_A||_{L^p}
\]
Next suppose you have a simple function \( s(\omega) = \sum_{k=1}^n X_{A_i}(\omega) c_i \) where we assume the \( A_i \) are disjoint. From the above,
\[
\int \left| \sum_{k=1}^n X_{A_i}(T^m(\omega)) c_i \right|^p d\mu = \int \sum_{k=1}^n X_{A_i}(T^m(\omega))^p |c_i|^p d\mu \leq \sum_{k=1}^n K\mu(A_i)|c_i|^p = K \int |s|^p d\mu
\]
and so
\[
||T^{*m}s|| \leq K^{1/p} ||s||
\]
and so the density of the simple functions implies that \( ||T^{*m}|| \leq K^{1/p} \).

Next let
\[
M = \left\{ g \in L^p(\Omega) : ||A_ng|| \to 0 \right\}
\]
It follows from 5.4.10 that \( M \) is a closed subspace of \( L^p(\Omega) \) containing \((I - T^*)(L^p(\Omega))\). This is shown next.

**Claim 1:** \( M \) is a closed subspace which contains \((I - T^{*m})(L^p(\Omega))\).

First it is shown that this is true if \( m = 1 \) and then it will be observed that the same argument would work for any positive integer \( m \).

\[
A_n(f - T^*f) = \frac{1}{n} \sum_{k=0}^{n-1} T^k f - T^{k+1} f = \frac{1}{n} \sum_{k=0}^{n-1} T^k f - \frac{1}{n} \sum_{k=1}^{n} T^k f = \frac{1}{n} (f - T^{*n}f)
\]
Hence
\[
||A_n(f - T^*f)||_p \leq \frac{1}{n} \left(||f||_p + ||T^{*n}f||_p\right) \leq \frac{1}{n} \left(||f||_p + K^{1/p} ||f||_p\right)
\]
and this clearly converges to 0. In fact, the same argument shows that \( M \) contains \((I - T^{*m})(L^p(\Omega))\) for any \( m \).

Now suppose \( g_n \in M \) and \( g_n \to g \). Does it follow that \( g \in M \) also? Note that \( T^{*m} \) is clearly linear. Thus
\[
||T^{*m}g|| \leq ||T^{*m}g - T^{*m}g_n|| + ||T^{*m}g_n|| \leq K^{1/p} ||g - g_n|| + ||T^{*m}g_n||
\]
Now pick \( n \) large enough that \( ||g_n - g|| < \varepsilon / (2K^{1/p}) \) so that
\[
||T^{*m}g|| \leq \frac{\varepsilon}{2} + ||T^{*m}g_n||
\]
Then for all \( m \) large enough, the right side of the above is less than \( \varepsilon \) and this shows that \( g \in M \). Note that \( M \) is also a subspace and so it is a closed subspace.

**Claim 2:** If \( A_n f \to g \) weakly and \( A_m f \to h \) weakly, then \( g = h \).

It is first shown that if \( \xi \in L^{p'}(\Omega) \) and \( \int \xi g d\mu = 0 \) for all \( g \in M \), then \( \int \xi (g-h) d\mu = 0 \).
If $\xi \in L^p' (\Omega)$ is such that $\int \xi g \, d\mu = 0$ for all $g \in M$, then since $M \supseteq (I - T^{*n}) (L^p (\Omega))$, it follows that for all $k \in L^p (\Omega)$,

$$\int \xi k \, d\mu = \int (\xi T^{*n}k + \xi (I - T^{*n}) k) \, d\mu = \int \xi T^{*n}k \, d\mu$$

and so from the definition of $A_n$ as an average, for such $\xi$,

$$\int \xi k \, d\mu = \int \xi A_n k \, d\mu. \quad (5.4.13)$$

Since $A_n f \to g$ weakly and $A_m k \to h$ weakly. Then $5.4.13$ shows that

$$\int \xi g \, d\mu = \lim_{k \to \infty} \int \xi A_n k f \, d\mu = \int \xi f \, d\mu = \lim_{k \to \infty} \int \xi A_m k f \, d\mu = \int \xi h \, d\mu. \quad (5.4.14)$$

Thus for these special $\xi$, it follows that

$$\int \xi (g - h) \, d\mu = 0. \quad (5.4.15)$$

Next observe that for each fixed $n$, if $n_k \to \infty$, $\lim_{k \to \infty} \|T^{*n} A_n f - A_{n_k} f\| = 0 \quad (5.4.16)$

this follows like the arguments given above in Claim 1. Note that if $L \in L(X, X)$ and $x_n \to x$ weakly in $X$, then for $\phi \in X'$

$$\langle \phi, L x_n \rangle = \langle L^* \phi, x_n \rangle \to \langle L^* \phi, x \rangle = \langle \phi, L x \rangle$$

and so $L x_n \to L x$ weakly. Therefore, this simple observation along with the above strong convergence $5.4.11$ implies

$$T^{*n} g = \text{ weak lim}_{k \to \infty} T^{*n} A_n f = \text{ weak lim}_{k \to \infty} A_{n_k} f = g.$$

Similarly $T^{*n} h = h$ where $A_{n_k} f \to h$ weakly. It follows that $A_n (g - h) = g - h$ so if $g \neq h$, then $g - h \notin M$ because

$$A_n (g - h) \to g - h \neq 0.$$ 

It follows that since $M$ is a closed subspace, there exists $\xi \in L^p' (\Omega)$ such that $\int \xi (g - h) \, d\mu \neq 0$ but $\int \xi k \, d\mu = 0$ for all $k \in M$, contradicting $5.4.13$. This verifies Claim 2.

Now

$$\| A_n f \|_p = \left( \frac{1}{n} \sum_{k=0}^{n-1} f (T^k \omega) \right)^p \, d\mu \leq \frac{1}{n} \sum_{k=0}^{n-1} \left( \int |f (T^k \omega)|^p \, d\mu \right)^{1/p}$$

$$= \frac{1}{n} \sum_{k=0}^{n-1} \| T^{*k} f \|_p \leq \frac{1}{n} \sum_{k=0}^{n-1} K^{1/p} \| f \|_p = K^{1/p} \| f \|_p.$$

Hence, by the Eberlein Smulian theorem, Theorem 1.3.12 in case $p > 1$, there is a subsequence for which $A_n f$ converges weakly in $L^p (\Omega)$. From the above, it follows that the original sequence must converge. That is, $A_n f$ converges weakly for each $f \in L^p (\Omega)$. Let $Af$ denote this weak limit. Then it is clear that $A$ is linear because this is true for each $A_n$. What of the claim about the estimate? From weak lower semicontinuity of the norm, Corollary 5.2.10,

$$\| Af \|_p \leq \liminf_{n \to \infty} \| A_n f \| \leq K^{1/p} \| f \|_p.$$
Chapter 6

Basic Theory Of Hilbert Spaces

**Definition 6.0.2** Let $X$ be a vector space. An inner product is a mapping from $X \times X$ to $\mathbb{C}$ if $X$ is complex and from $X \times X$ to $\mathbb{R}$ if $X$ is real, denoted by $(x, y)$ which satisfies the following.

\[
(x, x) \geq 0, \quad (x, x) = 0 \text{ if and only if } x = 0, \quad (6.0.1)
\]

\[
(x, y) = (y, x). \quad (6.0.2)
\]

For $a, b \in \mathbb{C}$ and $x, y, z \in X$,

\[
(ax + by, z) = a(x, z) + b(y, z). \quad (6.0.3)
\]

Note that $(6.0.1)$ and $(6.0.2)$ imply $(x, ay + bz) = a(x, y) + b(x, z)$. Such a vector space is called an inner product space.

The Cauchy Schwarz inequality is fundamental for the study of inner product spaces.

**Theorem 6.0.3** (*Cauchy Schwarz*) In any inner product space

\[
| (x, y) | \leq \| x \| \| y \| .
\]

**Proof:** Let $\omega \in \mathbb{C}$, $|\omega| = 1$, and $\mathcal{F}(x, y) = | (x, y) | = \Re(x, y\omega)$. Let

\[
F(t) = (x + ty\omega, x + t\omega y).
\]

If $y = 0$ there is nothing to prove because

\[
(x, 0) = (x, 0 + 0) = (x, 0) + (x, 0)
\]

and so $(x, 0) = 0$. Thus, it can be assumed $y \neq 0$. Then from the axioms of the inner product,

\[
F(t) = \| x \|^2 + 2t \Re(x, \omega y) + t^2 \| y \|^2 \geq 0.
\]

This yields

\[
\| x \|^2 + 2t |(x, y)| + t^2 \| y \|^2 \geq 0.
\]

Since this inequality holds for all $t \in \mathbb{R}$, it follows from the quadratic formula that

\[
4| (x, y) |^2 - 4\| x \|^2 \| y \|^2 \leq 0.
\]

This yields the conclusion and proves the theorem.

**Proposition 6.0.4** For an inner product space, $\| x \| \equiv (x, x)^{1/2}$ does specify a norm.

**Proof:** All the axioms are obvious except the triangle inequality. To verify this,

\[
\| x + y \|^2 \equiv (x + y, x + y) \equiv \| x \|^2 + \| y \|^2 + 2 \Re(x, y)
\]

\[
\leq \| x \|^2 + \| y \|^2 + 2 |(x, y)|
\]

\[
\leq \| x \|^2 + \| y \|^2 + 2 \| x \| \| y \| = (\| x \| + \| y \|)^2.
\]

The following lemma is called the parallelogram identity.
Lemma 6.0.5 In an inner product space,
\[ \|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2. \]

The proof, a straightforward application of the inner product axioms, is left to the reader.

Lemma 6.0.6 For \( x \in H \), an inner product space,
\[ \|x\| = \sup_{\|y\| \leq 1} |(x, y)| \quad (6.0.4) \]

Proof: By the Cauchy Schwarz inequality, if \( x \neq 0 \),
\[ \|x\| \geq \sup_{\|y\| \leq 1} |(x, y)| \geq \left( x, \frac{x}{\|x\|} \right) = \|x\|. \]
It is obvious that (6.0.4) holds in the case that \( x = 0 \).

Definition 6.0.7 A Hilbert space is an inner product space which is complete. Thus a Hilbert space is a Banach space in which the norm comes from an inner product as described above.

In Hilbert space, one can define a projection map onto closed convex nonempty sets.

Definition 6.0.8 A set, \( K \), is convex if whenever \( \lambda \in [0, 1] \) and \( x, y \in K \), \( \lambda x + (1 - \lambda) y \in K \).

Theorem 6.0.9 Let \( K \) be a closed convex nonempty subset of a Hilbert space, \( H \), and let \( x \in H \). Then there exists a unique point \( Px \in K \) such that \( \|Px - x\| \leq \|y - x\| \) for all \( y \in K \).

Proof: Consider uniqueness. Suppose that \( z_1 \) and \( z_2 \) are two elements of \( K \) such that for \( i = 1, 2 \),
\[ \|z_i - x\| \leq \|y - x\| \quad (6.0.5) \]
for all \( y \in K \). Also, note that since \( K \) is convex,
\[ \frac{z_1 + z_2}{2} \in K. \]

Therefore, by the parallelogram identity,
\[
\|z_1 - x\|^2 \leq \frac{\|z_1 + z_2\|^2}{2} - x\|^2 = \frac{\|z_1 - x\|^2 + \frac{z_2 - x}{2}}{2}
\]
\[ = 2(\|z_2 - x\|^2 + \|z_2 - x\|^2) - \|z_1 - z_2\|^2
\]
\[ = \frac{1}{2}\|z_1 - x\|^2 + \frac{1}{2}\|z_2 - x\|^2 - \|z_1 - z_2\|^2
\]
\[ \leq \|z_1 - x\|^2 - \|z_1 - z_2\|^2, \]
where the last inequality holds because of (6.0.4) letting \( z_1 = z_2 \) and \( y = z_1 \). Hence \( z_1 = z_2 \) and this shows uniqueness.

Now let \( \lambda = \inf\{\|x - y\| : y \in K\} \) and let \( y_n \) be a minimizing sequence. This means \( \{y_n\} \subseteq K \) satisfies \( \lim_{n \to \infty} \|x - y_n\| = \lambda \). Now the following follows from properties of the norm.
\[ \|y_n - x + y_m - x\|^2 = 4(\frac{y_n + y_m}{2} - x\|^2) \]

Then by the parallelogram identity, and convexity of \( K \), \( \frac{y_n + y_m}{2} \in K \), and so
\[
\|y_n - x - (y_m - x)\|^2 = 2(\|y_n - x\|^2 + \|y_m - x\|^2) - 4(\frac{y_n + y_m}{2} - x\|^2)
\]
\[ \leq 2(\|y_n - x\|^2 + \|y_m - x\|^2) - 4\lambda^2. \]

Since \( \|x - y_n\| \to \lambda \), this shows \( \{y_n - x\} \) is a Cauchy sequence. Thus also \( \{y_n\} \) is a Cauchy sequence. Since \( H \) is complete, \( y_n \to y \) for some \( y \in H \) which must be in \( K \) because \( K \) is closed. Therefore
\[ \|x - y\| = \lim_{n \to \infty} \|x - y_n\| = \lambda. \]

Let \( Px = y \).
Corollary 6.0.10 Let $K$ be a closed, convex, nonempty subset of a Hilbert space, $H$, and let $x \in H$. Then for $z \in K$, $z = Px$ if and only if

$$\text{Re}(x - z, y - z) \leq 0$$

(6.0.6)

for all $y \in K$.

Before proving this, consider what it says in the case where the Hilbert space is $\mathbb{R}^n$.

Condition (6.0.6) says the angle, $\theta$, shown in the diagram is always obtuse. Remember from calculus, the sign of $x \cdot y$ is the same as the sign of the cosine of the included angle between $x$ and $y$. Thus, in finite dimensions, the conclusion of this corollary says that $z = Px$ exactly when the angle of the indicated angle is obtuse. Surely the picture suggests this is reasonable.

The inequality (6.0.6) is an example of a variational inequality and this corollary characterizes the projection of $x$ onto $K$ as the solution of this variational inequality.

Proof of Corollary: Let $z \in K$ and let $y \in K$ also. Since $K$ is convex, it follows that if $t \in [0,1]$,

$$z + t(y - z) = (1 - t)z + ty \in K.$$

Furthermore, every point of $K$ can be written in this way. (Let $t = 1$ and $y \in K$.) Therefore, $z = Px$ if and only if for all $y \in K$ and $t \in [0,1]$,

$$||x - (z + t(y - z))||^2 = ||(x - z) - t(y - z)||^2 \geq ||x - z||^2$$

for all $t \in [0,1]$ and $y \in K$ if and only if for all $t \in [0,1]$ and $y \in K$

$$||x - z||^2 + t^2||y - z||^2 - 2t\text{Re}(x - z, y - z) \geq ||x - z||^2$$

If and only if for all $t \in [0,1],

$$t^2||y - z||^2 - 2t\text{Re}(x - z, y - z) \geq 0.$$ (6.0.7)

Now this is equivalent to (6.0.6) holding for all $t \in (0,1)$. Therefore, dividing by $t \in (0,1)$, (6.0.7) is equivalent to

$$t||y - z||^2 - 2\text{Re}(x - z, y - z) \geq 0$$

for all $t \in (0,1)$ which is equivalent to (6.0.6). This proves the corollary.

Corollary 6.0.11 Let $K$ be a nonempty convex closed subset of a Hilbert space, $H$. Then the projection map, $P$ is continuous. In fact,

$$|Px - Py| \leq |x - y|.$$

Proof: Let $x, x' \in H$. Then by Corollary 6.0.10,

$$\text{Re} (x' - Px', Px - Px') \leq 0, \text{Re} (x - Px, Px' - Px) \leq 0$$

Hence

$$0 \leq \text{Re} (x - Px, Px - Px') - \text{Re} (x' - Px', Px - Px') = \text{Re} (x - x', Px - Px') - |Px - Px'|^2$$

and so

$$|Px - Px'|^2 \leq |x - x'| |Px - Px'|.$$

This proves the corollary.

The next corollary is a more general form for the Brouwer fixed point theorem.

Corollary 6.0.12 Let $f : K \to K$ where $K$ is a convex compact subset of $\mathbb{R}^n$. Then $f$ has a fixed point.
Proof: Let $K \subseteq B(0, R)$ and let $P$ be the projection map onto $K$. Then consider the map $f \circ P$ which maps $B(0, R)$ to $B(0, R)$ and is continuous. By the Brouwer fixed point theorem for balls, this map has a fixed point. Thus there exists $x$ such that

$$f \circ P(x) = x$$

Now the equation also requires $x \in K$ and so $P(x) = x$. Hence $f(x) = x$.

Definition 6.0.13 Let $H$ be a vector space and let $U$ and $V$ be subspaces. $U \oplus V = H$ if every element of $H$ can be written as a sum of an element of $U$ and an element of $V$ in a unique way.

The case where the closed convex set is a closed subspace is of special importance and in this case the above corollary implies the following.

Corollary 6.0.14 Let $K$ be a closed subspace of a Hilbert space, $H$, and let $x \in H$. Then for $z \in K$, $z = Px$ if and only if

$$(x - z, y) = 0$$

for all $y \in K$. Furthermore, $H = K \oplus K^\perp$ where

$$K^\perp \equiv \{x \in H : (x, k) = 0 \text{ for all } k \in K\}$$

and

$$||x||^2 = ||x - Px||^2 + ||Px||^2.$$  

Proof: Since $K$ is a subspace, the condition implies $\Re(x - z, y) \leq 0$ for all $y \in K$. Replacing $y$ with $-y$, it follows $\Re(x - z, -y) \leq 0$ which implies $\Re(x - z, y) \geq 0$ for all $y$. Therefore, $\Re(x - z, y) = 0$ for all $y \in K$. Now let $|\alpha| = 1$ and $\alpha(x - z, y) = |(x - z, y)|$. Since $K$ is a subspace, it follows $\alpha y \in K$ for all $y \in K$. Therefore,

$$0 = \Re(x - z, \alpha y) = (x - z, \alpha y) = \alpha (x - z, y) = |(x - z, y)|.$$

This shows that $z = Px$, if and only if.

For $x \in H$, $x = x - Px + Px$ and from what was just shown, $x - Px \in K^\perp$ and $Px \in K$. This shows that $K^\perp + K = H$. Is there only one way to write a given element of $H$ as a sum of a vector in $K$ with a vector in $K^\perp$? Suppose $y + z = y_1 + z_1$ where $z, z_1 \in K^\perp$ and $y, y_1 \in K$. Then $(y - y_1) = (z_1 - z)$ and so from what was just shown, $(y - y_1, y - y_1) = (y - y_1, z_1 - z) = 0$ which shows $y_1 = y$ and consequently $z_1 = z$. Finally, letting $z = Px$,

$$||x||^2 = (x - z + z, x - z + z) = ||x - z||^2 + (x - z, z) + (z, x - z) + ||z||^2$$

This proves the corollary.

The following theorem is called the Riesz representation theorem for the dual of a Hilbert space. If $z \in H$ then define an element $f \in H'$ by the rule $(x, z) \equiv f(x)$. It follows from the Cauchy Schwarz inequality and the properties of the inner product that $f \in H'$. The Riesz representation theorem says that all elements of $H'$ are of this form.

Theorem 6.0.15 Let $H$ be a Hilbert space and let $f \in H'$. Then there exists a unique $z \in H$ such that

$$f(x) = (x, z)$$

for all $x \in H$.

Proof: Letting $y, w \in H$ the assumption that $f$ is linear implies

$$f(yf(w) - f(y)w) = f(w)f(y) - f(y)f(w) = 0$$

which shows that $yf(w) - f(y)w \in f^{-1}(0)$, which is a closed subspace of $H$ since $f$ is continuous. If $f^{-1}(0) = H$, then $f$ is the zero map and $z = 0$ is the unique element of $H$ which satisfies $f(x) = 0$. If $f^{-1}(0) \neq H$, pick $u \notin f^{-1}(0)$ and let $w \equiv u - Pu \neq 0$. Thus Corollary 6.0.14 implies $(y, w) = 0$ for all $y \in f^{-1}(0)$. In particular, let $y = xf(w) - f(x)w$ where $x \in H$ is arbitrary. Therefore,

$$0 = (f(w)x - f(x)w, w) = f(w)(x, w) - f(x)||w||^2.$$
Thus, solving for $f(x)$ and using the properties of the inner product,

$$f(x) = (x, \frac{f(w)w}{\|w\|^2})$$

Let $z = \overline{f(w)w}/\|w\|^2$. This proves the existence of $z$. If $f(x) = (x, z_i) i = 1, 2$, for all $x \in H$, then for all $x \in H$, then $(x, z_1 - z_2) = 0$ which implies, upon taking $x = z_1 - z_2$ that $z_1 = z_2$. This proves the theorem.

If $R : H \to H'$ is defined by $Rx(y) = (y, x)$, the Riesz representation theorem above states this map is onto. This map is called the Riesz map. It is routine to show $R$ is linear and $|Rx| = |x|$.

### 6.1 The Hilbert Space $L(U)$

Let $L \in \mathcal{L}(U, H)$. Then one can consider the image of $L, L(U)$ as a Hilbert space. This is another interesting application of Theorem 6.0.9. First here is a definition which involves abominable and atrociously misleading notation which nevertheless seems to be well accepted.

**Definition 6.1.1** Let $L \in \mathcal{L}(U, H)$, the bounded linear maps from $U$ to $H$ for $U, H$ Hilbert spaces. For $y \in L(U)$, let $L^{-1}y$ denote the unique vector in

$$\{x : Lx = y\} \equiv M_y$$

which is closest in $U$ to 0.

Note this is a good definition because $\{x : Lx = y\}$ is closed thanks to the continuity of $L$ and it is obviously convex. Thus Theorem 6.0.9 applies. With this definition define an inner product on $L(U)$ as follows. For $y, z \in L(U)$,

$$(y, z)_{L(U)} \equiv (L^{-1}y, L^{-1}z)_U$$

The notation is abominable because $L^{-1}(y)$ is the normal notation for $M_y$.

In terms of linear algebra, this $L^{-1}$ is the Moore Penrose inverse. There you obtain the least squares solution $x$ to $Lx = y$ which has smallest norm. Here there is an actual solution and among those solutions you get the one which has least norm. Of course a real honest solution is also a least squares solution so this is the Moore Penrose inverse restricted to $L(U)$.

First I want to understand $L^{-1}$ better. It is actually fairly easy to understand in terms of geometry. Here is a picture of $L^{-1}(y)$ for $y \in L(U)$.

As indicated in the picture, here is a lemma which gives a description of the situation.
Lemma 6.1.2 In the context of the above definition, \( L^{-1}(y) \) is characterized by

\[
(L^{-1}(y), x)_U = 0 \quad \text{for all } x \in \ker(L)
\]

\[
L(L^{-1}(y)) = y, \quad (L^{-1}(y)) \in M_y
\]

In addition to this, \( L^{-1} \) is linear and the above definition does define an inner product.

Proof: The point \( L^{-1}(y) \) is well defined as noted above. I claim it is characterized by the following for \( y \in L(U) \)

\[
(L^{-1}(y), x)_U = 0 \quad \text{for all } x \in \ker(L)
\]

\[
L(L^{-1}(y)) = y, \quad (L^{-1}(y)) \in M_y
\]

Let \( w \in M_y \) and suppose

\[
(v, x)_U = 0, L(v) = y
\]

Then from the above characterization,

\[
||w||^2 = \left( \frac{\in \ker(L)}{w = v + v} \right)^2 = ||w - v||^2 + ||v||^2
\]

which shows that \( w = L^{-1}(y) \) if and only if \( w = v \) just described. From this characterization, it is clear that \( L^{-1} \) is linear. Then it is also obvious that

\[
(y, z)_{L(U)} = (L^{-1}y, L^{-1}z)_U
\]

also specifies an inner product. The algebraic axioms are all obvious because \( L^{-1} \) is linear. If \((y, y)_{L(U)} = 0, \) then \( |L^{-1}y|_U^2 = 0 \) and so \( L^{-1}y = 0 \) which requires \( y = L(L^{-1}y) = 0. \]

With the above definition, here is the main result.

Theorem 6.1.3 Let \( U, H \) be Hilbert spaces and let \( L \in \mathcal{L}(U, H) \). Then Definition \( [\text{\ref{def:inverse}}] \) makes \( L(U) \) into a Hilbert space. Also \( L : U \rightarrow L(U) \) is continuous and \( L^{-1} : L(U) \rightarrow U \) is continuous. Also,

\[
\|L\|_{\mathcal{L}(U,H)} \|Lx\|_{L(U)} \geq \|Lx\|_H \quad (\text{6.1.11})
\]

If \( U \) is separable, so is \( L(U) \). Also \( (L^{-1}(y), x) = 0 \) for all \( x \in \ker(L) \), and \( L^{-1} : L(U) \rightarrow U \) is linear. Also, in case that \( L \) is one to one, both \( L \) and \( L^{-1} \) preserve norms.

Proof: First consider the claim that \( L : U \rightarrow L(U) \) is continuous and \( L^{-1} : L(U) \rightarrow U \) is also continuous. Why is \( L \) continuous? Say \( u_n \rightarrow 0 \) in \( U \). Then

\[
\|Lu_n\|_{L(U)} = \|L^{-1}(L(u_n))\|_U
\]

Now \( \|L^{-1}(L(u_n))\|_U \leq \|u_n\|_U \) and so it converges to 0. (Recall that \( L^{-1}(Lu_n) \) is the smallest vector in \( U \) which maps to \( Lu_n \). Since \( u_n \) is mapped by \( L \) to \( Lu_n \), it follows that \( \|L^{-1}(L(u_n))\|_U \leq \|u_n\|_U \). Hence \( L \) is continuous.

Next, why is \( L^{-1} \) continuous? Let \( \|y_n\|_{L(U)} \rightarrow 0 \). This requires \( \|L^{-1}(y_n)\|_U \rightarrow 0 \) by definition of the norm in \( L(U) \). Thus \( L^{-1} \) is continuous.

Why is \( L(U) \) a Hilbert space? Let \( \{y_n\} \) be a Cauchy sequence in \( L(U) \). Then from what was just observed, it follows that \( L^{-1}(y_n) \) is a Cauchy sequence in \( U \). Hence \( L^{-1}(y_n) \rightarrow x \in U \). It follows that \( y_n = L(L^{-1}(y_n)) \rightarrow Lx \) in \( L(U) \). This is in the norm of \( L(U) \). It was just shown that \( L \) is continuous as a map from \( U \) to \( L(U) \). This shows that \( L(U) \) is a Hilbert space. It was already shown that it is an inner product space and this has shown that it is complete.

If \( x \in U \), then \( \|Lx\|_H \leq \|L\|_{\mathcal{L}(U,H)} \|x\|_U \). It follows that

\[
\|L(x)\|_H = \|L(L^{-1}(L(x)))\|_H \leq \|L\|_{\mathcal{L}(U,H)} \|L^{-1}(L(x))\|_U
\]

\[
= \|L\|_{\mathcal{L}(U,H)} \|L(x)\|_{L(U)}.
\]

This verifies \( \text{\ref{def:inverse}} \).
If $U$ is separable, then letting $D$ be a countable dense subset, it follows from the continuity of the operators $L, L^{-1}$ discussed above that $L(D)$ is separable in. To see this, note that

$$
\|Lx_n - Lx\|_{L(U)} = \|L \left( L^{-1}(Lx_n) - Lx \right) \| \leq \|L\|_{\mathcal{L}(U,H)} \|L^{-1}(L(x_n-x))\|_U \leq \|L\|_{\mathcal{L}(U,H)} \|x_n-x\|_U
$$

As before, $L^{-1}(L(x_n-x))$ is the smallest vector which maps onto $L(x_n-x)$ and so its norm is no larger than $\|x_n-x\|_U$.

Consider the last claim. If $L$ is one to one, then for $y \in L(U)$, there is only one vector which maps to $y$. Therefore,

$$
L^{-1}(L(x)) = x.
$$

Hence for $y \in L(U)$,

$$
\|y\|_{L(U)} = \|L^{-1}(y)\|_U
$$

Also,

$$
\|Lu\|_{L(U)} = \|L^{-1}(L(u))\|_U = \|u\|_U
$$

Now here is another argument for various continuity claims.

$$
\|Lx\|_{L(U)} = \|L^{-1}(Lx)\|_U \leq \|x\|_U
$$

because $L^{-1}(Lx)$ is the smallest thing in $U$ which maps to $Lx$ and $x$ is something which maps to $Lx$ so it follows that the inequality holds. Hence $L \in \mathcal{L}(U, L(U))$ and in fact, $\|L\|_{\mathcal{L}(U,L(U))} = 1$. Next, letting $y \in L(U)$,

$$
\|L^{-1}y\|_U = \|y\|_{L(U)}
$$

and so $\|L^{-1}\|_{\mathcal{L}(L(U), U)} = 1$ and this shows that $L \in \mathcal{L}(U, L(U))$ while $L^{-1} \in \mathcal{L}(L(U), U)$ and both have norm equal to $1$.

Now

$$
\|Lx\|_H = \|L \left( L^{-1}(Lx) \right) \|_H \leq \|L\|_{\mathcal{L}(U,H)} \|L^{-1}(Lx)\|_U \equiv \|L\|_{\mathcal{L}(U,H)} \|Lx\|_{L(U)}
$$

Now here are some other very interesting results. I am following Lemma 6.1.4 on Page 66.

**Lemma 6.1.4** Let $L \in \mathcal{L}(U, H)$. Then $L \left( \overline{B(0,r)} \right)$ is closed and convex.

**Proof**: It is clear this is convex since $L$ is linear. Why is it closed? $\overline{B(0,r)}$ is compact in the weak topology by the Banach Alaoglu theorem, Theorem 6.3.2 on Page 66. Furthermore, $L$ is continuous with respect to the weak topologies on $U$ and $H$. Here is why this is so. Suppose $u_n \to u$ weakly in $U$. Then if $h \in H$,

$$
(Lu_n, h) = (u_n, L^*h) \to (u, L^*h) = (Lu, h)
$$

which shows $Lu_n \to Lu$ weakly. Therefore, $L \left( \overline{B(0,r)} \right)$ is weakly compact because it is the continuous image of a compact set. Therefore, it must also be weakly closed because the weak topology is a Hausdorff space. (See Lemma 6.3.4 on Page 66 and so you can apply the separation theorem, Theorem 6.3.4 on Page 66 to obtain a separating functional. Thus if $x \neq y$, there exists $f \in H'$ such that $Re f(y) > c > Re f(x)$ and so taking

$$
2r < \min (c - Re f(x), Re f(y) - c),
$$

$$
B_f(x, r) \cap B_f(y, r) = \emptyset
$$

where

$$
B_f(x, r) = \{ y \in H : |f(x-y)| < r \}
$$

is an example of a basic open set in the weak topology.)

Now suppose $p \notin L \left( \overline{B(0,r)} \right)$. Since the set is weakly closed and convex, it follows by Theorem 6.3.4 and the Riesz representation theorem for Hilbert space that there exists $z \in H$ such that

$$
Re (p, z) > c > Re (Lx, z)
$$

for all $x \in \overline{B(0,r)}$. Therefore, $p$ cannot be a strong limit point because if it were, there would exist $x_n \in \overline{B(0,r)}$ such that $Lx_n \to p$ which would require $Re (Lx_n, z) \to Re (p, z)$ which is prevented by the above inequality. This proves the lemma.
Thus for all $x \in H$ the latter set being a closed convex nonempty set thanks to Lemma 6.1.2. Then from Theorem 6.1.15, if it is not so, then there exists $u_0, ||u_0||_1 \leq 1$ but

$$T_1 (u_0) \notin T_2 \left( B \left( 0, c \right) \right)$$

the latter set being a closed convex nonempty set thanks to Lemma 6.1.2. Then by the separation theorem, Theorem 6.1.4, there exists $z \in H$ such that

$$\text{Re} \left( T_1 (u_0) , z \right)_H > 1 > \text{Re} \left( T_2 (v) , z \right)_H$$

for all $||v||_2 \leq c$. Therefore, replacing $v$ with $v\theta$ where $\theta$ is a suitable complex number having modulus 1, it follows

$$||T_1^* z|| > 1 > |(v, T_2^* z)_{U_2}|$$

for all $||v||_2 \leq c$. If $c = 0$, 6.1.12 gives a contradiction immediately because of 6.1.2. Assume then that $c > 0$. From 6.1.12, if $||v||_2 \leq 1$, then

$$|(v, T_2^* z)_{U_2}| < \frac{1}{c} < \frac{1}{c} ||T_1^* z||$$

Then from 6.1.12,

$$||T_2^* z||_{U_2} = \sup_{||v||_2 \leq 1} |(v, T_2^* z)_{U_2}| \leq \frac{1}{c} < \frac{1}{c} ||T_1^* z||$$

which contradicts 6.1.12. Therefore, it is clear that $T_1 (U_1) \subseteq T_2 (U_2)$.

Now consider the second claim. The first part shows $T_1 (U_1) = T_2 (U_2)$. Denote by $u_i \in U_i$, the point $T_i^{-1} x$. Without loss of generality, it can be assumed $x \neq 0$ because if $x = 0$, then the definition of $T_i^{-1}$ gives $T_i^{-1} (x) = 0$. Thus for $x \neq 0$ neither $u_i$ can equal 0. I need to verify that $||u_1||_1 = ||u_2||_2$. Suppose then that this is not so. Say $||u_1||_1 > ||u_2||_2 > 0$.

$$\frac{x}{||u_2||_2} = T_2 \left( \frac{u_2}{||u_2||_2} \right) \in T_2 \left( B \left( 0, 1 \right) \right)$$

But from the first part of the theorem this equals $T_1 \left( B \left( 0, 1 \right) \right)$ and so there exists $u'_1 \in B \left( 0, 1 \right)$ such that

$$\frac{x}{||u_2||_2} = T_1 u'_1$$

Hence

$$T_1 \left( u'_1 - \frac{u_1}{||u_2||_2} \right) = \frac{x}{||u_2||_2} - \frac{x}{||u_2||_2} = 0.$$

From Theorem 6.1.2 this implies

$$0 = \left( u'_1, u'_1 - \frac{u_1}{||u_2||_2} \right) \leq ||u_1||_1 ||u'_1||_1 - ||u_1||_1 \frac{||u_1||_1}{||u_2||_2}$$

$$= ||u_1||_1 \left( ||u'_1||_1 - \frac{||u_1||_1}{||u_2||_2} \right) \leq ||u_1||_1 \left( 1 - \frac{||u_1||_1}{||u_2||_2} \right)$$

which is a contradiction because it was assumed $\frac{||u_1||_1}{||u_2||_2} > 1$. This proves the theorem.
6.2 Compact Operators

6.2.1 Compact Operators In Hilbert Space

Definition 6.2.1 Define $v \otimes u \in \mathcal{L}(H, H)$ by

$$v \otimes u(x) = (x, u)v.$$ 

$A \in \mathcal{L}(H, H)$ is a compact operator if whenever $\{x_k\}$ is a bounded sequence, there exists a convergent subsequence of $\{Ax_k\}$. Equivalently, $A$ maps bounded sets to sets whose closures are compact or to use other terminology, $A$ maps bounded sets to sets which are precompact.

Lemma 6.2.2 Let $H$ be a separable Hilbert space and suppose $A \in \mathcal{L}(H, H)$ is a compact operator. Let $B$ denote the closed unit ball in $H$. Then $A$ is continuous as a map from $B$ with the weak topology into $H$ with the strong topology. For $u, v \in H$, $v \otimes u : H \to H$ is a compact operator. If $A$ is self adjoint and compact, the function

$$x \to (Ax, x)$$

is continuous on $B$ with respect to the weak topology on $B$. The function,

$$x \to (v \otimes u(x), x)$$

is continuous and the operator $u \otimes u$ is self adjoint.

Proof: Since $H$ is separable, it follows from Corollary [4] on Page 17 that $B$ can be considered as a metric space. Therefore, showing continuity reduces to showing convergent sequences are taken to convergent sequences. Let $x_n \to x$ weakly in $B$. Suppose $Ax_n$ does not converge to $Ax$. Then there exists a subsequence, still denoted by $\{x_n\}$ such that

$$\|Ax_n - Ax\| \geq \varepsilon > 0 \quad (6.2.16)$$

for all $n$. Then since $A$ maps bounded sets to compact sets, there is a further subsequence, still denoted by $\{x_n\}$ such that $Ax_n$ converges to some $y \in H$. Therefore,

$$(y, w) = \lim_{n \to \infty} (Ax_n, w) = \lim_{n \to \infty} (x_n, A^*w)$$

$$= (x, A^*w) = (Ax, w)$$

which shows $Ax = y$ since $w$ is arbitrary. However, this contradicts (6.2.16).

Next consider the claim about $v \otimes u$. Letting $\{x_n\}$ be a bounded sequence,

$$v \otimes u(x_n) = (x_n, u)v.$$ 

There exists a weakly convergent subsequence of $\{x_n\}$ say $\{x_{n_k}\}$ converging weakly to $x \in H$. Therefore,

$$\|v \otimes u(x_{n_k}) - v \otimes u(x)\| = \|(x_{n_k}, u) - (x, u)\| \|v\|$$

which converges to 0. Thus $v \otimes u$ is compact as claimed. It takes bounded sets to precompact sets.

To verify the assertion about $x \to (Ax, x)$, let $x_n \to x$ weakly. Then

$$|(Ax_n, x_n) - (Ax, x)|$$

$$\leq |(Ax_n, x_n) - (Ax, x_n)| + |(Ax, x_n) - (Ax, x)|$$

$$\leq |(Ax_n, x_n) - (Ax, x_n)| + |(Ax_n, x) - (Ax, x)|$$

$$\leq ||Ax_n - Ax|| \|x_n\| + ||Ax_n - Ax|| \|x\| \leq 2||Ax_n - Ax||$$

which converges to 0.

$$|(v \otimes u(x_n), x_n) - (v \otimes u(x), x)|$$

$$= |(x_n, u)(v, x_n) - (x, u)(v, x)|$$

and this converges to 0 by weak convergence. It follows from the definition that $u \otimes u$ is self adjoint. ■
Observation 6.2.3  Note that if \( A \) is any self adjoint operator,
\[
(Ax, x) = (x, Ax) = (Ax, x).
\]
so \((Ax, x)\) is real valued.

The big result is called the Hilbert Schmidt theorem. It is a generalization to arbitrary Hilbert spaces of standard finite dimensional results having to do with diagonalizing a symmetric matrix.

Lemma 6.2.4  Let \( A \in \mathcal{L}(H, H) \) and suppose it is self adjoint and compact. Let \( B \) denote the closed unit ball in \( H \). Let \( e \in B \) be such that
\[
|(Ae, e)| = \max_{x \in B} |(Ax, x)|.
\]
Then letting \( \lambda = (Ae, e) \), it follows \( Ae = \lambda e \). If \( \lambda \neq 0 \), then \(|e| = 1 \) and if \( \lambda = 0 \), it can be assumed \( e = 0 \) so it is still the case \( Ae = \lambda e \).

Proof:  From the above observation, \((Ax, x)\) is always real and since \( A \) is compact, \(|(Ax, x)|\) achieves a maximum at \( e \). It remains to verify \( e \) is an eigenvector. Note that \(|e| = 1\) whenever \( \lambda \neq 0 \) since otherwise \(|(Ae, e)|\) could be made larger by replacing \( e \) with \( e/||e|| \).

Suppose \( \lambda = (Ae, e) > 0 \). Then it is easy to verify that \( \lambda I - A \) is a nonnegative \((((\lambda I - A)x, x) \geq 0 \) for all \( x \)) and self adjoint operator. Therefore, the Cauchy-Schwarz inequality can be applied to write
\[
((\lambda I - A)e, x) \leq ((\lambda I - A)e, e)^{1/2}((\lambda I - A)x, x)^{1/2} = 0
\]
Since this is true for all \( x \) it follows \( Ae = \lambda e \).

Next suppose \( \lambda = (Ae, e) < 0 \). Then \(-\lambda = (-Ae, e)\) and the previous result can be applied to \(-A\) and \(-\lambda\). Thus \(-\lambda e = -Ae\) and so \( Ae = \lambda e \).

Finally consider the case where \( \lambda = 0 \). Then \( 0 = (A0, 0) \) and so it suffices to take \( e = 0 \) as claimed. This proves the lemma.

With these lemmas here is a major theorem, the Hilbert Schmidt theorem.

Theorem 6.2.5  Let \( A \in \mathcal{L}(H, H) \) be a compact self adjoint operator on a Hilbert space. Then there exist real numbers \( \{\lambda_k\}_{k=1}^{\infty} \) and vectors \( \{e_k\}_{k=1}^{\infty} \) such that
\[
\begin{align*}
|e_k| & = 1 \text{ if } \lambda_k \neq 0, \\
|e_k| & = 0 \text{ if } \lambda_k = 0, \\
(e_k, e_j)_H & = 0 \text{ if } k \neq j, \\
Ae_k & = \lambda_k e_k, \\
|\lambda_n| & \geq |\lambda_{n+1}| \text{ for all } n, \\
\lim_{n \to \infty} \lambda_n & = 0, \\
\lim_{n \to \infty} \left\| \sum_{k=1}^{n} \lambda_k (e_k \otimes e_k) \right\|_{\mathcal{L}(H,H)} & = 0. \quad (6.2.17)
\end{align*}
\]

Proof:  This is done by considering a sequence of compact self adjoint operators, \( A, A_1, A_2, \cdots \). Here is how these are defined. Using Lemma 6.2.4 let \( e_1, \lambda_1 \) be given by that lemma such that
\[
|(Ae_1, e_1)| = \max_{x \in B} |(Ax, x)|, \ \lambda_1 = (Ae_1, e_1).
\]
Then by that lemma, \( Ae_1 = \lambda_1 e_1 \) and \(|e_1| = 1 \) if \( \lambda_1 \neq 0 \) while \( e_1 = 0 \) if \( \lambda_1 = 0 \).

If \( A_n \) has been obtained, use Lemma 6.2.4 to obtain \( e_{n+1} \) and \( \lambda_{n+1} \) such that
\[
|(A_n e_{n+1}, e_{n+1})| = \max_{x \in B} |(A_n x, x)|, \ \lambda_{n+1} = (A_n e_{n+1}, e_{n+1}).
\]
By that lemma again, \( A_n e_{n+1} = \lambda_{n+1} e_{n+1} \) and \(|e_{n+1}| = 1 \) if \( \lambda_{n+1} \neq 0 \) while \( e_{n+1} = 0 \) if \( \lambda_{n+1} = 0 \). Then
\[
A_{n+1} = A_n - \lambda_{n+1} e_{n+1} \otimes e_{n+1}
\]
Thus
\[ A_n = A - \sum_{k=1}^{n} \lambda_k e_k \otimes e_k. \] (6.2.18)

**Claim 1:** If \( k < n + 1 \) then \( (e_{n+1}, e_k) = 0 \). Also \( Ae_k = \lambda_k e_k \) for all \( k \).

**Proof of claim:** From the above,
\[ \lambda_{n+1} e_{n+1} = A_n e_{n+1} = Ae_{n+1} - \sum_{k=1}^{n} \lambda_k (e_{n+1}, e_k) e_k. \]

If \( \lambda_{n+1} = 0 \), then \( (e_{n+1}, e_k) = 0 \) because \( e_{n+1} = 0 \). If \( \lambda_{n+1} \neq 0 \), then from the above and an induction hypothesis
\[
\lambda_{n+1} (e_{n+1}, e_j) = (Ae_{n+1}, e_j) - \sum_{k=1}^{n} \lambda_k (e_{n+1}, e_k) (e_k, e_j)
\]
\[
= (e_{n+1}, Ae_j) - \sum_{k=1}^{n} \lambda_k (e_{n+1}, e_k) (e_k, e_j)
\]
\[
= \lambda_j (e_{n+1}, e_j) - \lambda_j (e_{n+1}, e_j) = 0.
\]

To verify the second part of this claim,
\[ \lambda_{n+1} e_{n+1} = A_n e_{n+1} = Ae_{n+1} - \sum_{k=1}^{n} \lambda_k e_k (e_{n+1}, e_k) = Ae_{n+1} \]

This proves the claim.

**Claim 2:** \( |\lambda_n| \geq |\lambda_{n+1}| \).

**Proof of claim:** From (6.2.18) and the definition of \( A_n \) and \( e_k \otimes e_k \),
\[
\lambda_{n+1} = (A_n e_{n+1}, e_{n+1})
\]
\[
= (A_{n-1} e_{n+1}, e_{n+1}) - \lambda_n |(e_n, e_{n+1})|^2
\]
\[
= (A_{n-1} e_{n+1}, e_{n+1})
\]
By the previous claim, Therefore,
\[
|\lambda_{n+1}| = |(A_{n-1} e_{n+1}, e_{n+1})| \leq |(A_{n-1} e_n, e_n)| = |\lambda_n|
\]
by the definition of \( |\lambda_n| \). \( e_n \) makes \( |(A_{n-1} x, x)| \) as large as possible, not necessarily \( e_{n+1} \).

**Claim 3:** \( \lim_{n \to \infty} \lambda_n = 0 \).

**Proof of claim:** If for some \( n \), \( \lambda_n = 0 \), then \( \lambda_k = 0 \) for all \( k > n \) by claim 2. Assume then that \( \lambda_k \neq 0 \) for any \( k \). Then if \( \lim_{k \to \infty} |\lambda_k| = \varepsilon > 0 \), contrary to the claim, \( ||e_k|| = 1 \) for all \( k \) and
\[
||Ae_n - Ae_m||^2 = ||\lambda_n e_n - \lambda_m e_m||^2
\]
\[
= \lambda_n^2 + \lambda_m^2 \geq 2\varepsilon^2
\]
which shows there is no Cauchy subsequence of \( \{Ae_n\}_{n=1}^{\infty} \), which contradicts the compactness of \( A \). This proves the claim.

**Claim 4:** \( ||A_n|| \to 0 \)

**Proof of claim:** Let \( x, y \in B \)
\[
|\lambda_{n+1}| \geq \left| \left( A_n \frac{x + y}{2}, \frac{x + y}{2} \right) \right|
\]
\[
= \frac{1}{4} (A_n x, x) + \frac{1}{4} (A_n y, y) + \frac{1}{2} (A_n x, y)
\]
\[
\geq \frac{1}{2} |(A_n x, y)| - \frac{1}{4} |(A_n x, x) + (A_n y, y)|
\]
\[
\geq \frac{1}{2} |(A_n x, y)| - \frac{1}{4} (|(A_n x, x)| + |(A_n y, y)|)
\]
\[
\geq \frac{1}{2} |(A_n x, y)| - \frac{1}{2} |\lambda_{n+1}|
\]
and so
\[ 3|\lambda_{n+1}| \geq |(A_n x, y)|. \]
It follows \(|A_n| \leq 3|\lambda_{n+1}|\). This proves the claim.

By this proves and completes the proof.

It is convenient to write the main result of the above theorem in the form described in the following corollary.

**Corollary 6.2.6** The main conclusion of the above theorem can be written as
\[ A = \sum_{k=1}^{\infty} \lambda_k e_k \otimes e_k \]
where the convergence of the partial sums takes place in the operator norm and the \(e_k\) are orthonormal.

### 6.2.2 Nuclear (Trace Class) Operators

**Definition 6.2.7** A self adjoint operator \(A \in \mathcal{L}(H, H)\) for \(H\) a separable Hilbert space is called a nuclear operator if for some complete orthonormal set, \(\{e_k\}\),
\[ \sum_{k=1}^{\infty} |(A e_k, e_k)| < \infty \]

This is also called a trace class operator. Essentially you can take the trace of the infinite matrix whose \(ij^{th}\) entry is \((A e_i, e_j)\).

To begin with here is an interesting lemma.

**Lemma 6.2.8** Suppose \(\{A_n\}\) is a sequence of compact operators in \(\mathcal{L}(X, Y)\) for two Banach spaces, \(X\) and \(Y\) and suppose \(A \in \mathcal{L}(X, Y)\) and
\[ \lim_{n \to \infty} ||A - A_n|| = 0. \]
Then \(A\) is also compact.

**Proof:** Let \(B\) be a bounded set in \(X\) such that \(||b|| \leq C\) for all \(b \in B\). I need to verify \(AB\) is totally bounded. Suppose then it is not. Then there exists \(\varepsilon > 0\) and a sequence, \(\{Ab_i\}\) where \(b_i \in B\) and
\[ ||Ab_i - Ab_j|| \geq \varepsilon \]
whenever \(i \neq j\). Then let \(n\) be large enough that
\[ ||A - A_n|| \leq \frac{\varepsilon}{4C}. \]
Then
\[ ||A_n b_i - A_n b_j|| = ||Ab_i - Ab_j + (A_n - A) b_i - (A_n - A) b_j|| \geq ||Ab_i - Ab_j|| - ||(A_n - A) b_i|| - ||(A_n - A) b_j|| \geq ||Ab_i - Ab_j|| - \frac{\varepsilon}{4C} - \frac{\varepsilon}{4C} \geq \frac{\varepsilon}{2}, \]
a contradiction to \(A_n\) being compact. This proves the lemma.

Then one can prove the following lemma. In this lemma, \(A \geq 0\) will mean \((Ax, x) \geq 0\).

**Lemma 6.2.9** Let \(A \geq 0\) be a nuclear operator defined on a separable Hilbert space, \(H\). Then \(A\) is compact and also, whenever \(\{e_k\}\) is a complete orthonormal set,
\[ A = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} (A e_i, e_j) e_i \otimes e_j. \]
6.2. COMPACT OPERATORS

**Proof:** First consider the formula. Since $A$ is given to be continuous,

$$Ax = A \left( \sum_{j=1}^{\infty} (x, e_j) e_j \right) = \sum_{j=1}^{\infty} (x, e_j) A e_j,$$

the series converging because

$$x = \sum_{j=1}^{\infty} (x, e_j) e_j$$

Then also since $A$ is self adjoint,

$$\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} (A e_i, e_j) e_i \otimes e_j (x) = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} (A e_i, e_j) e_i$$

$$= \sum_{j=1}^{\infty} (x, e_j) \sum_{i=1}^{\infty} (A e_i, e_j) e_i$$

$$= \sum_{j=1}^{\infty} (x, e_j) \sum_{i=1}^{\infty} (A e_i, e_j) e_i$$

$$= \sum_{j=1}^{\infty} (x, e_j) A e_j$$

Next consider the claim that $A$ is compact. Let $C_A \equiv \left( \sum_{j=1}^{\infty} |(A e_j, e_j)| \right)^{1/2}$. Let $A_n$ be defined by

$$A_n = \sum_{j=1}^{\infty} \sum_{i=1}^{n} (A e_i, e_j) (e_i \otimes e_j).$$

Then $A_n$ has values in $\text{span}(e_1, \cdots, e_n)$ and so it must be a compact operator because bounded sets in a finite dimensional space must be precompact. Then

$$|(Ax - A_n x, y)| = \left| \sum_{j=1}^{\infty} \sum_{i=n+1}^{\infty} (A e_i e_j) (y, e_j) (e_i, x) \right|$$

$$= \left| \sum_{j=1}^{\infty} (y, e_j) \sum_{i=n+1}^{\infty} (A e_i e_j) (e_i, x) \right|$$

$$\leq \left| \sum_{j=1}^{\infty} |(y, e_j)| (A e_j, e_j)^{1/2} \sum_{i=n+1}^{\infty} (A e_i e_i)^{1/2} |(e_i, x)| \right|$$

$$\leq \left( \sum_{j=1}^{\infty} |(y, e_j)|^2 \right)^{1/2} \left( \sum_{j=1}^{\infty} |(A e_j, e_j)| \right)^{1/2}$$

$$\cdot \left( \sum_{i=n+1}^{\infty} |(x, e_i)|^2 \right)^{1/2} \left( \sum_{i=n+1}^{\infty} |(A e_i e_i)| \right)^{1/2}$$

$$\leq |y| |x| C_A \left( \sum_{i=n+1}^{\infty} |(A e_i, e_i)| \right)^{1/2}$$

and this shows that if $n$ is sufficiently large,

$$|(A - A_n) x, y) | \leq \varepsilon |x| |y|.$$
Therefore,
\[
\lim_{n \to \infty} \|A - A_n\| = 0
\]
and so \(A\) is the limit in operator norm of finite rank bounded linear operators, each of which is compact. Therefore, \(A\) is also compact.

**Definition 6.2.10** The trace of a nuclear operator \(A \in \mathcal{L}(H, H)\) such that \(A \geq 0\) is defined to equal
\[
\sum_{k=1}^{\infty} (Ae_k, e_k)
\]
where \(\{e_k\}\) is an orthonormal basis for the Hilbert space, \(H\).

**Theorem 6.2.11** Definition 6.2.10 is well defined and equals \(\sum_{j=1}^{\infty} \lambda_j\) where the \(\lambda_j\) are the eigenvalues of \(A\).

**Proof:** Suppose \(\{u_k\}\) is some other orthonormal basis. Then
\[
e_k = \sum_{j=1}^{\infty} u_j (e_k, u_j)
\]
By Lemma 6.2.9 \(A\) is compact and so
\[
A = \sum_{k=1}^{\infty} \lambda_k u_k \otimes u_k
\]
where the \(u_k\) are the orthonormal eigenvectors of \(A\) which form a complete orthonormal set. Then
\[
\begin{align*}
\sum_{k=1}^{\infty} (Ae_k, e_k) &= \sum_{k=1}^{\infty} \left( A \left( \sum_{j=1}^{\infty} u_j (e_k, u_j) \right), \sum_{j=1}^{\infty} u_j (e_k, u_j) \right) \\
&= \sum_{k=1}^{\infty} \sum_{ij} (Au_j, u_i) (e_k, u_j) (u_i, e_k) \\
&= \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} (Au_j, u_j) |(e_k, u_j)|^2 \\
&= \sum_{j=1}^{\infty} (Au_j, u_j) \sum_{k=1}^{\infty} |(e_k, u_j)|^2 = \sum_{j=1}^{\infty} (Au_j, u_j) |u_j|^2 \\
&= \sum_{j=1}^{\infty} (Au_j, u_j) = \sum_{j=1}^{\infty} \lambda_j
\end{align*}
\]
and this proves the theorem.

This is just like it is for a matrix. Recall the trace of a matrix is the sum of the eigenvalues.

It is also easy to see that in any separable Hilbert space, there exist nuclear operators. Let \(\sum_{k=1}^{\infty} |\lambda_k| < \infty\). Then let \(\{e_k\}\) be a complete orthonormal set of vectors. Let
\[
A = \sum_{k=1}^{\infty} \lambda_k e_k \otimes e_k.
\]
It is not too hard to verify this works.

Much more can be said about nuclear operators.

### 6.2.3 Hilbert Schmidt Operators

**Definition 6.2.12** Let \(H\) and \(G\) be two separable Hilbert spaces and let \(T\) map \(H\) to \(G\) be linear. Then \(T\) is called a Hilbert Schmidt operator if there exists some orthonormal basis for \(H\), \(\{e_j\}\) such that
\[
\sum_{j} \|Te_j\|^2 < \infty.
\]
The collection of all such linear maps will be denoted by \(\mathcal{L}_2(H, G)\).
Theorem 6.2.13 \( \mathcal{L}_2(H,G) \subseteq \mathcal{L}(H,G) \) and \( \mathcal{L}_2(H,G) \) is a separable Hilbert space with norm given by

\[
||T||_{\mathcal{L}_2} = \left( \sum_k ||Te_k||^2 \right)^{1/2}
\]

where \( \{e_k\} \) is some orthonormal basis for \( H \). Also \( \mathcal{L}_2(H,G) \subseteq \mathcal{L}(H,G) \) and

\[
||T|| \leq ||T||_{\mathcal{L}_2}.
\]

All Hilbert Schmidt operators are compact. Also for \( X \in H \) and \( Y \in G \), \( X \otimes Y \in \mathcal{L}_2(H,G) \) and

\[
||X \otimes Y||_{\mathcal{L}_2} = ||X||_H ||Y||_G
\]

An orthonormal complete basis \( \{f_j \otimes e_k\} \) where the \( f_j \) and \( e_k \) are orthonormal bases.

Proof: First I want to show \( \mathcal{L}_2(H,G) \subseteq \mathcal{L}(H,G) \) and \( ||T|| \leq ||T||_{\mathcal{L}_2} \). Pick an orthonormal basis for \( H \), \( \{e_k\} \) and an orthonormal basis for \( G \), \( \{f_k\} \). Then letting

\[
x = \sum_{k=1}^n x_k e_k,
\]

\[
Tx = T \left( \sum_{k=1}^n x_k e_k \right) = \sum_{k=1}^n x_k T(e_k)
\]

where \( x_k \equiv (x,e_k) \). Therefore using Minkowski’s inequality,

\[
||Tx|| = \left( \sum_{k=1}^\infty |(Tx,f_k)|^2 \right)^{1/2}
\]

\[
= \left( \sum_{k=1}^\infty \left| \sum_{j=1}^n x_j T(e_j,f_k) \right|^2 \right)^{1/2}
\]

\[
= \left( \sum_{k=1}^\infty \sum_{j=1}^n (x_j T(e_j,e_k))^2 \right)^{1/2}
\]

\[
\leq \sum_{j=1}^n \left( \sum_{k=1}^\infty |x_j|^2 ||T(e_j,e_k)||^2 \right)^{1/2}
\]

\[
= \sum_{j=1}^n |x_j| \left( \sum_{k=1}^\infty ||T(e_j,e_k)||^2 \right)^{1/2}
\]

\[
= \sum_{j=1}^n |x_j| ||T(e_j)|| \leq \left( \sum_{j=1}^n |x_j|^2 \right)^{1/2} ||T||_{\mathcal{L}_2} = ||x|| \ ||T||_{\mathcal{L}_2}
\]

Therefore, since finite sums of the form \( \sum_{k=1}^n x_k e_k \) are dense in \( H \), it follows \( T \in \mathcal{L}(H,G) \) and \( ||T|| \leq ||T||_{\mathcal{L}_2} \).

Next consider the norm. I need to verify the norm does not depend on the choice of orthonormal basis. Let \( \{f_k\} \) be an orthonormal basis for \( G \). Then for \( \{e_k\} \) an orthonormal basis for \( H \),

\[
\sum_k ||Te_k||^2 = \sum_k \sum_j |(Te_k,f_j)|^2 = \sum_k \sum_j |(e_k,T^* f_j)|^2
\]

\[
= \sum_j \sum_k |(e_k,T^* f_j)|^2 = \sum_j ||T^* f_j||^2.
\]
CHAPTER 6. BASIC THEORY OF HILBERT SPACES

The above computation makes sense because it was just shown that \( T \) is continuous. The same result would be obtained for any other orthonormal basis \( \{ e_j' \} \) and this shows the norm is at least well defined. It is clear this does indeed satisfy the axioms of a norm and this proves the above claims.

It only remains to verify \( L_2(H,G) \) is a separable Hilbert space. It is clear it is an inner product space because you only have to pick an orthonormal basis, \( \{ e_k \} \) and define the inner product as

\[
(S,T) \equiv \sum_k (Se_k, Te_k).
\]

The only remaining issue is the completeness. Suppose then that \( \{ T_n \} \) is a Cauchy sequence in \( L_2(H,G) \). Then from \( 6.2.19 \) \( \{ T_n \} \) is a Cauchy sequence in \( L(H,G) \) and so there exists a unique \( T \) such that \( \lim_{n \to \infty} ||T_n - T|| = 0 \). Then it only remains to verify \( T \in L_2(H,G) \). But by Fatou’s lemma,

\[
\sum_k ||T_n e_k||^2 \leq \liminf_{n \to \infty} \sum_k ||T_n e_k||^2 = \liminf_{n \to \infty} |||T_n|||_{L_2}^2 < \infty.
\]

All that remains is to verify \( L_2(H,G) \) is separable and these Hilbert Schmidt operators are compact. I will show an orthonormal basis for \( L_2(H,G) \) is \( \{ f_j \otimes e_k \} \) where \( \{ f_k \} \) is an orthonormal basis for \( G \) and \( \{ e_k \} \) is an orthonormal basis for \( H \).

I need to show \( f_j \otimes e_k \in L_2(H,G) \) and that it is an orthonormal basis for \( L_2(H,G) \) as claimed.

\[
\sum_k ||f_j \otimes e_i (e_k)||^2 = \sum_k ||f_j e_k||^2 = ||f_j||^2 = 1 < \infty
\]

so each of these operators is in \( L_2(H,G) \). Next I show they are orthonormal.

\[
(f_j \otimes e_k, f_s \otimes e_r) = \sum_p (f_j \otimes e_k (e_p), f_s \otimes e_r (e_p)) = \sum_p \delta_{rp} \delta_{kp} (f_j, f_s) = \sum_p \delta_{rp} \delta_{kp} \delta_{js}
\]

If \( j = s \) and \( k = r \) this reduces to 1. Otherwise, this gives 0. Thus these operators are orthonormal. Now let \( T \in L_2(H,G) \). Consider

\[
T_n = \sum_{i=1}^{n} \sum_{j=1}^{n} (Te_i, f_j) f_j \otimes e_i
\]

Then

\[
T_n e_k = \sum_{i=1}^{n} \sum_{j=1}^{n} (Te_i, f_j) (e_k, e_i) f_j = \sum_{j=1}^{n} (Te_k, f_j) f_j
\]

It follows

\[
||T_n e_k|| \leq ||Te_k||
\]

and

\[
\lim_{n \to \infty} T_n e_k = Te_k.
\]

Therefore, from the dominated convergence theorem,

\[
\lim_{n \to \infty} ||T - T_n||_{L_2}^2 = \lim_{n \to \infty} \sum_k ||(T - T_n) e_k||^2 = 0.
\]

Therefore, the linear combinations of the \( f_j \otimes e_i \) are dense in \( L_2(H,G) \) and this proves completeness of the orthonormal basis.
This also shows $L_2 (H, G)$ is separable. From \ref{sec:6.2}, it also shows that every $T \in L_2 (H, G)$ is the limit in the operator norm of a sequence of compact operators. This follows because each of the $f_j \otimes e_i$ is easily seen to be a compact operator because if $x_m \to x$ weakly, then

$$f_j \otimes e_i (x_m) = (x_m, e_i) f_j \to (x, e_i) f_j = f_j \otimes e_i (x)$$

and since if $\{x_m\}$ is any bounded sequence, there exists a subsequence, $\{x_{n_k}\}$ which converges weakly and by the above, $f_j \otimes e_i (x_{n_k}) \to f_j \otimes e_i (x)$ showing bounded sets are mapped to precompact sets. Therefore, each $T \in L_2 (H, G)$ must also be a compact operator. Here is why.

Let $B$ be a bounded set in which $||x|| < M$ for all $x \in B$ and consider $TB$. I need to show $TB$ is totally bounded. Let $\varepsilon > 0$ be given. Then let $||T_m - T|| < \frac{\varepsilon}{3M}$ where $T_m$ is a compact operator like those described above and let $\{T_m x_j\}_{j=1}^N$ be an $\varepsilon/3$ net for $T_m (B)$. Then

$$||Tx_j - T_m x_j|| < \frac{\varepsilon}{3}$$

and so letting $x \in B$, pick $x_j$ such that $||T_m x - T_m x_j|| < \varepsilon/3$. Then

$$||Tx - Tx_j|| \leq ||Tx - T_m x|| + ||T_m x - T_m x_j|| + ||T_m x_j - Tx_j|| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$$

showing $\{Tx_j\}_{j=1}^N$ is an $\varepsilon$ net for $TB$.

Finally, consider \ref{sec:6.2}. Let $\{e_k\}$ be an orthonormal basis for $H$ and consider the following computation which establishes this equation.

$$||Y \otimes X||_{L_2}^2 = \sum_{k=1}^\infty ||Y \otimes X (e_k)||^2 = \sum_{k=1}^\infty ||(e_k, X) Y||^2 = ||Y||_G^2 \sum_{k=1}^\infty |(e_k, X)|^2 = ||Y||_G^2 ||X||_H^2 < \infty.$$

(6.2.21)

This proves the theorem.

### 6.3 Square Roots

In this section, $H$ will be a Hilbert space, real or complex, and $T$ will denote an operator which satisfies the following definition. A useful theorem about the existence of square roots of certain operators is presented. This proof is very elementary. I found it in \ref{sec:6.2}.

**Definition 6.3.1** Let $T \in L (H, H)$ satisfy $T = T^*$ (Hermitian) and for all $x \in H$,

$$(Tx, x) \geq 0 \quad (6.3.22)$$

Such an operator is referred to as positive and self adjoint. It is probably better to refer to such an operator as “nonnegative” since the possibility that $Tx = 0$ for some $x \neq 0$ is not being excluded. Instead of “self adjoint” you can also use the term, Hermitian. To save on notation, write

$$T \geq 0$$

to mean $T$ is positive, satisfying \ref{sec:6.2}.

With the above definition here is a fundamental result about positive self adjoint operators.

**Proposition 6.3.2** Let $S, T$ be positive and self adjoint such that $ST = TS$. Then $ST$ is also positive and self adjoint.
Proof: It is obvious that $ST$ is self adjoint. The only problem is to show that $ST$ is positive. To show this, first suppose $S \leq I$. The idea is to write

$$S = S_{n+1} + \sum_{k=0}^{n} S_k^2$$

where $S_0 = S$ and the operators $S_k$ are self adjoint. This is a useful idea because it is then obvious that the sum is positive. If we want such a representation as above, then it follows that $S_0 \equiv S$ and

$$S_{n+1} \equiv S_n - S_n^2.$$ 

Thus it is obvious that the $S_k$ are all self adjoint. Also, the following claim holds.

Claim: $I \geq S_n \geq 0$.

Proof of the claim: This is true if $n = 0$. Assume true for $n$. Then from the definition,

$$S_{n+1} = S_n^2 (I - S_n) + (I - S_n)^2 S_n$$

and it is obvious from the definition that the sum of positive operators is positive. Therefore, it suffices to show the two terms in the above are both positive. It is clear from the definition that each $S_n$ is Hermitian (self adjoint) because they are just polynomials in $S$. Also each must commute with $T$ for the same reason. Therefore,

$$\left( S_n^2 (I - S_n) x, x \right) = \left( (I - S_n) S_n x, S_n x \right) \geq 0$$

and also

$$\left( (I - S_n)^2 S_n x, x \right) = \left( S_n (I - S_n) x, (I - S_n) x \right) \geq 0$$

This proves the claim.

Now each $S_k$ commutes with $T$ because this is true of $S_0$ and succeeding $S_k$ are polynomials in terms of $S_0$. Therefore,

$$\left( ST x, x \right) = \left( \left( S_{n+1} + \sum_{k=0}^{n} S_k^2 \right) T x, x \right)$$

$$= \left( S_{n+1} T x, x \right) + \sum_{k=0}^{n} \left( S_k^2 T x, x \right)$$

$$= \left( T x, S_{n+1} x \right) + \sum_{k=0}^{n} \left( T S_k x, S_k x \right) \quad (6.3.23)$$

Consider $S_{n+1} x$. From the claim,

$$(S, x) = (S_{n+1} x, x) + \sum_{k=0}^{n} |S_k x|^2 \geq \sum_{k=0}^{n} |S_k|^2$$

and so $\lim_{n \to \infty} S_n x = 0$. Hence from (6.3.23),

$$\lim_{n \to \infty} \inf \left( ST x, x \right) = \left( ST x, x \right) = \lim_{n \to \infty} \sum_{k=0}^{n} (T S_k x, S_k x) \geq 0.$$ 

All this was based on the assumption that $S \leq I$. The next task is to remove this assumption. Let $ST = TS$ where $T$ and $S$ are positive self adjoint operators. Then consider $S/\|S\|$. This is still a positive self adjoint operator and it commutes with $T$ just like $S$ does. Therefore, from the first part,

$$0 \leq \left( \frac{S}{\|S\|} T x, x \right) = \frac{1}{\|S\|} \left( ST x, x \right).$$

The proposition is like the familiar statement about real numbers which says that when you multiply two non-negative real numbers the result is a nonnegative real number. The next lemma is a generalization of the familiar fact that if you have an increasing sequence of real numbers which is bounded above, then the sequence converges.
Lemma 6.3.3 Let \( \{T_n\} \) be a sequence of self adjoint operators on a Hilbert space, \( H \) and let \( T_n \leq T_{n+1} \) for all \( n \). Also suppose there exists \( K \), a self adjoint operator such that for all \( n, T_n \leq K \). Suppose also that each operator commutes with all the others and that \( K \) commutes with all the \( T_n \). Then there exists a self adjoint continuous operator, \( T \) such that for all \( x \in H \),

\[
T_n x \rightarrow Tx,
\]

\( T \leq K \) and \( T \) commutes with all the \( T_n \) and with \( K \).

**Proof:** Consider \( K - T_n \equiv S_n \). Then the \( \{S_n\} \) are decreasing, that is,

\[
\{(S_n x, x)\}
\]

is a decreasing sequence and from the hypotheses, \( S_n \geq 0 \) so the above sequence is bounded below by 0. Therefore, \( \lim_{n \to \infty} (S_n x, x) \) exists. By Proposition 6.3.2, if \( n > m \),

\[
S_m^2 - S_m S_n = S_m (S_m - S_n) \geq 0
\]

and similarly,

\[
S_n S_m - S_n^2 = S_n (S_m - S_n) \geq 0.
\]

Therefore, since \( S_n \) is self adjoint,

\[
\begin{align*}
|T_n x - T_m x|^2 &= |S_n x - S_m x|^2 = \left( (S_n - S_m)^2 x, x \right) \\
&= ((S_m^2 - 2S_n S_m + S_n^2) x, x) + ((S_n^2 - S_m S_n) x, x) \\
&\leq ((S_m^2 - S_m S_n) x, x) + ((S_n^2 - S_n^2) x, x) \\
&= ((S_m - S_n) (S_m + S_n) x, x) \leq 2 ((S_m - S_n) K x, x) \\
&\leq 2 ((S_m - S_n) K x, K x)^{1/2} ((S_n - S_n) x, x)^{1/2}
\end{align*}
\]

The last step follows from an application of the Cauchy Schwarz inequality along with the fact \( S_m - S_n \geq 0 \). The last expression converges to 0 because \( \lim_{n \to \infty} (S_n x, x) \) exists for each \( x \). It follows \( \{T_n x\} \) is a Cauchy sequence. Let \( Tx \) be the thing to which it converges. \( T \) is obviously linear and

\[
(T x, x) = \lim_{n \to \infty} (T_n x, x) \leq (K x, x).
\]

Also

\[
(K T x, y) = \lim_{n \to \infty} (K T_n x, y) = \lim_{n \to \infty} (T_n K x, y) = (T K x, y)
\]

and so \( TK = KT \). Similarly, \( T \) commutes with all \( T_n \).

In order to show \( T \) is continuous, apply the uniform boundedness principle, Theorem 4.1.8. The convergence of \( \{T_n x\} \) implies there exists a uniform bound on the norms, \( ||T_n|| \) and so

\[
||(T_n x, y)|| \leq C ||x|| ||y||.
\]

Now take the limit as \( n \to \infty \) to conclude

\[
||(T x, y)|| \leq C ||x|| ||y||
\]

which shows \( ||T|| \leq C \). This proves the lemma.

With this preparation, here is the theorem about square roots.

**Theorem 6.3.4** Let \( T \in \mathcal{L}(H, H) \) be a positive self adjoint linear operator. Then there exists a unique square root, \( A \) with the following properties. \( A^2 = T \), \( A \) is positive and self adjoint, \( A \) commutes with every operator which commutes with \( T \).

**Proof:** First suppose \( T \leq I \). Then define

\[
A_0 \equiv 0, A_{n+1} = A_n + \frac{1}{2} (T - A_n^2).
\]

From this it follows that every \( A_n \) is a polynomial in \( T \). Therefore, \( A_n \) commutes with \( T \) and with every operator which commutes with \( T \).
Claim 1: \( A_n \leq I \).

**Proof of Claim 1:** This is true if \( n = 0 \). Suppose it is true for \( n \). Then by the assumption that \( T \leq I \),

\[
I - A_{n+1} = I - A_n + \frac{1}{2} (A_n^2 - T) \\
\geq I - A_n + \frac{1}{2} (A_n^2 - I) \\
= I - A_n - \frac{1}{2} (I - A_n) (I + A_n) \\
= (I - A_n) \left( I - \frac{1}{2} (I + A_n) \right) \\
= (I - A_n) (I - A_n) \frac{1}{2} \geq 0.
\]

Claim 2: \( A_n \leq A_{n+1} \)

**Proof of Claim 2:** From the definition of \( A_n \), this is true if \( n = 0 \) because

\[ A_1 = T \geq 0 = A_0. \]

Suppose true for \( n \). Then from Claim 1,

\[
A_{n+1} - A_{n+2} = A_{n+1} + \frac{1}{2} (T - A_{n+1}^2) - \left[ A_n + \frac{1}{2} (T - A_n^2) \right] \\
= A_{n+1} - A_n + \frac{1}{2} (A_n^2 - A_{n+1}^2) \\
= (A_{n+1} - A_n) \left( I - \frac{1}{2} (A_n + A_{n+1}) \right) \\
\geq (A_{n+1} - A_n) \left( I - \frac{1}{2} (2I) \right) = 0.
\]

Claim 3: \( A_n \geq 0 \)

**Proof of Claim 3:** This is true if \( n = 0 \). Suppose it is true for \( n \).

\[
(A_{n+1} x, x) = (A_n x, x) + \frac{1}{2} (T x, x) - \frac{1}{2} (A_n^2 x, x) \\
\geq (A_n x, x) + \frac{1}{2} (T x, x) - \frac{1}{2} (A_n x, x) \geq 0
\]

because \( A_n - A_n^2 = A_n (I - A_n) \geq 0 \) by Proposition 6.3.3.

Now \( \{A_n\} \) is a sequence of positive self adjoint operators which are bounded above by \( I \) such that each of these operators commutes with every operator which commutes with \( T \). By Lemma 6.3.3, there exists a bounded linear operator, \( A \) such that for all \( x \),

\[ A_n x \rightarrow A x \]

Then \( A \) commutes with every operator which commutes with \( T \) because each \( A_n \) has this property. Also \( A \) is a positive operator because each \( A_n \) is. From passing to the limit in the definition of \( A_n \),

\[ A x = A x + \frac{1}{2} (T x - A^2 x) \]

and so \( T x = A^2 x \). This proves the theorem in the case that \( T \leq I \).

In the general case, consider \( T/||T|| \). Then

\[
\left( \frac{T}{||T||} x, x \right) = \frac{1}{||T||} (T x, x) \leq ||x||^2 = (I x, x)
\]

and so \( T/||T|| \leq I \). Therefore, it has a square root, \( B \). Let \( A = \sqrt{||T||} B \). Then \( A \) has all the right properties and \( A^2 = ||T|| B^2 = ||T|| (T/||T||) = T \). This proves the existence part of the theorem.

Next suppose both \( A \) and \( B \) are square roots of \( T \) having all the properties stated in the theorem. Then \( AB = BA \) because both \( A \) and \( B \) commute with every operator which commutes with \( T \).

\[
(A (A - B) x, (A - B) x), (B (A - B) x, (A - B) x) \geq 0 \quad (6.3.24)
\]
Therefore, on adding these,
\[
((A^2 - AB + BA - B^2) x, (A - B) x) = ((A^2 - B^2) x, (A - B) x) = ((T - T) x, (A - B) x) = 0.
\]
It follows both expressions in (6.4.24) equal 0 since both are nonnegative and when they are added the result is 0. Now applying the existence part of the theorem to \(A\), there exists a positive square root of \(A\) which is self-adjoint. Thus
\[
\left(\sqrt{A} (A - B) x, \sqrt{A} (A - B) x\right) = 0
\]
so \(\sqrt{A} (A - B) x = 0\) which implies \((A - B) x = 0\). Similarly, \(B (A - B) x = 0\). Subtracting these and taking the inner product with \(x\),
\[
0 = ((A (A - B) - B (A - B)) x, x) = \left((A - B)^2 x, x\right) = |(A - B) x|^2
\]
and so \(Ax = Bx\) which shows \(A = B\) since \(x\) was arbitrary. This proves the theorem.

### 6.4 Radon Nikodym Theorem

The Radon Nikodym Theorem, is a representation theorem for one measure in terms of another. The approach given here is due to Von Neumann and depends on the Riesz representation theorem for Hilbert space, Theorem 6.0.15 on Page 86.

**Definition 6.4.1** Let \(\mu\) and \(\lambda\) be two measures defined on a \(\sigma\)-algebra, \(S\), of subsets of a set, \(\Omega\). \(\lambda\) is absolutely continuous with respect to \(\mu\), written as \(\lambda \ll \mu\), if \(\lambda(E) = 0\) whenever \(\mu(E) = 0\).

It is not hard to think of examples which should be like this. For example, suppose one measure is volume and the other is mass. If the volume of something is zero, it is reasonable to expect the mass of it should also be equal to zero. In this case, there is a function called the density which is integrated over volume to obtain mass. The Radon Nikodym theorem is an abstract version of this notion. Essentially, it gives the existence of the density function.

**Theorem 6.4.2 (Radon Nikodym)** Let \(\lambda\) and \(\mu\) be finite measures defined on a \(\sigma\)-algebra, \(S\), of subsets of \(\Omega\). Suppose \(\lambda \ll \mu\). Then there exists a unique \(f \in L^1(\Omega, \mu)\) such that \(f(x) \geq 0\) and
\[
\lambda(E) = \int_E f \, d\mu.
\]

If it is not necessarily the case that \(\lambda \ll \mu\), there are two measures, \(\lambda_{\perp}\) and \(\lambda||\) such that \(\lambda = \lambda_{\perp} + \lambda||\), \(\lambda|| \ll \mu\) and there exists a set of \(\mu\) measure zero, \(N\) such that for all \(E\) measurable, \(\lambda_{\perp}(E) = \lambda(E \cap N) = \lambda_{\perp}(E \cap N)\). In this case the two measures, \(\lambda_{\perp}\) and \(\lambda||\) are unique and the representation of \(\lambda = \lambda_{\perp} + \lambda||\) is called the Lebesgue decomposition of \(\lambda\). The measure \(\lambda||\) is the absolutely continuous part of \(\lambda\) and \(\lambda_{\perp}\) is called the singular part of \(\lambda\).

**Proof:** Let \(\Lambda : L^2(\Omega, \mu + \lambda) \to \mathbb{C}\) be defined by
\[
\Lambda g = \int_\Omega g \, d\lambda.
\]
By Holder’s inequality,
\[
|\Lambda g| \leq \left(\int_\Omega 1^2 \, d\lambda\right)^{1/2} \left(\int_\Omega |g|^2 \, d(\mu + \lambda)\right)^{1/2} = \lambda(\Omega)^{1/2} ||g||_2
\]
where \(||g||_2\) is the \(L^2\) norm of \(g\) taken with respect to \(\mu + \lambda\). Therefore, since \(\Lambda\) is bounded, it follows from Theorem 6.0.15 on Page 86 that \(\Lambda \in (L^2(\Omega, \mu + \lambda))'\), the dual space \(L^2(\Omega, \mu + \lambda)\). By the Riesz representation theorem in Hilbert space, Theorem 6.0.15 there exists a unique \(h \in L^2(\Omega, \mu + \lambda)\) with
\[
\Lambda g = \int_\Omega g \, d\lambda = \int_\Omega hgd(\mu + \lambda).
\]
(6.4.25)
The plan is to show $h$ is real and nonnegative at least a.e. Therefore, consider the set where $\text{Im } h$ is positive.

$$E = \{ x \in \Omega : \text{Im } h(x) > 0 \},$$

Now let $g = \mathcal{X}_E$ and use (6.4.25) to get

$$\lambda(E) = \int_E (\text{Re } h + i \text{ Im } h) d(\mu + \lambda). \quad (6.4.26)$$

Since the left side of (6.4.26) is real, this shows

$$0 = \int_E (\text{Im } h) d(\mu + \lambda) \geq \int_{E_n} (\text{Im } h) d(\mu + \lambda) \geq \frac{1}{n} (\mu + \lambda)(E_n)$$

Thus $(\mu + \lambda)(E_n) = 0$ and since $E = \bigcup_{n=1}^{\infty} E_n$, it follows $(\mu + \lambda)(E) = 0$. A similar argument shows that for

$$E = \{ x \in \Omega : \text{Im } h(x) < 0 \},$$

$(\mu + \lambda)(E) = 0$. Thus there is no loss of generality in assuming $h$ is real-valued.

The next task is to show $h$ is nonnegative. This is done in the same manner as above. Define the set where it is negative and then show this set has measure zero.

Let $E = \{ x : h(x) < 0 \}$ and let $E_n = \{ x : h(x) < -\frac{1}{n} \}$. Then let $g = \mathcal{X}_{E_n}$. Since $E = \bigcup_n E_n$, it follows that if $(\mu + \lambda)(E) > 0$ then this is also true for $(\mu + \lambda)(E_n)$ for all $n$ large enough. Then from (6.4.26)

$$\lambda(E_n) = \int_{E_n} h d(\mu + \lambda) \leq -\frac{1}{n} (\mu + \lambda)(E_n) < 0,$$

a contradiction. Thus it can be assumed $h \geq 0$.

At this point the argument splits into two cases.

**Case Where $\lambda \ll \mu$.** In this case, $h < 1$.

Let $E = \{ h \geq 1 \}$ and let $g = \mathcal{X}_E$. Then

$$\lambda(E) = \int_E h d(\mu + \lambda) \geq \mu(E) + \lambda(E).$$

Therefore $\mu(E) = 0$. Since $\lambda \ll \mu$, it follows that $\lambda(E) = 0$ also. Thus it can be assumed

$$0 \leq h(x) < 1$$

for all $x$.

From (6.4.25), whenever $g \in L^2(\Omega, \mu + \lambda),$

$$\int_{\Omega} g(1 - h) d\lambda = \int_{\Omega} h g d\mu. \quad (6.4.27)$$

Now let $E$ be a measurable set and define

$$g(x) \equiv \sum_{i=0}^{n} h^i(x) \mathcal{X}_E(x)$$

in (6.4.24). This yields

$$\int_E (1 - h^{n+1}(x)) d\lambda = \int_E \sum_{i=1}^{n+1} h^i(x) d\mu. \quad (6.4.28)$$
Let \( f(x) = \sum_{i=1}^{\infty} h^i(x) \) and use the Monotone Convergence theorem in \( L^1(\Omega, \mu) \) to let \( n \to \infty \) and conclude
\[
\lambda(E) = \int_E f \, d\mu.
\]

\( f \in L^1(\Omega, \mu) \) because \( \lambda \) is finite.

The function, \( f \) is unique \( \mu \) a.e. because, if \( g \) is another function which also serves to represent \( \lambda \), consider for each \( n \in \mathbb{N} \) the set,
\[
E_n \equiv \left\{ f - g > \frac{1}{n} \right\}
\]
and conclude that
\[
0 = \int_{E_n} (f - g) \, d\mu \geq \frac{1}{n} \mu(E_n).
\]
Therefore, \( \mu(E_n) = 0 \). It follows that
\[
\mu(|f - g|) \leq \sum_{n=1}^{\infty} \mu(E_n) = 0
\]
Similarly, the set where \( g \) is larger than \( f \) has measure zero. This proves the theorem.

**Case where it is not necessarily true that \( \lambda \ll \mu \)**

In this case, let \( N = \{ h \geq 1 \} \) and let \( g = \lambda |\lambda| N \). Then
\[
\lambda(N) = \int_N h \, d(\mu + \lambda) \geq \mu(N) + \lambda(N).
\]
and so \( \mu(N) = 0 \). Now define a measure, \( \lambda_\perp \) by
\[
\lambda_\perp (E) \equiv \lambda(E \cap N)
\]
so \( \lambda_\perp (E \cap N) = \lambda(E \cap N \cap N) \equiv \lambda_\perp (E) \) and let \( \lambda_\parallel \equiv \lambda - \lambda_\perp \). Therefore,
\[
\mu(E) = \mu(E \cap N^C)
\]
Also,
\[
\lambda_\parallel (E) = \lambda(E) - \lambda_\perp (E) \equiv \lambda(E) - \lambda(E \cap N) = \lambda(E \cap N^C).
\]
Suppose \( \lambda_\parallel (E) > 0 \). Therefore, since \( h < 1 \) on \( N^C \)
\[
\lambda_\parallel (E) = \lambda(E \cap N^C) = \int_{E \cap N^C} h \, d(\mu + \lambda) < \mu(E \cap N^C) + \lambda(E \cap N^C) = \mu(E) + \lambda_\parallel (E),
\]
which is a contradiction unless \( \mu(E) > 0 \). Therefore, \( \lambda_\parallel \ll \mu \) because if \( \mu(E) = 0 \), the above inequality cannot hold.

It only remains to verify the two measures \( \lambda_\perp \) and \( \lambda_\parallel \) are unique. Suppose then that \( \nu_1 \) and \( \nu_2 \) play the roles of \( \lambda_\perp \) and \( \lambda_\parallel \) respectively. Let \( N_1 \) play the role of \( N \) in the definition of \( \nu_1 \) and let \( g_1 \) play the role of \( g \) for \( \nu_2 \). I will show that \( g = g_1 \) \( \mu \) a.e. Let \( E_k \equiv \{ g_1 - g > 1/k \} \) for \( k \in \mathbb{N} \). Then on observing that \( \lambda_\perp - \nu_1 = \nu_2 - \lambda_\parallel \)
\[
0 = (\lambda_\perp - \nu_1) \left( E_n \cap (N_1 \cup N)^C \right) = \int_{E_n \cap (N_1 \cup N)^C} (g_1 - g) \, d\mu
\]
\[
\geq \frac{1}{k} \mu(E_k \cap (N_1 \cup N)^C) = \frac{1}{k} \mu(E_k).
\]
and so \( \mu(E_k) = 0 \). Therefore, \( \mu(|g_1 - g > 0|) = 0 \) because \( [g_1 - g > 0] = \bigcup_{k=1}^{\infty} E_k \). It follows \( g_1 \leq g \) \( \mu \) a.e. Similarly, \( g_1 \geq g \) \( \mu \) a.e. Therefore, \( \nu_2 = \lambda_\parallel \) and so \( \lambda_\perp = \nu_1 \) also. This proves the theorem.
Chapter 7

Integrals And Derivatives

7.1 The Fundamental Theorem Of Calculus

The version of the fundamental theorem of calculus found in Calculus says that if \( f \) is a Riemann integrable function, the function

\[
x \to \int_a^x f(t) \, dt,
\]

has a derivative at every point where \( f \) is continuous. It is natural to ask what occurs for \( f \) in \( L^1 \). It is an amazing fact that the same result is obtained aside from a set of measure zero even though \( f \), being only in \( L^1 \), may fail to be continuous anywhere. Proofs of this result are based on some form of the Vitali covering theorem presented above.

In what follows, the measure space is \((\mathbb{R}^n, S, m)\) where \( m \) is \( n \)-dimensional Lebesgue measure although the same theorems can be proved for arbitrary Radon measures. To save notation, \( m \) is written in place of \( m_n \).

Recall

\[
B(p, r) = \{ x : |x - p| < r \}.
\]

(7.1.1)

Also define the following.

\[
B(p, r) = \{ x : |x - p| < r \}.
\]

(7.1.2)

The first version of the Vitali covering theorem presented above will now be used to establish the fundamental theorem of calculus. The space of locally integrable functions is the most general one for which the maximal function defined below makes sense.

**Definition 7.1.1** \( f \in L^1_{\text{loc}}(\mathbb{R}^n) \) means \( f \chi_{B(0, R)} \in L^1(\mathbb{R}^n) \) for all \( R > 0 \). For \( f \in L^1_{\text{loc}}(\mathbb{R}^n) \), the Hardy Littlewood Maximal Function, \( Mf \), is defined by

\[
Mf(x) \equiv \sup_{r > 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y)| \, dy.
\]

**Theorem 7.1.2** If \( f \in L^1(\mathbb{R}^n) \), then for \( \alpha > 0 \),

\[
m([Mf > \alpha]) \leq \frac{5^n}{\alpha} ||f||_1.
\]

(Here and elsewhere, \([Mf > \alpha] \equiv \{ x \in \mathbb{R}^n : Mf(x) > \alpha \} \) with other occurrences of \([ \ ] \) being defined similarly.)

**Proof:** Let \( S \equiv [Mf > \alpha] \). For \( x \in S \), choose \( r_x > 0 \) with

\[
\frac{1}{m(B(x, r_x))} \int_{B(x, r_x)} |f| \, dm > \alpha.
\]

The \( r_x \) are all bounded because

\[
m(B(x, r_x)) < \frac{1}{\alpha} \int_{B(x, r_x)} |f| \, dm < \frac{1}{\alpha} ||f||_1.
\]

By the Vitali covering theorem, there are disjoint balls \( B(x_i, r_i) \) such that

\[
S \subseteq \bigcup_{i=1}^\infty B(x_i, 5r_i)
\]
and
\[ \frac{1}{m(B(x_i, r_i))} \int_{B(x_i, r_i)} |f| \, dm > \alpha. \]

Therefore
\[ m(S) \leq \sum_{i=1}^{\infty} m(B(x_i, 5r_i)) = 5^n \sum_{i=1}^{\infty} m(B(x_i, r_i)) \]
\[ \leq \frac{5^n}{\alpha} \sum_{i=1}^{\infty} \int_{B(x_i, r_i)} |f| \, dm \]
\[ \leq \frac{5^n}{\alpha} \int_{\mathbb{R}^n} |f| \, dm, \]

the last inequality being valid because the balls \( B(x_i, r_i) \) are disjoint. This proves the theorem.

Note that at this point it is unknown whether \( S \) is measurable. This is why \( m(S) \) and not \( m(S) \) is written.

The following is the fundamental theorem of calculus from elementary calculus.

**Lemma 7.1.3** Suppose \( g \) is a continuous function. Then for all \( x \),
\[ \lim_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} g(y) \, dy = g(x). \]

**Proof:** Note that
\[ g(x) = \frac{1}{m(B(x, r))} \int_{B(x, r)} g(y) \, dy \]
and so
\[ \left| g(x) - \frac{1}{m(B(x, r))} \int_{B(x, r)} g(y) \, dy \right| \]
\[ = \left| \frac{1}{m(B(x, r))} \int_{B(x, r)} (g(y) - g(x)) \, dy \right| \]
\[ \leq \frac{1}{m(B(x, r))} \int_{B(x, r)} |g(y) - g(x)| \, dy. \]

Now by continuity of \( g \) at \( x \), there exists \( r > 0 \) such that if \( |x - y| < r \), \( |g(y) - g(x)| < \varepsilon \). For such \( r \), the last expression is less than
\[ \frac{1}{m(B(x, r))} \int_{B(x, r)} \varepsilon \, dy < \varepsilon. \]

This proves the lemma.

**Definition 7.1.4** Let \( f \in L^1(\mathbb{R}^k, m) \). A point, \( x \in \mathbb{R}^k \) is said to be a Lebesgue point if
\[ \limsup_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, dm = 0. \]

Note that if \( x \) is a Lebesgue point, then
\[ \lim_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} f(y) \, dm = f(x). \]

and so the symmetric derivative exists at all Lebesgue points.

**Theorem 7.1.5** (Fundamental Theorem of Calculus) Let \( f \in L^1(\mathbb{R}^k) \). Then there exists a set of measure \( 0, N \), such that if \( x \not\in N \), then
\[ \lim_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, dy = 0. \]
Proof: Let $\lambda > 0$ and let $\varepsilon > 0$. By density of $C_c(\mathbb{R}^k)$ in $L^1(\mathbb{R}^k, m)$ there exists $g \in C_c(\mathbb{R}^k)$ such that $\|g - f\|_{L^1(\mathbb{R}^k)} < \varepsilon$. Now since $g$ is continuous,

$$\limsup_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, dm$$

$$= \limsup_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, dm$$

$$- \limsup_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} |g(y) - g(x)| \, dm$$

$$= \limsup_{r \to 0} \left( \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, dm \right)$$

$$\leq \limsup_{r \to 0} \left( \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, dm \right)$$

$$\leq \limsup_{r \to 0} \left( \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - g(y) - (f(x) - g(x))| \, dm \right)$$

$$\leq \limsup_{r \to 0} \left( \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - g(y)| \, dm \right) + |f(x) - g(x)|$$

$$\leq M(|f - g|(x) + |f(x) - g(x)|).$$

Therefore,

$$\left[ x : \limsup_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, dm > \lambda \right]$$

$$\subseteq \left[ M(|f - g|) > \frac{\lambda}{2} \right] \cup \left[ |f - g| > \frac{\lambda}{2} \right]$$

Now

$$\varepsilon > \int |f - g| \, dm \geq \int_{|f - g| > \frac{\lambda}{2}} |f - g| \, dm$$

$$\geq \frac{\lambda}{2} m \left( \left[ |f - g| > \frac{\lambda}{2} \right] \right)$$

This along with the weak estimate of Theorem 7.1 implies

$$m \left( \left[ x : \limsup_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, dm > \lambda \right] \right)$$

$$< \left( \frac{2}{\lambda} \delta^k + \frac{2}{\lambda} \right) \|f - g\|_{L^1(\mathbb{R}^k)}$$

$$< \left( \frac{2}{\lambda} \delta^k + \frac{2}{\lambda} \right) \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows

$$m_n \left( \left[ x : \limsup_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, dm > \lambda \right] \right) = 0.$$

Now let

$$N = \left[ x : \limsup_{r \to 0} \frac{1}{m(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, dm > 0 \right].$$
and
\[ N_n = \left\{ x : \limsup_{r \to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| \, dm > \frac{1}{n} \right\} \]

It was just shown that \( m(N_n) = 0 \). Also, \( N = \bigcup_{n=1}^{\infty} N_n \). Therefore, \( m(N) = 0 \) also. It follows that for \( x \notin N \),
\[ \limsup_{r \to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| \, dm = 0 \]
and this proves a.e. point is a Lebesgue point.

Of course it is sufficient to assume \( f \) is only in \( L^1_{loc}(\mathbb{R}^k) \).

**Corollary 7.1.6 (Fundamental Theorem of Calculus)** Let \( f \in L^1_{loc}(\mathbb{R}^k) \). Then there exists a set of measure 0, \( N \), such that if \( x \notin N \), then
\[ \lim_{r \to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| \, dy = 0. \]

**Proof:** Consider \( B(0,n) \) where \( n \) is a positive integer. Then \( f_n = f \chi_{B(0,n)} \in L^1(\mathbb{R}^k) \) and so there exists a set of measure 0, \( N_n \) such that if \( x \in B(0,n) \setminus N_n \), then
\[ \lim_{r \to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f_n(y) - f_n(x)| \, dy = 0. \]
Let \( N = \bigcup_{n=1}^{\infty} N_n \). Then if \( x \notin N \), the above equation holds.

**Corollary 7.1.7** If \( f \in L^1_{loc}(\mathbb{R}^n) \), then
\[ \lim_{r \to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} f(y) \, dy = f(x) \text{ a.e. } x. \]

**Proof:**
\[ \left| \frac{1}{m(B(x,r))} \int_{B(x,r)} f(y) \, dy - f(x) \right| \leq \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| \, dy \]
and the last integral converges to 0 a.e. \( x \).

**Definition 7.1.8** For \( N \) the set of Theorem 7.1.3 or Corollary 7.1.4, \( N^C \) is called the Lebesgue set or the set of Lebesgue points.

The next corollary is a one dimensional version of what was just presented.

**Corollary 7.1.9** Let \( f \in L^1(\mathbb{R}) \) and let
\[ F(x) = \int_{-\infty}^{x} f(t) \, dt. \]
Then for a.e. \( x \), \( F'(x) = f(x) \).

**Proof:** For \( h > 0 \)
\[ \frac{1}{h} \int_{x}^{x+h} |f(y) - f(x)| \, dy \leq 2 \left( \frac{1}{2h} \right) \int_{x-h}^{x+h} |f(y) - f(x)| \, dy \]
By Theorem 7.1.3, this converges to 0 a.e. Similarly
\[ \frac{1}{h} \int_{x-h}^{x} |f(y) - f(x)| \, dy \]
converges to 0 a.e. \( x \).

\[
\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| \leq \frac{1}{h} \int_{x}^{x+h} |f(y) - f(x)| \, dy \quad (7.1.4)
\]

and

\[
\left| \frac{F(x) - F(x-h)}{h} - f(x) \right| \leq \frac{1}{h} \int_{x-h}^{x} |f(y) - f(x)| \, dy. \quad (7.1.5)
\]

Now the expression on the right in (7.1.4) and (7.1.5) converges to zero for a.e. \( x \). Therefore, by (7.1.5), for a.e. \( x \) the derivative from the right exists and equals \( f(x) \) while from (7.1.4) the derivative from the left exists and equals \( f(x) \) a.e. It follows

\[
\lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = f(x) \text{ a.e.} \quad x.
\]

This proves the corollary.

### 7.2 Absolutely Continuous Functions

**Definition 7.2.1** Let \([a, b]\) be a closed and bounded interval and let \( f : [a, b] \to \mathbb{R} \). Then \( f \) is said to be absolutely continuous if for every \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that if \( \sum_{i=1}^{m} |y_i - x_i| < \delta \), then \( \sum_{i=1}^{m} |f(y_i) - f(x_i)| < \varepsilon \).

**Definition 7.2.2** A finite subset, \( P \) of \([a, b]\) is called a partition of \([x, y]\) if \( P = \{x_0, x_1, \cdots, x_n\} \) where

\[
x = x_0 < x_1 < \cdots < x_n = y.
\]

For \( f : [a, b] \to \mathbb{R} \) and \( P = \{x_0, x_1, \cdots, x_n\} \) define

\[
V_{P} [x, y] = \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|.
\]

Denoting by \( \mathcal{P} [x, y] \) the set of all partitions of \([x, y]\), define

\[
V [x, y] = \sup_{P \in \mathcal{P}[x, y]} V_{P}[x, y].
\]

For simplicity, \( V[a, x] \) will be denoted by \( V(x) \). It is called the total variation of the function, \( f \).

There are some simple facts about the total variation of an absolutely continuous function, \( f \) which are contained in the next lemma.

**Lemma 7.2.3** Let \( f \) be an absolutely continuous function defined on \([a, b]\) and let \( V \) be its total variation function as described above. Then \( V \) is an increasing bounded function. Also if \( P \) and \( Q \) are two partitions of \([x, y]\) with \( P \subseteq Q \), then \( V_{P}[x, y] \leq V_{Q}[x, y] \) and if \([x, y] \subseteq [z, w]\),

\[
V [x, y] \leq V [z, w] \quad (7.2.6)
\]

If \( P = \{x_0, x_1, \cdots, x_n\} \) is a partition of \([x, y]\), then

\[
V [x, y] = \sum_{i=1}^{n} V [x_i, x_{i-1}] \quad (7.2.7)
\]

Also if \( y > x \),

\[
V(y) - V(x) \geq |f(y) - f(x)| \quad (7.2.8)
\]

and the function, \( x \to V(x) - f(x) \) is increasing. The total variation function, \( V \) is absolutely continuous.

**Proof:** The claim that \( V \) is increasing is obvious as is the next claim about \( P \subseteq Q \) leading to \( V_{P}[x, y] \leq V_{Q}[x, y] \). To verify this, simply add in one point at a time and verify that from the triangle inequality, the sum involved gets no smaller. The claim that \( V \) is increasing consistent with set inclusion of intervals is also clearly true and follows directly from the definition.

Now let \( t < V [x, y] \) where \( P_0 = \{x_0, x_1, \cdots, x_n\} \) is a partition of \([x, y]\). There exists a partition, \( P \) of \([x, y]\) such that \( t < V_{P}[x, y] \). Without loss of generality it can be assumed that \( \{x_0, x_1, \cdots, x_n\} \subseteq P \) since if not, you can
simply add in the points of \( P_0 \) and the resulting sum for the total variation will get no smaller. Let \( P_i \) be those points of \( P \) which are contained in \([x_{i-1}, x_i]\). Then

\[
t < V_p [x, y] = \sum_{i=1}^{n} V_{P_i} [x_{i-1}, x_i] \leq \sum_{i=1}^{n} V [x_{i-1}, x_i].
\]

Since \( t < V [x, y] \) is arbitrary,

\[
V [x, y] \leq \sum_{i=1}^{n} V [x_i, x_{i-1}]
\]  \hspace{1em} (7.2.9)

Note that \( V [a, b] \) does not depend on \( f \) being absolutely continuous. Suppose now that \( f \) is absolutely continuous. Let \( \delta \) correspond to \( \varepsilon = 1 \). Then if \( [x, y] \) is an interval of length no larger than \( \delta \), the definition of absolute continuity implies

\[
V [x, y] < 1.
\]

Then from (7.2.9)

\[
V [a, n\delta] \leq \sum_{i=1}^{n} V [a + (i - 1) \delta, a + i\delta] < \sum_{i=1}^{n} 1 = n.
\]

Thus \( V \) is bounded on \([a, b]\). Now let \( P_i \) be a partition of \([x_{i-1}, x_i]\) such that

\[
V_{P_i} [x_{i-1}, x_i] > V [x_{i-1}, x_i] - \frac{\varepsilon}{n}
\]

Then letting \( P = \bigcup P_i \),

\[
-\varepsilon + \sum_{i=1}^{n} V [x_{i-1}, x_i] < \sum_{i=1}^{n} V_{P_i} [x_{i-1}, x_i] = V_P [x, y] \leq V [x, y].
\]

Since \( \varepsilon \) is arbitrary, (7.2.9) follows from this and (7.2.1).

Now let \( x < y \)

\[
V (y) - f (y) - (V (x) - f (x)) = V (y) - V (x) - (f (y) - f (x)) \geq V (y) - V (x) - |f (y) - f (x)| \geq 0.
\]

It only remains to verify that \( V \) is absolutely continuous.

Let \( \varepsilon > 0 \) be given and let \( \delta \) correspond to \( \varepsilon/2 \) in the definition of absolute continuity applied to \( f \). Suppose \( \sum_{i=1}^{n} |y_i - x_i| < \delta \) and consider \( \sum_{i=1}^{n} |V (y_i) - V (x_i)| \). By (7.2.9) this last equals \( \sum_{i=1}^{n} V [x_i, y_i] \). Now let \( P_i \) be a partition of \([x_i, y_i]\) such that \( V_{P_i} [x_i, y_i] + \frac{\varepsilon}{2n} > V [x_i, y_i] \). Then by the definition of absolute continuity,

\[
\sum_{i=1}^{n} |V (y_i) - V (x_i)| = \sum_{i=1}^{n} V [x_i, y_i]
\]

\[
\leq \sum_{i=1}^{n} V_{P_i} [x_i, y_i] + \eta < \varepsilon/2 + \varepsilon/2 = \varepsilon.
\]

and shows \( V \) is absolutely continuous as claimed.

**Lemma 7.2.4** Suppose \( f : [a, b] \to \mathbb{R} \) is absolutely continuous and increasing. Then \( f' \) exists a.e., is in \( L^1 ([a, b]) \), and

\[
f (x) = f (a) + \int_{a}^{x} f' (t) \, dt.
\]

**Proof:** Define \( L \), a positive linear functional on \( C ([a, b]) \) by

\[
L g \equiv \int_{a}^{b} g df
\]

where this integral is the Riemann Stieltjes integral with respect to the integrating function, \( f \). By the Riesz representation theorem for positive linear functionals, there exists a unique Radon measure, \( \mu \) such that \( L g = \int g d\mu \).

Now consider the following picture for \( g_n \in C ([a, b]) \) in which \( g_n \) equals 1 for \( x \) between \( x + 1/n \) and \( y \).
7.2. ABSOLUTELY CONTINUOUS FUNCTIONS

Then \( g_n(t) \to \mathcal{A}_{[x,y]}(t) \) pointwise. Therefore, by the dominated convergence theorem,

\[
\mu((x,y]) = \lim_{n \to \infty} \int g_n \, d\mu.
\]

However,

\[
\left( f(y) - f\left(x + \frac{1}{n}\right) \right) \\
\leq \int g_n \, d\mu = \int_a^b g_n \, df \leq \left( f\left(y + \frac{1}{n}\right) - f(y) \right) \\
+ \left( f(y) - f\left(x + \frac{1}{n}\right) \right) + \left( f\left(x + \frac{1}{n}\right) - f(x) \right)
\]

and so as \( n \to \infty \) the continuity of \( f \) implies

\[
\mu((x,y]) = f(y) - f(x).
\]

Similarly, \( \mu((x,y] = f(y) - f(x) \) and \( \mu([x,y]) = f(y) - f(x) \), the argument used to establish this being very similar to the above. It follows in particular that

\[
f(x) - f(a) = \int_{[a,x]} \, d\mu.
\]

Note that up till now, no reference has been made to the absolute continuity of \( f \). Any increasing continuous function would be fine.

Now if \( E \) is a Borel set such that \( m(E) = 0 \). Then the outer regularity of \( m \) implies there exists an open set, \( V \) containing \( E \) such that \( m(V) < \delta \) where \( \delta \) corresponds to \( \varepsilon \) in the definition of absolute continuity of \( f \). Then letting \( \{I_k\} \) be the connected components of \( V \) it follows \( E \subseteq \bigcup_{k=1}^\infty I_k \) with \( \sum_k m(I_k) = m(V) < \delta \). Therefore, from absolute continuity of \( f \), it follows that for \( I_k = (a_k,b_k) \) and each \( n \)

\[
\mu(\bigcup_{k=1}^n I_k) = \sum_{k=1}^n \mu(I_k) = \sum_{k=1}^n |f(b_k) - f(a_k)| < \varepsilon
\]

and so letting \( n \to \infty \),

\[
\mu(E) \leq \mu(V) = \sum_{k=1}^\infty |f(b_k) - f(a_k)| \leq \varepsilon.
\]

Since \( \varepsilon \) is arbitrary, it follows \( \mu(E) = 0 \). Therefore, \( \mu \ll m \) and so by the Radon Nikodym theorem there exists a unique \( h \in L^1([a,b]) \) such that

\[
\mu(E) = \int_E h \, dm.
\]

In particular,

\[
\mu([a,x]) = f(x) - f(a) = \int_{[a,x]} \, h \, dm.
\]

From the fundamental theorem of calculus \( f'(x) = h(x) \) at every Lebesgue point of \( h \). Therefore, writing in usual notation,

\[
f(x) = f(a) + \int_a^x f'(t) \, dt
\]
as claimed. This proves the lemma.

With the above lemmas, the following is the main theorem about absolutely continuous functions.

**Theorem 7.2.5** Let \( f : [a, b] \to \mathbb{R} \) be absolutely continuous if and only if \( f'(x) \) exists a.e., \( f' \in L^1 ([a, b]) \) and

\[
    f(x) = f(a) + \int_a^x f'(t) \, dt.
\]

**Proof:** Suppose first that \( f \) is absolutely continuous. By Lemma 7.2.3, the total variation function, \( V \) is absolutely continuous and \( f(x) = V(x) - (V(x) - f(x)) \) where both \( V \) and \( V - f \) are increasing and absolutely continuous. By Lemma 7.2.4,

\[
    f(x) - f(a) = V(x) - V(a) - [(V(x) - f(x)) - (V(a) - f(a))]
\]

\[
    = \int_a^x V'(t) \, dt - \int_a^x (V - f)'(t) \, dt.
\]

Now \( f' \) exists and is in \( L^1 \) because \( f = V - (V - f) \) and \( V \) and \( V - f \) have derivatives in \( L^1 \). Therefore, \( (V - f)' = V' - f' \) and so the above reduces to

\[
    f(x) - f(a) = \int_a^x f'(t) \, dt.
\]

This proves one half of the theorem.

Now suppose \( f' \in L^1 \) and \( f(x) = f(a) + \int_a^x f'(t) \, dt \). It is necessary to verify that \( f \) is absolutely continuous. But this follows easily from the observation that a single function, \( f' \) is uniformly integrable. Thus if \( \sum_i |y_i - x_i| \) is sufficiently small then

\[
    \sum_i \left| \int_{x_i}^{y_i} f'(t) \, dt \right| = \sum_i |f(y_i) - f(x_i)| < \varepsilon.
\]

The following simple corollary is a case of Rademacher’s theorem.

**Corollary 7.2.6** Suppose \( f : [a, b] \to \mathbb{R} \) is Lipschitz continuous,

\[
    |f(x) - f(y)| \leq K |x - y|.
\]

Then \( f'(x) \) exists a.e. and

\[
    f(x) = f(a) + \int_a^x f'(t) \, dt.
\]

**Proof:** It is easy to see that \( f \) is absolutely continuous. Therefore, Theorem 7.2.5 applies.
Chapter 8

Differentiation With Respect To General Radon Measures

This is a brief chapter on certain important topics on the differentiation theory for general Radon measures. For different proofs and some results which are not discussed here, a good source is [35] which is where I first read some of these things.

8.1 Besicovitch Covering Theorem

When dealing with probability distribution functions or some other Radon measure, it is necessary to have a better covering theorem than the Vitali covering theorem which works well for Lebesgue measure. However, for a Radon measure, if you enlarge the ball by making the radius larger, you don’t know what happens to the measure of the enlarged ball except that its measure does not get smaller. Thus the thing required is a covering theorem which does not depend on enlarging balls.

The first fundamental observation is found in the following lemma which holds for the context illustrated by the following picture. This picture is drawn such that the balls come from the usual Euclidean norm, but the norm could be any norm on $\mathbb{R}^n$.

LEMMA 8.1.1 Let the balls $B_a$, $B_x$, $B_y$ be as shown having radii $r, r_x, r_y$ respectively. Suppose the centers of $B_x$ and $B_y$ are not both in any of the balls shown, and suppose $r_y \geq r_x \geq \alpha r$ where $\alpha$ is a number larger than 1. Also let $P_x \equiv a + r \frac{x-a}{||x-a||}$ with $P_y$ being defined similarly. Then it follows that $||P_x - P_y|| \geq \frac{\alpha - 1}{\alpha + 1} r$. There exists a constant $L(n, \alpha)$ depending on $\alpha$ and the dimension, such that if $B_1, \cdots , B_m$ are all balls such that any pair are in the same situation relative to $B_a$ as $B_x$, and $B_y$, then $m \leq L(n, \alpha)$.

Proof: From the definition,

$$||P_x - P_y|| = r \left( \frac{x-a}{||x-a||} - \frac{y-a}{||y-a||} \right)$$

$$= r \left( \frac{(x-a)||y-a| - (y-a)||x-a||}{||x-a||||y-a||} \right)$$

$$= r \left( \frac{||y-a|| (x-y) + (y-a)(||y-a|| - ||x-a||)}{||x-a||||y-a||} \right)$$

Intersections with big balls
\[
\geq r \frac{\|x - y\|}{\|x - a\|} - \frac{\alpha r}{\|x - a\|} \left( \|y - a\| - \|x - a\| \right). \tag{8.1.1}
\]

There are two cases. First suppose that \(\|y - a\| - \|x - a\| \geq 0\). Then this reduces to
\[
r \frac{\|x - y\|}{\|x - a\|} - \frac{\alpha r}{\|x - a\|} |y - a| + r.
\]

From the assumptions, this is no larger than
\[
\geq r \left( \frac{r_y}{\|x - a\|} - \frac{r + r_y}{\|x - a\|} + 1 \right) \geq r \left( 1 - \frac{r}{\|x - a\|} \right) \geq r \left( 1 - \frac{r}{r_x} \right).
\]

The other case is that \(\|y - a\| - \|x - a\| < 0\). Then in this case it reduces to
\[
r \frac{\|x - y\|}{\|x - a\|} - \frac{\alpha r}{\|x - a\|} (\|x - a\| - \|y - a\|) \geq \frac{r}{\|x - a\|} (\|x - y\| - \|x - a\| + \|y - a\|)
\]
\[
\geq \frac{r}{r_x + r} (r_y - (r + r_x) + r_y) \geq \frac{r}{r_x + r} (r_y - r) \geq \frac{r}{r_x + r} (r_x - r) \geq \frac{r}{r_x + \frac{1}{\alpha} r_x} \left( r_x - \frac{1}{\alpha} r_x \right) = \frac{\alpha - 1}{\alpha + 1} r.
\]

This proves the estimate between \(P_x\) and \(P_y\).

Finally, in the case of the balls \(B_i\) having centers at \(x_i\), let \(P_{x_i}\) be the expression \(a + r \frac{x_i - a}{\|x_i - a\|}\). Then \((P_{x_i} - a) r^{-1}\) is on the unit sphere having center 0. Furthermore,
\[
\| (P_{x_i} - a) r^{-1} - (P_{x_j} - a) r^{-1} \| = r^{-1} \| P_{x_i} - P_{x_j} \| \geq r^{-1} \frac{\alpha - 1}{\alpha + 1} = \frac{\alpha - 1}{\alpha + 1},
\]

How many points on the unit sphere can be pairwise this far apart? This set is compact and so there exists an 
\(\frac{1}{4} \left( \frac{\alpha - 1}{\alpha + 1} \right)\) net having \(L(n, \alpha)\) points. Thus \(m\) cannot be any larger than \(L(n, \alpha)\) because if it were, then by the pigeon hole principal, two of the points \((P_{x_i} - a) r^{-1}\) would lie in a single ball \(B \left( p, \frac{1}{4} \left( \frac{\alpha - 1}{\alpha + 1} \right) \right)\) so they could not be \(\frac{\alpha - 1}{\alpha + 1}\) apart.

The above lemma has do do with balls which are relatively large intersecting a given ball. Next is a lemma which has to do with relatively small balls intersecting a given ball. Note that in the statement of this lemma, the radii are smaller than \(\alpha r\) in contrast to the above lemma in which the radii of the balls are larger than \(\alpha r\).

**Intersections with small but comparable balls**

**Lemma 8.1.2** Let \(B\) be a ball having radius \(r\) and suppose \(B\) has nonempty intersection with the balls \(B_1, \ldots, B_m\) having radii \(r_1, \ldots, r_m\) respectively. Suppose \(\alpha, \gamma > 1\) and the \(r_i\) are comparable with \(r\) in the sense that
\[
\frac{1}{\gamma} r \leq r_i \leq \alpha r.
\]

Let \(B'_i\) have the same center as \(B_i\) with radius equal to \(r'_i = \beta r_i\) for some \(\beta < 1\). If the \(B'_i\) are disjoint, then there exists a constant \(M(n, \alpha, \beta, \gamma)\) such that \(m \leq M(n, \alpha, \beta, \gamma)\). Letting \(\alpha = 10, \beta = 1/3, \gamma = 4/3\), it follows that \(m \leq 60^n\).
8.1. BESICOVITCH COVERING THEOREM

**Proof:** Let the volume of a ball of radius \( r \) be given by \( \alpha(n) r^n \) where \( \alpha(n) \) depends on the norm used and on the dimension \( n \) as indicated. The idea is to enlarge \( B \), till it swallows all the \( B'_i \). Then, since they are disjoint and their radii are not too small, there can’t be too many of them.

This can be done for a single \( B'_i \) by enlarging the radius of \( B \) to \( r + r_i + r'_i \).

Then to get all the \( B_i \), you would just enlarge the radius of \( B \) to \( r + \alpha \gamma r\). Then, using the inequality which makes \( r_i \) comparable to \( r \), it follows that

\[
\sum_{i=1}^{m} \alpha(n) \left( \frac{\beta}{\gamma} r \right)^n \leq \sum_{i=1}^{m} \alpha(n) (\beta r_i)^n \leq \alpha(n) (1 + \alpha \beta)^n r^n
\]

Therefore,

\[
m \left( \frac{\beta}{\gamma} \right)^n \leq (1 + \alpha \beta)^n
\]

and so \( m \leq (1 + \alpha \beta)^n \left( \frac{2}{\gamma} \right)^{-n} \equiv M(n, \alpha, \beta, \gamma). \)

From now on, let \( \alpha = 10 \) and let \( \beta = 1/3 \) and \( \gamma = 4/3 \). Then

\[
M(n, \alpha, \beta, \gamma) \leq \left( \frac{172}{3} \right)^n \leq 60^n
\]

Thus \( m \leq 60^n \). ■

The next lemma gives a construction which yields balls which are comparable as described in the above lemma.

**A construction of a sequence of balls**

**Lemma 8.1.3** Let \( \mathcal{F} \) be a nonempty set of nonempty balls in \( \mathbb{R}^n \) with

\[
sup \{ \text{diam}(B) : B \in \mathcal{F} \} \leq D < \infty
\]

and let \( A \) denote the set of centers of these balls. Suppose \( A \) is bounded. Define a sequence of balls from \( \mathcal{F} \), \( \{ B_j \}_{j=1}^{J} \) where \( J \leq \infty \) such that

\[
r(B_1) \geq \frac{3}{4} sup \{ r(B) : B \in \mathcal{F} \}
\]

and if

\[
A_m \equiv A \setminus (\bigcup_{i=1}^{m} B_i) \neq \emptyset,
\]

then \( B_{m+1} \in \mathcal{F} \) is chosen with center in \( A_m \) such that

\[
r_{m+1} \equiv r(B_{m+1}) \geq \frac{3}{4} sup \{ r(B(a, r) : B(a, r) \in \mathcal{F}, a \in A_m \}.
\]

Then letting \( B_j = B(a_j, r_j) \), this sequence satisfies

\[
r(B_k) \leq \frac{4}{3} r(B_j) \text{ for } j < k,
\]

\[
\{ B(a_j, r_j/3) \}_{j=1}^{J} \text{ are disjoint},
\]

\[
A \subseteq \bigcup_{i=1}^{J} B_i.
\]
Proof: Consider Lemma 8.1.2. First note the sets \( A_m \) form a decreasing sequence. Thus from the definition of \( B_j \), for \( j < k \),

\[
\begin{align*}
    r(B_k) &\leq \sup \{ r : B(a,r) \in \mathcal{F}, a \in A_{k-1} \} \\
    &\leq \sup \{ r : B(a,r) \in \mathcal{F}, a \in A_{j-1} \} \leq \frac{4}{3} r(B_j)
\end{align*}
\]

because the construction gave

\[
r(B_j) \geq \frac{3}{4} \sup \{ r : B(a,r) \in \mathcal{F}, a \in A_{j-1} \}
\]

Next consider Lemma 8.1.3. If \( x \in B(a_j, r_j/3) \cap B(a_i, r_i/3) \) where these balls are two which are chosen by the above scheme such that \( j > i \), then from what was just shown

\[
||a_j - a_i|| \leq ||a_j - x|| + ||x - a_i|| \leq \frac{r_j}{3} + \frac{r_i}{3} = \left( \frac{4}{9} + \frac{1}{3} \right) r_i = \frac{7}{9} r_i < r_i
\]

and this contradicts the construction because \( a_j \) is not covered by \( B(a_i, r_i) \).

Finally consider the claim that \( A \subseteq \bigcup_{i=1}^m B_i \). Pick \( B_1 \) satisfying Lemma 8.1.2. If \( B_1, \ldots, B_m \) have been chosen, and \( A_m \) is given in Lemma 8.1.3, then if it equals \( \emptyset \), it follows \( A \subseteq \bigcup_{i=1}^m B_i \). Set \( J = m \). Now let \( a \) be the center of \( B_n \in \mathcal{F} \). If \( a \in A_m \) for all \( n \) (That is \( a \) does not get covered by the \( B_i \)) then \( r_{m+1} \geq \frac{3}{4} r(B_n) \) for all \( m \), a contradiction since the balls \( B(a_j, \frac{r_j}{3}) \) are disjoint and \( A \) is bounded, implying that \( r_j \to 0 \). Thus \( a \) must fail to be in some \( A_m \) which means it got covered by some ball in the sequence. ■

Note that in this sequence of balls from the above lemma, if \( j < k \),

\[
\frac{3}{4} r(B_k) \leq r(B_j)
\]

Then there are two cases to consider,

\[
r(B_j) \geq 10 r(B_k), r(B_j) \leq 10 r(B_k)
\]

In the first case, we use Lemma 8.1.2 to estimate the number of intersections of \( B_k \) with \( B_j \) for \( j < k \). In the second case, we use Lemma 8.1.2 to estimate the number of intersections of \( B_k \) with \( B_j \) for \( j < k \).

Now here is the Besicovitch covering theorem.

**Theorem 8.1.4** There exists a constant \( N_n, \) depending only on \( n \) with the following property. If \( \mathcal{F} \) is any collection of nonempty balls in \( \mathbb{R}^n \) with

\[
\sup \{ \text{diam}(B) : B \in \mathcal{F} \} < D < \infty
\]

and if \( A \) is the set of centers of the balls in \( \mathcal{F} \), then there exist subsets of \( \mathcal{F}, \mathcal{H}_1, \ldots, \mathcal{H}_N, \) such that each \( \mathcal{H}_i \) is a countable collection of disjoint balls from \( \mathcal{F} \) (possibly empty) and

\[
A \subseteq \bigcup_{i=1}^N \bigcup \{ B : B \in \mathcal{H}_i \}.
\]

**Proof:** To begin with, suppose \( A \) is bounded. Let \( L(n,10) \) be the constant of Lemma 8.1.2 and let \( M_n = L(n,10) + 60^n + 1 \). Define the following sequence of subsets of \( \mathcal{F}, \mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_{M_n} \). Referring to the sequence \( \{ B_k \} \) just considered, let \( B_1 \in \mathcal{G}_1 \) and if \( B_1, \ldots, B_m \) have been assigned, to each a \( \mathcal{G}_i \), place \( B_{m+1} \) in the first \( \mathcal{G}_j \) such that \( B_{m+1} \) intersects no set already in \( \mathcal{G}_j \). The existence of such a \( j \) follows from Lemmas 8.1.2 and 8.1.3. Here is why. \( B_{m+1} \) can intersect at most \( L(n,10) \) sets of \( \{ B_1, \ldots, B_m \} \) which have radii at least as large as \( 10B_{m+1} \) thanks to Lemma 8.1.3. It can intersect at most \( 60^n \) sets of \( \{ B_1, \ldots, B_m \} \) which have radius smaller than \( 10B_{m+1} \) thanks to Lemma 8.1.4. Thus each \( \mathcal{G}_j \) consists of disjoint sets of \( \mathcal{F} \) and the set of centers is covered by the union of these \( \mathcal{G}_j \).

This proves the theorem in case the set of centers is bounded.

Now let \( R_1 = B(0,5D) \) and if \( R_m \) has been chosen, let

\[
R_{m+1} = B(0, (m+1)5D) \setminus R_m
\]

Thus, if \( |k - m| \geq 2 \), no ball from \( \mathcal{F} \) having nonempty intersection with \( R_m \) can intersect any ball from \( \mathcal{F} \) which has nonempty intersection with \( R_k \). This is because all these balls have radius less than \( D \). Now let \( A_m \equiv A \cap R_m \) and apply the above result for a bounded set of centers to those balls of \( \mathcal{F} \) which intersect \( R_m \) to obtain sets of disjoint balls \( \mathcal{G}_1(R_m), \mathcal{G}_2(R_m), \ldots, \mathcal{G}_{M_n}(R_m) \) covering \( A_m \). Then simply define \( \mathcal{G}_j(R_1) = \bigcup_{k=1}^\infty \mathcal{G}_j(R_{2k}), \mathcal{G}_j(R_2) = \bigcup_{k=1}^\infty \mathcal{G}_j(R_{2k-1}) \).

Let \( N_n = 2M_n \) and

\[
\{ \mathcal{H}_1, \ldots, \mathcal{H}_{N_n} \} = \{ \mathcal{G}_1, \ldots, \mathcal{G}_{M_n}, \mathcal{G}_1, \ldots, \mathcal{G}_{M_n} \}
\]

Note that the balls in \( \mathcal{G}_j \) are disjoint. This is because those in \( \mathcal{G}_j(R_{2k}) \) are disjoint and if you consider any ball in \( \mathcal{G}_j(R_{2k}) \), it cannot intersect a ball of \( \mathcal{G}_j(R_{2m}) \) for \( m \neq k \) because \( |2k - 2m| \geq 2 \). Similar considerations apply to the balls of \( \mathcal{G}_j \). ■
8.2 Fundamental Theorem Of Calculus For Radon Measures

In this section the Besicovitch covering theorem will be used to give a generalization of the Lebesgue differentiation theorem to general Radon measures. In what follows, \( \mu \) will be a Radon measure,

\[
Z \equiv \{ x \in \mathbb{R}^n : \mu(B(x,r)) = 0 \text{ for some } r > 0 \},
\]

**Lemma 8.2.1** \( Z \) is measurable and \( \mu(Z) = 0 \).

**Proof:** For each \( x \in Z \), there exists a ball \( B(x,r) \) with \( \mu(B(x,r)) = 0 \). Let \( \mathcal{C} \) be the collection of these balls. Since \( \mathbb{R}^n \) has a countable basis, a countable subset, \( \tilde{\mathcal{C}} \), of \( \mathcal{C} \) also covers \( Z \). Let

\[
\tilde{\mathcal{C}} = \{ B_i \}_{i=1}^{\infty}.
\]

Then letting \( \bar{\mu} \) denote the outer measure determined by \( \mu \),

\[
\bar{\mu}(Z) \leq \sum_{i=1}^{\infty} \bar{\mu}(B_i) = \sum_{i=1}^{\infty} \mu(B_i) = 0
\]

Therefore, \( Z \) is measurable and has measure zero as claimed. \( \blacksquare \)

Let \( Mf : \mathbb{R}^n \to [0, \infty] \) by

\[
Mf(x) = \begin{cases} 
\sup_{r \leq 1} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f| \, d\mu & \text{if } x \notin Z \\
0 & \text{if } x \in Z 
\end{cases}
\]

**Theorem 8.2.2** Let \( \mu \) be a Radon measure and let \( f \in L^1(\mathbb{R}^n, \mu) \). Then for a.e. \( x \),

\[
\lim_{r \to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| \, d\mu(y) = 0
\]

**Proof:** First consider the following claim which is a weak type estimate of the same sort used when differentiating with respect to Lebesgue measure.

**Claim 1:** The following inequality holds for \( N_n \) the constant of the Besicovitch covering theorem.

\[
\bar{\mu}([Mf > \varepsilon]) \leq N_n \varepsilon^{-1} \|f\|_1
\]

**Proof:** First note \([Mf > \varepsilon] \cap Z = \emptyset \) and without loss of generality, you can assume \( \bar{\mu}([Mf > \varepsilon]) > 0 \). Next, for each \( x \in [Mf > \varepsilon] \) there exists a ball \( B_x = B(x,r_x) \) with \( r_x \leq 1 \) and

\[
\mu(B_x)^{-1} \int_{B(x,r_x)} |f| \, d\mu > \varepsilon.
\]

Let \( \mathcal{F} \) be this collection of balls so that \([Mf > \varepsilon] \) is the set of centers of balls of \( \mathcal{F} \). By the Besicovitch covering theorem,

\[
[Mf > \varepsilon] \subseteq \bigcup_{i=1}^{N_n} \{ B : B \in \mathcal{G}_i \}
\]

where \( \mathcal{G}_i \) is a collection of disjoint balls of \( \mathcal{F} \). Now for some \( i \),

\[
\bar{\mu}([Mf > \varepsilon]) / N_n \leq \mu(\bigcup \{ B : B \in \mathcal{G}_i \})
\]

because if this is not so, then

\[
\bar{\mu}([Mf > \varepsilon]) \leq \sum_{i=1}^{N_n} \mu(\bigcup \{ B : B \in \mathcal{G}_i \})
\]

\[
< \sum_{i=1}^{N_n} \frac{\bar{\mu}([Mf > \varepsilon])}{N_n} = \bar{\mu}([Mf > \varepsilon]),
\]

a contradiction. Therefore for this \( i \),

\[
\frac{\bar{\mu}([Mf > \varepsilon])}{N_n} \leq \mu(\bigcup \{ B : B \in \mathcal{G}_i \}) = \sum_{B \in \mathcal{G}_i} \mu(B) \leq \sum_{B \in \mathcal{G}_i} \varepsilon^{-1} \int_B |f| \, d\mu
\]

\[
\leq \varepsilon^{-1} \int_{\mathbb{R}^n} |f| \, d\mu = \varepsilon^{-1} \|f\|_1.
\]
This shows Claim 1.

**Claim 2:** If \( g \) is any continuous function defined on \( \mathbb{R}^n \), then for \( x \notin Z \),

\[
\lim_{r \to 0} \frac{1}{\mu(B(x, r))} \int_{B(x, r)} |g(y) - g(x)| \, d\mu(y) = 0
\]

and

\[
\lim_{r \to 0} \frac{1}{\mu(B(x, r))} \int_{B(x, r)} g(y) \, d\mu(y) = g(x). \quad (8.2.8)
\]

**Proof:** Since \( g \) is continuous at \( x \), whenever \( r \) is small enough,

\[
\frac{1}{\mu(B(x, r))} \int_{B(x, r)} |g(y) - g(x)| \, d\mu(y) \leq \frac{1}{\mu(B(x, r))} \int_{B(x, r)} \varepsilon \, d\mu(y) = \varepsilon.
\]

Claim 2 follows from the above and the triangle inequality. This proves the claim.

Now let \( g \in C_c(\mathbb{R}^n) \) and \( x \notin Z \). Then from the above observations about continuous functions,

\[
\overline{\mu} \left( \left\{ x \notin Z : \limsup_{r \to 0} \frac{1}{\mu(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, d\mu(y) > \varepsilon \right\} \right) \tag{8.2.9}
\]

\[
\leq \overline{\mu} \left( \left\{ x \notin Z : \limsup_{r \to 0} \frac{1}{\mu(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, d\mu(y) > \frac{\varepsilon}{2} \right\} \right)
+ \overline{\mu} \left( \left\{ x \notin Z : |g(x) - f(x)| > \frac{\varepsilon}{2} \right\} \right).
\]

\[
\leq \overline{\mu} \left( \left\{ M(f-g) > \frac{\varepsilon}{2} \right\} \right) + \overline{\mu} \left( \left\{ |f-g| > \frac{\varepsilon}{2} \right\} \right). \quad (8.2.10)
\]

Now

\[
\int_{|f-g|>\frac{\varepsilon}{2}} |f-g| \, d\mu \geq \frac{\varepsilon}{2} \overline{\mu} \left( \left\{ |f-g| > \frac{\varepsilon}{2} \right\} \right)
\]

and so from Claim 1 and hence is dominated by

\[
\left( \frac{2}{\varepsilon} + \frac{N_n}{\varepsilon} \right) \left\| f - g \right\|_{L^1(\mathbb{R}^n, \mu)}.
\]

But by regularity of Radon measures, \( C_c(\mathbb{R}^n) \) is dense in \( L^1(\mathbb{R}^n, \mu) \), and so since \( g \) in the above is arbitrary, this shows equals 0. Now

\[
\overline{\mu} \left( \left\{ x \notin Z : \limsup_{r \to 0} \frac{1}{\mu(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, d\mu(y) > 0 \right\} \right)
\leq \sum_{k=1}^{\infty} \overline{\mu} \left( \left\{ x \notin Z : \limsup_{r \to 0} \frac{1}{\mu(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, d\mu(y) > \frac{1}{k} \right\} \right) = 0
\]

By completeness of \( \mu \) this implies

\[
\left\{ x \notin Z : \limsup_{r \to 0} \frac{1}{\mu(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, d\mu(y) > 0 \right\}
\]

is a set of \( \mu \) measure zero. ■

The following corollary is the main result referred to as the Lebesgue Besicovich Differentiation theorem.

**Corollary 8.2.3** If \( f \in L^1_{\text{loc}}(\mathbb{R}^n, \mu) \), then for a.e. \( x \notin Z \),

\[
\lim_{r \to 0} \frac{1}{\mu(B(x, r))} \int_{B(x, r)} |f(y) - f(x)| \, d\mu(y) = 0. \quad (8.2.11)
\]
8.3. SLICING MEASURES

Theorem 8.3.1 The space $C_c(\mathbb{R}^m)$ with the norm

$$||f|| = \sup \{|f(y)| : y \in \mathbb{R}^m\}$$

is separable.

Proof: Let $D_l$ consist of all functions which are of the form

$$\sum_{|\alpha| \leq N} a_{\alpha} y^\alpha \left( \text{dist} \left( y, B(0,l+1)^C \right) \right)^{n_{\alpha}}$$

where $a_{\alpha} \in \mathbb{Q}$, $\alpha$ is a multi-index, and $n_{\alpha}$ is a positive integer. Consider $D \equiv \cup_l D_l$. Then $D$ is countable. If $f \in C_c(\mathbb{R}^n)$, then choose $l$ large enough that $\text{spt} (f) \subseteq B(0,l+1)$, a locally compact space, $f \in C_0(B(0,l+1))$. Then since $D_l$ separates the points of $B(0,l+1)$ is closed with respect to conjugates, and annihilates no point, it is dense in $C_0(B(0,l+1))$ by the Stone Weierstrass theorem. Alternatively, $D$ is dense in $C_0(\mathbb{R}^n)$ by Stone Weierstrass and $C_c(\mathbb{R}^n)$ is a subspace so it is also separable. So is $C_c(\mathbb{R}^n)^+$, the nonnegative functions in $C_c(\mathbb{R}^n)$. ■

From the regularity of Radon measures, the following lemma follows.

Lemma 8.3.2 If $\mu$ and $\nu$ are two Radon measures defined on $\sigma$ algebras, $S_\mu$ and $S_\nu$, of subsets of $\mathbb{R}^n$ and if $\mu(V) = \nu(V)$ for all $V$ open, then $\mu = \nu$ and $S_\mu = S_\nu$.

Proof: Every compact set is a countable intersection of open sets so the two measures agree on every compact set. Hence it is routine that the two measures agree on every $G_\delta$ and $F_\sigma$ set. (Recall $G_\delta$ sets are countable intersections of open sets and $F_\sigma$ sets are countable unions of closed sets.) Now suppose $E \in S_\nu$ is a bounded set. Then by regularity of $\nu$ there exists $G$ a $G_\delta$ set and $F$, an $F_\sigma$ set such that $F \subseteq E \subseteq G$ and $\nu (G \setminus F) = 0$. Then it is also true that $\mu(G \setminus F) = 0$. Hence $E = F \cup (E \setminus F)$ and $E \setminus F$ is a subset of $G \setminus F$, a set of $\mu$ measure zero. By completeness of $\mu$, it follows $E \in S_\mu$ and

$$\mu(E) = \mu(F) = \nu(F) = \nu(E).$$

If $E \in S_\nu$ not necessarily bounded, let $E_m = E \cap B(0,m)$ and then $E_m \in S_\mu$ and $\mu(E_m) = \nu(E_m)$. Letting $m \to \infty$, $E \in S_\mu$ and $\mu(E) = \nu(E)$. Similarly, $S_\mu \subseteq S_\nu$ and the two measures are equal on $S_\mu$.

The main result in the section is the following theorem.
Theorem 8.3.3 Let $\mu$ be a finite Radon measure on $\mathbb{R}^{n+m}$ defined on a $\sigma$ algebra, $\mathcal{F}$. Then there exists a unique finite Radon measure $\alpha$, defined on a $\sigma$ algebra $\mathcal{S}$, of sets of $\mathbb{R}^n$ which satisfies

$$\alpha (E) = \mu (E \times \mathbb{R}^m)$$

(8.3.12)

for all $E$ Borel. There also exists a Borel set of $\alpha$ measure zero $N$, such that for each $x \notin N$, there exists a Radon probability measure $\nu_x$ such that if $f$ is a nonnegative $\mu$ measurable function or a $\mu$ measurable function in $L^1(\mu)$,

$$y \to f(x,y) \text{ is } \nu_x \text{ measurable } \alpha \text{ a.e.}$$

$$x \to \int_{\mathbb{R}^m} f(x,y) \ d\nu_x(y) \text{ is } \alpha \text{ measurable}$$

(8.3.13)

and

$$\int_{\mathbb{R}^{n+m}} f(x,y) \ d\mu = \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}^m} f(x,y) \ d\nu_x(y) \right) \ d\alpha(x).$$

(8.3.14)

If $\tilde{\nu}_x$ is any other collection of Radon measures satisfying 8.3.12 and 8.3.13, then $\tilde{\nu}_x = \nu_x$ for $\alpha$ a.e. $x$.

Proof:

Existence and uniqueness of $\alpha$

First consider the uniqueness of $\alpha$. Suppose $\alpha_1$ is another Radon measure satisfying 8.3.12. Then in particular, $\alpha_1$ and $\alpha$ agree on open sets and so the two measures are the same by Lemma 8.3.2.

To establish the existence of $\alpha$, define $\alpha_0$ on Borel sets by

$$\alpha_0 (E) = \mu (E \times \mathbb{R}^m).$$

Thus $\alpha_0$ is a finite Borel measure and so it is finite on compact sets. Lemma 2.3.2 on Page 110 implies the existence of the Radon measure $\alpha$ extending $\alpha_0$.

Uniqueness of $\nu_x$

Next consider the uniqueness of $\nu_x$. Suppose $\nu_x$ and $\tilde{\nu}_x$ satisfy all conclusions of the theorem with exceptional sets denoted by $N$ and $\tilde{N}$ respectively. Then, enlarging $N$ and $\tilde{N}$, one may also assume, using Lemma 8.2.1, that for $x \notin N \cup \tilde{N}$, $\alpha (B(x,r)) > 0$ whenever $r > 0$. Now let

$$A = \prod_{i=1}^{m} (a_i, b_i)$$

where $a_i$ and $b_i$ are rational. Thus there are countably many such sets. Then from the conclusion of the theorem, if $x_0 \notin N \cup \tilde{N}$,

$$\frac{1}{\alpha (B(x_0,r))} \int_{B(x_0,r)} \int_{\mathbb{R}^m} X_A (y) d\nu_x (y) \ d\alpha = \frac{1}{\alpha (B(x_0,r))} \int_{B(x_0,r)} \int_{\mathbb{R}^m} X_A (y) d\tilde{\nu}_x (y) \ d\alpha,$$

and by the Lebesgue Besicovitch Differentiation theorem, there exists a set of $\alpha$ measure zero, $E_A$, such that if $x_0 \notin E_A \cup N \cup \tilde{N}$, then the limit in the above exists as $r \to 0$ and yields

$$\nu_{x_0} (A) = \tilde{\nu}_{x_0} (A).$$

Letting $E$ denote the union of all the sets $E_A$ for $A$ as described above, it follows that $E$ is a set of measure zero and if $x_0 \notin E \cup N \cup \tilde{N}$ then $\nu_{x_0} (A) = \tilde{\nu}_{x_0} (A)$ for all such sets $A$. But every open set can be written as a disjoint union of sets of this form and so for all such $x_0$, $\nu_{x_0} (V) = \tilde{\nu}_{x_0} (V)$ for all $V$ open. By Lemma 8.3.2 this shows the two measures are equal and proves the uniqueness assertion for $\nu_x$. It remains to show the existence of the measures $\nu_x$. 

Existence of $\nu_x$
For \( f \geq 0, f, g \in C_c(\mathbb{R}^m) \) and \( C_c(\mathbb{R}^n) \) respectively, define
\[
g \to \int_{\mathbb{R}^{n+m}} g(x) f(y) \, d\mu
\]
Since \( f \geq 0 \), this is a positive linear functional on \( C_c(\mathbb{R}^n) \). Therefore, there exists a unique Radon measure \( \nu_f \) such that for all \( g \in C_c(\mathbb{R}^n) \),
\[
\int_{\mathbb{R}^{n+m}} g(x) f(y) \, d\mu = \int_{\mathbb{R}^n} g(x) \, d\nu_f.
\]
I claim that \( \nu_f \ll \alpha \), the two being considered as measures on \( B(\mathbb{R}^n) \). Suppose then that \( K \) is a compact set and \( \alpha(K) = 0 \). Then let \( K \prec \prec V \) where \( V \) is open.
\[
\nu_f(K) = \int_{\mathbb{R}^n} \mathcal{X}_K(x) \, d\nu_f(x) \leq \int_{\mathbb{R}^n} g(x) \, d\nu_f(x) = \int_{\mathbb{R}^n+m} g(x) f(y) \, d\mu
\]
\[
\leq \int_{\mathbb{R}^{n+m}} \mathcal{X}_{V \times \mathbb{R}^m} (x, y) f(y) \, d\mu \leq ||f||_\infty \mu(V \times \mathbb{R}^m) = ||f||_\infty \alpha(V)
\]
Then for any \( \varepsilon > 0 \), one can choose \( V \) such that the right side is less than \( \varepsilon \). Therefore, \( \nu_f(K) = 0 \) also. By regularity considerations, \( \nu_f \ll \alpha \) as claimed.

It follows from the Radon Nikodym theorem the existence of a function \( \hat{h}_f \in L^1(\alpha) \) such that for all \( g \in C_c(\mathbb{R}^n) \),
\[
\int_{\mathbb{R}^n+m} g(x) f(y) \, d\mu = \int_{\mathbb{R}^n} g(x) \, d\nu_f = \int_{\mathbb{R}^n} g(x) \hat{h}_f(x) \, d\alpha.
\]
(8.3.15)
It is obvious from the formula that the map from \( f \in C_c(\mathbb{R}^m) \) to \( L^1(\alpha) \) given by \( f \to \hat{h}_f \) is linear. However, this is not sufficiently specific because functions in \( L^1(\alpha) \) are only determined a.e. However, for \( h_f \in L^1(\alpha) \), you can specify a particular representative \( \alpha \) a.e. By the fundamental theorem of calculus,
\[
\hat{h}_f(x) \equiv \lim_{r \to 0} \frac{1}{\alpha(B(x, r))} \int_{B(x, r)} h_f(z) \, d\alpha(z)
\]
(8.3.16)
exists off some set of measure zero \( Z_f \). Note that since this involves the integral over a ball, it does not matter which representative of \( h_f \) is placed in the formula. Therefore, \( \hat{h}_f(x) \) is well defined pointwise for all \( x \) not in some set of measure zero \( Z_f \). Since \( \hat{h}_f = h_f \) a.e. it follows that \( \hat{h}_f \) is well defined and will work in the formula. Let
\[
Z = \cup \{Z_f : f \in \mathcal{D}\}
\]
where \( \mathcal{D} \) is a countable dense subset of \( C_c(\mathbb{R}^m) \). Of course it is desired to have the limit \( \text{Squeez} \) hold for all \( f \), not just \( f \in \mathcal{D} \). We will show that this limit holds for all \( x \notin Z \). Thus, we will have \( x \to \hat{h}_f(x) \) defined by the above limit off \( Z \) and so, since \( \hat{h}_f(x) = h_f(x) \) a.e., it follows that
\[
\int_{\mathbb{R}^{n+m}} g(x) f(y) \, d\mu = \int_{\mathbb{R}^n} g(x) \, d\nu_f = \int_{\mathbb{R}^n} g(x) \hat{h}_f(x) \, d\alpha
\]
One could then take \( \hat{h}_f(x) \) to be defined as 0 for \( x \notin Z \).

For \( f \) an arbitrary function in \( C_c(\mathbb{R}^m) \) and \( f' \in \mathcal{D} \), a dense countable subset of \( C_c(\mathbb{R}^n) \), it follows from \( \text{Squeez} \),
\[
\left| \int_{\mathbb{R}^n} g(x) (h_f(x) - h_{f'}(x)) \, d\alpha \right| \leq ||f - f'||_\infty \int_{\mathbb{R}^{n+m}} |g(x)| \, d\mu
\]
Let \( g_k(x) \uparrow \mathcal{X}_{B(z, r)}(x) \) where \( z \notin Z \). Then by the dominated convergence theorem, the above implies
\[
\left| \int_{B(z, r)} (h_f(x) - h_{f'}(x)) \, d\alpha \right| \leq ||f - f'||_\infty \int_{B(z, r) \times \mathbb{R}^m} d\mu = ||f - f'||_\infty \alpha(B(z, r)).
\]
Dividing by \( \alpha(B(z, r)) \), it follows that if \( \alpha(B(z, r)) > 0 \) for all \( r > 0 \), then for all \( r > 0 \),
\[
\left| \frac{1}{\alpha(B(z, r))} \int_{B(z, r)} (h_f(x) - h_{f'}(x)) \, d\alpha \right| \leq ||f - f'||_\infty.
It follows that for \( f \in C_c(\mathbb{R}^m)^+ \) arbitrary and \( z \notin Z \),
\[
\limsup_{r \to 0} \frac{1}{\alpha(B(z, r))} \int_{B(z, r)} h_f(x) \, d\alpha - \liminf_{r \to 0} \frac{1}{\alpha(B(z, r))} \int_{B(z, r)} h_f(x) \, d\alpha = \limsup_{r \to 0} \frac{1}{\alpha(B(z, r))} \int_{B(z, r)} (h_f(x) - h_{f'}(x)) \, d\alpha(x) - \liminf_{r \to 0} \frac{1}{\alpha(B(z, r))} \int_{B(z, r)} (h_f(x) - h_{f'}(x)) \, d\alpha(x) \\
\leq \left| \limsup_{r \to 0} \frac{1}{\alpha(B(z, r))} \int_{B(z, r)} (h_f(x) - h_{f'}(x)) \, d\alpha(x) \right| + \left| \liminf_{r \to 0} \frac{1}{\alpha(B(z, r))} \int_{B(z, r)} (h_f(x) - h_{f'}(x)) \, d\alpha(x) \right| \\
\leq 2 \| f - f' \|_\infty
\]

and since \( f' \) is arbitrary, it follows that the limit of \( 8.3.10 \) holds for all \( f \in C_c(\mathbb{R}^m)^+ \) whenever \( z \notin Z \), the above set of measure zero.

Now for \( f \) an arbitrary real valued function of \( C_c(\mathbb{R}^n) \), simply apply the above result to positive and negative parts to obtain \( h_f \equiv h_{f^+} - h_{f^-} \) and \( \widehat{h}_f \equiv \widehat{h}_{f^+} - \widehat{h}_{f^-} \). Then it follows that for all \( f \in C_c(\mathbb{R}^m) \) and \( g \in C_c(\mathbb{R}^m) \)
\[
\int_{\mathbb{R}^{n+m}} g(x) f(y) \, d\mu = \int_{\mathbb{R}^n} g(x) \widehat{h}_f(x) \, d\alpha.
\]

It is obvious from the description given above that that for each \( x \notin Z \), the set of measure zero given above, that \( f \to \widehat{h}_f(x) \) is a positive linear functional. It is clear that it acts like a linear map for nonnegative \( f \) and so the usual trick just described above is well defined and delivers a positive linear functional. Hence by the Riesz representation theorem, there exists a unique \( \nu_x \) such that for all \( x \)
\[
\widehat{h}_f(x) = \int_{\mathbb{R}^m} f(y) \, d\nu_x(y).
\]

It follows that
\[
\int_{\mathbb{R}^{n+m}} g(x) f(y) \, d\mu = \int_{\mathbb{R}^m} g(x) \int_{\mathbb{R}^m} f(y) \, d\nu_x(y) \, d\alpha(x) \quad (8.3.17)
\]
and \( x \to \int_{\mathbb{R}^m} f(y) \, d\nu_x \) is \( \alpha \) measurable and \( \nu_x \) is a Radon measure.

Now let \( f_k \uparrow X_{\mathbb{R}^m} \) and \( g \geq 0 \). Then by monotone convergence theorem,
\[
\int_{\mathbb{R}^{n+m}} g(x) \, d\mu = \int_{\mathbb{R}^m} g(x) \int_{\mathbb{R}^m} d\nu_x \, d\alpha
\]
If \( g_k \uparrow X_{\mathbb{R}^n} \), the monotone convergence theorem shows that \( x \to \int_{\mathbb{R}^m} d\nu_x \) is \( L^1(\alpha) \).

Next let \( g_k \uparrow X_{B(x, r)} \) and use monotone convergence theorem to write
\[
\alpha(B(x, r)) \equiv \int_{B(x, r) \times \mathbb{R}^m} d\mu = \int_{B(x, r)} \int_{\mathbb{R}^m} d\nu_x \, d\alpha
\]

Then dividing by \( \alpha(B(x, r)) \) and taking a limit as \( r \to 0 \), it follows that for \( \alpha \) a.e. \( x \), \( 1 = \nu_x(\mathbb{R}^m) \), so these \( \nu_x \) are probability measures off a set of \( \alpha \) measure zero. Letting \( g_k(x) \uparrow X_A(x) \), \( f_k(y) \uparrow X_B(y) \) for \( A, B \) open, it follows that \( 8.3.14 \) is valid for \( g(x) \) replaced with \( X_A(x) \) and \( f(y) \) replaced with \( X_B(y) \).

Now let \( \mathcal{G} \) denote the Borel sets \( F \) of \( \mathbb{R}^{n+m} \) such that
\[
\int_{\mathbb{R}^{n+m}} X_F(x,y) \, d\mu(x,y) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} X_F(x,y) \, d\nu_x(y) \, d\alpha(x)
\]
and that all the integrals make sense. As just explained, this includes all Borel sets of the form \( F = A \times B \) where \( A, B \) are open. It is clear that \( \mathcal{G} \) is closed with respect to countable disjoint unions and complements, while sets of the form \( A \times B \) for \( A, B \) open form a \( \pi \) system. Therefore, by Lemma \( 8.3.3 \), \( \mathcal{G} \) contains the Borel sets which is the
8.3. SLICING MEASURES

smallest σ algebra which contains such products of open sets. It follows from the usual approximation with simple functions that if \( f \geq 0 \) and is Borel measurable, then

\[
\int_{\mathbb{R}^{n+m}} f(x, y) \, d\mu(x, y) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} f(x, y) \, d\nu_x(y) \, d\alpha(x)
\]

with all the integrals making sense.

This proves the theorem in the case where \( f \) is Borel measurable and nonnegative. It just remains to extend this to the case where \( f \) is only \( \mu \)-measurable. However, from regularity of \( \mu \) there exist Borel measurable functions \( g, h, g \leq f \leq h \) such that

\[
\int_{\mathbb{R}^{n+m}} f(x, y) \, d\mu(x, y) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} g(x, y) \, d\mu(x, y)
\]

\[
= \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} h(x, y) \, d\mu(x, y)
\]

It follows

\[
\int_{\mathbb{R}^n} \int_{\mathbb{R}^m} g(x, y) \, d\nu_x(y) \, d\alpha(x) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} h(x, y) \, d\nu_x(y) \, d\alpha(x)
\]

and so, since for \( \alpha \text{ a.e. } x, y \rightarrow g(x, y) \) and \( y \rightarrow h(x, y) \) are \( \nu_x \) measurable with

\[
0 = \int_{\mathbb{R}^m} (h(x, y) - g(x, y)) \, d\nu_x(y)
\]

and \( \nu_x \) is a Radon measure, hence complete, it follows for \( \alpha \text{ a.e. } x, y \rightarrow f(x, y) \) must be \( \nu_x \) measurable because it is equal to \( y \rightarrow g(x, y), \nu_x \text{ a.e.} \). Therefore, for \( \alpha \text{ a.e. } x \), it makes sense to write

\[
\int_{\mathbb{R}^m} f(x, y) \, d\nu_x(y).
\]

Similar reasoning applies to the above function of \( x \) being \( \alpha \)-measurable due to \( \alpha \) being complete. It follows

\[
\int_{\mathbb{R}^{n+m}} f(x, y) \, d\mu(x, y) = \int_{\mathbb{R}^{n+m}} g(x, y) \, d\mu(x, y)
\]

\[
= \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} g(x, y) \, d\nu_x(y) \, d\alpha(x)
\]

\[
= \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} f(x, y) \, d\nu_x(y) \, d\alpha(x)
\]

with everything making sense. ■
Chapter 9

The Bochner Integral

9.1 Strong And Weak Measurability

In this chapter \((\Omega, S, \mu)\) will be a \(\sigma\) finite measure space and \(X\) will be a Banach space which contains the values of either a function or a measure. The Banach space will be either a real or a complex Banach space but the field of scalars does not matter and so it is denoted by \(F\) with the understanding that \(F = \mathbb{C}\) unless otherwise stated. The theory presented here includes the case where \(X = \mathbb{R}^n\) or \(\mathbb{C}^n\) but it does not include the situation where \(f\) could have values in a space like \([0, \infty]\). To begin with here is a definition.

**Definition 9.1.1** A function, \(x: \Omega \to X\), for \(X\) a Banach space, is a simple function if it is of the form

\[
x(s) = \sum_{i=1}^{n} a_i \chi_{B_i}(s)
\]

where \(B_i \in S\) and \(\mu(B_i) < \infty\) for each \(i\). A function \(x\) from \(\Omega\) to \(X\) is said to be strongly measurable if there exists a sequence of simple functions \(\{x_n\}\) converging pointwise to \(x\). The function \(x\) is said to be weakly measurable if, for each \(f \in X'\),

\[
f \circ x
\]

is a scalar valued measurable function.

Earlier, a function was measurable if inverse images of open sets were measurable. Something similar holds here. The difference is that another condition needs to hold.

**Theorem 9.1.2** \(x\) is strongly measurable if and only if \(x^{-1}(U)\) is measurable for all \(U\) open in \(X\) and \(x(\Omega)\) is separable.

**Proof:** Suppose first \(x^{-1}(U)\) is measurable for all \(U\) open in \(X\) and \(x(\Omega)\) is separable. Let \(\{a_n\}_{n=1}^{\infty}\) be the dense subset of \(x(\Omega)\). It follows \(x^{-1}(B)\) is measurable for all \(B\) Borel because

\[
\{B: x^{-1}(B) \text{ is measurable}\}
\]

is a \(\sigma\) algebra containing the open sets. Let

\[
U^n_k \equiv \{z \in X : ||z - a_k|| \leq \min\{||z - a_l||_{l=1}^{n}\}\}.
\]

In words, \(U^n_k\) is the set of points of \(X\) which are as close to \(a_k\) as they are to any of the \(a_l\) for \(l \leq n\).

\[
B^n_k \equiv x^{-1}(U^n_k), \ D^n_k \equiv B^n_k \setminus \left(\bigcup_{l=1}^{k-1} B^n_l\right), \ D^n_1 \equiv B^n_1,
\]

and

\[
x_n(s) = \sum_{k=1}^{n} a_k \chi_{D^n_k}(s).
\]

Thus \(x_n(s)\) is a closest approximation to \(x(s)\) from \(\{a_k\}_{k=1}^{n}\) and so \(x_n(s) \to x(s)\) because \(\{a_n\}_{n=1}^{\infty}\) is dense in \(x(\Omega)\). Furthermore, \(x_n\) is measurable because each \(D^n_k\) is measurable.
Since \((\Omega, \mathcal{S}, \mu)\) is \(\sigma\) finite, there exists \(\Omega_n \uparrow \Omega\) with \(\mu(\Omega_n) < \infty\). Let
\[
y_n(s) = \mathcal{X}_{\Omega_n}(s) x_n(s).
\]
Then \(y_n(s) \to x(s)\) for each \(s\) because for any \(s, s \in \Omega_n\) if \(n\) is large enough. Also \(y_n\) is a simple function because it equals 0 off a set of finite measure.

Now suppose that \(x\) is strongly measurable. Then some sequence of simple functions, \(\{x_n\}\), converges pointwise to \(x\). Then \(x_n^{-1}(W)\) is measurable for every open set \(W\) because it is just a finite union of measurable sets. Thus, \(x_n^{-1}(W)\) is measurable for every Borel set \(W\). This follows by considering
\[
\{W : x_n^{-1}(W) \text{ is measurable}\}
\]
and observing this is a \(\sigma\) algebra which contains the open sets. Since \(X\) is a metric space, it follows that if \(U\) is an open set in \(X\), there exists a sequence of open sets, \(\{V_n\}\) which satisfies
\[
V_n \subseteq U, V_n \subseteq V_{n+1}, U = \bigcup_{n=1}^{\infty} V_n.
\]
Then
\[
x^{-1}(V_m) \subseteq \bigcup_{n < k \geq n} x_k^{-1}(V_m) \subseteq x^{-1}(V_n).
\]
This implies
\[
x^{-1}(U) = \bigcup_{m < \infty} x^{-1}(V_m)
\]
\[
\subseteq \bigcup_{m < \infty} \bigcup_{n < \infty} \bigcap_{k \geq n} x_k^{-1}(V_m) \subseteq \bigcup_{m < \infty} x^{-1}(V_n) \subseteq x^{-1}(U).
\]
Since
\[
x^{-1}(U) = \bigcup_{m < \infty} \bigcup_{n < \infty} \bigcap_{k \geq n} x_k^{-1}(V_m),
\]
it follows that \(x^{-1}(U)\) is measurable for every open \(U\). It remains to show \(x(\Omega)\) is separable. Let
\[
D = \text{all values of the simple functions } x_n
\]
which converge to \(x\) pointwise. Then \(D\) is clearly countable and dense in \(\overline{D}\), a set which contains \(x(\Omega)\).

**Claim:** \(x(\Omega)\) is separable.

**Proof of claim:** For \(n \in \mathbb{N}\), let \(B_n \equiv \{B(d,r) : 0 < r < \frac{1}{n}, r \text{ rational}, d \in D\}\). Thus \(B_n\) is countable. Let \(z \in \overline{D}\). Consider \(B(z, \frac{1}{n})\). Then there exists \(d \in D \cap B(z, \frac{1}{m})\). Now pick \(r \in \mathbb{Q} \cap (\frac{1}{m}, \frac{1}{n})\) so that \(B(d,r) \in B_n\). Now \(z \in B(d,r)\) and so this shows that \(x(\Omega) \subseteq \overline{D} \subseteq \bigcup B_n\) for each \(n\). Now let \(B'_n\) denote those sets of \(B_n\) which have nonempty intersection with \(x(\Omega)\). Say \(B'_n = B_n^\infty \cap x(\Omega)\). Then if \(z \in x(\Omega)\), \(z\) is contained in some set of \(B'_n\) which also contains a point of \(\{x_n^m\}_{m=1}^{\infty}\). Therefore, \(z\) is at least as close as \(2/n\) to some point of \(\{x_n^m\}_{m=1}^{\infty}\) which shows \(\{x_n^m\}_{m=1}^{\infty}\) is a countable dense subset of \(x(\Omega)\). Therefore \(x(\Omega)\) is separable. 

The last part also shows that a subset of a separable metric space is also separable. Therefore, the following simple corollary is obtained.

**Corollary 9.1.3** If \(X\) is a separable Banach space then \(x\) is strongly measurable if and only if \(x^{-1}(U)\) is measurable for all \(U\) open in \(X\).

The next lemma is interesting for its own sake. Roughly it says that if a Banach space is separable, then the unit ball in the dual space is weak * separable. This will be used to prove Pettis’s theorem, one of the major theorems in this subject which relates weak measurability to strong measurability.

**Lemma 9.1.4** If \(X\) is a separable Banach space with \(B'\) the closed unit ball in \(X'\), then there exists a sequence \(\{f_n\}_{n=1}^{\infty} \subseteq D' \subseteq B'\) with the property that for every \(x \in X\),
\[
||x|| = \sup_{f \in D'} |f(x)|
\]
If \(H\) is a dense subset of \(X'\) then \(D'\) may be chosen to be contained in \(H\).
9.1. STRONG AND WEAK MEASURABILITY

**Proof:** Let \( \{a_k\} \) be a countable dense set in \( X \), and consider the mapping 
\[
\phi_n : B' \to \mathbb{R}^n
\]
given by 
\[
\phi_n(f) = (f(a_1), \ldots, f(a_n)).
\]

Then \( \phi_n(B') \) is contained in a compact subset of \( \mathbb{R}^n \) because \( |f(a_k)| \leq \|a_k\| \). Therefore, there exists a countable dense subset of \( \phi_n(B') \), \( \{\phi_n(f_k)\}_{k=1}^{\infty} \). Then pick \( h_k^j \in B \cap B' \) such that \( \lim_{j \to \infty} \|f_k - h_k^j\| = 0 \). Then \( \{\phi_n(h_k^j), k, j\} \) must also be dense in \( \phi_n(B') \). Let \( D_n' = \{h_k^j, k, j\} \). Define 
\[
D' = \bigcup_{k=1}^{\infty} D_k'.
\]

Note that for each \( x \in X \), there exists \( f_x \in B' \) such that \( f_x(x) = \|x\| \). From the construction, 
\[
\|a_m\| = \sup \{|f(a_m)| : f \in D'\}
\]
because \( f_m(a_m) \) is the limit of numbers \( f(a_m) \) for \( f \in D_m' \subseteq D' \). Therefore, for \( x \) arbitrary, 
\[
\|x\| \leq \|x - a_m\| + \|a_m\| = \sup \{|f(a_m)| : f \in D'\} + \|x - a_m\|
\]
\[
\leq \sup \{|f(a_m - x) + f(x)| : f \in D'\} + \|x - a_m\|
\]
\[
\leq \sup \{|f(x)| : f \in D'\} + 2\|x - a_m\| \leq \|x\| + 2\|x - a_m\|.
\]

Since \( a_m \) is arbitrary and the \( \{a_m\}_{m=1}^{\infty} \) are dense, this establishes the claim of the lemma. \( \blacksquare \)

The next theorem is one of the most important results in the subject. It is due to Pettis and appeared in 1938.

**Theorem 9.1.5** If \( x \) has values in a separable Banach space \( X \). Then \( x \) is weakly measurable if and only if \( x \) is strongly measurable.

**Proof:** It is necessary to show \( x^{-1}(U) \) is measurable whenever \( U \) is open. Since every open set is a countable union of balls, it suffices to show \( x^{-1}(B(a,r)) \) is measurable for any ball, \( B(a,r) \). Since every open ball is the countable union of closed balls, it suffices to verify \( x^{-1} \left(B(a,r)^{\ast}\right) \) is measurable. From Lemma 9.1.4
\[
x^{-1} \left(B(a,r)^{\ast}\right) = \{s : ||x(s) - a|| \leq r\}
\]
\[
= \{s : \sup_{f \in D'} |f(x(s) - a)| \leq r\}
\]
\[
= \cap_{f \in D'} \{s : |f(x(s) - a)| \leq r\}
\]
\[
= \cap_{f \in D'} \{s : |f(x(s)) - f(a)| \leq r\}
\]
\[
= \cap_{f \in D'} (\cap f \circ x)^{-1} B(f(a),r)
\]
which equals a countable union of measurable sets because it is assumed that \( f \circ x \) is measurable for all \( f \in X' \).

Next suppose \( x \) is strongly measurable. Then there exists a sequence of simple functions \( x_n \) which converges to \( x \) pointwise. Hence for all \( f \in X' \), \( f \circ x_n \) is measurable and \( f \circ x_n \to f \circ x \) pointwise. Thus \( x \) is weakly measurable. \( \blacksquare \)

The same method of proof yields the following interesting corollary.

**Corollary 9.1.6** Let \( X \) be a separable Banach space and let \( \mathcal{B}(X) \) denote the \( \sigma \)-algebra of Borel sets. Let \( H \) be a dense subset of \( X' \). Then \( \mathcal{B}(X) = \sigma(H) = \mathcal{F} \), the smallest \( \sigma \)-algebra of subsets of \( X \) which has the property that every function, \( x^* \in H \) is measurable.

**Proof:** First I need to show \( \mathcal{F} \) contains open balls because then \( \mathcal{F} \) will contain the open sets and hence the Borel sets. As noted above, it suffices to show \( \mathcal{F} \) contains closed balls. Let \( D' \) be those functionals in \( B' \) defined in Lemma 9.1.4. Then
\[
\{x : ||x - a|| \leq r\} = \{x : \sup_{x^* \in D'} |x^*(x - a)| \leq r\}
\]
\[
= \cap_{x^* \in D'} \{x : |x^*(x - a)| \leq r\}
\]
\[
= \cap_{x^* \in D'} \{x : |x^*(x) - x^*(a)| \leq r\}
\]
\[
= \cap_{x^* \in D'} (\cap f \circ x)^{-1} \left(B(x^*(a),r)\right) \in \sigma(H)\]
which is measurable because this is a countable intersection of measurable sets. Thus $\mathcal{F}$ contains open sets so

$$\sigma(H) \equiv \mathcal{F} \supseteq \mathcal{B}(X).$$

To show the other direction for the inclusion, note that each $x^*$ is $\mathcal{B}(X)$ measurable because $x^{*-1}$ (open set) $= \sigma(H)$). Therefore, $\mathcal{B}(X) \supseteq \sigma(H).$ ■

It is important to verify the limit of strongly measurable functions is itself strongly measurable. This happens under very general conditions. Suppose $X$ is any separable metric space and let $\tau$ denote the open sets of $X.$ Then it is routine to see that

$$\tau$$ has a countable basis, $\mathcal{B}.$ \hspace{1cm} (9.1.1)

Whenever $U \in \mathcal{B},$ there exists a sequence of open sets, $\{V_m\}_{m=1}^{\infty},$ such that

$$\cdots \subseteq V_m \subseteq V_{m+1} \subseteq \cdots, \quad U = \bigcup_{m=1}^{\infty} V_m.$$ \hspace{1cm} (9.1.2)

**Theorem 9.1.7** Let $f_n$ and $f$ be functions mapping $\Omega$ to $X$ where $\mathcal{F}$ is a $\sigma$ algebra of measurable sets of $\Omega$ and $(X, \tau)$ is a topological space satisfying $\mathcal{F} \subseteq \mathcal{W}$. Then if $f_n$ is measurable, and $f(\omega) = \lim_{n \to \infty} f_n(\omega),$ it follows that $f$ is also measurable. (Pointwise limits of measurable functions are measurable.)

**Proof:** Let $\mathcal{B}$ be the countable basis of $\Omega$ and let $U \in \mathcal{B}.$ Let $\{V_m\}$ be the sequence of $\mathcal{B}.$ Since $f$ is the pointwise limit of $f_n,$

$$f^{-1}(V_m) \subseteq \{\omega : f_k(\omega) \in V_m \text{ for all } k \text{ large enough} \} \subseteq f^{-1}(V_m).$$

Therefore,

$$f^{-1}(U) = \bigcup_{m=1}^{\infty} f^{-1}(V_m) \subseteq \bigcup_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=1}^{\infty} f_k^{-1}(V_m)$$

$$\subseteq \bigcup_{m=1}^{\infty} f^{-1}(V_m) = f^{-1}(U).$$

It follows $f^{-1}(U) \in \mathcal{F}$ because it equals the expression in the middle which is measurable. Now let $W \in \tau.$ Since $\mathcal{B}$ is countable, $W = \bigcup_{n=1}^{\infty} U_n$ for some sets $U_n \in \mathcal{B}.$ Hence

$$f^{-1}(W) = \bigcup_{n=1}^{\infty} f^{-1}(U_n) \in \mathcal{F}.$$ ■

Note that the same conclusion would hold for any topological space with the property that for any open set $U,$ it has such a sequence of $V_k$ attached to it as in $\Omega.$$\Box$

**Corollary 9.1.8** $x$ is strongly measurable if and only if $x(\Omega)$ is separable and $x$ is weakly measurable.

**Proof:** Strong measurability clearly implies weak measurability. If $x_n(s) \to x(s)$ where $x_n$ is simple, then $f(x_n(s)) \to f(x(s))$ for all $f \in X'.$ Hence $f \circ x$ is measurable by Theorem $\Omega.$ because it is the limit of a sequence of measurable functions. Let $D$ denote the set of all values of $x_n.$ Then $D$ is a separable set containing $x(\Omega).$ Thus $D$ is a separable metric space. Therefore $x(\Omega)$ is separable also by the last part of the proof of Theorem $\Omega.$$\Box$

Now suppose $D$ is a countable dense subset of $x(\Omega)$ and $x$ is weakly measurable. Let $Z$ be the subset consisting of all finite linear combinations of $D$ with the scalars coming from the set of rational points of $\mathbb{F}.$ Thus, $Z$ is countable. Letting $Y = Z,$ $Y$ is a separable Banach space containing $x(\Omega).$ If $f \in Y'$, $f$ can be extended to an element of $X'$ by the Hahn Banach theorem. Therefore, $x$ is a weakly measurable $Y$ valued function. Now use Theorem $\Omega.$ to conclude $x$ is strongly measurable. $\Box$

Weakly measurable as defined above means $s \to x^*(x(s))$ is measurable for every $x^* \in X'.$ The next lemma ties this weak measurability to the usual version of measurability in which a function is measurable when inverse images of open sets are measurable.

**Lemma 9.1.9** Let $X$ be a Banach space and let $x : (\Omega, \mathcal{F}) \to K \subseteq X$ where $K$ is weakly compact and $X'$ is separable. Then $x$ is weakly measurable if and only if $x^{-1}(U) \in \mathcal{F}$ whenever $U$ is a weakly open set.

**Proof:** By Corollary $\Omega.$ on Page $\Omega.$ there exists a metric $d,$ such that the metric space topology with respect to $d$ coincides with the weak topology. Since $K$ is compact, it follows that $K$ is also separable. Hence it is completely separable and so there exists a countable basis of open sets $\mathcal{B}$ for the weak topology on $K.$ It follows that if $U$ is any weakly open set, covered by basic sets of the form $B_A(x, r)$ where $A$ is a finite subset of $X',$ there exists a countable collection of these sets of the form $B_A(x, r)$ which covers $U.$
Suppose now that \( x \) is weakly measurable. To show \( x^{-1}(U) \in \mathcal{F} \) whenever \( U \) is weakly open, it suffices to verify \( x^{-1}(B_A(z,r)) \in \mathcal{F} \) for any set, \( B_A(z,r) \). Let \( A = \{x^*_1, \ldots, x^*_m\} \). Then
\[
x^{-1}(B_A(z,r)) = \{s \in \Omega : \rho_A(x(s) - z) < r\}
\]
\[
\equiv \{s \in \Omega : \max_{x^* \in A} |x^*(x(s) - z)| < r\}
\]
\[
= \bigcup_{i=1}^m \{s \in \Omega : |x^*_i(x(s) - z)| < r\}
\]
\[
= \bigcup_{i=1}^m \{s \in \Omega : |x^*_i(x(s)) - x^*_i(z)| < r\}
\]
which is measurable because each \( x^*_i \circ x \) is given to be measurable.

Next suppose \( x^{-1}(U) \in \mathcal{F} \) whenever \( U \) is weakly open. Then in particular this holds when \( U = B_{x^*}(z,r) \) for arbitrary \( x^* \). Hence
\[
\{s \in \Omega : x(s) \in B_{x^*}(z,r)\} \in \mathcal{F}.
\]

But this says the same as
\[
\{s \in \Omega : |x^*(x(s)) - x^*(z)| < r\} \in \mathcal{F}
\]
Since \( x^*(z) \) can be a completely arbitrary element of \( \mathcal{F} \), it follows \( x^* \circ x \) is an \( \mathcal{F} \) valued measurable function. In other words, \( x \) is weakly measurable according to the former definition. ■

One can also define weak * measurability and prove a theorem just like the Pettis theorem above. The next lemma is the analogue of Lemma 9.1.8.

**Lemma 9.1.10** Let \( B \) be the closed unit ball in \( X \). If \( X' \) is separable, there exists a sequence \( \{x_m\}_{m=1}^\infty \equiv D \subseteq B \) with the property that for all \( y^* \in X' \),
\[
||y^*|| = \sup_{x \in D} |y^*(x)|.
\]

**Proof:** Let
\[
\{x^*_k\}_{k=1}^\infty
\]
be the dense subspace of \( X' \). Define \( \phi_n : B \to \mathbb{F}^n \) by
\[
\phi_n(x) \equiv (x^*_1(x), \ldots, x^*_n(x)).
\]

Then \( |x^*_k(x)| \leq ||x^*_k|| \) and so \( \phi_n(B) \) is contained in a compact subset of \( \mathbb{F}^n \). Therefore, there exists a countable set, \( D_n \subseteq B \) such that \( \phi_n(D_n) \) is dense in \( \phi_n(B) \). Let
\[
D \equiv \bigcup_{n=1}^\infty D_n.
\]

It remains to verify this works. Let \( y^* \in X' \). Then there exists \( y \) such that
\[
|y^*(y)| > ||y^*|| - \varepsilon.
\]
By density, there exists one of the \( x^*_k \) from the countable dense subset of \( X' \) such that also
\[
|x^*_k(y)| > ||y^*|| - \varepsilon, \quad ||x^*_k - y^*|| < \varepsilon.
\]
Now \( x^*_k(y) \in \phi_k(B) \) and so there exists \( x \in D_k \subseteq D \) such that
\[
|x^*_k(x)| > ||y^*|| - \varepsilon.
\]
Then since \( ||x^*_k - y^*|| < \varepsilon \), this implies
\[
|y^*(x)| \geq ||y^*|| - 2\varepsilon.
\]
Since \( \varepsilon > 0 \) is arbitrary,
\[
||y^*|| \leq \sup_{x \in D} |y^*(x)| \leq ||y^*||.
\]
The next theorem is another version of the Pettis theorem. First here is a definition.

**Definition 9.1.11** A function \( y \) having values in \( X' \) is weak * measurable, when for each \( x \in X \), \( y(\cdot)(x) \) is a measurable scalar valued function.

**Theorem 9.1.12** If \( X' \) is separable and \( y : \Omega \to X' \) is weak * measurable, then \( y \) is strongly measurable.
Proof: It is necessary to show \( y^{-1} \left( B(a^*, r) \right) \) is measurable. This will suffice because the separability of \( X' \) implies every open set is the countable union of such balls of the form \( B(a^*, r) \). It also suffices to verify inverse images of closed balls are measurable because every open ball is the countable union of closed balls. From Lemma 9.1.10,

\[
y^{-1} \left( B(a^*, r) \right) = \{ s : ||y(s) - a^*|| \leq r \}
\]

\[
= \left\{ s : \sup_{x \in D} |(y(s) - a^*) (x)| \leq r \right\}
\]

\[
= \left\{ s : \sup_{x \in D} |y(s) (x) - a^* (x)| \leq r \right\}
\]

\[
= \bigcap_{x \in D} y\left( \cdot \right) (x)^{-1} \left( B(a^* (x), r) \right)
\]

which is a countable intersection of measurable sets by hypothesis.

The following are interesting consequences of the theory developed so far and are of interest independent of the theory of integration of vector valued functions.

**Theorem 9.1.13** If \( X' \) is separable, then so is \( X \).

**Proof:** Let \( D = \{x_m\} \subseteq B \), the unit ball of \( X \), be the sequence promised by Lemma 9.1.10. Let \( V \) be all finite linear combinations of elements of \( \{x_m\} \) with rational scalars. Thus \( V \) is a separable subspace of \( X \). The claim is that \( \overline{V} = X \). If not, there exists \( x_0 \in X \setminus \overline{V} \).

But by the Hahn Banach theorem there exists \( x_0^* \in X' \) satisfying \( x_0^* (x_0) \neq 0 \), but \( x_0^* (v) = 0 \) for every \( v \in \overline{V} \). Hence

\[ ||x_0^*|| = \sup_{x \in D} |x_0^* (x)| = 0, \]

a contradiction.

**Corollary 9.1.14** If \( X \) is reflexive, then \( X \) is separable if and only if \( X' \) is separable.

**Proof:** From the above theorem, if \( X' \) is separable, then so is \( X \). Now suppose \( X \) is separable with a dense subset equal to \( D \). Then since \( X \) is reflexive, \( J(D) \) is dense in \( X'' \) where \( J \) is the James map satisfying \( J (x^*) = x^* (x) \). Then since \( X'' \) is separable, it follows from the above theorem that \( X' \) is also separable.

### 9.2 The Essential Bochner Integral

**Definition 9.2.1** Let \( a_k \in X \), a Banach space and let

\[
x(s) = \sum_{k=1}^{n} a_k X_{E_k} (s) \quad (9.2.3)
\]

where for each \( k \), \( E_k \) is measurable and \( \mu (E_k) < \infty \). Then define

\[
\int_{\Omega} x(s) \, d\mu \equiv \sum_{k=1}^{n} a_k \mu (E_k).
\]

**Proposition 9.2.2** Definition 9.2.1 is well defined.

**Proof:** It suffices to verify that if

\[
\sum_{k=1}^{n} a_k X_{E_k} (s) = 0,
\]

then

\[
\sum_{k=1}^{n} a_k \mu (E_k) = 0.
\]
Let $f \in X'$. Then
\[
f \left( \sum_{k=1}^{n} a_k \chi_{E_k} (s) \right) = \sum_{k=1}^{n} f (a_k) \chi_{E_k} (s) = 0
\]
and, therefore,
\[
0 = \int_{\Omega} \left( \sum_{k=1}^{n} f (a_k) \chi_{E_k} (s) \right) d\mu = \sum_{k=1}^{n} f (a_k) \mu (E_k) = f \left( \sum_{k=1}^{n} a_k \mu (E_k) \right).
\]
Since $f \in X'$ is arbitrary, and $X'$ separates the points of $X$, it follows that
\[
\sum_{k=1}^{n} a_k \mu (E_k) = 0
\]
as claimed. This proves the proposition.

It follows easily from this proposition that $\int_{\Omega} d\mu$ is well defined and linear on simple functions.

**Definition 9.2.3** A strongly measurable function $x$ is Bochner integrable if there exists a sequence of simple functions $x_n$ converging to $x$ pointwise and satisfying
\[
\int_{\Omega} ||x_n (s) - x_m (s)|| d\mu \to 0 \text{ as } m, n \to \infty. \tag{9.2.4}
\]
If $x$ is Bochner integrable, define
\[
\int_{\Omega} x (s) d\mu \equiv \lim_{n \to \infty} \int_{\Omega} x_n (s) d\mu. \tag{9.2.5}
\]

**Theorem 9.2.4** The Bochner integral is well defined and if $x$ is Bochner integrable and $f \in X'$,
\[
f \left( \int_{\Omega} x (s) d\mu \right) = \int_{\Omega} f (x (s)) d\mu \tag{9.2.6}
\]
and
\[
\left\| \int_{\Omega} x (s) d\mu \right\| \leq \int_{\Omega} ||x (s)|| d\mu. \tag{9.2.7}
\]
Also, the Bochner integral is linear. That is, if $a, b$ are scalars and $x, y$ are two Bochner integrable functions, then
\[
\int_{\Omega} (ax (s) + by (s)) d\mu = a \int_{\Omega} x (s) d\mu + b \int_{\Omega} y (s) d\mu \tag{9.2.8}
\]

**Proof:** First it is shown that the triangle inequality holds on simple functions and that the limit in (9.2.3) exists. Thus, if $x$ is given by (9.2.3) with the $E_k$ disjoint,
\[
\left\| \int_{\Omega} x (s) d\mu \right\|
\]
\[
= \left\| \int_{\Omega} \sum_{k=1}^{n} a_k \chi_{E_k} (s) d\mu \right\| = \left\| \sum_{k=1}^{n} a_k \mu (E_k) \right\|
\]
\[
\leq \sum_{k=1}^{n} ||a_k|| \mu (E_k) = \int_{\Omega} \sum_{k=1}^{n} ||a_k|| \chi_{E_k} (s) d\mu = \int_{\Omega} ||x (s)|| d\mu
\]
which shows the triangle inequality holds on simple functions. This implies
\[
\left\| \int_{\Omega} x_n (s) d\mu - \int_{\Omega} x_m (s) d\mu \right\| = \left\| \int_{\Omega} (x_n (s) - x_m (s)) d\mu \right\|
\]
\[
\leq \int_{\Omega} ||x_n (s) - x_m (s)|| d\mu
\]
which verifies the existence of the limit in (9.2.3). This completes the first part of the argument.
Next it is shown the integral does not depend on the choice of the sequence satisfying (9.2.9) so that the integral is well defined. Suppose \( y_n, x_n \) both satisfy (9.2.9) and converge to \( x \) pointwise. By Fatou’s lemma,

\[
\left\| \int_{\Omega} y_n d\mu - \int_{\Omega} x_m d\mu \right\| \leq \int_{\Omega} \| y_n - x \| \, d\mu + \int_{\Omega} \| x - x_m \| \, d\mu \\
\leq \liminf_{k \to \infty} \int_{\Omega} \| y_n - y_k \| \, d\mu + \liminf_{k \to \infty} \int_{\Omega} \| x_k - x_m \| \\
\leq \varepsilon / 2 + \varepsilon / 2
\]

if \( m \) and \( n \) are chosen large enough. Since \( \varepsilon \) is arbitrary, this shows the limit is the same for both sequences and demonstrates the Bochner integral is well defined.

It remains to verify the triangle inequality on Bochner integral functions and the claim about passing a continuous linear functional inside the integral. Let \( x \) be Bochner integrable and let \( x_n \) be a sequence which satisfies the conditions of the definition. Define

\[
y_n (s) \equiv \begin{cases} 
\begin{align*}
x_n (s) & \text{if } \| x_n (s) \| \leq 2 \| x (s) \|, \\
0 & \text{if } \| x_n (s) \| > 2 \| x (s) \|. 
\end{align*}
\end{cases}
\]

(9.2.9)

Thus

\[
y_n (s) = x_n (s) \chi_{\| x_n \| \leq 2 \| x \|} (s).
\]

If \( x (s) = 0 \) then \( y_n (s) = 0 \) for all \( n \). If \( \| x (s) \| > 0 \) then for all \( n \) large enough,

\[
y_n (s) = x_n (s).
\]

Thus, \( y_n (s) \to x (s) \) and

\[
\| y_n (s) \| \leq 2 \| x (s) \|. 
\]

(9.2.10)

By Fatou’s lemma,

\[
\int_{\Omega} \| x \| \, d\mu \leq \liminf_{n \to \infty} \int_{\Omega} \| x_n \| \, d\mu. 
\]

(9.2.11)

Also from (9.2.9) and the triangle inequality on simple functions, \( \left\{ \int_{\Omega} \| x_n \| \, d\mu \right\}_{n=1}^{\infty} \) is a Cauchy sequence and so it must be bounded. Therefore, by (9.2.10), (9.2.9), and the dominated convergence theorem,

\[
0 = \lim_{n,m \to \infty} \int_{\Omega} \| y_n - y_m \| \, d\mu 
\]

(9.2.12)

and it follows \( x_n \) can be replaced with \( y_n \) in Definition (9.2.9).

From Definition (9.2.1),

\[
f \left( \int_{\Omega} y_n d\mu \right) = \int_{\Omega} f (y_n) d\mu.
\]

Thus,

\[
f \left( \int_{\Omega} x d\mu \right) = \lim_{n \to \infty} f \left( \int_{\Omega} y_n d\mu \right) = \lim_{n \to \infty} \int_{\Omega} f (y_n) d\mu = \int_{\Omega} f (x) d\mu,
\]

the last equation holding from the dominated convergence theorem and (9.2.10) and (9.2.9). This shows (9.2.10). To verify (9.2.4),

\[
\left\| \int_{\Omega} x (s) d\mu \right\| = \lim_{n \to \infty} \left\| \int_{\Omega} y_n (s) d\mu \right\| \\
\leq \lim_{n \to \infty} \int_{\Omega} \| y_n (s) \| \, d\mu = \int_{\Omega} \| x (s) \| \, d\mu
\]

where the last equation follows from the dominated convergence theorem and (9.2.9), (9.2.11).

It remains to verify (9.2.8). Let \( f \in X' \). Then from (9.2.4),

\[
f \left( \int_{\Omega} (ax (s) + by (s)) d\mu \right) = \int_{\Omega} (af (x (s)) + bf (y (s))) d\mu \\
= a \int_{\Omega} f (x (s)) d\mu + b \int_{\Omega} f (y (s)) d\mu \\
= f \left( a \int_{\Omega} x (s) d\mu + b \int_{\Omega} y (s) d\mu \right).
\]

(9.2.13)
Since $X'$ separates the points of $X$, it follows
\[
\int_{\Omega} (ax(s) + by(s)) \, d\mu = a \int_{\Omega} x(s) \, d\mu + b \int_{\Omega} y(s) \, d\mu
\]
and this proves 9.2.8. This proves the theorem.

**Theorem 9.2.5** An $X$ valued function, $x$, is Bochner integrable if and only if $x$ is strongly measurable and
\[
\int_{\Omega} ||x(s)|| \, d\mu < \infty. \tag{9.2.13}
\]
In this case there exists a sequence of simple functions $\{y_n\}$ satisfying 9.2.4, $y_n(s)$ converging pointwise to $x(s)$,
\[
||y_n(s)|| \leq 2 ||x(s)|| \tag{9.2.14}
\]
and
\[
\lim_{n \to \infty} \int_{\Omega} ||x(s) - y_n(s)|| \, d\mu = 0. \tag{9.2.15}
\]

**Proof:** Suppose $x$ is strongly measurable and condition 9.2.13 holds. Since $x$ is strongly measurable, there exists a sequence of simple functions, $\{x_n\}$ converging pointwise to $x$. As before, let
\[
y_n(s) = \begin{cases} 
  x_n(s) \text{ if } ||x_n(s)|| \leq 2 ||x(s)||, \\
  0 \text{ if } ||x_n(s)|| > 2 ||x(s)||.
\end{cases} \tag{9.2.16}
\]
Then 9.2.4 holds for $y_n$ and $y_n(s) \to x(s)$. Also
\[
0 = \lim_{m,n \to \infty} \int_{\Omega} ||y_n(s) - y_m(s)|| \, d\mu
\]
since otherwise, there would exist $\varepsilon > 0$ and $N_\varepsilon \to \infty$ as $\varepsilon \to 0$ and $n_\varepsilon, m_\varepsilon > N_\varepsilon$ such that
\[
\int_{\Omega} ||y_{n_\varepsilon}(s) - y_{m_\varepsilon}(s)|| \, d\mu \geq \varepsilon.
\]
But then taking a limit as $\varepsilon \to 0$ and using the dominated convergence theorem and 9.2.13 and 9.2.14, this would imply $0 \geq \varepsilon$. Therefore, $x$ is Bochner integrable. 9.2.15 follows from the dominated convergence theorem and 9.2.14.

Now suppose $x$ is Bochner integrable. Then it is strongly measurable and there exists a sequence of simple functions $\{x_n\}$ such that $x_n(s)$ converges pointwise to $x$ and
\[
\lim_{m,n \to \infty} \int_{\Omega} ||x_n(s) - x_m(s)|| \, d\mu = 0.
\]
Therefore, as before, since $\left\{ \int_{\Omega} x_n \, d\mu \right\}_{n=1}^{\infty}$ is a Cauchy sequence, it follows
\[
\left\{ \int_{\Omega} ||x_n|| \, d\mu \right\}_{n=1}^{\infty}
\]
is also a Cauchy sequence because
\[
\left| \int_{\Omega} ||x_n|| \, d\mu - \int_{\Omega} ||x_m|| \, d\mu \right| \leq \int_{\Omega} ||x_n|| - ||x_m|| \, d\mu
\]
\[
\leq \int_{\Omega} ||x_n - x_m|| \, d\mu.
\]
Thus
\[
\int_{\Omega} ||x|| \, d\mu \leq \lim \inf_{n \to \infty} \int_{\Omega} ||x_n|| \, d\mu < \infty
\]
Using 9.2.13 it follows $y_n$ satisfies 9.2.4, converges pointwise to $x$ and then from the dominated convergence theorem 9.2.14 holds. This proves the theorem.

Here is a simple corollary.
Corollary 9.2.6 Let an $X$ valued function $x$ be Bochner integrable and let $L \in \mathcal{L}(X, Y)$ where $Y$ is another Banach space. Then $Lx$ is a $Y$ valued Bochner integrable function and

$$L \left( \int_{\Omega} x(s) \, d\mu \right) = \int_{\Omega} Lx(s) \, d\mu$$

Proof: From Theorem 9.2.5 there is a sequence of simple functions $\{y_n\}$ having the properties listed in that theorem. Then consider $\{Ly_n\}$ which converges pointwise to $Lx$. Since $L$ is continuous and linear,

$$\int_{\Omega} \|Ly_n - Lx\|_Y \, d\mu \leq \|L\| \int_{\Omega} \|y_n - x\|_X \, d\mu$$

which converges to 0. This implies

$$\lim_{m, n \to \infty} \int_{\Omega} \|Ly_n - Ly_m\| \, d\mu = 0$$

and so by definition $Lx$ is Bochner integrable. Also

$$\int_{\Omega} x(s) \, d\mu = \lim_{n \to \infty} \int_{\Omega} y_n(s) \, d\mu$$

$$\int_{\Omega} Lx(s) \, d\mu = \lim_{n \to \infty} \int_{\Omega} Ly_n(s) \, d\mu$$

$$= \lim_{n \to \infty} L \int_{\Omega} y_n(s) \, d\mu$$

$$\left\|L \left( \int_{\Omega} x(s) \, d\mu \right) - \int_{\Omega} Lx(s) \, d\mu \right\|_Y \leq \left\|L \left( \int_{\Omega} x(s) \, d\mu \right) - L \int_{\Omega} y_n(s) \, d\mu \right\|_Y$$

$$+ \left\|\int_{\Omega} Ly_n(s) \, d\mu - \int_{\Omega} Lx(s) \, d\mu \right\|_Y < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

whenever $n$ large enough. This proves the corollary.

### 9.3 The Spaces $L^p(\Omega; X)$

**Definition 9.3.1** $x \in L^p(\Omega; X)$ for $p \in [1, \infty)$ if $x$ is strongly measurable and

$$\int_{\Omega} \|x(s)\|^p \, d\mu < \infty$$

Also

$$\|x\|_{L^p(\Omega; X)} \equiv \|x\|_p \equiv \left( \int_{\Omega} \|x(s)\|^p \, d\mu \right)^{1/p} . \quad (9.3.17)$$

As in the case of scalar valued functions, two functions in $L^p(\Omega; X)$ are considered equal if they are equal a.e. With this convention, and using the same arguments found in the presentation of scalar valued functions it is clear that $L^p(\Omega; X)$ is a normed linear space with the norm given by $\|\cdot\|_{L^p(\Omega; X)}$. In fact, $L^p(\Omega; X)$ is a Banach space. This is the main contribution of the next theorem.

**Lemma 9.3.2** If $x_n$ is a Cauchy sequence in $L^p(\Omega; X)$ satisfying

$$\sum_{n=1}^{\infty} \|x_{n+1} - x_n\|_p < \infty,$$

then there exists $x \in L^p(\Omega; X)$ such that $x_n(s) \to x(s)$ a.e. and

$$\|x - x_n\|_p \to 0.$$
9.3. THE SPACES $L^p (\Omega; X)$

Proof: Let

$$g_N (s) \equiv \sum_{n=1}^{N} ||x_{n+1} (s) - x_n (s)||_X$$

Then by the triangle inequality,

$$\left( \int_{\Omega} g_N (s)^p \, d\mu \right)^{1/p} \leq \sum_{n=1}^{N} \left( \int_{\Omega} ||x_{n+1} (s) - x_n (s)||^p \, d\mu \right)^{1/p}$$

$$\leq \sum_{n=1}^{\infty} ||x_{n+1} - x_n||_p < \infty.$$ 

Let

$$g (s) = \lim_{N \to \infty} g_N (s) = \sum_{n=1}^{\infty} ||x_{n+1} (s) - x_n (s)||_X.$$ 

By the monotone convergence theorem,

$$\left( \int_{\Omega} g (s)^p \, d\mu \right)^{1/p} = \lim_{N \to \infty} \left( \int_{\Omega} g_N (s)^p \, d\mu \right)^{1/p} < \infty.$$ 

Therefore, there exists a set of measure 0, $E$, such that for $s \notin E$, $g (s) < \infty$. Hence, for $s \notin E$,

$$\lim_{N \to \infty} x_{N+1} (s)$$

exists because

$$x_{N+1} (s) = x_{N+1} (s) - x_1 (s) + x_1 (s) = \sum_{n=1}^{N} (x_{n+1} (s) - x_n (s)) + x_1 (s).$$

Thus, if $N > M$, and $s$ is a point where $g (s) < \infty$,

$$||x_{N+1} (s) - x_{M+1} (s)||_X \leq \sum_{n=M+1}^{N} ||x_{n+1} (s) - x_n (s)||_X$$

$$\leq \sum_{n=M+1}^{\infty} ||x_{n+1} (s) - x_n (s)||_X$$

which shows that \(\{x_{n+1} (s)\}_{n=1}^{\infty}\) is a Cauchy sequence. Now let

$$x (s) \equiv \begin{cases} 
\lim_{N \to \infty} x_N (s) & \text{if } s \notin E, \\
0 & \text{if } s \in E.
\end{cases}$$

By Theorem \ref{thm:weak_measurable}, $x_n (\Omega)$ is separable for each $n$. Therefore, $x (\Omega)$ is also separable. Also, if $f \in X'$, then

$$f (x (s)) = \lim_{N \to \infty} f (x_N (s))$$

if $s \notin E$ and $f (x (s)) = 0$ if $s \in E$. Therefore, $f \circ x$ is measurable because it is the limit of the measurable functions,

$$f \circ x_N \in C(\Omega).$$

Since $x$ is weakly measurable and $x (\Omega)$ is separable, Corollary \ref{cor:strongly_measurable} shows that $x$ is strongly measurable. By Fatou’s lemma,

$$\int_{\Omega} ||x (s) - x_N (s)||^p \, d\mu \leq \liminf_{M \to \infty} \int_{\Omega} ||x_M (s) - x_N (s)||^p \, d\mu.$$ 

But if $N$ and $M$ are large enough with $M > N$,

$$\left( \int_{\Omega} ||x_M (s) - x_N (s)||^p \, d\mu \right)^{1/p} \leq \sum_{n=N}^{M} ||x_{n+1} - x_n||_p$$

$$\leq \sum_{n=N}^{\infty} ||x_{n+1} - x_n||_p < \varepsilon.$$
and this shows, since \( \varepsilon \) is arbitrary, that
\[
\lim_{N \to \infty} \int_{\Omega} ||x(s) - x_N(s)||^p d\mu = 0.
\]

It remains to show \( x \in L^p(\Omega; X) \). This follows from the above and the triangle inequality. Thus, for \( N \) large enough,
\[
\left( \int_{\Omega} ||x(s)||^p d\mu \right)^{1/p} \leq \left( \int_{\Omega} ||x_N(s)||^p d\mu \right)^{1/p} + \left( \int_{\Omega} ||x(s) - x_N(s)||^p d\mu \right)^{1/p} \leq \left( \int_{\Omega} ||x_N(s)||^p d\mu \right)^{1/p} + \varepsilon < \infty.
\]

This proves the lemma.

**Theorem 9.3.3** \( L^p(\Omega; X) \) is complete. Also every Cauchy sequence has a subsequence which converges pointwise.

**Proof:** If \( \{x_n\} \) is Cauchy in \( L^p(\Omega; X) \), extract a subsequence \( \{x_{n_k}\} \) satisfying
\[
||x_{n_{k+1}} - x_{n_k}||_p \leq 2^{-k}
\]
and apply Lemma 9.3.2. The pointwise convergence of this subsequence was established in the proof of this lemma. This proves the theorem because if a subsequence of a Cauchy sequence converges, then the Cauchy sequence must also converge.

**Observation 9.3.4** If the measure space is Lebesgue measure then you have continuity of translation in \( L^p(\mathbb{R}^n; X) \) in the usual way. More generally, for \( \mu \) a Radon measure on \( \Omega \) a locally compact Hausdorff space, \( C_c(\Omega; X) \) is dense in \( L^p(\Omega; X) \). Here \( C_c(\Omega; X) \) is the space of continuous \( X \) valued functions which have compact support in \( \Omega \). The proof of this little observation follows immediately from approximating with simple functions and then applying the appropriate considerations to the simple functions.

Clearly Fatou’s lemma and the monotone convergence theorem make no sense for functions with values in a Banach space but the dominated convergence theorem holds in this setting.

**Theorem 9.3.5** If \( x \) is strongly measurable and \( x_n(s) \to x(s) \) a.e. with \( ||x_n(s)|| \leq g(s) \) a.e.

where \( g \in L^1(\Omega) \), then \( x \) is Bochner integrable and
\[
\int_{\Omega} x(s) d\mu = \lim_{n \to \infty} \int_{\Omega} x_n(s) d\mu.
\]

**Proof:** \( ||x_n(s) - x(s)|| \leq 2g(s) \) a.e. so by the usual dominated convergence theorem,
\[
0 = \lim_{n \to \infty} \int_{\Omega} ||x_n(s) - x(s)|| d\mu.
\]

Also,
\[
\int_{\Omega} ||x_n(s) - x_m(s)|| d\mu \\
\leq \int_{\Omega} ||x_n(s) - x(s)|| d\mu + \int_{\Omega} ||x_m(s) - x(s)|| d\mu,
\]
and so \( \{x_n\} \) is a Cauchy sequence in \( L^1(\Omega; X) \). Therefore, by Theorem 9.3.3, there exists \( y \in L^1(\Omega; X) \) and a subsequence \( x_{n'} \) satisfying
\[
x_{n'}(s) \to y(s) \text{ a.e. and in } L^1(\Omega; X).
\]
But \( x(s) = \lim_{n' \to \infty} x_{n'}(s) \) a.e. and so \( x(s) = y(s) \) a.e. Hence
\[
\int_{\Omega} \|x(s)\| \, d\mu = \int_{\Omega} \|y(s)\| \, d\mu < \infty
\]
which shows that \( x \) is Bochner integrable. Finally, since the integral is linear,
\[
\left\| \int_{\Omega} x(s) \, d\mu - \int_{\Omega} x_n(s) \, d\mu \right\| = \left\| \int_{\Omega} (x(s) - x_n(s)) \, d\mu \right\| \\
\leq \int_{\Omega} \|x_n(s) - x(s)\| \, d\mu,
\]
and this last integral converges to 0. This proves the theorem.

One can also prove a version of the Vitali convergence theorem. To do this, here is a more general version of Egoroff’s theorem.

**Theorem 9.3.6 (Egoroff)** Let \( (\Omega, F, \mu) \) be a finite measure space, 
\[
(\mu(\Omega) < \infty)
\]
and let \( f_n, f \) be \( X \) valued measurable functions where \( X \) is a separable metric space and for all \( \omega \notin E \) where \( \mu(E) = 0 \)
\[
f_n(\omega) \to f(\omega)
\]
Then for every \( \varepsilon > 0 \), there exists a set, 
\[
F \supseteq E, \mu(F) < \varepsilon,
\]
such that \( f_n \) converges uniformly to \( f \) on \( F^c \).

**Proof:** First suppose \( E = \emptyset \) so that convergence is pointwise everywhere. Let 
\[
E_{km} = \{ \omega \in \Omega : d(f_n(\omega), f(\omega)) \geq 1/m \text{ for some } n > k \}.
\]

**Claim:** \([\omega : d(f_n(\omega), f(\omega)) \geq 1/m] \) is measurable.

**Proof of claim:** Let \( \{x_k\}_{k=1}^\infty \) be a countable dense subset of \( X \) and let \( r \) denote a positive rational number, \( \mathbb{Q}^+ \).
Then
\[
\bigcup_{k \in \mathbb{N}, r \in \mathbb{Q}} f_n^{-1}(B(x_k, r)) \cap f^{-1}\left(B\left(x_k, \frac{1}{m} - r\right)\right) = \left[d(f, f_n) < \frac{1}{m}\right]
\]
(9.3.18)

Here is why. If \( \omega \) is in the set on the left, then \( d(f_n(\omega), x_k) < r \) and 
\[
d(f(\omega), x_k) < \frac{1}{m} - r.
\]

Therefore,
\[
d(f(\omega), f_n(\omega)) < r + \frac{1}{m} - r = \frac{1}{m}.
\]

Thus the left side is contained in the right. Now let \( \omega \) be in the right side. That is \( d(f_n(\omega), f(\omega)) < \frac{1}{m} \). Choose 
\( 2r < \frac{1}{m} - d(f_n(\omega), f(\omega)) \) and pick \( x_k \in B(f_n(\omega), r) \). Then
\[
d(f(\omega), x_k) \leq d(f(\omega), f_n(\omega)) + d(f_n(\omega), x_k) < \frac{1}{m} - 2r + r = \frac{1}{m} - r
\]

Thus \( \omega \in f_n^{-1}(B(x_k, r)) \cap f^{-1}\left(B\left(x_k, \frac{1}{m} - r\right)\right) \) and so \( \omega \) is in the left side. Thus the two sets are equal. Now the set on the left in (9.3.18) is measurable because it is a countable union of measurable sets. This proves the claim since 
\[
\left[\omega : d(f_n(\omega), f(\omega)) \geq 1/m\right]
\]
is the complement of this measurable set.

Hence $E_{km}$ is measurable because

$$E_{km} = \bigcup_{n=k+1}^{\infty} \left[ \omega : d(f_n(\omega), f(\omega)) \geq \frac{1}{m} \right].$$

For fixed $m, \cap_{k=1}^{\infty} E_{km} = \emptyset$ because $f_n(\omega)$ converges to $f(\omega)$. Therefore, if $\omega \in \Omega$ there exists $k$ such that if $n > k$, $|f_n(\omega) - f(\omega)| < \frac{1}{m}$ which means $\omega \notin E_{km}$. Note also that

$$E_{km} \supseteq E_{(k+1)m}.$$

Since $\mu(E_{1m}) < \infty$,

$$0 = \mu(\cap_{k=1}^{\infty} E_{km}) = \lim_{k \to \infty} \mu(E_{km}).$$

Let $k(m)$ be chosen such that $\mu(E_{k(m)m}) < \varepsilon 2^{-m}$ and let

$$F = \bigcup_{m=1}^{\infty} E_{k(m)m}.$$ 

Then $\mu(F) < \varepsilon$ because

$$\mu(F) \leq \sum_{m=1}^{\infty} \mu(E_{k(m)m}) < \sum_{m=1}^{\infty} \varepsilon 2^{-m} = \varepsilon$$

Now let $\eta > 0$ be given and pick $m_0$ such that $m_0^{-1} < \eta$. If $\omega \in F^C$, then

$$\omega \in \bigcap_{m=1}^{\infty} E_{k(m)m}^C.$$ 

Hence $\omega \in F_{k(m_0)m_0}^C$ so

$$d(f(\omega), f_n(\omega)) < 1/m_0 < \eta$$

for all $n > k(m_0)$. This holds for all $\omega \in F^C$ and so $f_n$ converges uniformly to $f$ on $F^C$.

Now if $E \neq \emptyset$, consider $\{X_{E\cap F} f_n\}_{n=1}^{\infty}$. Then $X_{E\cap F} f_n$ is measurable and the sequence converges pointwise to $X_E f$ everywhere. Therefore, from the first part, there exists a set of measure less than $\varepsilon, F$ such that on $F^C, \{X_{E\cap F} f_n\}$ converges uniformly to $X_{E\cap F} f$. Therefore, on $(E \cup F)^C, \{f_n\}$ converges uniformly to $f$. This proves the theorem.

Now here is the Vitali convergence theorem and a definition.

**Definition 9.3.7** Let $A \subseteq L^1(\Omega; X)$. Then $A$ is said to be uniformly integrable if for every $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $\mu(E) < \delta$, it follows

$$\int_E ||f||_X \, d\mu < \varepsilon$$

for all $f \in A$. It is bounded if

$$\sup_{f \in A} \int_{\Omega} ||f||_X \, d\mu < \infty.$$

**Theorem 9.3.8** Let $(\Omega, F, \mu)$ be a measure space and let $X$ be a separable Banach space. Let $\{f_n\} \subseteq L^1(\Omega; X)$ be uniformly integrable and bounded such that $f_n(\omega) \to f(\omega)$ for each $\omega \in \Omega$. Then $f \in L^1(\Omega; X)$ and

$$\lim_{n \to \infty} \int_{\Omega} ||f_n - f||_X \, d\mu = 0.$$

**Proof:** Let $\varepsilon > 0$ be given. Then by uniform integrability there exists $\delta > 0$ such that if $\mu(E) < \delta$ then

$$\int_E ||f_n|| \, d\mu < \varepsilon/3.$$

By Fatou’s lemma the same inequality holds for $f$. Also Fatou’s lemma shows $f \in L^1(\Omega; X)$, $f$ being measurable because of Theorem 9.3.4.
By Egoroff’s theorem, Theorem 9.3.6, there exists a set of measure less than \( \delta \), \( E \) such that the convergence of \( \{f_n\} \) to \( f \) is uniform off \( E \). Therefore,

\[
\int_{\Omega} \|f - f_n\| \, d\mu \leq \int_{E} (\|f\|_X + \|f_n\|_X) \, d\mu + \int_{E^c} \|f - f_n\|_X \, d\mu < \frac{2\varepsilon}{3} + \int_{E^c} \frac{\varepsilon}{3} \, d\mu < \varepsilon
\]

if \( n \) is large enough. This proves the theorem.

Note that a convenient way to achieve uniform integrability is to simple say \( \{f_n\} \) is bounded in \( L^p(\Omega; X) \) for some \( p > 1 \). This follows from Holder’s inequality.

\[
\int_{E} \|f_n\| \, d\mu \leq \left( \int_{E} d\mu \right)^{1/p'} \left( \int_{\Omega} \|f_n\|^p \, d\mu \right)^{1/p}
\]

### 9.4 Measurable Representatives

In this section consider the special case where \( X = L^1(\mathcal{B}, \nu) \) where \( (\mathcal{B}, \mathcal{F}, \nu) \) is a \( \sigma \) finite measure space and \( x \in L^1(\Omega; X) \). Thus for each \( s \in \Omega \), \( x(s) \in L^1(\mathcal{B}, \nu) \). In general, the map

\[
(s, t) \to x(s)(t)
\]

will not be product measurable, but one can obtain a measurable representative. This is important because it allows the use of Fubini’s theorem on the measurable representative.

By Theorem 9.3.24 there exists a sequence of simple functions, \( \{x_n\} \), of the form

\[
x_n(s) = \sum_{k=1}^{m} a_k \chi_{E_k}(s)
\]

(9.4.19)

where \( a_k \in L^1(\mathcal{B}, \nu) \) which satisfy the conditions of Definition 9.4.2 and

\[
\|x_n - x_m\|_{L^1(\Omega, L^1(\mathcal{B}))} \to 0 \text{ as } m, n \to \infty
\]

(9.4.20)

For such a simple function, you can assume the \( E_k \) are disjoint and then

\[
\|x_n\|_{L^1(\Omega, L^1(\mathcal{B}))} = \sum_{k=1}^{m} \|a_k\|_{L^1(\mathcal{B})} \mu(E_k) = \sum_{k=1}^{m} \int_{\mathcal{B}} |a_k| \, d\nu(E_k)
\]

\[
= \int_{\Omega} \int_{\mathcal{B}} |a_k(t)| \, d\nu(t) \chi_{E_k}(s) \, d\mu(s)
\]

\[
= \int_{\Omega} \int_{\mathcal{B}} |x_n| \, d\nu d\mu
\]

Also, each \( x_n \) is product measurable. Thus from 9.2.5,

\[
\|x_n - x_m\|_{L^1(\Omega, L^1(\mathcal{B}))} = \int_{\Omega} \int_{\mathcal{B}} |x_n - x_m| \, d\nu d\mu
\]

which shows that \( \{x_n\} \) is a Cauchy sequence in \( L^1(\Omega \times B, \mu \times \lambda) \). Then there exists \( y \in L^1(\Omega \times B, \mu \times \lambda) \) and a subsequence still called \( \{x_n\} \) such that

\[
\lim_{n \to \infty} \int_{\Omega} \int_{\mathcal{B}} |x_n - y| \, d\nu d\mu = \lim_{n \to \infty} \int_{\Omega} \|x_n - y\|_{L^1(\mathcal{B})} \, d\mu = \|x_n - y\|_{L^1(\Omega, L^1(\mathcal{B}))} = 0.
\]

Now consider 9.2.5. Since \( \lim_{m \to \infty} x_m(s) = x(s) \) in \( L^1(B) \), it follows from Fatou’s lemma that

\[
\|x_n - x\|_{L^1(\Omega, L^1(\mathcal{B}))} \leq \liminf_{m \to \infty} \|x_n - x_m\|_{L^1(\Omega, L^1(\mathcal{B}))} < \varepsilon
\]

for all \( n \) large enough. Hence

\[
\lim_{n \to \infty} \|x_n - x\|_{L^1(\Omega, L^1(\mathcal{B}))} = 0
\]
and so
\[ x(s) = y(s) \text{ in } L^1(B) \text{ } \mu \text{ a.e. } s \]
In particular, for a.e. \( s \), it follows that
\[ x(s)(t) = y(s,t) \text{ for a.e. } t. \]

Now \( \int x(s) \, d\mu \in X = L^1(B,\nu) \) so it makes sense to ask for \( (\int x(s) \, d\mu)(t) \), at least \( \mu \text{ a.e. } t \). To find what this is, note
\[
\left\| \int x_n(s) \, d\mu - \int x(s) \, d\mu \right\|_X \leq \int \|x_n(s) - x(s)\|_X \, d\mu.
\]
Therefore, since the right side converges to 0,
\[
\lim_{n \to \infty} \left\| \int x_n(s) \, d\mu - \int x(s) \, d\mu \right\|_X = 0.
\]
But
\[
\left( \int x_n(s) \, d\mu \right)(t) = \int x_n(s,t) \, d\mu \text{ a.e. } t.
\]
Therefore
\[
\lim_{n \to \infty} \int_B \left| \int x_n(s) \, d\mu - \left( \int x(s) \, d\mu \right) \right| \, d\nu = 0. \tag{9.4.21}
\]
Also, since \( x_n \to y \) in \( L^1(\Omega \times B) \),
\[
0 = \lim_{n \to \infty} \int_B \int_\Omega |x_n(s,t) - y(s,t)| \, d\mu d\nu \geq \lim_{n \to \infty} \int_B \int_\Omega |x_n(s,t) - y(s,t)| \, d\mu d\nu. \tag{9.4.22}
\]
From (9.4.21) and (9.4.22)
\[
\int y(s,t) \, d\mu = \left( \int x(s) \, d\mu \right)(t) \text{ a.e. } t.
\]
This proves the following theorem.

**Theorem 9.4.1** Let \( X = L^1(B) \) where \( (B,\mathcal{F},\nu) \) is a \( \sigma \)-finite measure space and let \( x \in L^1(\Omega;X) \). Then there exists a measurable representative, \( y \in L^1(\Omega \times B) \), such that
\[
x(s) = y(s,\cdot) \text{ a.e. } s \text{ in } \Omega, \text{ the equation in } L^1(B),
\]
and
\[
\int y(s,t) \, d\mu = \left( \int x(s) \, d\mu \right)(t) \text{ a.e. } t.
\]

### 9.5 Vector Measures

There is also a concept of vector measures.

**Definition 9.5.1** Let \( (\Omega,\mathcal{S}) \) be a set and a \( \sigma \) algebra of subsets of \( \Omega \). A mapping
\[
F : \mathcal{S} \to X
\]
is said to be a vector measure if
\[
F(\cup_{i=1}^\infty E_i) = \sum_{i=1}^\infty F(E_i)
\]
whenever \( \{E_i\}_{i=1}^\infty \) is a sequence of disjoint elements of \( \mathcal{S} \). For \( F \) a vector measure,
\[
|F|(A) \equiv \sup \{ \sum_{F \in \pi(A)} ||\mu(F)|| : \pi(A) \text{ is a partition of } A \}. 
\]
This is the same definition that was given in the case where $F$ would have values in $\mathbb{C}$, the only difference being the fact that now $F$ has values in a general Banach space $X$ as the vector space of values of the vector measure. Recall that a partition of $A$ is a finite set, $\{F_1, \cdots, F_m\} \subseteq \mathcal{S}$ such that $\cup_{i=1}^{m} F_i = A$. The same theorem about $|F|$ proved in the case of complex valued measures holds in this context with the same proof. For completeness, it is included here.

**Theorem 9.5.2** If $|F|(\Omega) < \infty$, then $|F|$ is a measure on $\mathcal{S}$.

**Proof:** Let $E_1$ and $E_2$ be sets of $\mathcal{S}$ such that $E_1 \cap E_2 = \emptyset$ and let $\{A_1 \cdots A_n\} = \pi(E_i)$, a partition of $E_i$ which is chosen such that

$$|F|(E_i) - \varepsilon < \sum_{j=1}^{n_i} \|F(A_j^i)\| \quad i = 1, 2.$$  

Consider the sets which are contained in either of $\pi(E_1)$ or $\pi(E_2)$, it follows this collection of sets is a partition of $E_1 \cup E_2$ which is denoted here by $\pi(E_1 \cup E_2)$. Then by the definition of total variation,

$$|F|(E_1 \cup E_2) \geq \sum_{F \in \pi(E_1 \cup E_2)} \|F(F)\| > |F|(E_1) + |F|(E_2) - 2\varepsilon,$$  

which shows that since $\varepsilon > 0$ was arbitrary,

$$|F|(E_1 \cup E_2) \geq |F|(E_1) + |F|(E_2). \quad (9.5.23)$$

Let $\{E_j\}_{j=1}^{\infty}$ be a sequence of disjoint sets of $\mathcal{S}$ and let $E_{\infty} = \cup_{j=1}^{\infty} E_j$. Then by the definition of total variation there exists a partition of $E_{\infty}$, $\pi(E_{\infty}) = \{A_1, \cdots, A_n\}$ such that

$$|F|(E_{\infty}) - \varepsilon < \sum_{i=1}^{n} \|F(A_i)\|.$$  

Also,

$$A_i = \cup_{j=1}^{\infty} A_i \cap E_j$$  

and so by the triangle inequality, $\|F(A_i)\| \leq \sum_{j=1}^{\infty} \|F(A_i \cap E_j)\|$. Therefore, by the above,

$$|F|(E_{\infty}) - \varepsilon < \sum_{i=1}^{n} \sum_{j=1}^{\infty} \|F(A_i \cap E_j)\|$$

$$= \sum_{j=1}^{\infty} \sum_{i=1}^{n} \|F(A_i \cap E_j)\|$$

$$\leq \sum_{j=1}^{\infty} |F|(E_j)$$

because $\{A_i \cap E_j\}_{i=1}^{n}$ is a partition of $E_j$.

Since $\varepsilon > 0$ is arbitrary, this shows

$$|F|(\cup_{j=1}^{\infty} E_j) \leq \sum_{j=1}^{\infty} |F|(E_j).$$

Also, implies that whenever the $E_i$ are disjoint, $|F|(\cup_{j=1}^{n} E_j) \geq \sum_{j=1}^{n} |F|(E_j)$. Therefore,

$$\sum_{j=1}^{\infty} |F|(E_j) \geq |F|(\cup_{j=1}^{\infty} E_j) \geq |F|(\cup_{j=1}^{n} E_j) \geq \sum_{j=1}^{n} |F|(E_j).$$

Since $n$ is arbitrary,

$$|F|(\cup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} |F|(E_j)$$

which shows that $|F|$ is a measure as claimed. This proves the theorem.
**Definition 9.5.3** A Banach space is said to have the Radon Nikodym property if whenever 
\[(\Omega, \mathcal{S}, \mu)\] is a finite measure space 
\[F : \mathcal{S} \to X\] is a vector measure with \(|F| (\Omega) < \infty\) 
\[F \ll \mu\] 
then one may conclude there exists \(g \in L^1 (\Omega; X)\) such that 
\[F (E) = \int_E g (s) \, d\mu\] for all \(E \in \mathcal{S}\). 

Some Banach spaces have the Radon Nikodym property and some don’t. No attempt is made to give a complete answer to the question of which Banach spaces have this property but the next theorem gives examples of many spaces which do.

**Theorem 9.5.4** Suppose \(X'\) is a separable dual space. Then \(X'\) has the Radon Nikodym property.

**Proof:** Let \(F \ll \mu\) and let \(|F| (\Omega) < \infty\) for \(F : \mathcal{S} \to X'\), a vector measure. Pick \(x \in X\) and consider the map 
\[E \to F (E) (x)\] for \(E \in \mathcal{S}\). This defines a complex measure which is absolutely continuous with respect to \(|F|\). Therefore, by the Radon Nikodym theorem, there exists \(f_x \in L^1 (\Omega, |F|)\) such that 
\[F (E) (x) = \int_E f_x (s) \, d|F|, \quad (9.5.24)\]

**Claim:** \(|f_x (s)| \leq ||x||\) for \(|F|\) a.e. \(s\).

**Proof of claim:** Consider the closed ball in \(F, B (0, ||x||)\) and let \(B \equiv B (p, r)\) be an open ball contained in its complement. Let \(f_x^{-1} (B) \equiv E \in \mathcal{S}\). I want to argue that \(|F| (E) = 0\) so suppose \(|F| (E) > 0\). then 
\[|F| (E) ||x|| \geq ||F (E)|| ||x|| \geq |F (E) (x)|\] and so from [Eq. 24], 
\[\frac{1}{|F| (E)} \int_E f_x (s) \, d|F| \leq ||x||. \quad (9.5.25)\]

But on \(E, |f_x (s) - p| < r\) and so 
\[\left| \frac{1}{|F| (E)} \int_E f_x (s) \, d|F| - p \right| < r\] which contradicts [Eq. 24] because \(B (p, r)\) was given to have empty intersection with \(B (0, ||x||)\). Therefore, \(|F| (E) = 0\) as hoped. Now \(F \setminus B (0, ||x||)\) can be covered by countably many such balls and so \(|F| (F \setminus B (0, ||x||)) = 0\).

Denote the exceptional set of measure zero by \(N_x\). By Theorem [Eq. 24], \(X\) is separable. Letting \(D\) be a dense, countable subset of \(X\), define 
\[N_D \equiv \cup_{x \in D} N_x.\] Thus 
\[|F| (N_1) = 0.\]

For any \(E \in \mathcal{S}, x, y \in D,\) and \(a, b \in F,\) 
\[\int_E f_{ax + by} (s) \, d|F| = F (E) (ax + by) = aF (E) (x) + bF (E) (y) = \int_E (af_x (s) + bf_y (s)) \, d|F|. \quad (9.5.26)\]
Since [Eq. 26] holds for all \(E \in \mathcal{S}\), it follows 
\[f_{ax + by} (s) = af_x (s) + bf_y (s)\]
for $|F|$ a.e. $s$ and $x, y \in D$. Let $\tilde{D}$ consist of all finite linear combinations of the form $\sum_{i=1}^{m} a_i x_i$ where $a_i$ is a rational point of $F$ and $x_i \in D$. If
\[ \sum_{i=1}^{m} a_i x_i \in \tilde{D}, \]
the above argument implies
\[ f_{\sum_{i=1}^{m} a_i x_i} (s) = \sum_{i=1}^{m} a_i f_{x_i} (s) \text{ a.e.} \]

Since $\tilde{D}$ is countable, there exists a set, $N_2$, with
\[ |F| (N_2) = 0 \]
such that for $s \notin N_2$,
\[ f_{\sum_{i=1}^{m} a_i x_i} (s) = \sum_{i=1}^{m} a_i f_{x_i} (s) \quad (9.5.27) \]
whenever $\sum_{i=1}^{m} a_i x_i \in \tilde{D}$. Let
\[ N = N_1 \cup N_2 \]
and let
\[ \tilde{h}_x (s) \equiv \chi_{N^c} (s) f_x (s) \]
for all $x \in \tilde{D}$. Now for $x \in X$ define
\[ h_x (s) \equiv \lim_{x' \to x} \{ \tilde{h}_{x'} (s) : x' \in \tilde{D} \}. \]
This is well defined because if $x'$ and $y'$ are elements of $\tilde{D}$, the above claim and (9.5.29) imply
\[ |\tilde{h}_{x'} (s) - \tilde{h}_{y'} (s)| = |\tilde{h}_{(x'-y')} (s)| \leq ||x' - y'||. \]
Using (9.5.27), the dominated convergence theorem may be applied to conclude that for $x_n \to x$, with $x_n \in \tilde{D}$,
\[ \int_{E} h_x (s) d |F| = \lim_{n \to \infty} \int_{E} \tilde{h}_{x_n} (s) d |F| = \lim_{n \to \infty} F (E) (x_n) = F (E) (x). \quad (9.5.28) \]
It follows from the density of $\tilde{D}$ that for all $x, y \in X$ and $a, b \in F$,
\[ |h_x (s)| \leq ||x||, \quad h_{ax + by} (s) = ah_x (s) + bh_y (s), \quad (9.5.29) \]
for all $s$ because if $s \in N$, both sides of the equation in (9.5.29) equal 0.

Let $\theta (s)$ be given by
\[ \theta (s) (x) = h_x (s). \]
By (9.5.29) it follows that $\theta (s) \in X'$ for each $s$. Also
\[ \theta (s) (x) = h_x (s) \in L^1 (\Omega) \]
so $\theta (\cdot)$ is weak * measurable. Since $X'$ is separable, Theorem 4.12 implies that $\theta$ is strongly measurable. Furthermore, by (9.5.29),
\[ ||\theta (s)|| = \sup_{||x|| \leq 1} |\theta (s) (x)| \leq \sup_{||x|| \leq 1} |h_x (s)| \leq 1. \]
Therefore,
\[ \int_{\Omega} ||\theta (s)|| d |F| < \infty \]
so $\theta \in L^1 (\Omega; X')$. By (9.5.29), if $E \in \mathcal{S}$,
\[ \int_{E} h_x (s) d |F| = \int_{E} \theta (s) (x) d |F| = \left( \int_{E} \theta (s) d |F| \right) (x). \quad (9.5.30) \]
From (9.5.28) and (9.5.30),
\[ \left( \int_{E} \theta (s) d |F| \right) (x) = F (E) (x) \]
for all \( x \in X \) and therefore,
\[
\int_E \theta(s) \, d|F| = F(E).
\]
Finally, since \( F \ll \mu, |F| \ll \mu \) also and so there exists \( k \in L^1(\Omega) \) such that
\[
|F|(E) = \int_E k(s) \, d\mu
\]
for all \( E \in \mathcal{S} \), by the Radon Nikodym Theorem. It follows
\[
F(E) = \int_E \theta(s) \, d|F| = \int_E \theta(s) k(s) \, d\mu.
\]
Letting \( g(s) = \theta(s) k(s) \), this has proved the theorem.

**Corollary 9.5.5** Any separable reflexive Banach space has the Radon Nikodym property.

It is not necessary to assume separability in the above corollary. For the proof of a more general result, consult *Vector Measures* by Diestal and Uhl. [30].

### 9.6 The Riesz Representation Theorem

The Riesz representation theorem for the spaces \( L^p(\Omega; X) \) holds under certain conditions. The proof follows the proofs given earlier for scalar valued functions.

**Definition 9.6.1** If \( X \) and \( Y \) are two Banach spaces, \( X \) is isometric to \( Y \) if there exists \( \theta \in \mathcal{L}(X, Y) \) such that
\[
\|\theta x\|_Y = \|x\|_X.
\]
This will be written as \( X \cong Y \). The map \( \theta \) is called an isometry.

The next theorem says that \( L^{p'}(\Omega; X') \) is always isometric to a subspace of \( (L^p(\Omega; X))' \) for any Banach space, \( X \).

**Theorem 9.6.2** Let \( X \) be any Banach space and let \((\Omega, \mathcal{S}, \mu)\) be a finite measure space. Let \( p \geq 1 \) and let \( 1/p + 1/p' = 1 \). (If \( p = 1, p' \equiv \infty \)) Then \( L^{p'}(\Omega; X') \) is isometric to a subspace of \( (L^p(\Omega; X))' \). Also, for \( g \in L^{p'}(\Omega; X') \),
\[
\sup_{\|f\|_p \leq 1} \left| \int\Omega \, g(s) (f(s)) \, d\mu \right| = \|g\|_{p'}.
\]

**Proof:** First observe that for \( f \in L^p(\Omega; X) \) and \( g \in L^{p'}(\Omega; X') \),
\[
s \rightarrow g(s)(f(s))
\]
is a function in \( L^1(\Omega) \). (To obtain measurability, write \( f \) as a limit of simple functions. Holder’s inequality then yields the function is in \( L^1(\Omega) \).) Define
\[
\theta : L^{p'}(\Omega; X') \rightarrow (L^p(\Omega; X))'
\]
by
\[
\theta g(f) = \int\Omega \, g(s)(f(s)) \, d\mu.
\]
Holder’s inequality implies
\[
\|\theta g\| \leq \|g\|_{p'} \quad \text{(9.6.31)}
\]
and it is also clear that \( \theta \) is linear. Next it is required to show
\[
\|\theta g\| = \|g\|.
\]
9.6. THE RIESZ REPRESENTATION THEOREM

This will first be verified for simple functions. Let

\[ g(s) = \sum_{i=1}^{m} c_i \chi_{E_i}(s) \]

where \( c_i \in X' \), the \( E_i \) are disjoint and

\[ \bigcup_{i=1}^{m} E_i = \Omega. \]

Then \( ||g|| \in L^{p'}(\Omega) \). Let \( \varepsilon > 0 \) be given. By the scalar Riesz representation theorem, there exists \( h \in L^{p}(\Omega) \) such that \( ||h||_{p} = 1 \) and

\[ \int_{\Omega} ||g(s)||_{X'} h(s) \, d\mu \geq ||g||_{L^{p'}(\Omega; X')} - \varepsilon. \]

Now let \( d_i \) be chosen such that

\[ c_i(d_i) \geq ||c_i||_{X'} - \varepsilon / ||h||_{L^{1}(\Omega)} \]

and \( ||d_i||_{X} \leq 1 \). Let

\[ f(s) = \sum_{i=1}^{m} d_i h(s) \chi_{E_i}(s). \]

Thus \( f \in L^{p}(\Omega; X) \) and \( ||f||_{L^{p}(\Omega; X)} \leq 1 \). This follows from

\[ ||f||_{p}^{p} = \int_{\Omega} \sum_{i=1}^{m} ||d_i||_{X}^{p} ||h(s)||_{X}^{p} \chi_{E_i}(s) \, d\mu \]

\[ = \sum_{i=1}^{m} \left( \int_{E_i} ||h(s)||_{X}^{p} \, d\mu \right) ||d_i||_{X}^{p} \leq \int_{\Omega} ||h||_{X}^{p} \, d\mu = 1. \]

Also

\[ ||\theta g|| \geq ||\theta g(f)|| = \left| \int_{\Omega} g(s)(f(s)) \, d\mu \right| \geq \left| \int_{\Omega} \sum_{i=1}^{m} \left( ||c_i||_{X'} - \varepsilon / ||h||_{L^{1}(\Omega)} \right) h(s) \chi_{E_i}(s) \, d\mu \right| \]

\[ \geq \left| \int_{\Omega} ||g(s)||_{X'} h(s) \, d\mu \right| - \varepsilon \int_{\Omega} h(s) / ||h||_{L^{1}(\Omega)} \, d\mu \]

\[ \geq ||g||_{L^{p'}(\Omega; X')} - 2\varepsilon. \]

Since \( \varepsilon \) was arbitrary,

\[ ||\theta g|| \geq ||g|| \] (9.6.32)

and from (9.6.31) this shows equality holds in (9.6.32) whenever \( g \) is a simple function.

In general, let \( g \in L^{p'}(\Omega; X') \) and let \( g_n \) be a sequence of simple functions converging to \( g \) in \( L^{p'}(\Omega; X') \). Then

\[ ||\theta g|| = \lim_{n \to \infty} ||\theta g_n|| = \lim_{n \to \infty} ||g_n|| = ||g||. \]

This proves the theorem and shows \( \theta \) is the desired isometry.

**Theorem 9.6.3** If \( X \) is a Banach space and \( X' \) has the Radon Nikodym property, then if \((\Omega, \mathcal{S}, \mu)\) is a finite measure space,

\[ (L^{p}(\Omega; X))' \cong L^{p'}(\Omega; X') \]

and in fact the mapping \( \theta \) of Theorem 9.6.2 is onto.

**Proof:** Let \( l \in (L^{p}(\Omega; X))' \) and define \( F(E) \in X' \) by

\[ F(E)(x) = l(\chi_{E}(\cdot) x). \]

**Lemma 9.6.4** \( F \) defined above is a vector measure with values in \( X' \) and \( |F|(\Omega) < \infty \).
**Proof of the lemma:** Clearly $F(E)$ is linear. Also

$$
||F(E)|| = \sup_{||x|| \leq 1} ||F(E)(x)||
$$

$$
\leq ||l|| \sup_{||x|| \leq 1} ||\chi_E(\cdot)x||_{L^p(\Omega;X)} \leq ||l|| \mu(E)^{1/p}.
$$

Let $\{E_i\}_{i=1}^{\infty}$ be a sequence of disjoint elements of $S$ and let $E = \cup_{n<\infty}E_n$.

$$
F(E)(x) - \sum_{k=1}^{n} F(E_k)(x) = l(\chi_E(\cdot)x) - \sum_{i=1}^{n} l(\chi_{E_i}(\cdot)x)
$$

Thus

$$
\leq ||l|| \sum_{i=1}^{n} \chi_{E_i}(\cdot)x \leq ||l|| \mu(\Omega)^{1/p} X(E(\cdot)x).
$$

Since $\mu(\Omega) < \infty$,

$$
\lim_{n \to \infty} \mu(\bigcup_{k>n} E_k)^{1/p} = 0
$$

and so inequality (9.6.33) shows that

$$
\lim_{n \to \infty} \left| F(E) - \sum_{k=1}^{n} F(E_k) \right|_{X'} = 0.
$$

To show $|F| (\Omega) < \infty$, let $\varepsilon > 0$ be given, let $\{H_1, \cdots, H_n\}$ be a partition of $\Omega$, and let $||x_i|| \leq 1$ be chosen in such a way that

$$
F(H_i)(x_i) > ||F(H_i)|| - \varepsilon/n.
$$

Thus

$$
-\varepsilon + \sum_{i=1}^{n} ||F(H_i)|| \leq \sum_{i=1}^{n} l(\chi_{H_i}(\cdot)x_i) \leq ||l|| \sum_{i=1}^{n} \chi_{H_i}(\cdot)x_i
$$

$$
\leq ||l|| \left( \int_{\Omega} \sum_{i=1}^{n} \chi_{H_i}(s) d\mu \right)^{1/p} = ||l|| \mu(\Omega)^{1/p}.
$$

Since $\varepsilon > 0$ was arbitrary,

$$
\sum_{i=1}^{n} ||F(H_i)|| < ||l|| \mu(\Omega)^{1/p}.
$$

Since the partition was arbitrary, this shows $|F| (\Omega) \leq ||l|| \mu(\Omega)^{1/p}$ and this proves the lemma.

Continuing with the proof of Theorem (9.6.34), note that

$$
F \ll \mu.
$$

Since $X'$ has the Radon Nikodym property, there exists $g \in L^1 (\Omega; X')$ such that

$$
F(E) = \int_{E} g(s) d\mu.
$$

Also, from the definition of $F(E)$,

$$
l \left( \sum_{i=1}^{n} x_i \chi_{E_i}(\cdot) \right) = \sum_{i=1}^{n} l(\chi_{E_i}(\cdot)x_i)
$$

$$
= \sum_{i=1}^{n} F(E_i)(x_i) = \sum_{i=1}^{n} \int_{E_i} g(s)(x_i) d\mu.
$$

(9.6.34)
It follows from (9.6.34) that whenever \( h \) is a simple function,
\[
 l(h) = \int_{\Omega} g(s) (h(s)) \, d\mu. \tag{9.6.35}
\]

Let
\[
 G_n = \{ s : \|g(s)\|_{X'} \leq n \}
\]
and let
\[
 j : L^p(G_n; X) \to L^p(\Omega; X)
\]
be given by
\[
 jh(s) = \begin{cases} 
 h(s) & \text{if } s \in G_n, \\
 0 & \text{if } s \notin G_n.
\end{cases}
\]

Letting \( h \) be a simple function in \( L^p(G_n; X) \),
\[
 j^*l(h) = l(jh) = \int_{G_n} g(s) (h(s)) \, d\mu. \tag{9.6.36}
\]

Since the simple functions are dense in \( L^p(G_n; X) \), and \( g \in L^{p'}(G_n; X') \), it follows (9.6.36) holds for all \( h \in L^p(G_n; X) \).

By Theorem 9.6.2,
\[
 \|g\|_{L^{p'}(G_n; X')} = \|j^*l\|_{L^p(G_n; X)} \leq \|l\|_{L^p(\Omega; X)}'.
\]

By the monotone convergence theorem,
\[
 \|g\|_{L^{p'}(\Omega; X')} = \lim_{n \to \infty} \|g\|_{L^{p'}(G_n; X')} \leq \|l\|_{L^p(\Omega; X)}'.
\]

Therefore \( g \in L^{p'}(\Omega; X') \) and since simple functions are dense in \( L^p(\Omega; X) \), (9.6.36) holds for all \( h \in L^p(\Omega; X) \). Thus \( l = \theta g \) and the theorem is proved because, by Theorem 9.6.2, \( \|l\| = \|g\| \) and the mapping \( \theta \) is onto because \( l \) was arbitrary.

As in the scalar case, everything generalizes to the case of \( \sigma \) finite measure spaces. The proof is almost identical.

**Lemma 9.6.5** Let \((\Omega, \mathcal{S}, \mu)\) be a \( \sigma \) finite measure space and let \( X \) be a Banach space such that \( X' \) has the Radon Nikodym property. Then there exists a measurable function, \( r \) such that \( r(x) > 0 \) for all \( x \), such that \( |r(x)| < M \) for all \( x \), and \( \int r \, d\mu < \infty \). For
\[
 \Lambda \in (L^p(\Omega; X))', \ p \geq 1,
\]
there exists a unique \( h \in L^{p'}(\Omega; X'), \ L^{\infty}(\Omega; X') \) if \( p = 1 \) such that
\[
 \Lambda f = \int h(f) \, d\mu.
\]

Also \( \|h\| = |\Lambda|, (\|h\| = \|h\|_{p'} \text{ if } p > 1, \|h\|_{\infty} \text{ if } p = 1). \) Here
\[
 \frac{1}{p} + \frac{1}{p'} = 1.
\]

**Proof:** First suppose \( r \) exists as described. Also, to save on notation and to emphasize the similarity with the scalar case, denote the norm in the various spaces by \(|.|\). Define a new measure \( \tilde{\mu} \), according to the rule
\[
 \tilde{\mu}(E) = \int_E r \, d\mu. \tag{9.6.37}
\]

Thus \( \tilde{\mu} \) is a finite measure on \( \mathcal{S} \). Now define a mapping, \( \eta : L^p(\Omega; X, \mu) \to L^p(\Omega; X, \tilde{\mu}) \) by
\[
 \eta f = r^{-\frac{1}{p}} f.
\]

Then
\[
 \|\eta f\|_{L^p(\tilde{\mu})}^p = \int |r^{-\frac{1}{p}} f|^p \, r \, d\mu = \|f\|_{L^p(\mu)}^p.
\]

Thus \( \eta f \) is defined on \( \mathcal{S} \).
and so \( \eta \) is one to one and in fact preserves norms. I claim that also \( \eta \) is onto. To see this, let \( g \in L^p(\Omega; \bar{\mu}) \) and consider the function, \( r^{\frac{1}{p}} g \). Then
\[
\int |r^{\frac{1}{p}} g|^p d\mu = \int |g|^p r d\mu = \int |g|^p d\bar{\mu} < \infty
\]
Thus \( r^{\frac{1}{p}} g \in L^p(\Omega; X, \mu) \) and \( \eta \left( r^{\frac{1}{p}} g \right) = g \) showing that \( \eta \) is onto as claimed. Thus \( \eta \) is one to one, onto, and preserves norms. Consider the diagram below which is descriptive of the situation in which \( \eta^* \) must be one to one and onto.

\[
\begin{array}{ccc}
h, L^p(\bar{\mu}) & \xrightarrow{\eta^*} & \bar{\Lambda} \\
L^p(\bar{\mu}) & \xleftarrow{\eta} & L^p(\mu)
\end{array}
\]

Then for \( \Lambda \in L^p(\mu)' \), there exists a unique \( \bar{\Lambda} \in L^p(\bar{\mu})' \) such that \( \eta^* \bar{\Lambda} = \Lambda, \left| |\bar{\Lambda}| \right| = ||\Lambda|| \). By the Riesz representation theorem for finite measure spaces, there exists a unique \( h \in L^p(\bar{\mu}) \equiv L^{p'}(\Omega; X', \bar{\mu}) \) which represents \( \bar{\Lambda} \) in the manner described in the Riesz representation theorem. Thus \( ||h||_{L^{p'}(\bar{\mu})} = \left| |\bar{\Lambda}| \right| = ||\Lambda|| \) and for all \( f \in L^p(\mu) \),
\[
\Lambda (f) = \eta^* \bar{\Lambda} (f) = \bar{\Lambda}(\eta f) = \int h(\eta f) d\bar{\mu} = \int r h \left( r^{\frac{1}{p}} f \right) d\mu
\]
Now
\[
\int |r^{\frac{1}{p}} h|^p d\mu = \int |h|^p r d\mu = ||h||_{L^p(\mu)} < \infty.
\]
Thus \( \left| |r^{\frac{1}{p}} h| \right|_{L^p(\mu)} = ||h||_{L^{p'}(\bar{\mu})} = \left| |\bar{\Lambda}| \right| = ||\Lambda|| \) and represents \( \Lambda \) in the appropriate way. If \( p = 1 \), then \( 1/p' = 0 \).

Now consider the existence of \( r \). Since the measure space is \( \sigma \) finite, there exist \( \{\Omega_n\} \) disjoint, each having positive measure and their union equals \( \Omega \). Then define
\[
r (\omega) \equiv \sum_{n=1}^{\infty} \frac{1}{n^2} \mu(\Omega_n)^{-1} \chi_{\Omega_n} (\omega)
\]
This proves the Lemma.

**Theorem 9.6.6** *(Riesz representation theorem)* Let \( (\Omega, \mathcal{S}, \mu) \) be \( \sigma \) finite and let \( X' \) have the Radon Nikodym property. Then for
\[
\Lambda \in (L^p(\Omega; X, \mu))', \ p \geq 1
\]
there exists a unique \( h \in L^q(\Omega, X', \mu), \ L^\infty(\Omega, X', \mu) \) if \( p = 1 \) such that
\[
\Lambda f = \int h(f) d\mu.
\]
Also \( ||h|| = ||\Lambda||, \ (||h|| = ||h||_q \text{ if } p > 1, \ ||h||_\infty \text{ if } p = 1) \). Here
\[
\frac{1}{p} + \frac{1}{q} = 1.
\]

**Proof:** The above lemma gives the existence part of the conclusion of the theorem. Uniqueness is done as before.

**Corollary 9.6.7** If \( X' \) is separable, then for \( (\Omega, \mathcal{S}, \mu) \) a \( \sigma \) finite measure space,
\[
(L^p(\Omega; X))' \cong L^{p'}(\Omega; X').
\]

**Corollary 9.6.8** If \( X \) is separable and reflexive, then for \( (\Omega, \mathcal{S}, \mu) \) a \( \sigma \) finite measure space,
\[
(L^p(\Omega; X))' \cong L^{p'}(\Omega; X').
\]

**Corollary 9.6.9** If \( X \) is separable and reflexive and \( (\Omega, \mathcal{S}, \mu) \) a \( \sigma \) finite measure space, then if \( p \in (1, \infty) \), then \( L^p(\Omega; X) \) is reflexive.

**Proof:** This is just like the scalar valued case.
9.7 Pointwise Behavior Of Weakly Convergent Sequences

There is an interesting little result which relates to weak limits in $L^2(\Gamma, E)$ for $E$ a Banach space. I am not sure where to put this thing but think that this would be a good place for it. It obviously generalizes to $L^p$ spaces.

**Proposition 9.7.1** Let $E$ be a Banach space and let $\{u_n\}$ be a sequence in $L^2(\Gamma, E)$ and let $G(x)$ be a weakly compact set in $E$, and $u_n(x) \in G(x)$ a.e. for each $n$. Let $\limsup \{u_n(x)\}$ denote the set of all weak limits of subsequences of $\{u_n(x)\}$ and let $H(x)$ be the closure of the convex hull of $\limsup \{u_n(x)\}$. Then if $u_n \to u$ weakly in $L^2(\Gamma, E)$, then $u(x) \in H(x)$ for a.e. $x$.

**Proof:** Let $H = \{w \in L^2(\Gamma, E) : w(x) \in H(x) \text{ a.e.}\}$. Then $H$ is convex. If you have $w_1 \in H$, then since each $H(x)$ is convex, it follows that $\lambda w_1(x) + (1 - \lambda) w_2(x) \in H$ for a.e. $x$ and $\lambda \in [0, 1]$. Is $H$ closed? Suppose you have $w_n \in H$ and $w_n \to w$ in $L^2(\Gamma, E)$. Then there is a subsequence such that pointwise convergence happens a.e. and so since $H$ is closed, you have $w(x) \in H$ for a.e. $x$. Hence $H$ is also weakly closed in $L^2(\Gamma, H)$. Thus if $u$ is the weak limit of $\{u_n\}$ in $L^2(\Gamma, E)$, it must be the case that $u(x) \in H(x)$ a.e. ■

As a case of this which might be pretty interesting, suppose $G(x)$ is not just weakly compact but also convex. Then $H(x) = G(x)$ and you can say that $u(x) \in H(x)$ a.e. whenever it is a weak limit in $L^2(\Gamma, E)$ of functions $u_n$ for which $u_n(x) \in G(x)$. 
Chapter 10

Basic Probability

Caution: This material on probability and stochastic processes may be half baked in places. I have not yet rewritten it several times. This is not to say that nothing else is half baked. However, the probability is higher here.

10.1 Random Variables And Independence

Recall Lemma 2.3.3 on Page 33 which is stated here for convenience.

Lemma 10.1.1 Let $M$ be a metric space with the closed balls compact and suppose $\lambda$ is a measure defined on the Borel sets of $M$ which is finite on compact sets. Then there exists a unique Radon measure, $\overline{\lambda}$ which equals $\lambda$ on the Borel sets. In particular $\lambda$ must be both inner and outer regular on all Borel sets.

Also important is the following fundamental result which is called the Borel Cantelli lemma.

Lemma 10.1.2 Let $(\Omega,\mathcal{F},\lambda)$ be a measure space and let $\{A_i\}$ be a sequence of measurable sets satisfying

$$\sum_{i=1}^{\infty} \lambda(A_i) < \infty.$$ 

Then letting $S$ denote the set of $\omega \in \Omega$ which are in infinitely many $A_i$, it follows $S$ is a measurable set and $\lambda(S) = 0$.

Proof: $S = \cap_{k=1}^{\infty} \cup_{m=k}^{\infty} A_m$. Therefore, $S$ is measurable and also

$$\lambda(S) \leq \lambda(\cup_{m=k}^{\infty} A_m) \leq \sum_{m=k}^{\infty} \lambda(A_k)$$

and this converges to 0 as $k \to \infty$ because of the convergence of the series.

Here is another nice observation.

Proposition 10.1.3 Suppose $E_i$ is a separable Banach space. Then if $B_i$ is a Borel set of $E_i$, it follows $\prod_{i=1}^{n} B_i$ is a Borel set in $\prod_{i=1}^{n} E_i$.

Proof: An easy way to do this is to consider the projection maps.

$$\pi_i x \equiv x_i$$

Then these projection maps are continuous. Hence for $U$ an open set,

$$\pi_i^{-1}(U) \equiv \prod_{j=1}^{n} A_j, \quad A_j = E_j \text{ if } j \neq i \text{ and } A_i = U.$$ 

Thus $\pi_i^{-1}(open)$ equals an open set. Let

$$S \equiv \{ V \subseteq \mathbb{R} : \pi_i^{-1}(V) \text{ is Borel} \}$$
Then $S$ contains all the open sets and is clearly a $\sigma$ algebra. Therefore, $S$ contains the Borel sets. Let $B_i$ be a Borel set in $E_i$. Then

$$\prod_{i=1}^n B_i = \cap_{i=1}^n \pi_i^{-1}(B_i),$$

a finite intersection of Borel sets. □

**Definition 10.1.4** A probability space is a measure space, $(\Omega, F, \mathbb{P})$ where $\mathbb{P}$ is a measure satisfying $\mathbb{P}(\Omega) = 1$. A random vector (variable) is a measurable function, $X : \Omega \to Z$ where $Z$ is some topological space. It is often the case that $Z$ will equal $\mathbb{R}^p$. Assume $Z$ is a separable Banach space. Define the following $\sigma$ algebra.

$$\sigma(X) \equiv \{X^{-1}(E): E \text{ is Borel in } Z\}$$

Thus $\sigma(X) \subseteq F$. For $E$ a Borel set in $Z$ define

$$\lambda_X(E) \equiv \mathbb{P}(X^{-1}(E)).$$

This is called the distribution of the random variable, $X$. If

$$\int_\Omega |X(\omega)| \, d\mathbb{P} < \infty$$

then define

$$E(X) = \int_\Omega X \, d\mathbb{P}$$

where the integral is defined as the Bochner integral.

Recall the following fundamental result which was proved earlier but which I will give a short proof of now.

**Proposition 10.1.5** Let $(\Omega, S, \mu)$ be a measure space and let $X : \Omega \to Z$ where $Z$ is a separable Banach space. Then $X$ is strongly measurable if and only if $X^{-1}(U) \in S$ for all $U$ open in $Z$.

**Proof:** To begin with, let $D(a, r)$ be the closure of the open ball $B(a, r)$. By Lemma 10.1.4, there exists $\{f_i\} \subseteq B'$, the unit ball in $Z'$ such that

$$\|z\|_Z = \sup_i \{|f_i(z)|\}$$

Then

$$D(a, r) = \{z: \|a - z\| \leq r\} = \bigcap_i \{z: |f_i(z) - f_i(a)| \leq r\}$$

$$= \bigcap_i f_i^{-1}(B(f_i(a), r))$$

It follows that

$$X^{-1}(D(a, r)) = \bigcap_i X^{-1}(f_i^{-1}(B(f_i(a), r)))$$

$$= \bigcap_i (f_i \circ X)^{-1}(B(f_i(a), r))$$

If $X$ is strongly measurable, then it is weakly measurable and so each $f_i \circ X$ is a real (complex) valued measurable function. Hence the expression on the right in the above is measurable. Now if $U$ is any open set in $Z$, then it is the countable union of such closed disks $U = \cup_i D_i$. Therefore, $X^{-1}(U) = \bigcap_i X^{-1}(D_i) \in S$. It follows that strongly measurable implies inverse images of open sets are in $S$.

Conversely, suppose $X^{-1}(U) \in S$ for every open $U$. Then for $f \in Z'$, $f \circ X$ is real valued and measurable. Therefore, $X$ is weakly measurable. By the Pettis theorem, it follows that $f \circ X$ is strongly measurable. □

**Proposition 10.1.6** If $X : \Omega \to Z$ is measurable, then $\sigma(X)$ equals the smallest $\sigma$ algebra such that $X$ is measurable with respect to it. Also if $X_i$ are random variables having values in separable Banach spaces $Z_i$, then $\sigma(X) = \sigma(X_1, \ldots, X_n)$ where $X$ is the vector mapping $\Omega$ to $\prod_{i=1}^n Z_i$ and $\sigma(X_1, \ldots, X_n)$ is the smallest $\sigma$ algebra such that each $X_i$ is measurable with respect to it.
10.1. RANDOM VARIABLES AND INDEPENDENCE

Proof: Let $\mathcal{G}$ denote the smallest $\sigma$ algebra such that $X$ is measurable with respect to this $\sigma$ algebra. By definition $X^{-1}(\text{open}) \in \mathcal{G}$. Furthermore, the set of all $E$ such that $X^{-1}(E) \in \mathcal{G}$ is a $\sigma$ algebra. Hence it includes all the Borel sets. Hence $X^{-1}(\text{Borel}) \in \mathcal{G}$ and so $\mathcal{G} \supseteq \sigma(X)$. However, $\sigma(X)$ defined above is a $\sigma$ algebra such that $X$ is measurable with respect to $\sigma(X)$. Therefore, $\mathcal{G} = \sigma(X)$.

Letting $B_i$ be a Borel set in $Z$, $\prod_{i=1}^n B_i$ is a Borel set by Proposition 10.1.6 and so

$$X^{-1}(\prod_{i=1}^n B_i) = \cap_{i=1}^n X_i^{-1}(B_i) \in \sigma(X_1, \ldots, X_n)$$

If $\mathcal{G}$ denotes the Borel sets $F \subseteq \prod_{i=1}^n Z_i$ such that $X^{-1}(F) \subseteq \sigma(X_1, \ldots, X_n)$, then $\mathcal{G}$ is clearly a $\sigma$ algebra which contains the open sets. Hence $\mathcal{G} = \mathcal{B}$ the Borel sets of $\prod_{i=1}^n Z_i$. This shows that $\sigma(X) \subseteq \sigma(X_1, \ldots, X_n)$. Next we observe that $\sigma(X)$ is a $\sigma$ algebra with the property that each $X_i$ is measurable with respect to $\sigma(X)$. This follows from $X_i^{-1}(B_i) = X^{-1}(\prod_{i=1}^n A_i) \in \sigma(X)$, where each $A_j = Z_j$ except for $A_i = B_i$. Since $\sigma(X_1, \ldots, X_n)$ is defined as the smallest such $\sigma$ algebra, it follows that $\sigma(X) \supseteq \sigma(X_1, \ldots, X_n)$.

For random variables having values in a separable Banach space or even more generally for a separable metric space, much can be said about regularity of $\lambda_X$.

**Definition 10.1.7** A measure, $\mu$ defined on $\mathcal{B}(E)$ will be called inner regular if for all $F \in \mathcal{B}(E)$,

$$\mu(F) = \sup \{ \mu(K) : K \subseteq F \text{ and } K \text{ is closed} \}$$

A measure, $\mu$ defined on $\mathcal{B}(E)$ will be called outer regular if for all $F \in \mathcal{B}(E)$,

$$\mu(F) = \inf \{ \mu(V) : V \supseteq F \text{ and } V \text{ is open} \}$$

When a measure is both inner and outer regular, it is called regular.

For probability measures, the above definition of regularity tends to come free. Note it is a little weaker than the usual definition of regularity because $K$ is only assumed to be closed, not compact. The following lemmas and corollaries were proved earlier. The following is Lemma 10.1.8.

**Lemma 10.1.8** Let $\mu$ be a finite measure defined on $\mathcal{B}(E)$ where $E$ is a metric space. Then $\mu$ is regular.

One can say more if the metric space is complete and separable. In fact in this case the above definition of inner regularity can be shown to imply the usual one. This is Lemma 10.1.9 proved earlier.

**Lemma 10.1.9** Let $\mu$ be a finite measure on a $\sigma$ algebra containing $\mathcal{B}(E)$, the Borel sets of $E$, a separable complete metric space. Then if $C$ is a closed set,

$$\mu(C) = \sup \{ \mu(K) : K \subseteq C \text{ and } K \text{ is compact} \}$$

**Corollary 10.1.10** Let $X$ be a random variable with values in a separable complete metric space, $Z$. Then $\lambda_X$ is an inner and outer regular measure defined on $\mathcal{B}(Z)$.

**Proposition 10.1.11** For $X$ a random vector defined above, $X$ having values in a complete separable metric space $Z$, then $\lambda_X$ is inner and outer regular and Borel.

$$(\Omega, \mathcal{F}, P) \overset{X}{\rightarrow} (Z, \lambda_X) \overset{h}{\rightarrow} E$$

If $h$ is Borel measurable and $h \in L^1(Z, \lambda_X; E)$ for $E$ a Banach space, then

$$\int_{\Omega} h(X(\omega)) \, dP = \int_{Z} h(x) \, d\lambda_X.$$  \hspace{1cm} (10.1.1)

In the case where $Z = E$, a separable Banach space, if $X$ is measurable then $X \in L^1(\Omega; E)$ if and only if the identity map on $E$ is in $L^1(\Omega; \lambda_X)$ and

$$\int_{\Omega} X(\omega) \, dP = \int_{E} x \, d\lambda_X(x).$$  \hspace{1cm} (10.1.2)
Proof: The regularity claims are established above. It remains to verify (10.1.3). Since \( h \in L^1(Z, E) \), it follows there exists a sequence of simple functions \( \{h_n\} \) such that

\[
h_n(x) \to h(x), \quad \int_Z \|h_m - h_n\| \, d\lambda_X \to 0 \quad \text{as} \quad m, n \to \infty.
\]

The first convergence above implies

\[
h_n \circ X \to h \circ X \quad \text{pointwise on} \quad \Omega \quad (10.1.3)
\]

Then letting \( h_n(x) = \sum_{k=1}^n x_k \chi_{E_k}(x) \), where the \( E_k \) are disjoint and Borel, it follows easily that \( h_n \circ X \) is also a simple function of the form \( h_n \circ X(\omega) = \sum_{k=1}^n x_k \chi_{X^{-1}(E_k)}(\omega) \) and by assumption \( X^{-1}(E_k) \in \mathcal{F} \). From the definition of the integral, it is easily seen

\[
\int h_n \circ XdP = \int h_n d\lambda_X, \quad \int \|h_n\| \circ XdP = \int \|h_n\| d\lambda_X
\]

Also, \( h_n \circ X - h_m \circ X \) is a simple function and so

\[
\int \|h_n \circ X - h_m \circ X\| dP = \int \|h_n - h_m\| d\lambda_X \quad (10.1.4)
\]

It follows from the definition of the Bochner integral and (10.1.3), and (10.1.4) that \( h \circ X \) is in \( L^1(\Omega; E) \) and

\[
\int h \circ XdP = \lim_{n \to \infty} \int h_n \circ XdP = \lim_{n \to \infty} \int h_n d\lambda_X = \int h d\lambda_X.
\]

Finally consider the case that \( E = Z \) and suppose \( X \in L^1(\Omega; E) \). Then letting \( h \) be the identity map on \( E \), it follows \( h \) is obviously separably valued and \( h^{-1}(U) \in \mathcal{B}(E) \) for all \( U \) open and so \( h \) is measurable. Why is it in \( L^1(E; E) \)?

\[
\int_E \|h(x)\| \, d\lambda_X = \int_0^\infty \lambda_X(\|\cdot\| > t) \, dt = \int_0^\infty P(X \in \{\|X\| > t\}) \, dt
\]

\[
= \int_0^\infty P(\|X\| > t) \, dt = \int_\Omega \|X\| \, dP < \infty
\]

Thus the identity map on \( E \) is in \( L^1(E; \lambda_X) \). Next suppose the identity map \( h \) is in \( L^1(\Omega; E) \). Then \( X(\omega) = h \circ X(\omega) \) and so from the first part, \( X \in L^1(\Omega; E) \) and from (10.1.1), (10.1.2) follows. \( \blacksquare \)

## 10.2 Kolmogorov Extension Theorem Review

Let \( M_t \) be a complete separable metric space. This is called a Polish space. \( I \) will denote a totally ordered index set, (Like \( \mathbb{R} \)) and the interest will be in building a measure on the product space, \( \prod_{t \in I} M_t \). By the well ordering principle, you can always put an order on any index set so this order is no restriction, but we do not insist on a well order and in fact, index sets of great interest are \( \mathbb{R} \) or \([0, \infty)\). Also for \( X \) a topological space, \( \mathcal{B}(X) \) will denote the Borel sets.

**Notation 10.2.1** The symbol \( J \) will denote a finite subset of \( I, \ J = (t_1, \cdots, t_n) \), the \( t_i \) taken in order. \( E_J \) will denote a set which has a set \( E_t \) of \( \mathcal{B}(M_t) \) in the \( t \)th position for \( t \in J \) and for \( t \notin J \), the set in the \( t \)th position will be \( M_t \). \( K_J \) will denote a set which has a compact set in the \( t \)th position for \( t \in J \) and for \( t \notin J \), the set in the \( t \)th position will be \( M_t \). Also denote by \( \mathcal{R}_J \) the sets \( E_J \) and \( \mathcal{R} \) the union of the \( \mathcal{R}_J \). Let \( \mathcal{E}_J \) denote finite disjoint unions of sets of \( \mathcal{R}_J \) and let \( \mathcal{E} \) denote finite disjoint unions of sets of \( \mathcal{R} \). Thus if \( F \) is a set of \( \mathcal{E} \), there exists \( J \) such that \( F \) is a finite disjoint union of sets of \( \mathcal{R}_J \). For \( F \in \mathcal{E} \), denote by \( \pi_J(F) \) the set \( \prod_{t \in J} F_t \) where \( F = \prod_{t \in I} F_t \).

Recall the following lemma about these things which have just been defined.

**Lemma 10.2.2** The sets, \( \mathcal{E}, \mathcal{E}_J \) defined above form an algebra of sets of \( \prod_{t \in I} M_t \).

With this preparation, here is a review of the Kolmogorov extension theorem, Theorem 10.2.1. In the statement and proof of the theorem, \( F_{\alpha}, G_{\alpha} \), and \( E_0 \) will denote Borel sets. Any list of indices from \( I \) will always be assumed to be taken in order. Thus, if \( J \subseteq I \) and \( J = (t_1, \cdots, t_n) \), it will always be assumed \( t_1 < t_2 < \cdots < t_n \).
Theorem 10.2.3 For each finite set

\[ J = (t_1, \cdots, t_n) \subseteq I, \]

suppose there exists a Borel probability measure, \( \nu_J = \nu_{t_1,\cdots,t_n} \) defined on the Borel sets of \( \prod_{t \in J} M_t \) such that the following consistency condition holds. If

\[ (t_1,\cdots,t_n) \subseteq (s_1,\cdots,s_p), \]

then

\[ \nu_{t_1,\cdots,t_n} (F_{t_1} \times \cdots \times F_{t_n}) = \nu_{s_1,\cdots,s_p} (G_{s_1} \times \cdots \times G_{s_p}) \]

where if \( s_i = t_j \), then \( G_{s_i} = F_{t_j} \) and if \( s_i \) is not equal to any of the indices, \( t_k \), then \( G_{s_i} = M_{s_i} \). Then for \( E \) defined in Notation 10.2.2, there exists a probability measure, \( P \) and a \( \sigma \) algebra \( \mathcal{F} = \sigma(E) \) such that

\[ \left( \prod_{t \in I} M_t, P, \mathcal{F} \right) \]

is a probability space. Also there exist measurable functions, \( X_s : \prod_{t \in I} M_t \rightarrow M_s \) defined as

\[ X_s x \equiv x_s \]

for each \( s \in I \) such that for each \( (t_1 \cdots t_n) \subseteq I \),

\[ \nu_{t_1,\cdots,t_n} (F_{t_1} \times \cdots \times F_{t_n}) = P (\cap_{k=1}^n [X_{t_k} \in F_{t_k}]) \]

\[ = P \left( (X_{t_1},\cdots,X_{t_n}) \in \prod_{j=1}^n F_{t_j} \right) = P \left( \prod_{t \in I} F_t \right) \]

where \( F_t = M_t \) for every \( t \notin \{t_1 \cdots t_n\} \) and \( F_{t_k} \) is a Borel set. Also if \( f \) is a nonnegative function of finitely many variables, \( x_{t_1},\cdots,x_{t_n} \), measurable with respect to \( \mathcal{B} \left( \prod_{j=1}^n M_{t_j} \right) \), then \( f \) is also measurable with respect to \( \mathcal{F} \) and

\[ \int_{M_{t_1} \times \cdots \times M_{t_n}} f (x_{t_1},\cdots,x_{t_n}) \, d\nu_{t_1,\cdots,t_n} \]

\[ = \int_{\prod_{t \in I} M_t} f (x_{t_1},\cdots,x_{t_n}) \, dP \]

10.2.1 Independence

The concept of independence is probably the main idea which separates probability from analysis and causes some of us to struggle to understand what is going on.

Definition 10.2.4 Let \( (\Omega, \mathcal{F}, P) \) be a probability space. The sets in \( \mathcal{F} \) are called events. A set of events, \( \{A_t\}_{t \in I} \) is called independent if whenever \( \{A_{i_k}\}_{k=1}^m \) is a finite subset

\[ P \left( \bigcap_{k=1}^m A_{i_k} \right) = \prod_{k=1}^m P (A_{i_k}). \]

Each of these events defines a rather simple \( \sigma \) algebra, \( (A_i, A_i^C, \emptyset, \Omega) \) denoted by \( \mathcal{F}_i \). Now the following lemma is interesting because it motivates a more general notion of independent \( \sigma \) algebras.

Lemma 10.2.5 Suppose \( B_i \in \mathcal{F}_i \) for \( i \in I \). Then for any \( m \in \mathbb{N} \)

\[ P \left( \bigcap_{i=1}^m B_{i_k} \right) = \prod_{k=1}^m P (B_{i_k}). \]
Proof: The proof is by induction on the number \( l \) of the \( B_{ik} \) which are not equal to \( A_{ik} \). First suppose \( l = 0 \). Then the above assertion is true by assumption. Suppose it is so for some \( l \) and there are \( l+1 \) sets not equal to \( A_{ik} \). If any equals \( \emptyset \) there is nothing to show. Both sides equal 0. If any equals \( \Omega \), there is also nothing to show. You can ignore that set in both sides and then you have by induction the two sides are equal because you have no more than \( l \) sets different than \( A_{ik} \). The only remaining case is where some \( B_{ik} = A_{ik}^C \). Say \( B_{i_{m+1}} = A_{i_{m+1}}^C \) for simplicity.

\[
P\left(\cap_{k=1}^{m+1} B_{ik}\right) = P\left(A_{i_{m+1}}^C \cap \cap_{k=1}^{m} B_{ik}\right)
\]

\[
= P\left(\cap_{k=1}^{m} B_{ik}\right) - P\left(A_{i_{m+1}} \cap \cap_{k=1}^{m} B_{ik}\right)
\]

Then by induction,

\[
= \prod_{k=1}^{m} P\left(B_{ik}\right) - \prod_{k=1}^{m} P\left(A_{i_{m+1}} \cap B_{ik}\right) = \prod_{k=1}^{m} P\left(B_{ik}\right)\left(1 - P\left(A_{i_{m+1}}\right)\right)
\]

thus proving it for \( l + 1 \). □

This motivates a more general notion of independence in terms of \( \sigma \) algebras.

Definition 10.2.6 If \( \{\mathcal{F}_i\}_{i \in I} \) is any set of \( \sigma \) algebras contained in \( \mathcal{F} \), they are said to be independent if whenever \( A_{ik} \in \mathcal{F}_{ik} \) for \( k = 1, 2, \ldots, m \), then

\[
P\left(\cap_{k=1}^{m} A_{ik}\right) = \prod_{k=1}^{m} P\left(A_{ik}\right).
\]

A set of random variables \( \{X_i\}_{i \in I} \) is independent if the \( \sigma \) algebras \( \{\sigma(X_i)\}_{i \in I} \) are independent \( \sigma \) algebras. Here \( \sigma(X) \) denotes the smallest \( \sigma \) algebra such that \( X \) is measurable. Thus \( \sigma(X) = \{X^{-1}(U) : U \text{ is a Borel set}\} \). More generally, \( \sigma(X_i : i \in I) \) is the smallest \( \sigma \) algebra such that each \( X_i \) is measurable.

Note that by Lemma 10.2.5 you can consider independent events in terms of independent \( \sigma \) algebras. That is, a set of independent events can always be considered as events taken from a set of independent \( \sigma \) algebras. This is a more general notion because here the \( \sigma \) algebras might have infinitely many sets in them.

Lemma 10.2.7 Suppose the set of random variables, \( \{X_i\}_{i \in I} \), is independent. Also suppose \( I_1 \subseteq I \) and \( j \notin I_1 \). Then the \( \sigma \) algebras \( \sigma(X_i : i \in I_1) \), \( \sigma(X_j) \) are independent \( \sigma \) algebras.

Proof: Let \( B \in \sigma(X_i) \). I want to show that for any \( A \in \sigma(X_i : i \in I_1) \), it follows that \( P(A \cap B) = P(A) P(B) \). Let \( \mathcal{K} \) consist of finite intersections of sets of the form \( X_k^{-1}(B_k) \) where \( B_k \) is a Borel set and \( k \in I_1 \). Thus \( \mathcal{K} \) is a \( \pi \) system and \( \sigma(\mathcal{K}) = \sigma(X_i : i \in I_1) \). Now if you have one of these sets of the form \( A = \cap_{k=1}^{m} X_k^{-1}(B_k) \) where without loss of generality, it can be assumed the \( k \) are distinct since \( X_k^{-1}(B_k) \cap X_k^{-1}(B_k') = X_k^{-1}(B_k \cap B_k') \), then

\[
P(A \cap B) = P\left(\cap_{k=1}^{m} X_k^{-1}(B_k) \cap B\right) = P\left(B\right) \prod_{k=1}^{m} P\left(X_k^{-1}(B_k)\right)
\]

Thus \( \mathcal{K} \) is contained in

\[
\mathcal{G} = \{A \in \sigma(X_i : i \in I_1) : P(A \cap B) = P(A) P(B)\}.
\]

Now \( \mathcal{G} \) is closed with respect to complements and countable disjoint unions. Here is why: If each \( A_i \in \mathcal{G} \) and the \( A_i \) are disjoint,

\[
P\left(\cup_{i=1}^{\infty} A_i \cap B\right) = P\left(\cup_{i=1}^{\infty} (A_i \cap B)\right)
\]

\[
= \sum_{i} P(A_i \cap B) = \sum_{i} P(A_i) P(B)
\]

\[
= P(B) \sum_{i} P(A_i) = P(B) P(\cup_{i=1}^{\infty} A_i)
\]
10.2. KOLMOGOROV EXTENSION THEOREM REVIEW

If \( A \in \mathcal{G} \),
\[
P(A^C \cap B) + P(A \cap B) = P(B)
\]
and so
\[
P(A^C \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = P(B)(1 - P(A)) = P(B)P(A^C).
\]

Therefore, from the lemma on \( \pi \) systems, Lemma 10.2.8 on Page 159 it follows \( \mathcal{G} \supseteq \sigma(\mathcal{K}) = \sigma(\mathcal{X}_i : i \in I_1). \) ■

**Lemma 10.2.8** If \( \{X_k\}_{k=1}^T \) are independent random variables having values in \( Z \) a separable metric space, and if \( g_k \) is a Borel measurable function, then \( \{g_k(X_k)\}_{k=1}^T \) is also independent. Furthermore, if the random variables have values in \( \mathbb{R} \), and they are all bounded, then
\[
E\left(\prod_{i=1}^r X_i\right) = \prod_{i=1}^r E(X_i).
\]

More generally, the above formula holds if it is only known that each \( X_i \in L^1(\Omega;\mathbb{R}) \) and
\[
\prod_{i=1}^r X_i \in L^1(\Omega;\mathbb{R}).
\]

**Proof:** First consider the claim about \( \{g_k(X_k)\}_{k=1}^T \). Letting \( O \) be an open set in \( Z \),
\[
(g_k \circ X_k)^{-1}(O) = X_k^{-1}(g_k^{-1}(O)) = X_k^{-1}(\text{Borel set}) \in \sigma(X_k).
\]

It follows \( (g_k \circ X_k)^{-1}(E) \) is in \( \sigma(X_k) \) whenever \( E \) is Borel because the sets whose inverse images are measurable includes the Borel sets. Thus \( \sigma(g_k \circ X_k) \subseteq \sigma(X_k) \) and this proves the first part of the lemma.

Let \( X_1 = \sum_{i=1}^m c_iX_{E_i}, X_2 = \sum_{j=1}^m d_jX_{F_j} \) where \( P(E_{i}F_{j}) = P(E_{i})P(F_{j}) \). Then
\[
\int X_1X_2dP = \sum_{i,j} d_jc_iP(E_{i})P(F_{j}) = \left(\int X_1dP\right) \left(\int X_2dP\right)
\]
In general for \( X_1, X_2 \) independent, there exist sequences of bounded simple functions \( \{s_n\}, \{t_n\} \) measurable with respect to \( \sigma(X_1) \) and \( \sigma(X_2) \) respectively such that \( s_n \to X_1 \) pointwise and \( t_n \to X_2 \) pointwise. Then from the above and the dominated convergence theorem,
\[
\int X_1X_2dP = \lim_{n \to \infty} \int s_nt_ndP = \lim_{n \to \infty} \left(\int s_ndP\right) \left(\int t_ndP\right) = \left(\int X_1dP\right) \left(\int X_2dP\right)
\]

Next suppose there are \( m \) of these independent bounded random variables. Then \( \prod_{i=2}^m X_i \in \sigma(X_2, \cdots, X_m) \) and by Lemma 10.2.7 the two random variables \( X_1 \) and \( \prod_{i=2}^m X_i \) are independent. Hence from the above and induction,
\[
\int \prod_{i=1}^m X_idP = \int X_1 \prod_{i=2}^m X_idP = \int X_1dP \int \prod_{i=2}^m X_idP = \prod_{i=1}^m \int X_idP
\]

Now consider the last claim. Replace each \( X_i \) with \( X_i^n \) where this is just a truncation of the form
\[
X_i^n = \begin{cases} 
X_i & \text{if } |X_i| \leq n \\
- n & \text{if } X_i > n \\
- n & \text{if } X_i < n
\end{cases}
\]
Then by the first part
\[
E\left(\prod_{i=1}^r X_i^n\right) = \prod_{i=1}^r E(X_i^n)
\]
Now $|\prod_{i=1}^n X_i^{|x^*}| \leq |\prod_{i=1}^n X_i| \in L^1$ and so by the dominated convergence theorem, you can pass to the limit in both sides to get the desired result. ■

Maybe this would be a good place to put a really interesting result known as the Doob Dynkin lemma. This amazing result is illustrated with the following diagram in which $X = \{X_1, \cdots, X_m\}$. By Proposition 10.10.10 $\sigma (X) = \sigma (X_1, \cdots, X_m)$.

$$
\begin{array}{c}
\Omega \\
\downarrow
\end{array}
\xrightarrow{X}
\begin{array}{c}
\prod_{i=1}^m E_i, B (\prod_{i=1}^m E_i) \\
\downarrow
\end{array}
F
$$

You start with $X$ and can write it as the composition $g \circ X$ provided $X$ is $\sigma (X)$ measurable.

**Lemma 10.2.9** Let $(\Omega, F)$ be a measure space and let $X_i : \Omega \to E_i$ where $E_i$ is a separable Banach space. Suppose also that $X : \Omega \to F$ where $F$ is a separable Banach space. Then $X$ is $\sigma (X_1, \cdots, X_m)$ measurable if and only if there exists a Borel measurable function $g : \prod_{i=1}^m E_i \to F$ such that $X = g (X_1, \cdots, X_m)$.

**Proof:** First suppose $X (\omega) = fX_W (\omega)$ where $f \in F$ and $W \in \sigma (X_1, \cdots, X_m)$. Then by Proposition 10.10.10, $W$ is of the form $(X_1, \cdots, X_m)^{-1} (B) \equiv X^{-1} (B)$ where $B$ is Borel in $\prod_{i=1}^m E_i$. Therefore,

$$X (\omega) = fX_{X^{-1} (B)} (\omega) = fX_B (X (\omega)).$$

Now suppose $X$ is measurable with respect to $\sigma (X_1, \cdots, X_m)$. Then there exist simple functions

$$X_n (\omega) = \sum_{k=1}^m \sum_{j=1}^n f_k X_{B_k} (X (\omega)) = g_n (X (\omega))$$

where the $B_k$ are Borel sets in $\prod_{i=1}^m E_i$, such that $X_n (\omega) \to X (\omega)$, each $g_n$ being Borel. Thus $g_n$ converges on $X (\Omega)$. Furthermore, the set on which $g_n$ does converge is a Borel set equal to

$$\cap_{n=1}^\infty \cup_{m=1}^\infty \cap_{p, q \geq m} \left\{ \| g_p - g_q \| < \frac{1}{n} \right\}$$

which contains $X (\Omega)$. Therefore, modifying $g_n$ by multiplying it by the indicator function of this Borel set containing $X (\Omega)$, we can conclude that $g_n$ converges to a Borel function $g$ and, passing to a limit in the above,

$$X (\omega) = g (X (\omega))$$

Conversely, suppose $X (\omega) = g (X (\omega))$. Why is $X \in \sigma (X)$ measurable?

$$X^{-1} (\text{open}) = X^{-1} (g^{-1} (\text{open})) = X^{-1} (\text{Borel}) \in \sigma (X)$$ ■

### 10.2.2 Independence For Banach Space Valued Random Variables

Recall that for $X$ a random variable, $\sigma (X)$ is the smallest $\sigma$ algebra containing all the sets of the form $X^{-1} (F)$ where $F$ is Borel. Since such sets, $X^{-1} (F)$ for $F$ Borel form a $\sigma$ algebra it follows $\sigma (X) = \{ X^{-1} (F) : F \text{ is Borel} \}$.

Next consider the case where you have a set of $\sigma$ algebras. The following lemma is helpful when you try to verify such a set of $\sigma$ algebras is independent. It says you only need to check things on $\pi$ systems contained in the $\sigma$ algebras. This is really nice because it is much easier to consider the smaller $\pi$ systems than the whole $\sigma$ algebra.

**Lemma 10.2.10** Suppose $\{ F_i \}_{i \in I}$ is a set of $\sigma$ algebras contained in $F$ where $F$ is a $\sigma$ algebra of sets of $\Omega$. Suppose that $\mathcal{K}_j \subseteq F_i$ is a $\pi$ system and $\mathcal{F}_i = \sigma (\mathcal{K}_i)$. Suppose also that whenever $J$ is a finite subset of $I$ and $A_j \in \mathcal{K}_j$ for $j \in J$, it follows

$$P (\cap_{j \in J} A_j) = \prod_{j \in J} P (A_j).$$

Then $\{ F_i \}_{i \in I}$ is independent.

**Proof:** I need to verify that under the given conditions, if $\{ j_1, j_2, \cdots, j_n \} \subseteq I$ and $A_{j_k} \subseteq \mathcal{F}_{j_k}$, then

$$P (\cap_{k=1}^n A_{j_k}) = \prod_{k=1}^n P (A_{j_k}).$$
By hypothesis, this is true if each \( A_{jk} \in K_{jk} \). Suppose it is true whenever there are at most \( r - 1 \geq 0 \) of the \( A_{jk} \) which are not in the corresponding \( K_{jk} \). Without loss of generality, say there are at most \( r - 1 \) sets in the first \( n - 1 \) which are not in the corresponding \( K_{jk} \).

Pick \( (A_{j1}, \ldots, A_{jn-1}) \) let

\[
G(A_{j1} \ldots A_{jn-1}) = \left\{ B \in F_{jn} : P\left(\bigcap_{k=1}^{n-1} A_{jk} \cap B\right) = \prod_{k=1}^{n-1} P\left(A_{jk}\right) P\left(B\right) \right\}
\]

I am going to show \( G(A_{j1} \ldots A_{jn-1}) \) is closed with respect to complements and countable disjoint unions and then apply the Lemma on \( \pi \) systems. By the induction hypothesis, \( K_{jn} \subseteq G(A_{j1} \ldots A_{jn-1}) \). If \( B \in G(A_{j1} \ldots A_{jn-1}) \),

\[
\prod_{k=1}^{n-1} P\left(A_{jk}\right) = P\left(\bigcap_{k=1}^{n-1} A_{jk}\right)
\]

\[
= P\left((\bigcap_{k=1}^{n-1} A_{jk} \cap B^C) \cup (\bigcap_{k=1}^{n-1} A_{jk} \cap B)\right)
\]

\[
= P\left(\bigcap_{k=1}^{n-1} A_{jk} \cap B^C\right) + P\left(\bigcap_{k=1}^{n-1} A_{jk} \cap B\right)
\]

\[
= P\left(\bigcap_{k=1}^{n-1} A_{jk} \cap B^C\right) + \prod_{k=1}^{n-1} P\left(A_{jk}\right) P\left(B\right)
\]

and so

\[
P\left(\bigcap_{k=1}^{n-1} A_{jk} \cap B^C\right) = \prod_{k=1}^{n-1} P\left(A_{jk}\right) (1 - P\left(B\right))
\]

\[
= \prod_{k=1}^{n-1} P\left(A_{jk}\right) P\left(B^C\right)
\]

showing if \( B \in G(A_{j1} \ldots A_{jn-1}) \), then so is \( B^C \). It is clear that \( G(A_{j1} \ldots A_{jn-1}) \) is closed with respect to disjoint unions also. Here is why. If \( \{B_{j}\}_{j=1}^{\infty} \) are disjoint sets in \( G(A_{j1} \ldots A_{jn-1}) \),

\[
P\left(\bigcup_{i=1}^{\infty} B_i \cap \bigcap_{k=1}^{n-1} A_{jk}\right) = \sum_{i=1}^{\infty} P\left(B_i \cap \bigcap_{k=1}^{n-1} A_{jk}\right)
\]

\[
= \sum_{i=1}^{\infty} P\left(B_i\right) \prod_{k=1}^{n-1} P\left(A_{jk}\right)
\]

\[
= \prod_{k=1}^{n-1} P\left(A_{jk}\right) \sum_{i=1}^{\infty} P\left(B_i\right)
\]

\[
= \prod_{k=1}^{n-1} P\left(A_{jk}\right) P\left(\bigcup_{i=1}^{\infty} B_i\right)
\]

Therefore, by the \( \pi \) system lemma, Lemma \( G(A_{j1} \ldots A_{jn-1}) = F_{jn} \). This proves the induction step in going from \( r - 1 \) to \( r \).

What is a useful \( \pi \) system for \( \mathcal{B}(E) \), the Borel sets of \( E \) where \( E \) is a Banach space?

Recall the fundamental lemma used to prove the Pettis theorem. It was proved on Page 128.

**Lemma 10.2.11** Let \( E \) be a separable real Banach space. Sets of the form

\[
\{x \in E : x_i^* (x) \leq \alpha_i, i = 1, 2, \ldots, m\}
\]

where \( x_i^* \in M \), a dense subspace of \( E^* \) and \( \alpha_i \in [-\infty, \infty] \) are a \( \pi \) system, and denoting this \( \pi \) system by \( \mathcal{K} \), it follows \( \sigma(\mathcal{K}) = \mathcal{B}(E) \). The sets of \( \mathcal{K} \) are examples of cylindrical sets.
Proof: The sets described are obviously a $\pi$ system. I want to show $\sigma(K)$ contains the closed balls because then $\sigma(K)$ contains the open sets and hence the open sets and the result will follow. Let $D' \subseteq B' \cap M$ be described in Lemma 10.2.13. Then

$$\{ x \in E : ||x - a|| \leq r \}$$

$$= \left\{ x \in E : \sup_{f \in D'} |f(x) - f(a)| \leq r \right\}$$

$$= \left\{ x \in E : \sup_{f \in D'} |f(x) - f(a)| \leq r \right\}$$

$$= \cap_{f \in D'} \{ x \in E : f(1) - r \leq f(x) \leq f(a) + r \}$$

$$= \cap_{f \in D'} \{ x \in E : f(x) \leq f(a) + r \text{ and } (-f)(x) \leq r - f(a) \}$$

which equals a countable intersection of sets of the given $\pi$ system. Therefore, every closed ball is contained in $\sigma(K)$. It follows easily that every open ball is also contained in $\sigma(K)$ because

$$B(a,r) = \cup_{n=1}^{\infty} B(a, r - \frac{1}{n}).$$

Since the Banach space is separable, it is completely separable and so every open set is the countable union of balls. This shows the open sets are in $\sigma(K)$ and so $\sigma(K) \supseteq B(E)$. However, all the sets in the $\pi$ system are closed hence Borel because they are inverse images of closed sets. Therefore, $\sigma(K) \subseteq B(E)$ and so $\sigma(K) = B(E)$. ■

Observation 10.2.12 Denote by $C_{\alpha,n}$ the set $\{ \beta \in \mathbb{R}^n : \beta_i \leq \alpha_i \}$. Also denote by $g_n$ an element of $M^n$ with the understanding that $g_n : E \to \mathbb{R}^n$ according to the rule

$$g_n(x) \equiv (g_1(x), \cdots, g_n(x)).$$

Then the sets in the above lemma can be written as $g_n^{-1}(C_{\alpha,n})$. In other words, sets of the form $g_n^{-1}(C_{\alpha,n})$ form a $\pi$ system for $B(E)$.

Next suppose you have some random variables having values in a separable Banach space, $E$, $\{X_i\}_{i \in I}$. How can you tell if they are independent? To show they are independent, you need to verify that

$$P(\cap_{k=1}^{n} X_{ik}^{-1}(F_{ik})) = \prod_{k=1}^{n} P(X_{ik}^{-1}(F_{ik}))$$

whenever the $F_{ik}$ are Borel sets in $E$. It is desirable to find a way to do this easily.

Lemma 10.2.13 Let $K$ be a $\pi$ system of sets of $E$, a separable real Banach space and let $(\Omega, F, P)$ be a probability space and $X : \Omega \to E$ be a random variable. Then

$$X^{-1}(\sigma(K)) = \sigma(X^{-1}(K))$$

Proof: First note that $X^{-1}(\sigma(K))$ is a $\sigma$ algebra which contains $X^{-1}(K)$ and so it contains $\sigma(X^{-1}(K))$. Thus

$$X^{-1}(\sigma(K)) \supseteq \sigma(X^{-1}(K))$$

Now let

$$G \equiv \{ A \in \sigma(K) : X^{-1}(A) \in \sigma(X^{-1}(K)) \}$$

Then $G \supseteq K$. If $A \in G$, then

$$X^{-1}(A) \in \sigma(X^{-1}(K))$$

and so

$$X^{-1}(A)^C = X^{-1}(A^C) \in \sigma(X^{-1}(K))$$

because $\sigma(X^{-1}(K))$ is a $\sigma$ algebra. Hence $A^C \in G$. Finally suppose $\{A_i\}$ is a sequence of disjoint sets of $G$. Then

$$X^{-1}(\cup_{i=1}^{\infty} A_i) = \cup_{i=1}^{\infty} X^{-1}(A_i) \in \sigma(X^{-1}(K))$$

again because $\sigma(X^{-1}(K))$ is a $\sigma$ algebra. It follows from Lemma 10.2.13 on Page 1 that $G \supseteq \sigma(K)$ and this shows that whenever

$$A \in \sigma(K), X^{-1}(A) \in \sigma(X^{-1}(K)).$$

Thus $X^{-1}(\sigma(K)) \subseteq \sigma(X^{-1}(K))$. ■

With this lemma, here is the desired result about independent random variables. Essentially, you can reduce to the case of random vectors having values in $\mathbb{R}^n$. 


10.2.3 Reduction To Finite Dimensions

Let $E$ be a Banach space and let $\mathbf{g} \in (E')^n$. Then for $x \in E$, $\mathbf{g} \otimes x$ is the vector in $\mathbb{F}^n$ which equals $(g_1(x), g_2(x), \ldots, g_n(x))$.

**Theorem 10.2.14** Let $X_i$ be a random variable having values in $E$ a real separable Banach space. The random variables $\{X_i\}_{i \in I}$ are independent if whenever

$$\{i_1, \ldots, i_n\} \subseteq I,$$

$m_{i_1}, \ldots, m_{i_n}$ are positive integers, and $\mathbf{g}_{m_{i_1}}, \ldots, \mathbf{g}_{m_{i_n}}$ are respectively in

$$(M)^{m_{i_1}}, \ldots, (M)^{m_{i_n}}$$

for $M$ a dense subspace of $E'$, $\left\{\mathbf{g}_{m_{i_j}} \circ X_{i_j}\right\}_{j=1}^n$ are independent random vectors having values in $\mathbb{R}^{m_{i_1}}, \ldots, \mathbb{R}^{m_{i_n}}$ respectively.

**Proof:** It is necessary to show that the events $X_{i_j}^{-1}(B_{i_j})$ are independent events whenever $B_{i_j}$ are Borel sets. By Lemma 10.2.11 and the above Lemma 10.2.14 it suffices to verify that the events

$$X_{i_j}^{-1}\left(\mathbf{g}_{m_{i_j}}^{-1}\left(C_{\tilde{\alpha}, m_{i_j}}\right)\right) = \left(\mathbf{g}_{m_{i_j}} \circ X_{i_j}\right)^{-1}\left(C_{\tilde{\alpha}, m_{i_j}}\right)$$

are independent where $C_{\tilde{\alpha}, m_{i_j}}$ are the cones described in Lemma 10.2.11. Thus

$$\tilde{\alpha} = (\alpha_{k_1}, \ldots, \alpha_{k_m})$$

$$C_{\tilde{\alpha}, m_{i_j}} = \prod_{i=1}^{m_{i_j}} (-\infty, \alpha_{k_i}]$$

But this condition is implied when the finite dimensional valued random vectors $\mathbf{g}_{m_{i_j}} \circ X_{i_j}$ are independent. ■

The above assertion also goes the other way as you may want to show.

10.2.4 0,1 Laws

I am following [32] for the proof of many of the following theorems. Recall the set of $\omega$ which are in infinitely many of the sets $\{A_n\}$ is

$$\cap_{n=1}^{\infty} \cup_{m=n}^{\infty} A_m.$$  

This is because $\omega$ is in the above set if and only if for every $n$ there exists $m \geq n$ such that it is in $A_m$.

**Theorem 10.2.15** Suppose $A_n \in \mathcal{F}_n$ where the $\sigma$ algebras $\{\mathcal{F}_n\}_{n=1}^{\infty}$ are independent. Suppose also that

$$\sum_{k=1}^{\infty} P(A_k) = \infty.$$  

Then

$$P\left(\cap_{n=1}^{\infty} \cup_{m=n}^{\infty} A_m\right) = 1.$$  

**Proof:** It suffices to verify that

$$P\left(\cup_{n=1}^{\infty} \cap_{m=n}^{\infty} A_m^c\right) = 0$$

which can be accomplished by showing

$$P\left(\cap_{m=n}^{\infty} A_m^c\right) = 0$$

for each $n$. The sets $\{A_k^c\}$ satisfy $A_k^c \in \mathcal{F}_k$. Therefore, noting that $e^{-x} \geq 1 - x$,

$$P\left(\cap_{m=n}^{\infty} A_m^c\right) = \lim_{N \to \infty} P\left(\cap_{m=n}^{N} A_m^c\right) = \lim_{N \to \infty} \prod_{m=n}^{N} P\left(A_m^c\right)$$

$$= \lim_{N \to \infty} \prod_{m=n}^{N} (1 - P(A_m)) \leq \lim_{N \to \infty} \prod_{m=n}^{N} e^{-P(A_m)}$$

$$= \lim_{N \to \infty} \exp \left( - \sum_{m=n}^{N} P(A_m) \right) = 0. \quad ■$$

The Kolmogorov zero one law follows next. It has to do with something called a tail event.
Definition 10.2.16 Let \( \{ F_n \} \) be a sequence of \( \sigma \) algebras. Then \( T_n \equiv \sigma ( \bigcup_{k=n}^{\infty} F_k ) \) where this means the smallest \( \sigma \) algebra which contains each \( F_k \) for \( k \geq n \). Then a tail event is a set which is in the \( \sigma \) algebra, \( T \equiv \cap_{n=1}^{\infty} T_n \).

As usual, \((\Omega, \mathcal{F}, P)\) is the underlying probability space such that all \( \sigma \) algebras are contained in \( \mathcal{F} \).

Lemma 10.2.17 Suppose \( \{ F_n \}_{n=1}^{\infty} \) are independent \( \sigma \) algebras and suppose \( A \) is a tail event and \( A_k \in \mathcal{F}_{k_i} \), \( i = 1, \cdots, m \) are given sets. Then

\[
P ( A_{k_1} \cap \cdots \cap A_{k_m} \cap A ) = P ( A_{k_1} \cap \cdots \cap A_{k_m} ) P ( A )
\]

Proof: Let \( \mathcal{K} \) be the \( \pi \) system consisting of finite intersections of the form

\[
B_{m_1} \cap B_{m_2} \cap \cdots \cap B_{m_j}
\]

where \( m \in \mathcal{F}_{k_i} \) for \( k_i > \max \{ k_1, \cdots, k_m \} \equiv N \). Thus \( \sigma ( \mathcal{K} ) = \sigma ( \bigcup_{i=N+1}^{\infty} F_i ) \). Now let

\[
\mathcal{G} \equiv \{ B \in \sigma ( \mathcal{K} ) : P ( A_{k_1} \cap \cdots \cap A_{k_m} \cap B ) = P ( A_{k_1} \cap \cdots \cap A_{k_m} ) P ( B ) \}
\]

Then clearly \( \mathcal{K} \subseteq \mathcal{G} \). It is also true that \( \mathcal{G} \) is closed with respect to complements and countable disjoint unions. By the lemma on \( \pi \) systems, \( \mathcal{G} = \sigma ( \mathcal{K} ) = \sigma ( \bigcup_{i=N+1}^{\infty} F_i ) \). Since \( A \) is in \( \sigma ( \bigcup_{i=N+1}^{\infty} F_i ) \) due to the assumption that it is a tail event, it follows that

\[
P ( A_{k_1} \cap \cdots \cap A_{k_m} \cap A ) = P ( A_{k_1} \cap \cdots \cap A_{k_m} ) P ( A )
\]

Theorem 10.2.18 Suppose the \( \sigma \) algebras, \( \{ F_n \}_{n=1}^{\infty} \) are independent and suppose \( A \) is a tail event. Then \( P ( A ) \) either equals 0 or 1.

Proof: Let \( A \in T \). I want to show that \( P ( A ) = P ( A )^2 \). Let \( \mathcal{K} \) denote sets of the form \( A_{k_1} \cap \cdots \cap A_{k_m} \) for some \( m, A_{k_j} \in \mathcal{F}_{k_j} \) where each \( k_j > n \). Thus \( \mathcal{K} \) is a \( \pi \) system and

\[
\sigma ( \mathcal{K} ) = \sigma ( \bigcup_{k=n+1}^{\infty} F_k ) \equiv T_{n+1}
\]

Let

\[
\mathcal{G} \equiv \{ B \in T_{n+1} : P ( A \cap B ) = P ( A ) P ( B ) \}
\]

Thus \( \mathcal{K} \subseteq \mathcal{G} \) because

\[
P ( A_{k_1} \cap \cdots \cap A_{k_m} \cap A ) = P ( A_{k_1} \cap \cdots \cap A_{k_m} ) P ( A )
\]

by Lemma 10.2.17. However, \( \mathcal{G} \) is closed with respect to countable disjoint unions and complements. Here is why. If \( B \in \mathcal{G} \),

\[
P ( A \cap B^C ) + P ( A \cap B ) = P ( A )
\]

and so

\[
P ( A \cap B^C ) = P ( A ) - P ( A \cap B ) = P ( A ) ( 1 - P ( B ) ) = P ( A ) P ( B^C ).
\]

and so \( B^C \in \mathcal{G} \). If \( \{ B_i \}_{i=1}^{\infty} \) are disjoint sets in \( \mathcal{G} \),

\[
P ( A \cap \bigcup_{k=1}^{\infty} B_k ) = \sum_{k=1}^{\infty} P ( A \cap B_k ) = P ( A ) \sum_{k=1}^{\infty} P ( B_k )
\]

and so \( \bigcup_{k=1}^{\infty} B_k \in \mathcal{G} \). Therefore by the Lemma on \( \pi \) systems Lemma 10.2.13 on Page \( \alpha \) it follows \( \mathcal{G} = \sigma ( \mathcal{K} ) = \sigma ( \bigcup_{k=n+1}^{\infty} F_k ) \).

Thus for any \( B \in \sigma ( \bigcup_{k=n+1}^{\infty} F_k ) = T_{n+1} \), \( P ( A \cap B ) = P ( A ) P ( B ) \). However, \( A \) is in all of these \( T_{n+1} \) and so \( P ( A \cap A ) = P ( A ) = P ( A )^2 \) so \( P ( A ) \) equals either 0 or 1.

What sorts of things are tail events of independent \( \sigma \) algebras?

Theorem 10.2.19 Let \( \{ X_k \} \) be a sequence of independent random variables having values in \( Z \) a Banach space. Then

\[
A \equiv \{ \omega : \{ X_k ( \omega ) \} \text{ converges} \}
\]

is a tail event of the independent \( \sigma \) algebras \( \{ \sigma ( X_k ) \} \). So is

\[
B \equiv \{ \omega : \left\{ \sum_{k=1}^{\infty} X_k ( \omega ) \right\} \text{ converges} \}.
\]
The set where the sum converges is then
\[ \cap_{n=1}^{\infty} \cap_{p=1}^{\infty} \cup_{m=p}^{\infty} \cap_{l,k \geq m} \{ \omega : ||X_k(\omega) - X_l(\omega)|| < 1/n \} \]

Note that
\[ \cup_{m=p}^{\infty} \cap_{l,k \geq m} \{ \omega : ||X_k(\omega) - X_l(\omega)|| < 1/n \} \in \sigma \left( \cup_{j=p}^{\infty} \sigma (X_j) \right) \]
for every \( p \). Thus the set where ultimately any pair of \( X_k, X_l \) are closer together than \( 1/n \),
\[ \cap_{p=1}^{\infty} \cup_{m=p}^{\infty} \cap_{l,k \geq m} \{ \omega : ||X_k(\omega) - X_l(\omega)|| < 1/n \} \]
is a tail event. The set where \( \{X_k(\omega)\} \) is a Cauchy sequence is the intersection of all these and is therefore, also a tail event.

Now consider \( B \). This set is the same as the set where the partial sums are Cauchy sequences. Let \( S_n \equiv \sum_{k=1}^{n} X_k \).
The set where the sum converges is then
\[ \cap_{n=1}^{\infty} \cap_{p=2}^{\infty} \cup_{m=p}^{\infty} \cap_{l,k \geq m} \{ \omega : ||S_k(\omega) - S_l(\omega)|| < 1/n, \ k \geq m \} \]
Say \( k < l \) and consider for \( m \geq p \)
\[ \{ \omega : ||S_k(\omega) - S_l(\omega)|| < 1/n, \ k \geq m \} \]
This is the same as
\[ \left\{ \omega : \left| \sum_{j=k-1}^{l} X_j(\omega) \right| < 1/n, k \geq m \right\} \in \sigma \left( \cup_{j=p-1}^{\infty} \sigma (X_j) \right) \]
Thus
\[ \cup_{m=p}^{\infty} \cap_{l,k \geq m} \{ \omega : ||S_k(\omega) - S_l(\omega)|| < 1/n \} \in \sigma \left( \cup_{j=p-1}^{\infty} \sigma (X_j) \right) \]
and so the intersection for all \( p \) of these is a tail event. Then the intersection over all \( n \) of these tail events is a tail event.

From this it can be concluded that if you have a sequence of independent random variables, \( \{X_k\} \) the set where it converges is either of probability \( 1 \) or probability \( 0 \). A similar conclusion holds for the set where the infinite sum of these random variables converges. This is stated in the next corollary. This incredible assertion is the next corollary.

**Corollary 10.2.20** Let \( \{X_k\} \) be a sequence of random variables having values in a Banach space. Then
\[ \lim_{n \to \infty} X_n(\omega) \]
either exists for a.e. \( \omega \) or the convergence fails to take place for a.e. \( \omega \). Also if
\[ A \equiv \left\{ \omega : \sum_{k=1}^{\infty} X_k(\omega) \text{ converges} \right\}, \]
then \( P(A) = 0 \) or \( 1 \).

**10.2.5 Kolmogorov’s Inequality, Strong Law Of Large Numbers**
Kolmogorov’s inequality is a very interesting inequality which depends on independence of a set of random vectors. The random vectors have values in \( \mathbb{R}^n \) or more generally some real separable Hilbert space.

**Lemma 10.2.21** If \( Y, X \) are independent random variables having values in a real separable Hilbert space, \( H \) with \( E( |X|^2 ), E( |Y|^2 ) < \infty \), then
\[ \int_{\Omega} (X, Y) dP = \left( \int_{\Omega} X dP \right) \left( \int_{\Omega} Y dP \right). \]

**Proof:** Let \( \{e_k\} \) be a complete orthonormal basis. Thus
\[ \int_{\Omega} (X, Y) dP = \int_{\Omega} \sum_{k=1}^{\infty} (X, e_k) (Y, e_k) dP \]
Now
\[
\int_\Omega \sum_{k=1}^{\infty} \left\| (X, e_k) (\varepsilon, e_k) \right\| dP \leq \int_\Omega \left( \sum_{k} \left\| (X, e_k) \right\|^2 \right)^{1/2} \left( \sum_{k} \left\| (\varepsilon, e_k) \right\|^2 \right)^{1/2} dP
\]
\[= \int_\Omega \left\| X \right\| \left\| Y \right\| dP \leq \left( \int_\Omega \left\| X \right\|^2 dP \right)^{1/2} \left( \int_\Omega \left\| Y \right\|^2 dP \right)^{1/2} < \infty\]

and so by Fubini’s theorem,
\[
\int_\Omega (X, Y) dP = \int_\Omega \sum_{k=1}^{\infty} (X, e_k) (\varepsilon, e_k) dP = \sum_{k=1}^{\infty} \int_\Omega (X, e_k) (\varepsilon, e_k) dP
\]
\[= \sum_{k=1}^{\infty} \int_\Omega (X, e_k) dP \int_\Omega (\varepsilon, e_k) dP = \sum_{k=1}^{\infty} \left( \int_\Omega X dP, e_k \right) \left( \int_\Omega \varepsilon dP, e_k \right) dP
\]
\[= \left( \int_\Omega X dP, \int_\Omega \varepsilon dP \right) \quad \Box
\]

Now here is Kolmogorov’s inequality.

**Theorem 10.2.22** Suppose \( \{X_k\}_{k=1}^{n} \) are independent with \( E(\|X_k\|) < \infty, E(X_k) = 0 \). Then for any \( \varepsilon > 0 \),

\[
P \left( \max_{1 \leq k \leq n} \left| \sum_{j=1}^{k} X_j \right| \geq \varepsilon \right) \leq \frac{1}{\varepsilon^2} \sum_{j=1}^{n} E \left( \|X_k\|^2 \right).
\]

**Proof:** Let
\[
A = \left\{ \max_{1 \leq k \leq n} \left| \sum_{j=1}^{k} X_j \right| \geq \varepsilon \right\}
\]

Now let \( A_1 = \{\left| X_1 \right| \geq \varepsilon \} \) and if \( A_1, \ldots, A_m \) have been chosen,
\[
A_{m+1} = \left( \sum_{j=1}^{m+1} X_j \right) \geq \varepsilon \right\} \cap \bigcap_{r=1}^{m} \left( \sum_{j=1}^{r} X_j \right) < \varepsilon \right\}
\]

Thus the \( A_k \) partition \( A \) and \( \omega \in A_k \) means
\[
\left| \sum_{j=1}^{k} X_j \right| \geq \varepsilon
\]

but this did not happen for \( \sum_{j=1}^{r} X_j \) for any \( r < k \). Note also that \( A_k \in \sigma(X_1, \ldots, X_k) \). Then from algebra,
\[
\left( \sum_{j=1}^{n} X_j \right)^2 = \left( \sum_{i=1}^{k} X_i + \sum_{j=k+1}^{n} X_j \right)^2 = \left( \sum_{i=1}^{k} X_i \right)^2 + \sum_{i \leq k \leq j \geq k} (X_i, X_j) + \sum_{i \geq k \leq j} (X_j, X_i)
\]

Written more succinctly,
\[
\left( \sum_{j=1}^{n} X_j \right)^2 = \left( \sum_{j=1}^{k} X_j \right)^2 + \sum_{j > k \text{ or } i > k} (X_i, X_j)
\]

Now multiply both sides by \( \mathcal{X}_{A_k} \) and integrate. Suppose \( i \leq k \) for one of the terms in the second sum. Then by Lemma 10.4 and \( A_k \in \sigma(X_1, \ldots, X_k) \), the two random vectors \( \mathcal{X}_k X_i, X_j \) are independent,
\[
\int_\Omega \mathcal{X}_{A_k} (X_i, X_j) dP = \left( \int_\Omega \mathcal{X}_{A_k} X_i dP, \int_\Omega X_j dP \right) = 0
\]
the last equality holding because by assumption $E(X_j) = 0$. Therefore, it can be assumed both $i, j$ are larger than $k$ and

$$
\int_\Omega \mathcal{X}_k \left| \sum_{j=1}^n X_j \right|^2 dP = \int_\Omega \mathcal{X}_k \left| \sum_{j=1}^k X_j \right|^2 dP + \sum_{j>k, i>k} \int_\Omega \mathcal{X}_k (X_i, X_j) dP \tag{10.2.5}
$$

The last term on the right is interesting. Suppose $i > j$. The integral inside the sum is of the form

$$
\int_\Omega (X_i, \mathcal{X}_k, X_j) dP
$$

The second factor in the inner product is in $\sigma(X_1, \ldots, X_k, X_j)$ and $X_i$ is not included in the list of random vectors. Thus by Lemma [10.2.7], the two random vectors $X_i, \mathcal{X}_k, X_j$ are independent and so (10.2.6) reduces to

$$
\left( \int_\Omega X_idP, \int_\Omega \mathcal{X}_k X_j dP \right) = (0, \int_\Omega \mathcal{X}_k X_j dP) = 0.
$$

A similar result holds if $j > i$. Thus the mixed terms in the last term of (10.2.5) are all equal to 0. Hence (10.2.6) reduces to

$$
\int_\Omega \mathcal{X}_k \left| \sum_{j=1}^n X_j \right|^2 dP = \int_\Omega \mathcal{X}_k \left| \sum_{j=1}^k X_j \right|^2 dP + \sum_{i>k} \int_\Omega \mathcal{X}_k |X_i|^2 dP
$$

and so

$$
\int_\Omega \mathcal{X}_k \left| \sum_{j=1}^n X_j \right|^2 dP \geq \int_\Omega \mathcal{X}_k \left| \sum_{j=1}^k X_j \right|^2 dP \geq \varepsilon^2 P(A_k).
$$

Now, summing these yields

$$
\varepsilon^2 P(A) \leq \int_\Omega \mathcal{X}_k \left| \sum_{j=1}^n X_j \right|^2 dP \leq \int_\Omega \left| \sum_{j=1}^n X_j \right|^2 dP = \sum_{i,j} \int_\Omega (X_i, X_j) dP
$$

By independence of the random vectors the mixed terms of the above sum equal zero and so it reduces to

$$
\sum_{i=1}^n \int_\Omega |X_i|^2 dP \tag{\blacksquare}
$$

This theorem implies the following amazing result.

**Theorem 10.2.23** Let $\{X_k\}_{k=1}^\infty$ be independent random vectors having values in a separable real Hilbert space and suppose $E(|X_k|) < \infty$ for each $k$ and $E(X_k) = 0$. Suppose also that

$$
\sum_{j=1}^\infty E\left(|X_j|^2\right) < \infty.
$$

Then

$$
\sum_{j=1}^\infty X_j
$$

converges a.e.
**Proof:** Let \( \varepsilon > 0 \) be given. By Kolmogorov’s inequality, Theorem 10.2.22, it follows that for \( p \leq m < n \)

\[
P \left( \left\{ \max_{m \leq k \leq n} \left| \sum_{j=m}^{k} X_j \right| \geq \varepsilon \right\} \right) \leq \frac{1}{\varepsilon^2} \sum_{j=p}^{n} E \left( |X_j|^2 \right).
\]

Therefore, letting \( n \to \infty \) it follows that for all \( m, n \) such that \( p \leq m \leq n \)

\[
P \left( \left\{ \max_{p \leq m \leq n} \left| \sum_{j=m}^{n} X_j \right| \geq \varepsilon \right\} \right) \leq \frac{1}{\varepsilon^2} \sum_{j=p}^{\infty} E \left( |X_j|^2 \right).
\]

It follows from the assumption

\[
\sum_{j=1}^{\infty} E \left( |X_j|^2 \right) < \infty
\]

there exists a sequence, \( \{p_n\} \) such that if \( m \geq p_n \)

\[
P \left( \left\{ \max_{k \geq m \geq p_n} \left| \sum_{j=m}^{k} X_j \right| \geq 2^{-n} \right\} \right) \leq 2^{-n}.
\]

By the Borel Cantelli lemma, Lemma 10.1.2, there is a set of measure 0, \( N \) such that for \( \omega \notin N \), \( \omega \) is in only finitely many of the sets,

\[
\left[ \max_{k \geq m \geq p_n} \left| \sum_{j=m}^{k} X_j \right| \geq 2^{-n} \right]
\]

and so for \( \omega \notin N \), it follows that for large enough \( n \),

\[
\left[ \max_{k \geq m \geq p_n} \left| \sum_{j=m}^{k} X_j(\omega) \right| < 2^{-n} \right]
\]

However, this says the partial sums \( \left\{ \sum_{j=1}^{k} X_j(\omega) \right\}_{k=1}^{\infty} \) are a Cauchy sequence. Therefore, they converge. 

With this amazing result, there is a simple proof of the strong law of large numbers. In the following lemma, \( s_k \) and \( a_j \) could have values in any normed linear space.

**Lemma 10.2.24** Suppose \( s_k \to s \). Then

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} s_k = s.
\]

Also if

\[
\sum_{j=1}^{\infty} \frac{a_j}{j}
\]

converges, then

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} a_j = 0.
\]

**Proof:** Consider the first part. Since \( s_k \to s \), it follows there is some constant, \( C \) such that \( |s_k| < C \) for all \( k \) and \( |s| < C \) also. Choose \( K \) so large that if \( k \geq K \), then for \( n > K \),

\[
|s - s_k| < \varepsilon/2.
\]

\[
\left| s - \frac{1}{n} \sum_{k=1}^{n} s_k \right| \leq \frac{1}{n} \sum_{k=1}^{n} |s_k - s|
\]

Therefore, by the Borel Cantelli lemma, there is a set of measure 0, \( N \) such that for \( \omega \notin N \), \( \omega \) is in only finitely many of the sets,

\[
\left[ \max_{k \geq m \geq p_n} \left| \sum_{j=m}^{k} X_j \right| \geq 2^{-n} \right]
\]

and so for \( \omega \notin N \), it follows that for large enough \( n \),

\[
\left[ \max_{k \geq m \geq p_n} \left| \sum_{j=m}^{k} X_j(\omega) \right| < 2^{-n} \right]
\]

However, this says the partial sums \( \left\{ \sum_{j=1}^{k} X_j(\omega) \right\}_{k=1}^{\infty} \) are a Cauchy sequence. Therefore, they converge. 

With this amazing result, there is a simple proof of the strong law of large numbers. In the following lemma, \( s_k \) and \( a_j \) could have values in any normed linear space.

**Lemma 10.2.24** Suppose \( s_k \to s \). Then

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} s_k = s.
\]

Also if

\[
\sum_{j=1}^{\infty} \frac{a_j}{j}
\]

converges, then

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} a_j = 0.
\]

**Proof:** Consider the first part. Since \( s_k \to s \), it follows there is some constant, \( C \) such that \( |s_k| < C \) for all \( k \) and \( |s| < C \) also. Choose \( K \) so large that if \( k \geq K \), then for \( n > K \),

\[
|s - s_k| < \varepsilon/2.
\]

\[
\left| s - \frac{1}{n} \sum_{k=1}^{n} s_k \right| \leq \frac{1}{n} \sum_{k=1}^{n} |s_k - s|
\]

However, this says the partial sums \( \left\{ \sum_{j=1}^{k} X_j(\omega) \right\}_{k=1}^{\infty} \) are a Cauchy sequence. Therefore, they converge. 

With this amazing result, there is a simple proof of the strong law of large numbers. In the following lemma, \( s_k \) and \( a_j \) could have values in any normed linear space.
\[
\begin{align*}
= \frac{1}{n} \sum_{k=1}^{K} |s_k - s| + \frac{1}{n} \sum_{k=K}^{n} |s_k - s| \\
\leq 2CK + \frac{\varepsilon n - K}{n} < 2CK + \frac{\varepsilon}{2}
\end{align*}
\]

Therefore, whenever \( n \) is large enough,
\[
|s - \frac{1}{n} \sum_{k=1}^{n} s_k| < \varepsilon.
\]

Now consider the second claim. Let
\[
s_k = \sum_{j=1}^{k} \frac{a_j}{j}
\]
and \( s = \lim_{k \to \infty} s_k \) Then by the first part,
\[
s = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} s_k = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \sum_{j=1}^{k} \frac{a_j}{j}
\]
\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} \frac{a_j}{j} \sum_{k=j}^{n} 1 = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} \frac{a_j}{j} (n - j)
\]
\[
= \lim_{n \to \infty} \left( \sum_{j=1}^{n} \frac{a_j}{j} - \frac{1}{n} \sum_{j=1}^{n} a_j \right) = s - \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} a_j
\]

Now here is the strong law of large numbers.

**Theorem 10.2.25** Suppose \( \{X_k\} \) are independent random variables and \( E(|X_k|) < \infty \) for each \( k \) and \( E(X_k) = m_k \). Suppose also
\[
\sum_{j=1}^{\infty} \frac{1}{j^2} E\left(|X_j - m_j|^2\right) < \infty.
\]

Then
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} (X_j - m_j) = 0
\]

**Proof:** Consider the sum
\[
\sum_{j=1}^{\infty} \frac{X_j - m_j}{j}.
\]

This sum converges a.e. because of (10.2.7) and Theorem 10.2.23 applied to the random vectors \( \{\frac{X_j - m_j}{j}\} \). Therefore, from Lemma 10.2.24 it follows that for a.e. \( \omega \),
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} (X_j(\omega) - m_j) = 0 \quad \blacksquare
\]

The next corollary is often called the strong law of large numbers. It follows immediately from the above theorem.

**Corollary 10.2.26** Suppose \( \{X_j\}_{j=1}^{\infty} \) are independent random vectors, having mean \( m \) and variance equal to
\[
\sigma^2 = \int_{\Omega} |X_j - m|^2 \, dP < \infty.
\]

Then for a.e. \( \omega \in \Omega \)
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} X_j(\omega) = m
\]
10.3 Tight Measures

Now here is a definition of what it means for a set of measures to be tight.

**Definition 10.3.1** Let $\Lambda$ be a set of probability measures defined on the Borel sets of a topological space. Then $\Lambda$ is “tight” if for all $\varepsilon > 0$ there exists a compact set, $K_\varepsilon$ such that

$$\mu \left( \left\{ x \notin K_\varepsilon \right\} \right) < \varepsilon$$

for all $\mu \in \Lambda$.

Lemma 10.3.1 implies a single probability measure on the Borel sets of a separable metric space is tight. The proof of that lemma generalizes slightly to give a simple criterion for a set of measures to be tight.

**Lemma 10.3.2** Let $E$ be a separable complete metric space and let $\Lambda$ be a set of Borel probability measures. Then $\Lambda$ is tight if and only if for every $\varepsilon > 0$ and $r > 0$ there exists a finite collection of balls, $\{B(a_i, r)\}_{i=1}^m$ such that

$$\mu \left( \bigcup_{i=1}^m B(a_i, r) \right) > 1 - \varepsilon$$

for every $\mu \in \Lambda$.

**Proof:** If $\Lambda$ is tight, then there exists a compact set, $K_\varepsilon$ such that

$$\mu(K_\varepsilon) > 1 - \varepsilon$$

for all $\mu \in \Lambda$. Then consider the open cover, $\{B(x, r) : x \in K_\varepsilon\}$. Finitely many of these cover $K_\varepsilon$ and this yields the above condition.

Now suppose the above condition and let

$$C_n = \bigcup_{m=1}^n B(a_i^m, 1/n)$$

satisfy $\mu(C_n) > 1 - \varepsilon/2^n$ for all $\mu \in \Lambda$. Then let $K_\varepsilon = \bigcap_{n=1}^\infty C_n$. This set $K_\varepsilon$ is a compact set because it is a closed subset of a complete metric space and is therefore complete, and it is also totally bounded by construction. For $\mu \in \Lambda$,

$$\mu(K^c_\varepsilon) = \mu(\bigcup_n C^c_n) \leq \sum_n \mu(C^c_n) < \sum_n \frac{\varepsilon}{2^n} = \varepsilon$$

Therefore, $\Lambda$ is tight. ■

Prokhorov’s theorem is an important result which also involves tightness. In order to give a proof of this important theorem, it is necessary to consider some simple results from topology which are interesting for their own sake.

**Theorem 10.3.3** Let $H$ be a compact metric space. Then there exists a compact subset of $[0, 1]$, $K$ and a continuous function, $\theta$ which maps $K$ onto $H$.

**Proof:** Without loss of generality, it can be assumed $H$ is an infinite set since otherwise the conclusion is trivial. You could pick finitely many points of $[0, 1]$ for $K$.

Since $H$ is compact, it is totally bounded. Therefore, there exists a 1 net for $H \{h_i\}_{i=1}^{m_1}$. Letting $H^1_i = B(h_i, 1)$, it follows $H^1_i$ is also a compact metric space and so there exists a 1/2 net for each $H^1_i$, $\{h_j^i\}_{j=1}^{m_i}$. Then taking the intersection of $B(h^i_j, 1/2)$ with $H^1_i$ to obtain sets denoted by $H^2_j$ and continuing this way, one can obtain compact subsets of $H$, $\{H^i_j\}$ which satisfies: each $H^i_j$ is contained in some $H^i_{k-1}$, each $H^i_j$ is compact with diameter less than $i^{-1}$, each $H^i_{j}$ is the union of sets of the form $H^{i+1}_k$ which are contained in it. Denoting by $\{H^i_j\}_{j=1}^{m_i}$ those sets corresponding to a superscript of $i$, it can also be assumed $m_i < m_{i+1}$. If this is not so, simply add in another point to the $i^{-1}$ net. Now let $\{I^i_j\}_{j=1}^{m_i}$ be disjoint closed intervals in $[0, 1]$ each of length no longer than $2^{-m_i}$ which have the property that $I^i_j$ is contained in $I^{i-1}_k$ for some $k$. Letting $K_i \equiv \bigcup_{j=1}^{m_i} I^i_j$, it follows $K_i$ is a sequence of nested compact sets. Let $K = \bigcap_{i=1}^\infty K_i$. Then each $x \in K$ is the intersection of a unique sequence of these closed intervals, $\{I^i_j\}_{j=1}^\infty$. Define $\theta x = \bigcap_{i=1}^\infty H^i_{j_i}$. Since the diameters of the $H^i_j$ converge to 0 as $i \to \infty$, this function is well defined. It is continuous because if $x_n \to x$, then ultimately $x_n$ and $x$ are both in $I^k_{j_k}$, the $k^{th}$ closed interval in the sequence whose intersection is $x$. Hence, $d(\theta x_n, \theta x) \leq \text{diameter}(H^k_{j_k}) \leq 1/k$. To see the map is onto, let $h \in H$. Then from the construction, there exists a sequence $\{H^k_{j_k}\}_{k=1}^\infty$ of the above sets whose intersection equals $h$. Then $\theta \left( \bigcap_{k=1}^\infty H^k_{j_k} \right) = h$.

Note $\theta$ is maybe not one to one.

As an important corollary, it follows that the continuous functions defined on any compact metric space is separable.
10.3. TIGHT MEASURES

Corollary 10.3.4 Let $H$ be a compact metric space and let $C(H)$ denote the continuous functions defined on $H$ with the usual norm,

$$\|f\|_\infty \equiv \max \{|f(x)| : x \in H\}$$

Then $C(H)$ is separable.

Proof: The proof is by contradiction. Suppose $C(H)$ is not separable. Let $\mathcal{H}_k$ denote a maximal collection of functions of $C(H)$ with the property that if $f, g \in \mathcal{H}_k$, then $\|f - g\|_\infty \geq 1/k$. The existence of such a maximal collection of functions is a consequence of a simple use of the Hausdorff maximality theorem. Then $\bigcup_{k=1}^\infty \mathcal{H}_k$ is dense. Therefore, it cannot be countable by the assumption that $C(H)$ is not separable. It follows that for some $k$, $\mathcal{H}_k$ is uncountable. Now by Theorem 10.3.3 there exists a continuous function $\theta$ defined on a compact subset, $K$ of $[0, 1]$ which maps $K$ onto $H$. Now consider the functions defined on $K$

$$\mathcal{G}_k \equiv \{ f \circ \theta : f \in \mathcal{H}_k \}.$$

Then $\mathcal{G}_k$ is an uncountable set of continuous functions defined on $K$ with the property that the distance between any two of them is at least as large as $1/k$. This contradicts separability of $C(K)$ which follows from the Weierstrass approximation theorem in which the separable countable set of functions is the restrictions of polynomials that involve only rational coefficients. ■

Now here is Prokhorov’s theorem.

Theorem 10.3.5 Let $\Lambda = \{\mu_n\}_{n=1}^\infty$ be a sequence of probability measures defined on $\mathcal{B}(E)$ where $E$ is a separable complete metric space. If $\Lambda$ is tight then there exists a probability measure, $\lambda$ and a subsequence of $\{\mu_n\}_{n=1}^\infty$, still denoted by $\{\mu_n\}_{n=1}^\infty$ such that whenever $\phi$ is a continuous bounded complex valued function defined on $E$,

$$\lim_{n \to \infty} \int \phi d\mu_n = \int \phi d\lambda.$$

Proof: By tightness, there exists an increasing sequence of compact sets, $\{K_n\}$ such that

$$\mu(K_n) > 1 - \frac{1}{n}$$

for all $\mu \in \Lambda$. Now letting $\mu \in \Lambda$ and $\phi \in C(K_n)$ such that $|\|\phi\|_\infty \leq 1$, it follows

$$\left| \int_{K_n} \phi d\mu \right| \leq \mu(K_n) \leq 1$$

and so the restrictions of the measures of $\Lambda$ to $K_n$ are contained in the unit ball of $C(K_n)^\prime$. Recall from the Riesz representation theorem, the dual space of $C(K_n)$ is a space of complex Borel measures. Theorem 4.3.2 on Page 66 implies the unit ball of $C(K_n)^\prime$ is weak * sequentially compact. This follows from the observation that $C(K_n)$ is separable which is proved in Corollary 10.3.4 and leads to the fact that the unit ball in $C(K_n)^\prime$ is actually metrizable by Theorem 4.3.7 on Page 66. Therefore, there exists a subsequence of $\Lambda$, $\{\mu_{1k}\}$ such that their restrictions to $K_1$ converge weak * to a measure, $\lambda_1 \in C(K_1)^\prime$. That is, for every $\phi \in C(K_1)$,

$$\lim_{k \to \infty} \int_{K_1} \phi d\mu_{1k} = \int_{K_1} \phi d\lambda_1.$$

By the same reasoning, there exists a further subsequence $\{\mu_{2k}\}$ such that the restrictions of these measures to $K_2$ converge weak * to a measure $\lambda_2 \in C(K_2)^\prime$ etc. Continuing this way,

$$\mu_{11}, \mu_{12}, \mu_{13}, \cdots \to \text{ Weak * in } C(K_1)^\prime$$

$$\mu_{21}, \mu_{22}, \mu_{23}, \cdots \to \text{ Weak * in } C(K_2)^\prime$$

$$\mu_{31}, \mu_{32}, \mu_{33}, \cdots \to \text{ Weak * in } C(K_3)^\prime$$

$$\vdots$$

Here the $j^{th}$ sequence is a subsequence of the $(j - 1)^{th}$. Let $\lambda_n$ denote the measure in $C(K_n)^\prime$ to which the sequence $\{\mu_{nk}\}_{k=1}^\infty$ converges weak*. Let $\{\mu_n\} \equiv \{\mu_{nn}\}$, the diagonal sequence. Thus this sequence is ultimately a subsequence of every one of the above sequences and so $\mu_n$ converges weak* in $C(K_m)^\prime$ to $\lambda_m$ for each $m$. Note that this is all happening on different sets so there is no contradiction with something converging to two different things.
Claim: For $p > n$, the restriction of $\lambda_p$ to the Borel sets of $K_n$ equals $\lambda_n$.

Proof of claim: Let $H$ be a compact subset of $K_n$. Then there are sets, $V_l$ open in $K_n$ which are decreasing and whose intersection equals $H$. This follows because this is a metric space. Then let $H \prec \phi_l \prec V_l$. It follows

$$\lambda_n (V_l) \geq \int_{K_n} \phi_l d\lambda_n = \lim_{k \to \infty} \int_{K_n} \phi_l d\mu_k$$

$$= \lim_{k \to \infty} \int_{K_p} \phi_l d\mu_k = \int_{K_p} \phi_l d\lambda_p \geq \lambda_p (H).$$

Now considering the ends of this inequality, let $l \to \infty$ and pass to the limit to conclude

$$\lambda_n (H) \geq \lambda_p (H).$$

Similarly,

$$\lambda_n (H) \leq \int_{K_n} \phi_l d\lambda_n = \lim_{k \to \infty} \int_{K_n} \phi_l d\mu_k$$

$$= \lim_{k \to \infty} \int_{K_p} \phi_l d\mu_k = \int_{K_p} \phi_l d\lambda_p \leq \lambda_p (V_l).$$

Then passing to the limit as $l \to \infty$, it follows

$$\lambda_n (H) \leq \lambda_p (H).$$

Thus the restriction of $\lambda_p, \lambda_p|_{K_n}$ to the compact sets of $K_n$ equals $\lambda_n$. Then by inner regularity it follows the two measures, $\lambda_p|_{K_n}$, and $\lambda_n$ are equal on all Borel sets of $K_n$. Recall that for finite measures on separable metric spaces, regularity is obtained for free.

It is fairly routine to exploit regularity of the measures to verify that $\lambda_m (F) \geq 0$ for all $F$ a Borel subset of $K_m$. Note that $\phi \to \int_{K_n} \phi d\lambda_n$ is a positive linear functional and so $\lambda_n \geq 0$. Also, letting $\phi \equiv 1$,

$$1 \geq \lambda_m (K_m) \geq 1 - \frac{1}{m} \quad (10.3.8)$$

Define for $F$ a Borel set,

$$\lambda (F) \equiv \lim_{n \to \infty} \lambda_n (F \cap K_n).$$

The limit exists because the sequence on the right is increasing due to the above observation that $\lambda_n = \lambda_m$ on the Borel subsets of $K_m$ whenever $n > m$. Thus for $n > m$

$$\lambda_n (F \cap K_n) \geq \lambda_n (F \cap K_m) = \lambda_m (F \cap K_m).$$

Now let $\{F_k\}$ be a sequence of disjoint Borel sets. Then

$$\lambda (\bigcup_{k=1}^{\infty} F_k) \equiv \lim_{n \to \infty} \lambda_n (\bigcup_{k=1}^{\infty} F_k \cap K_n) = \lim_{n \to \infty} \lambda_n (\bigcup_{k=1}^{\infty} (F_k \cap K_n))$$

$$= \lim_{n \to \infty} \sum_{k=1}^{\infty} \lambda_n (F_k \cap K_n) = \sum_{k=1}^{\infty} \lambda_n (F_k)$$

the last equation holding by the monotone convergence theorem.

It remains to verify

$$\lim_{k \to \infty} \int \phi d\mu_k = \int \phi d\lambda$$

for every $\phi$ bounded and continuous. This is where tightness is used again. Then as noted above,

$$\lambda_n (K_n) = \lambda (K_n)$$

because for $p > n, \lambda_p (K_n) = \lambda_n (K_n)$ and so letting $p \to \infty$, the above is obtained. Also, from [LUMA],

$$\lambda (K_n^C) = \lim_{p \to \infty} \lambda_p (K_n^C \cap K_p)$$

$$\leq \lim sup_{p \to \infty} (\lambda_p (K_p) - \lambda_p (K_n))$$

$$\leq \lim sup_{p \to \infty} (\lambda_p (K_p) - \lambda_n (K_n))$$

$$\leq \lim sup_{p \to \infty} \left( 1 - \left( 1 - \frac{1}{n} \right) \right) = \frac{1}{n}.$$
10.4 A Major Existence And Convergence Theorem

Suppose $||\phi||_\infty < M$. Then

$$\left| \int \phi d\mu_k - \int \phi d\lambda \right| \leq \left| \int_{K_n^c} \phi d\mu_k + \int_{K_n^c} \phi d\lambda \left( \int_{K_n^c} \phi d\lambda + \int_{K_n^c} \phi d\lambda \right) \right|$$

$$\leq \left| \int_{K_n^c} \phi d\mu_k - \int_{K_n^c} \phi d\lambda \right| + \left| \int_{K_n^c} \phi d\mu_k - \int_{K_n^c} \phi d\lambda \right|$$

$$\leq \left| \int_{K_n^c} \phi d\mu_k - \int_{K_n^c} \phi d\lambda \right| + \left| \int_{K_n^c} \phi d\mu_k \right| + \left| \int_{K_n^c} \phi d\lambda \right|$$

$$\leq \left| \int_{K_n^c} \phi d\mu_k - \int_{K_n^c} \phi d\lambda \right| + \frac{M}{n} + \frac{M}{n}$$

First let $n$ be so large that $2M/n < \varepsilon/2$ and then pick $k$ large enough that the above expression is less than $\varepsilon$. ■

**Definition 10.3.6** Let $E$ be a complete separable metric space and let $\mu$ and the sequence of probability measures, $\{\mu_n\}$ defined on $B(E)$ satisfy

$$\lim_{n \to \infty} \int \phi d\mu_n = \int \phi d\mu.$$ 

for every $\phi$ a bounded continuous function. Then $\mu_n$ is said to converge weakly to $\mu$.

10.4 A Major Existence And Convergence Theorem

This section is on Skorokhod’s theorem. The proof given follows the presentation in [27]. This is an incredible result. Here is an interesting lemma about weak convergence.

**Lemma 10.4.1** Let $\mu_n$ converge weakly to $\mu$ and let $U$ be an open set with $\mu(\partial U) = 0$. Then

$$\lim_{n \to \infty} \mu_n(U) = \mu(U).$$

**Proof:** Let $\{\psi_k\}$ be a sequence of bounded continuous functions which decrease to $\chi_{\overline{U}}$. Also let $\{\phi_k\}$ be a sequence of bounded continuous functions which increase to $\chi_{\overline{U}}$. For example, you could let

$$\psi_k(x) \equiv (1 - k \text{ dist } (x, U))^+,$$

$$\phi_k(x) \equiv 1 - (1 - k \text{ dist } (x, U^c))^+.$$ 

Let $\varepsilon > 0$ be given. Then since $\mu(\partial U) = 0$, the dominated convergence theorem implies there exists $\psi = \psi_k$ and $\phi = \phi_k$ such that

$$\varepsilon > \int \psi d\mu - \int \phi d\mu.$$

Next use the weak convergence to pick $N$ large enough that if $n \geq N$,

$$\int \psi d\mu_n \leq \int \psi d\mu + \varepsilon, \int \phi d\mu_n \geq \int \phi d\mu - \varepsilon.$$

Therefore, for $n$ this large,

$$\mu(U), \mu_n(U) \in \left[ \int \phi d\mu - \varepsilon, \int \psi d\mu + \varepsilon \right]$$

and so

$$|\mu(U) - \mu_n(U)| < 3\varepsilon.$$ 

since $\varepsilon$ is arbitrary, this proves the lemma.
Definition 10.4.2 Let \((\Omega, \mathcal{F}, P)\) be a probability space and let \(X : \Omega \to E\) be a random variable where here \(E\) is some topological space. Then one can define a probability measure, \(\lambda_X\) on \(\mathcal{B}(E)\) as follows:

\[
\lambda_X(F) \equiv P([X \in F])
\]

More generally, if \(\mu\) is a probability measure on \(\mathcal{B}(E)\), and \(X\) is a random variable defined on a probability space, \(\mathcal{L}(X) = \mu\) means

\[
\mu(F) \equiv P([X \in F]) .
\]

The following amazing theorem is due to Skorokhod. It starts with a measure, \(\mu\) on \(\mathcal{B}(E)\) and produces a random variable, \(X\) for which \(\mathcal{L}(X) = \mu\). It also has something to say about the convergence of a sequence of such random variables.

**Theorem 10.4.3** Let \(E\) be a separable complete metric space and let \(\{\mu_n\}\) be a sequence of Borel probability measures defined on \(\mathcal{B}(E)\) such that \(\mu_n\) converges weakly to \(\mu\) another probability measure on \(\mathcal{B}(E)\). Then there exist random variables, \(X_n, X\) defined on the probability space, \(((0,1), \mathcal{B}([0,1]), m)\) where \(m\) is one dimensional Lebesgue measure such that

\[
\mathcal{L}(X) = \mu, \quad \mathcal{L}(X_n) = \mu_n,
\]

(10.4.9)

each random variable, \(X, X_n\) is continuous off a set of measure zero, and

\[
X_n(\omega) \to X(\omega) \text{ m . a . e .}
\]

**Proof:** Let \(\{a_k\}\) be a countable dense subset of \(E\).

**Construction of sets in \(E\)**

First I will describe a construction. Letting \(C \in \mathcal{B}(E)\) and \(r > 0\),

\[
C^r_1 \equiv C \cap B(a_1, r), \quad C^r_2 \equiv B(a_2, r) \cap C \setminus C^r_1, \ldots
\]

\[
C^r_n \equiv B(a_n, r) \cap C \setminus \left(\bigcup_{k=1}^{n-1} C^r_k\right).
\]

Thus the sets, \(C^r_k\) for \(k = 1, 2, \cdots\) are disjoint Borel sets whose union is all of \(C\). Of course many may be empty.

\[
C_r^{\text{size}(C)}(\text{index of the \(a_k\) it is close to})
\]

Now let \(C = E\), the whole metric space. Also let \(\{r_k\}\) be a decreasing sequence of positive numbers which converges to 0. Let

\[
A_k \equiv E^r_k, \quad k = 1, 2, \cdots
\]

Thus \(\{A_k\}\) is a sequence of Borel sets, \(A_k \subseteq B(a_k, r_1)\), and the union of the \(A_k\) equals \(E\). For \((i_1, \cdots, i_m) \in \mathbb{N}^m\), suppose \(A_{i_1, \cdots, i_m}\) has been defined. Then for \(k \in \mathbb{N}\),

\[
A_{i_1, \cdots, i_{m_k}} \equiv (A_{i_1, \cdots, i_m})_{k}^{r_{m+1}}
\]

Thus \(A_{i_1, \cdots, i_{m_k}} \subseteq B(a_k, r_{m+1})\), is a Borel set, and

\[
\bigcup_{k=1}^{\infty} A_{i_1, \cdots, i_{m_k}} = A_{i_1, \cdots, i_m}.
\]

(10.4.10)

Also note that \(A_{i_1, \cdots, i_m}\) could be empty. This is because \(A_{i_1, \cdots, i_{m_k}} \subseteq B(a_k, r_{m+1})\) but \(A_{i_1, \cdots, i_m} \subseteq B(a_{i_m}, r_m)\) which might have empty intersection with \(B(a_k, r_{m+1})\). However, applying 10.4.10 repeatedly,

\[
E = \bigcup_{i_1} \cdots \bigcup_{i_m} A_{i_1, \cdots, i_m}
\]

and also, the construction shows the Borel sets, \(A_{i_1, \cdots, i_m}\) are disjoint. Note that to get \(A_{i_1, \cdots, i_{m_k}}\), you do to \(A_{i_1, \cdots, i_m}\) what was done for \(E\) but you consider smaller sized pieces.

**Construction of intervals depending on the measure**
Next I will construct intervals, $I_{i_1,\ldots,i_m}^\nu$ in $[0,1)$ corresponding to these $A_{i_1,\ldots,i_m}$. In what follows, $\nu = \mu_n$ or $\mu$. These intervals will depend on the measure chosen as indicated in the notation.

$$I_i^\nu = [0, \nu(A_i)), \ldots, I_j^\nu = \left[\sum_{k=1}^{j-1} \nu(A_k), \sum_{k=1}^{j} \nu(A_k)\right]$$

for $j = 1, 2, \cdots$. Note these are disjoint intervals whose union is $[0,1)$. Also note

$$m(I_j^\nu) = \nu(A_j).$$

The endpoints of these intervals as well as their lengths depend on the measures of the sets $A_k$. Now supposing $I_{i_1,\ldots,i_m} = [\alpha, \beta]$ where $\beta - \alpha = \nu(A_{i_1,\ldots,i_m})$, define

$$I_{i_1,\ldots,i_m,j}^\nu = \left[\alpha + \sum_{k=1}^{j-1} \nu(A_{i_1,\ldots,i_m,k}), \alpha + \sum_{k=1}^{j} \nu(A_{i_1,\ldots,i_m,k})\right].$$

Thus $m(I_{i_1,\ldots,i_m,j}^\nu) = \nu(A_{i_1,\ldots,i_m,j})$ and

$$\nu(A_{i_1,\ldots,i_m}) = \sum_{k=1}^{\infty} \nu(A_{i_1,\ldots,i_m,k}) = \sum_{k=1}^{\infty} m(I_{i_1,\ldots,i_m,k}) = \beta - \alpha,$$

the intervals, $I_{i_1,\ldots,i_m,j}^\nu$ being disjoint and

$$I_{i_1,\ldots,i_m}^\nu = \bigcup_{j=1}^{\infty} I_{i_1,\ldots,i_m,j}^\nu.$$ These intervals satisfy the same inclusion properties as the sets $\{A_{i_1,\ldots,i_m}\}$. They are just on $[0,1)$ rather than on $E$. The intervals $I_{i_1,\ldots,i_m,j}^\nu$ correspond to the sets $A_{i_1,\ldots,i_m,j}$ and in fact the Lebesgue measure of the interval is the same as $\nu(A_{i_1,\ldots,i_m,j})$.

### Choosing the sequence $\{r_k\}$ in an auspicious manner

There are at most countably many positive numbers, $r$ such that for $\nu = \mu_n$ or $\mu, \nu(\partial B(a_i, r)) > 0$. This is because $\nu$ is a finite measure. Taking the countable union of these countable sets, there are only countably many $r$ such that $\nu(\partial B(a_i, r)) > 0$ for some $a_i$. Let the sequence avoid all these bad values of $r$. Thus for

$$F = \bigcup_{m=1}^{\infty} \bigcup_{k=1}^{\infty} \partial B(a_k, r_m)$$

and $\nu = \mu$ or $\mu_n, \nu(F) = 0$. Here the $r_m$ are all good values such that for all $k, m, \partial B(a_k, r_m)$ has $\mu$ measure zero and $\mu_n$ measure zero. The next claim is illustrated in the following picture. In the picture, $A$ represents one of those $A_{i_1,\ldots,i_m}$ and $A_1$ and $A_2$ are two of the sets $A_{i_1,\ldots,i_m,k}$ which partition $A_{i_1,\ldots,i_m}$. 

![Diagram](image)

**Claim 1:** $\partial A_{i_1,\ldots,i_k} \subseteq F$. This really follows from the construction. However, the details follow.

**Proof of claim:** Suppose $C$ is a Borel set for which $\partial C \subseteq F$. I need to show $\partial C^{r_i}_k \in F$. First consider $k = 1$. Then $C^{r_i}_1 = B(a_1, r_1) \cap C$. If $x \in \partial C^{r_i}_1$, then $B(x, \delta)$ contains points of $B(a_1, r_1) \cap C$ and points of $B(a_1, r_1) \cap C^C$ for every $\delta > 0$. First suppose $x \in B(a_1, r_1)$. Then a small enough neighborhood of $x$ has no points of $B(a_1, r_1) \cap C$ and so every $B(x, \delta)$ has points of $C$ and points of $C^C$ so that $x \in \partial C \subseteq F$ by assumption. If $x \in \partial C^{r_i}_1$, then it can’t happen that $|x - a_i| > r_i$ because then there would be a neighborhood of $x$ having no points of $C^{r_i}_1$. The only other case to consider is that $|x - a_i| = r_i$ but this says $x \in F$. Now assume $\partial C^{r_i}_j \subseteq F$ for $j \leq k - 1$ and consider $\partial C^{r_i}_k$.

$$C^{r_i}_k = B(a_k, r_i) \cap C \setminus \bigcup_{j=1}^{k-1} C^{r_i}_j = B(a_k, r_i) \cap C \cap \left(\bigcap_{j=1}^{k-1} (C^{r_i}_j)^C\right)$$

(10.4.11)
Consider \( x \in \partial C^r_s \). If \( x \in \text{int} \left( (B(a_k, r_i) \cap C) \right) \) (int \( \equiv \text{interior} \)) then a small enough ball about \( x \) contains no points of \( (B(a_k, r_i) \cap C)^C \) and so every ball about \( x \) must contain points of
\[
\left( \bigcap_{j=1}^{k-1} (C^r_j)^C \right)^C = \bigcup_{j=1}^{k-1} C^r_j
\]
Since there are only finitely many sets in the union, there exists \( s \leq k-1 \) such that every ball about \( x \) contains points of \( C^r_s \) but from \( \partial C^r_k \) every ball about \( x \) contains points of \( (C^r_s)^C \) which implies \( x \in \partial C^r_s \subseteq F \) by induction. It is not possible that \( d(x, a_k) > r_i \) and yet have \( x \in \partial C^r_k \). This follows from the description in \( \partial C^r_k \). If \( d(x, a_k) = r_i \) then by definition, \( x \in F \). The only other case to consider is that \( x \notin \text{int} \left( (B(a_k, r_i) \cap C) \right) \) but \( x \in \text{int} (B(a_k, r_i)) \). From \( \partial C^r_k \) every ball about \( x \) contains points of \( C \). However, since \( x \in B(a_k, r_i) \), a small enough ball is contained in \( B(a_k, r_i) \). Therefore, every ball about \( x \) must also contain points of \( C \) since otherwise, \( x \in \text{int} (B(a_k, r_i) \cap C) \). Thus \( x \in \partial C \subseteq F \) by assumption. Now apply what was just shown to the case where \( C = E \), the whole space. In this case, \( \partial E \subseteq F \) because \( \partial E = \emptyset \). Then keep applying what was just shown to the \( A_{i_1, \ldots, i_n} \). This proves the claim.

From the claim, \( \nu \left( \text{int} \left( A_{i_1, \ldots, i_n} \right) \right) = \nu \left( A_{i_1, \ldots, i_n} \right) \) whenever \( \nu = \mu \) or \( \mu_n \). This is because that in \( A_{i_1, \ldots, i_n} \) which is not in \( \text{int} \left( A_{i_1, \ldots, i_n} \right) \) is in \( F \) which has measure zero.

**Some functions on \([0, 1]\)**

By the axiom of choice, there exists \( x_{i_1, \ldots, i_m} \in \text{int} \left( A_{i_1, \ldots, i_m} \right) \) whenever
\[
\text{int} \left( A_{i_1, \ldots, i_m} \right) \neq \emptyset.
\]
For \( \nu = \mu_n \) or \( \mu \), define the following functions. For \( \omega \in I^\nu_{i_1, \ldots, i_m} \),
\[
Z^\nu_m(\omega) \equiv x_{i_1, \ldots, i_m}.
\]
This defines the functions, \( Z^\nu_m \) and \( Z^\mu_m \). Note these functions have the same values but on slightly different intervals. Here is an important claim.

**Claim 2 (Limit on \( \mu_n \)):** For a.e. \( \omega \in [0, 1] \), \( \lim_{n \to \infty} Z^\nu_m(\omega) = Z^\mu_m(\omega) \).

**Proof of the claim:** This follows from the weak convergence of \( \mu_n \) to \( \mu \) and Lemma 2. This lemma implies \( \mu_n \left( \text{int} \left( A_{i_1, \ldots, i_m} \right) \right) \to \mu \left( \text{int} \left( A_{i_1, \ldots, i_m} \right) \right) \). Thus by the construction described above, \( \mu_n \left( A_{i_1, \ldots, i_m} \right) \to \mu \left( A_{i_1, \ldots, i_m} \right) \) because of claim 1 and the construction of \( F \) in which it is always a set of measure zero. It follows that if \( \omega \in \text{int} \left( I^\nu_{i_1, \ldots, i_m} \right) \), then for all \( n \) large enough, \( \omega \in \text{int} \left( I^\mu_{i_1, \ldots, i_m} \right) \) and so \( Z^\nu_m(\omega) = Z^\mu_m(\omega) \). Note this convergence is very far from being uniform.

**Claim 3 (Limit on size of sets, fixed measure):** For \( \nu = \mu_n \) or \( \mu, \{Z^\nu_m \}_{m=1}^\infty \) is uniformly Cauchy independent of \( \nu \).

**Proof of the claim:** For \( \omega \in I^\nu_{i_1, \ldots, i_m} \), then by the construction, \( \omega \in I^\nu_{i_1, \ldots, i_m, i_{m+1}, \ldots, i_n} \) for some \( i_{m+1}, \ldots, i_n \). Therefore, \( Z^\nu_m(\omega) \) and \( Z^\nu_n(\omega) \) are both contained in \( A_{i_1, \ldots, i_m} \) which is contained in \( B(a_{i_m}, r_{m}) \). Since \( \omega \in [0, 1] \) was arbitrary, and \( r_m \to 0 \), it follows these functions are uniformly Cauchy as claimed.

Let \( X^\nu(\omega) = \lim_{m \to \infty} Z^\nu_m(\omega) \). Since each \( Z^\nu_m \) is continuous off a set of measure zero, it follows from the uniform convergence that \( X^\nu \) is also continuous off a set of measure zero.

**Claim 4:** For a.e. \( \omega \),
\[
\lim_{n \to \infty} X^\mu_n(\omega) = X^\mu(\omega).
\]

**Proof of the claim:** From Claim 3 and letting \( \varepsilon > 0 \) be given, there exists \( m \) large enough that for all \( n \),
\[
\sup_{\omega} d(\mu_n(\omega), X^\mu_n(\omega)) < \varepsilon/3, \quad \sup_{\omega} d(\mu(\omega), X^\mu(\omega)) < \varepsilon/3.
\]
for \( \omega \) off a set of measure zero. Now pick \( \omega \in [0, 1] \) such that \( \omega \) is not equal to any of the end points of any of the intervals, \( \{I^\nu_{i_1, \ldots, i_m}\} \), this countable set of endpoints, a set of Lebesgue measure zero. Then by Claim 2, there exists \( N \) such that if \( n \geq N \), then \( d(\mu_n(\omega), \mu(\omega)) < \varepsilon/3 \). Therefore, for such \( n \) and this \( \omega \),
\[
d(\mu_n(\omega), \mu(\omega)) \leq d(\mu_n(\omega), Z^\mu_m(\omega)) + d(\mu(\omega), Z^\mu_m(\omega)) + d(Z^\mu_m(\omega), X^\mu(\omega)) \leq \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.
\]
This proves the claim.

**Showing \( \mathcal{L}(X^\nu) = \nu \).**
This has mostly proved the theorem except for the claim that $L(X^\nu) = \nu$ for $\nu = \mu_n$ and $\mu$. To do this, I will first show $m \left( (X^\nu)^{-1} (\partial A_{i_1, \ldots, i_m}) \right) = 0$. By the construction, $\nu(\partial A_{i_1, \ldots, i_m}) = 0$. Let $\varepsilon > 0$ be given and let $\delta > 0$ be small enough that

$$H_\delta \equiv \{ x \in E : \text{dist}(x, \partial A_{i_1, \ldots, i_m}) \leq \delta \}$$

is a set of measure less than $\varepsilon/2$. Denote by $G_k$ the sets of the form $A_{i_1, \ldots, i_k}$ where $(i_1, \ldots, i_k) \in \mathbb{N}^k$. Recall also that corresponding to $A_{i_1, \ldots, i_k}$ is an interval, $I^\nu_k$ having length equal to $\nu(A_{i_1, \ldots, i_k})$. Denote by $B_k$ those sets of $G_k$ which have nonempty intersection with $H_\delta$ and let the corresponding intervals be denoted by $I^\nu_k$. If $\omega \notin \cup I^\nu_k$, then from the construction, $Z^\nu_p(\omega)$ is at a distance of at least $\delta$ from $\partial A_{i_1, \ldots, i_m}$ for all $p \geq k$. (If $Z^\nu_p(\omega)$ were in some set of $B_k$, this would require $\omega$ to be in the corresponding $I^\nu_k$ and it is assumed this does not happen. Then for any $p > k$, $Z^\nu_p(\omega)$ cannot be in any set of $G_p$ which intersects $H_\delta$ either. If it did, you would need to have $\omega \notin \cup I^\nu_p$ but all of these intervals are inside the intervals $I^\nu_k$.) Passing to the limit as $p \to \infty$, it follows $X^\nu(\omega) \notin \partial A_{i_1, \ldots, i_m}$. Therefore,

$$(X^\nu)^{-1} (\partial A_{i_1, \ldots, i_m}) \subseteq \cup I^\nu_k$$

Recall that $A_{i_1, \ldots, i_k} \subseteq B(a_i, r_k)$ and the $r_k \to 0$. Therefore, if $k$ is large enough,

$$\nu(\cup B_k) \leq \varepsilon$$

because $\cup B_k$ approximates $H_\delta$ closely (In fact, $\cap_{k=1}^{\infty} (\cup B_k) = H_\delta$). Therefore,

$$m \left( (X^\nu)^{-1} (\partial A_{i_1, \ldots, i_m}) \right) \leq m \left( \cup I^\nu_k \right)$$

$$= \sum_{I^\nu_k \in T^\nu} m \left( I^\nu_k \right)$$

$$= \sum_{A_{i_1, \ldots, i_k} \in B_k} \nu(A_{i_1, \ldots, i_k})$$

$$= \nu(\cup B_k) \leq \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this shows $m \left( (X^\nu)^{-1} (\partial A_{i_1, \ldots, i_m}) \right) = 0$.

If $\omega \in I^\nu_p$, then from the construction, $Z^\nu_p(\omega) \in \text{int}(A_{i_1, \ldots, i_m})$ for all $p \geq k$. Therefore, taking a limit, as $p \to \infty$,

$$X^\nu(\omega) \in \text{int}(A_{i_1, \ldots, i_m}) \cup \partial A_{i_1, \ldots, i_m}$$

and so

$$I^\nu_{i_1, \ldots, i_m} \subseteq (X^\nu)^{-1} \left( \text{int}(A_{i_1, \ldots, i_m}) \cup \partial A_{i_1, \ldots, i_m} \right)$$

but also, if $X^\nu(\omega) \in \text{int}(A_{i_1, \ldots, i_m})$, then $Z^\nu_p(\omega) \in \text{int}(A_{i_1, \ldots, i_m})$ for all $p$ large enough and so

$$(X^\nu)^{-1} \left( \text{int}(A_{i_1, \ldots, i_m}) \right)$$

$$\subseteq I^\nu_{i_1, \ldots, i_m} \subseteq (X^\nu)^{-1} \left( \text{int}(A_{i_1, \ldots, i_m}) \cup \partial A_{i_1, \ldots, i_m} \right)$$

Therefore,

$$m \left( (X^\nu)^{-1} \left( \text{int}(A_{i_1, \ldots, i_m}) \right) \right)$$

$$\leq m \left( I^\nu_{i_1, \ldots, i_m} \right)$$

$$\leq m \left( (X^\nu)^{-1} \left( \text{int}(A_{i_1, \ldots, i_m}) \right) \right) + m \left( (X^\nu)^{-1} (\partial A_{i_1, \ldots, i_m}) \right)$$

$$= m \left( (X^\nu)^{-1} (\partial A_{i_1, \ldots, i_m}) \right)$$

which shows

$$m \left( (X^\nu)^{-1} (\text{int}(A_{i_1, \ldots, i_m})) \right) = m \left( I^\nu_{i_1, \ldots, i_m} \right) = \nu(A_{i_1, \ldots, i_m}).$$

(10.4.12)

Also

$$m \left( (X^\nu)^{-1} (\text{int}(A_{i_1, \ldots, i_m})) \right)$$

$$\leq m \left( (X^\nu)^{-1} (A_{i_1, \ldots, i_m}) \right)$$

$$\leq m \left( (X^\nu)^{-1} \left( \text{int}(A_{i_1, \ldots, i_m}) \cup \partial A_{i_1, \ldots, i_m} \right) \right)$$

$$= m \left( (X^\nu)^{-1} (\text{int}(A_{i_1, \ldots, i_m})) \right)$$
Hence from \(10.4.12\),
\[
\nu(A_{i_1, \ldots, i_m}) = m\left((X^\nu)^{-1}(\text{int}(A_{i_1, \ldots, i_m}))\right)
= m\left((X^\nu)^{-1}(A_{i_1, \ldots, i_m})\right) \tag{10.4.13}
\]

Now let \(U\) be an open set in \(E\). Then letting
\[
H_k = \{x \in U : \text{dist}(x, U^C) \geq r_k\}
\]
it follows
\[
\bigcup_k H_k = U.
\]
Next consider the sets of \(G_k\) which have nonempty intersection with \(H_k, H_k\). Then \(H_k\) is covered by \(H_k\) and every set of \(H_k\) is contained in \(U\), the sets of \(H_k\) also being disjoint. Then from \(10.4.12\),
\[
m\left((X^\nu)^{-1}(\bigcup H_k)\right) = \sum_{A \in H_k} m\left((X^\nu)^{-1}(A)\right)
= \sum_{A \in H_k} \nu(A) = \nu(\bigcup H_k).
\]
Therefore, letting \(k \to \infty\) and passing to the limit in the above,
\[
m\left((X^\nu)^{-1}(U)\right) = \nu(U).
\]
Since this holds for every open set, it is routine to verify using regularity that it holds for every Borel set and so \(\mathcal{L}(X^\nu) = \nu\) as claimed. \(\blacksquare\)
Chapter 11

Characteristic Functions

11.1 The Characteristic Function

One of the most important tools in probability is the characteristic function. To begin with, assume the random variables have values in \( \mathbb{R}^p \).

**Definition 11.1.1** Let \( X \) be a random variable as above. The characteristic function is

\[
\phi_X(t) \equiv E(e^{it \cdot X}) \equiv \int_{\Omega} e^{it \cdot X(\omega)} dP = \int_{\mathbb{R}^p} e^{it \cdot x} d\lambda_X
\]

the last equation holding by Proposition [10.1.11].

**Definition 11.1.2** Define \( G_1 \) to be the functions of the form \( p(x) e^{-a|x|^2} \) where \( a > 0 \) is some positive number and \( p(x) \) is a polynomial. Let \( G \) be all finite sums of functions in \( G_1 \). Thus \( G \) is an algebra of functions which has the property that if \( f \in G \) then \( \bar{f} \) and \( f \) are in the algebraic dual of \( G \).

Then the Fourier transform is defined on things in \( G^* \), the algebraic dual of \( G \) as follows.

**Definition 11.1.3** For \( T \in G^* \), define \( FT, F^{-1}T \in G^* \) by

\[
FT(\phi) \equiv T(F\phi) , \quad F^{-1}T(\phi) \equiv T(F^{-1}\phi)
\]

**Lemma 11.1.4** \( F \) and \( F^{-1} \) are both one to one, onto, and are inverses of each other.

The above observations are proved in Chapter 3. The main result on characteristic functions is the following.

**Theorem 11.1.5** Let \( X \) and \( Y \) be random vectors with values in \( \mathbb{R}^p \) and suppose \( E(e^{it \cdot X}) = E(e^{it \cdot Y}) \) for all \( t \in \mathbb{R}^p \). Then \( \lambda_X = \lambda_Y \).

**Proof:** For \( \psi \in G \), let \( \lambda_X(\psi) \equiv \int_{\mathbb{R}^p} \psi d\lambda_X \) and \( \lambda_Y(\psi) \equiv \int_{\mathbb{R}^p} \psi d\lambda_Y \). Thus both \( \lambda_X \) and \( \lambda_Y \) are in \( G^* \). Then letting \( \psi \in G \) and using Fubini’s theorem,

\[
\int_{\mathbb{R}^p} \int_{\mathbb{R}^p} e^{it \cdot Y} \psi(t) dtd\lambda_Y = \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} e^{it \cdot Y} d\lambda_Y \psi(t) dt
\]

\[
= \int_{\mathbb{R}^p} E(e^{it \cdot Y}) \psi(t) dt
\]

\[
= \int_{\mathbb{R}^p} E(e^{it \cdot X}) \psi(t) dt
\]

\[
= \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} e^{it \cdot X} d\lambda_X \psi(t) dt
\]

Thus \( \lambda_Y(F^{-1}\psi) = \lambda_X(F^{-1}\psi) \). Since \( \psi \in G \) is arbitrary and \( F^{-1} \) is onto, this implies \( \lambda_X = \lambda_Y \) in \( G^* \). But \( G \) is dense in \( C_0(\mathbb{R}^p) \) from the Stone Weierstrass theorem and so \( \lambda_X = \lambda_Y \) as measures. Recall from real analysis the dual space of \( C_0(\mathbb{R}^n) \) is the space of complex measures.

179
CHAPTER 11. CHARACTERISTIC FUNCTIONS

Alternatively, the above shows that since \( F^{-1} \) is onto, for all \( \psi \in \mathcal{G} \),
\[
\int_{\mathbb{R}^p} \psi d\lambda_Y = \int_{\mathbb{R}^p} \psi d\lambda_X
\]
and then, by a use of the Stone Weierstrass theorem, the above will hold for all \( \psi \in C_c(\mathbb{R}^n) \) and now, by the Riesz representation theorem for positive linear functionals, the two measures are equal. ■

You can also give a version of this theorem in which reference is made only to the probability distribution measures.

Definition 11.1.6 For \( \mu \) a probability measure on the Borel sets of \( \mathbb{R}^n \),
\[
\phi_\mu(t) \equiv \int_{\mathbb{R}^n} e^{ix \cdot t} d\mu.
\]

Theorem 11.1.7 Let \( \mu \) and \( \nu \) be probability measures on the Borel sets of \( \mathbb{R}^n \) and suppose \( \phi_\mu(t) = \phi_\nu(t) \). Then \( \mu = \nu \).

Proof: The proof is identical to the above. Just replace \( \lambda_X \) with \( \mu \) and \( \lambda_Y \) with \( \nu \). ■

11.2 Conditional Probability

Here I will consider the concept of conditional probability depending on the theory of differentiation of general Radon measures. This leads to a different way of thinking about independence.

If \( X, Y \) are two random vectors defined on a probability space having values in \( \mathbb{R}^{p_1} \) and \( \mathbb{R}^{p_2} \) respectively, and if \( E \) is a Borel set in the appropriate space, then \( (X, Y) \) is a random vector with values in \( \mathbb{R}^{p_1} \times \mathbb{R}^{p_2} \) and \( \lambda_{(X,Y)}(E \times \mathbb{R}^{p_2}) = \lambda_X(E), \; \lambda_{(X,Y)}(\mathbb{R}^{p_1} \times E) = \lambda_Y(E) \). Thus, by Theorem 8.3.3 on Page 122, there exist probability measures, denoted here by \( \lambda_{X|Y} \) and \( \lambda_{Y|X} \), such that whenever \( E \) is a Borel set in \( \mathbb{R}^{p_1} \times \mathbb{R}^{p_2}, \)
\[
\int_{\mathbb{R}^{p_1} \times \mathbb{R}^{p_2}} \chi_E(x,y) d\lambda_{(X,Y)} = \int_{\mathbb{R}^{p_1}} \int_{\mathbb{R}^{p_2}} \chi_E(x,y) d\lambda_{X|Y} d\lambda_X,
\]
and
\[
\int_{\mathbb{R}^{p_1} \times \mathbb{R}^{p_2}} \chi_E(x,y) d\lambda_{(X,Y)} = \int_{\mathbb{R}^{p_2}} \int_{\mathbb{R}^{p_1}} \chi_E(x,y) d\lambda_{Y|X} d\lambda_Y.
\]

Definition 11.2.1 Let \( X \) and \( Y \) be two random vectors defined on a probability space. The conditional probability measure of \( Y \) given \( X \) is the measure \( \lambda_{Y|X} \) in the above. Similarly the conditional probability measure of \( X \) given \( Y \) is the measure \( \lambda_{X|Y} \).

More generally, one can use the theory of slicing measures to consider any finite list of random vectors, \{\( X_i \)\}, defined on a probability space with \( X_i \in \mathbb{R}^{p_i} \), and write the following for \( E \) a Borel set in \( \prod_{i=1}^n \mathbb{R}^{p_i} \).
\[
\int_{\mathbb{R}^{p_1} \times \cdots \times \mathbb{R}^{p_n}} \chi_E(x_1,\ldots,x_n) d\lambda_{(X_1,\ldots,X_n)} = \int_{\mathbb{R}^{p_1} \times \cdots \times \mathbb{R}^{p-n-1}} \int_{\mathbb{R}^{p_n}} \chi_E(x_1,\ldots,x_{n-1}) d\lambda_{(x_1,\ldots,x_{n-1})}
\]
\[
= \int_{\mathbb{R}^{p_1} \times \cdots \times \mathbb{R}^{p-n-2}} \int_{\mathbb{R}^{p-n-1}} \int_{\mathbb{R}^{p_n}} \chi_E(x_1,\ldots,x_{n-1}) d\lambda_{x_{n-1}(x_1,\ldots,x_{n-2})} d\lambda_{(x_1,\ldots,x_{n-2})}
\]
\[
\vdots
\]
\[
\int_{\mathbb{R}^{p_1}} \cdots \int_{\mathbb{R}^{p_n}} \chi_E(x_1,\ldots,x_{n-1}) d\lambda_{x_{n-1}(x_1,\ldots,x_{n-2})} \cdots d\lambda_{x_2|x_1} d\lambda_{x_1}. \tag{11.2.1}\]

Obviously, this could have been done in any order in the iterated integrals by simply modifying the “given” variables, those occurring after the symbol \( \mid \), to be those which have been integrated in an outer level of the iterated integral. For simplicity, write
\[
\lambda_{X_n|(x_1,\ldots,x_{n-1})} = \lambda_{X_n|x_1,\ldots,x_{n-1}}
\]
Definition 11.2.2 Let \( \{X_1, \ldots, X_n\} \) be random vectors defined on a probability space having values in \( \mathbb{R}^p, \ldots, \mathbb{R}^p \) respectively. The random vectors are independent if for every \( E \) a Borel set in \( \mathbb{R}^p \times \cdots \times \mathbb{R}^p \),

\[
\int_{\mathbb{R}^p \times \cdots \times \mathbb{R}^p} \mathcal{X}_E \, d\lambda(x_1, \ldots, x_n)
= \int_{\mathbb{R}^p} \cdots \int_{\mathbb{R}^p} \mathcal{X}_E \, d\lambda_{x_{n-1}} \cdots d\lambda_{x_1} \lambda
\]

and the iterated integration may be taken in any order. If \( A \) is any set of random vectors defined on a probability space, \( A \) is independent if any finite set of random vectors from \( A \) is independent.

Thus, the random vectors are independent exactly when the dependence on the givens in \( 11.2.1 \) can be dropped. Does this amount to the same thing as discussed earlier? Suppose you have three random variables \( X, Y, Z \). Let \( A = X^{-1}(E), B = Y^{-1}(F), C = Z^{-1}(G) \) where \( E, F, G \) are Borel sets. Thus these inverse images are typical sets in \( \sigma(X), \sigma(Y), \sigma(Z) \) respectively. First suppose that the random variables are independent in the earlier sense. Then

\[
P(A \cap B \cap C) = P(A) P(B) P(C)
= \int_{\mathbb{R}^p} \mathcal{X}_E(x) \, d\lambda_X \int_{\mathbb{R}^p} \mathcal{X}_F(y) \, d\lambda_Y \int_{\mathbb{R}^p} \mathcal{X}_G(z) \, d\lambda_Z
= \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \mathcal{X}_E(x) \mathcal{X}_F(y) \mathcal{X}_G(z) \, d\lambda_Z \, d\lambda_Y \, d\lambda_X
\]

Also

\[
P(A \cap B \cap C) = \int_{\mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p} \mathcal{X}_E(x) \mathcal{X}_F(y) \mathcal{X}_G(z) \, d\lambda(x,y,z)
= \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \mathcal{X}_E(x) \mathcal{X}_F(y) \mathcal{X}_G(z) \, d\lambda_Z |_{xy} \, d\lambda_Y |_{x} \, d\lambda_X
\]

Thus

\[
\int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \mathcal{X}_E(x) \mathcal{X}_F(y) \mathcal{X}_G(z) \, d\lambda_Z \, d\lambda_Y \, d\lambda_X
= \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \mathcal{X}_E(x) \mathcal{X}_F(y) \mathcal{X}_G(z) \, d\lambda_Z |_{xy} \, d\lambda_Y |_{x} \, d\lambda_X
\]

Now letting \( G = \mathbb{R}^3 \), it follows that

\[
\int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \mathcal{X}_E(x) \mathcal{X}_F(y) \, d\lambda_Y \, d\lambda_X
= \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \mathcal{X}_E(x) \mathcal{X}_F(y) \, d\lambda_Y |_{x} \, d\lambda_X
\]

By uniqueness of the slicing measures or an application of the Besikovitch differentiation theorem, it follows that for \( \lambda_X \) a.e. \( x \),

\[
\lambda_Y = \lambda_Y |_x
\]

Thus, using this in the above,

\[
\int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \mathcal{X}_E(x) \mathcal{X}_F(y) \mathcal{X}_G(z) \, d\lambda_Z \, d\lambda_Y \, d\lambda_X
= \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \mathcal{X}_E(x) \mathcal{X}_F(y) \mathcal{X}_G(z) \, d\lambda_Z |_{xy} \, d\lambda_Y |_{x} \, d\lambda_X
\]

and also it reduces to

\[
\int_{\mathbb{R}^p \times \mathbb{R}^p} \int_{\mathbb{R}^p} \mathcal{X}_E(x) \mathcal{X}_F(y) \mathcal{X}_G(z) \, d\lambda_Z |_{x,y} \, d\lambda_{x,y}
= \int_{\mathbb{R}^p \times \mathbb{R}^p} \int_{\mathbb{R}^p} \mathcal{X}_E(x) \mathcal{X}_F(y) \mathcal{X}_G(z) \, d\lambda_Z |_{x,y} \, d\lambda_{x,y,}\mathcal{X}_E(x) \mathcal{X}_F(y) \mathcal{X}_G(z) \, d\lambda_Z |_{x,y} \, d\lambda_{x,y}}
\]
Now by uniqueness of the slicing measures again, for \( \lambda_{(X,Y)} \) a.e. \( (x,y) \), it follows that
\[
\lambda_Z = \lambda_{Z|_{XY}}
\]
Similar conclusions hold for \( \lambda_X, \lambda_Y \). In each case, off a set of measure zero the distribution measures equal the slicing measures.

Conversely, if the distribution measures equal the slicing measures off sets of measure zero as described above, then it is obvious that the random variables are independent. The same reasoning applies for any number of random variables.

Thus this gives a different and more analytical way to think of independence of finitely many random variables. Clearly, the argument given above will apply to any finite set of random variables.

**Proposition 11.2.3** Equations \((11.2.2)\) and \((11.2.3)\) hold with \( \lambda_E \) replaced by any nonnegative Borel measurable function and for any bounded continuous function or for any function in \( L^1 \).

**Proof:** The two equations hold for simple functions in place of \( \lambda_E \) and so an application of the monotone convergence theorem applied to an increasing sequence of simple functions converging pointwise to a given nonnegative Borel measurable function yields the conclusion of the proposition in the case of the nonnegative Borel function. For a bounded continuous function or one in \( L^1 \), one can apply the result just established to the positive and negative parts of the real and imaginary parts of the function.

**Lemma 11.2.4** Let \( X_1, \cdots, X_n \) be random vectors with values in \( \mathbb{R}^p_1, \cdots, \mathbb{R}^p_n \) respectively and let \( g : \mathbb{R}^p_1 \times \cdots \times \mathbb{R}^p_n \to \mathbb{R}^k \) be Borel measurable. Then \( g(X_1, \cdots, X_n) \) is a random vector with values in \( \mathbb{R}^k \) and if \( h : \mathbb{R}^k \to [0, \infty) \), then
\[
\int_{\mathbb{R}^k} h(y) \, d\lambda_g(x_1, \cdots, x_n)(y) = \\
\int_{\mathbb{R}^p_1 \times \cdots \times \mathbb{R}^p_n} h\left(g(x_1, \cdots, x_n)\right) \, d\lambda(x_1, \cdots, x_n).
\]
If \( X_i \) is a random vector with values in \( \mathbb{R}^p_i, \) \( i = 1, 2, \cdots \) and if \( g_i : \mathbb{R}^p_i \to \mathbb{R}^k_i \), where \( g_i \) is Borel measurable, then the random vectors \( g_i(X_i) \) are also independent whenever the \( X_i \) are independent.

**Proof:** First let \( E \) be a Borel set in \( \mathbb{R}^k \). From the definition,
\[
\lambda_{g(x_1, \cdots, x_n)}(E) = P(g(X_1, \cdots, X_n) \in E) = P((X_1, \cdots, X_n) \in g^{-1}(E)) = \lambda(x_1, \cdots, x_n)(g^{-1}(E))
\]
Thus
\[
\int_{\mathbb{R}^k} \lambda_E d\lambda_{g(x_1, \cdots, x_n)} = \\
\int_{\mathbb{R}^p_1 \times \cdots \times \mathbb{R}^p_n} \lambda_{g^{-1}(E)} d\lambda(x_1, \cdots, x_n)
\]
\[
= \\
\int_{\mathbb{R}^p_1 \times \cdots \times \mathbb{R}^p_n} \lambda_E(g(x_1, \cdots, x_n)) d\lambda(x_1, \cdots, x_n).
\]
This proves \((11.2.3)\) in the case when \( h = \lambda_E \). To prove it in the general case, approximate the nonnegative Borel measurable function with simple functions for which the formula is true, and use the monotone convergence theorem.

It remains to prove the last assertion that functions of independent random vectors are also independent random vectors. Let \( E \) be a Borel set in \( \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_n} \). Then for
\[
\pi_i(x_1, \cdots, x_n) \equiv x_i,
\]
\[
\int_{\mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_n}} \lambda_E d\lambda_{g_1(x_1) \cdots g_n(x_n)}
\]
\[
\equiv \\
\int_{\mathbb{R}^p_1 \times \cdots \times \mathbb{R}^p_n} \lambda_E \circ (g_1 \circ \pi_1, \cdots, g_n \circ \pi_n) d\lambda(x_1, \cdots, x_n)
\]
\[
= \\
\int_{\mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_n}} \lambda_E d\lambda_{g_1(x_1) \cdots g_n(x_n)}
\]
and this proves the last assertion.
11.3 Characteristic Functions And Independence

**Proposition 11.2.5** Let \( \nu_1, \ldots, \nu_n \) be Radon probability measures defined on \( \mathbb{R}^p \). Then there exists a probability space and independent random vectors \( \{X_1, \ldots, X_n\} \) defined on this probability space such that \( \lambda_{X_i} = \nu_i \).

**Proof:** Let \( (\Omega, \mathcal{S}, P) \equiv ((\mathbb{R}^p)^n, \mathcal{S}_1 \times \cdots \times \mathcal{S}_n, \nu_1 \times \cdots \times \nu_n) \) where this is just the product \( \sigma \) algebra and product measure which satisfies the following for measurable rectangles.

\[
(\nu_1 \times \cdots \times \nu_n) \left( \prod_{i=1}^{n} E_i \right) = \prod_{i=1}^{n} \nu_i (E_i).
\]

Now let \( X_i (x_1, \ldots, x_i, \ldots, x_n) = x_i \). Then from the definition, if \( E \) is a Borel set in \( \mathbb{R}^p \),

\[
\lambda_{X_i} (E) = P \{ X_i \in E \} = (\nu_1 \times \cdots \times \nu_n) (\mathbb{R}^p \times \cdots \times E \times \cdots \times \mathbb{R}^p) = \nu_i (E).
\]

Let \( \mathcal{M} \) consist of all Borel sets of \( (\mathbb{R}^p)^n \) such that

\[
\int_{\mathbb{R}^p} \cdots \int_{\mathbb{R}^p} \chi_E (x_1, \ldots, x_n) d\lambda_{X_1} \cdots d\lambda_{X_n} = \int_{(\mathbb{R}^p)^n} \chi_E d\lambda (x_1, \ldots, x_n).
\]

From what was just shown and the definition of \( (\nu_1 \times \cdots \times \nu_n) \) that \( \mathcal{M} \) contains all sets of the form \( \prod_{i=1}^{n} E_i \) where each \( E_i \) is Borel sets of \( \mathbb{R}^p \). Also \( \mathcal{M} \) is clearly closed with respect to complements and countable disjoint unions. Therefore, from Lemma 11.2.6, \( \mathcal{M} \) equals the Borel sets. Therefore, the given random vectors are independent and this proves the proposition.

The following Lemma was proved earlier in a different way.

**Lemma 11.2.6** If \( \{X_i\}_{i=1}^{n} \) are independent random variables having values in \( \mathbb{R} \),

\[
E \left( \prod_{i=1}^{n} X_i \right) = \prod_{i=1}^{n} E (X_i).
\]

**Proof:** By Lemma 11.2.6 and denoting by \( P \) the product, \( \prod_{i=1}^{n} X_i \),

\[
E \left( \prod_{i=1}^{n} X_i \right) = \int_{\mathbb{R}} z d\mu (z) = \int_{\mathbb{R}} \prod_{i=1}^{n} x_i d\lambda (x_1, \ldots, x_n)
= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{i=1}^{n} x_i d\lambda_{X_1} \cdots d\lambda_{X_n} = \prod_{i=1}^{n} E (X_i).
\]

11.3 Characteristic Functions And Independence

There is a way to tell if random vectors are independent by using their characteristic functions.

**Proposition 11.3.1** If \( X_i \) is a random vector having values in \( \mathbb{R}^p \), then the random vectors are independent if and only if

\[
E (e^{iP}) = \prod_{j=1}^{n} E (e^{i t_j X_j})
\]

where \( P = \sum_{j=1}^{n} t_j \cdot X_j \) for \( t_j \in \mathbb{R}^p \).

The proof of this proposition will depend on the following lemma.

**Lemma 11.3.2** Let \( Y \) be a random vector with values in \( \mathbb{R}^p \) and let \( f \) be bounded and measurable with respect to the Radon measure \( \lambda_Y \), and satisfy

\[
\int f(y) e^{i t \cdot y} d\lambda_Y = 0
\]

for all \( t \in \mathbb{R}^p \). Then \( f(y) = 0 \) for \( \lambda_Y \) a.e. \( y \).
Proof: You could write the following for \( \phi \in \mathcal{G} \)
\[
\int \phi(t) \int f(y) e^{it \cdot y} d\lambda_Y \, dt = 0 = \int f(y) \left( \int \phi(t) e^{it \cdot y} \, dt \right) d\lambda_Y
\]
and now recall that the inverse Fourier transform maps \( \mathcal{G} \) onto \( \mathcal{G} \). Hence
\[
\int f(y) \psi(y) \, d\lambda_Y = 0
\]
for all \( \psi \in \mathcal{G} \). Thus this is also so for every \( \psi \in C^\infty_0(\mathbb{R}^p) \) by an obvious application of the Stone Weierstrass theorem. Let \( \{ \phi_k \} \) be a sequence of functions in \( C^\infty_0(\mathbb{R}^p) \) which converges to
\[
\text{sgn}(f) \equiv \begin{cases} \frac{f}{|f|} & \text{if } f \neq 0 \\ 0 & \text{if } f = 0 \end{cases}
\]
pointwise a.e. and in \( L^1(\mathbb{R}^p, \lambda_Y) \), each \( |\phi_k| \leq 2 \). Then for any \( \psi \in C^\infty_0(\mathbb{R}^p) \),
\[
0 = \int f(y) \phi_n(y) \psi(y) \, d\lambda_Y \to \int |f(y)| \psi(y) \, d\lambda_Y
\]
Also, the above holds for any \( \psi \in C_c(\mathbb{R}^p) \) as can be seen by taking such a \( \psi \) and convolving with a mollifier. By the Riesz representation theorem, \( f(y) = 0 \lambda_Y \text{ a.e. (The measure } \mu(E) \equiv \int_E |f(y)| \, d\lambda_Y \text{ equals 0.)} \]

Proof of the proposition: If the \( X_j \) are independent, the formula follows from Lemma 11.2.41 and Lemma 11.2.43.

Now suppose the formula holds. Thus
\[
\prod_{j=1}^n E(e^{it_j \cdot X_j}) = \int_{\mathbb{R}^p} \cdots \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} e^{it_1 \cdot x_1} e^{it_2 \cdot x_2} \cdots e^{it_n \cdot x_n} d\lambda_{x_1} d\lambda_{x_2} \cdots d\lambda_{x_n} = E(e^{iP})
\]
\[
= \int_{\mathbb{R}^p} \cdots \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} e^{it_1 \cdot x_1} e^{it_2 \cdot x_2} \cdots e^{it_{n-1} \cdot x_{n-1}} d\lambda_{x_1} d\lambda_{x_2} \cdots d\lambda_{x_{n-1}}
\]
\[
\int_{\mathbb{R}^p} \cdots \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} e^{it_1 \cdot x_1} e^{it_2 \cdot x_2} \cdots e^{it_{n-1} \cdot x_{n-1}} d\lambda_{x_1} d\lambda_{x_2} \cdots d\lambda_{x_{n-1}} \cdot \cdots d\lambda_{x_n}
\]
Let \( t_i = 0 \) for \( i = 1, 2, \ldots, n-2 \). Then this implies
\[
\int_{\mathbb{R}^p} \cdots \int_{\mathbb{R}^p} e^{it_{n-1} \cdot x_{n-1}} d\lambda_{x_{n-1}} = \int_{\mathbb{R}^p} e^{it_{n-1} \cdot x_{n-1}} d\lambda_{x_{n-1}}
\]
By the fact that the characteristic function determines the distribution measure, Theorem 11.3.3, it follows that for these \( x_n \) off a set of \( \lambda_{x_n} \) measure zero, \( \lambda_{x_{n-1}} = \lambda_{x_{n-1} \cdot x_n} \). Returning to 11.3.3, one can replace \( \lambda_{x_{n-1} \cdot x_n} \) with \( \lambda_{x_{n-1}} \) to obtain
\[
\int_{\mathbb{R}^p} \cdots \int_{\mathbb{R}^p} e^{it_1 \cdot x_1} e^{it_2 \cdot x_2} \cdots e^{it_{n-1} \cdot x_{n-1}} d\lambda_{x_1} d\lambda_{x_2} \cdots d\lambda_{x_{n-1}} d\lambda_{x_n}
\]
\[
= \int_{\mathbb{R}^p} \cdots \int_{\mathbb{R}^p} e^{it_1 \cdot x_1} e^{it_2 \cdot x_2} \cdots e^{it_{n-1} \cdot x_{n-1}} d\lambda_{x_1} d\lambda_{x_2} \cdots d\lambda_{x_{n-1}} d\lambda_{x_n}
\]
Next let \( t_n = 0 \) and applying the above Lemma 11.3.4 again, this implies that for \( \lambda_{x_{n-1}} \) a.e. \( x_{n-1} \), the following equals 0.
\[
\int_{\mathbb{R}^p} \cdots \int_{\mathbb{R}^p} e^{it_1 \cdot x_1} e^{it_2 \cdot x_2} \cdots e^{it_{n-2} \cdot x_{n-2}} d\lambda_{x_1} d\lambda_{x_2} \cdots d\lambda_{x_{n-2}}
\]
\[
\int_{\mathbb{R}^p} \cdots \int_{\mathbb{R}^p} e^{it_1 \cdot x_1} e^{it_2 \cdot x_2} \cdots e^{it_{n-2} \cdot x_{n-2}} d\lambda_{x_1} d\lambda_{x_2} \cdots d\lambda_{x_{n-2}}
\]
Let \( t_i = 0 \) for \( i = 1, 2, \ldots, n - 3 \). Then you obtain
\[
\int_{\mathbb{R}^{n-2}} e^{i x_{n-2} \cdot \xi_{n-2}} d\lambda_{\xi_{n-2}} = \int_{\mathbb{R}^{n-2}} e^{i x_{n-2} \cdot \xi_{n-2}} d\lambda_{\xi_{n-2}|\xi_{n-1}}
\]
and so \( \lambda_{\xi_{n-2}} = \lambda_{\xi_{n-2}|\xi_{n-1}} \) for \( \xi_{n-1} \) off a set of \( \lambda_{\xi_{n-1}} \) measure zero. Continuing this way, it follows that
\[
\lambda_{\xi_{n-k}} = \lambda_{\xi_{n-k}|\xi_{n-1} \cdots \xi_{n-k+1}}
\]
for \( \xi_{n-k+1} \) off a set of \( \lambda_{\xi_{n-k+1}} \) measure zero. Thus if \( E \) is Borel in \( \mathbb{R}^{n-1} \times \cdots \times \mathbb{R}^1 \),
\[
\int_{\mathbb{R}^{n-1} \times \cdots \times \mathbb{R}^1} \mathcal{X}_E d\lambda(x_1, \ldots, x_n) =
\int_{\mathbb{R}^{n-1} \times \cdots \times \mathbb{R}^1} \mathcal{X}_E d\lambda_{x_1|x_2 \ldots x_n} d\lambda_{x_2|x_3 \ldots x_n} \cdots d\lambda_{x_{n-1}|x_n} d\lambda_{x_n}
\]
\[
= \int_{\mathbb{R}^n} \mathcal{X}_E d\lambda_{x_1} d\lambda_{x_2} \cdots d\lambda_{x_n}
\]
One could achieve this iterated integral in any order by similar arguments to the above. By Definition 11.2.2 and the discussion which follows, this implies that the random variables \( X_i \) are independent.

Here is another proof of the Doob Dynkin lemma based on differentiation theory.

**Lemma 11.3.3** Suppose \( X, Y_1, Y_2, \ldots, Y_k \) are random vectors \( X \) having values in \( \mathbb{R}^n \) and \( Y_j \) having values in \( \mathbb{R}^{p_j} \) and
\[
X, Y_j \in L^1(\Omega).
\]
Suppose \( X \) is \( \sigma(Y_1, \ldots, Y_k) \) measurable. Thus
\[
\{X^{-1}(E) : E \text{ Borel} \} \subseteq \left\{(Y_1, \ldots, Y_k)^{-1}(F) : F \text{ Borel in } \prod_{j=1}^k \mathbb{R}^{p_j} \right\}
\]
Then there exists a Borel function, \( g : \prod_{j=1}^k \mathbb{R}^{p_j} \to \mathbb{R}^n \) such that
\[
X = g(Y_1, Y_2, \ldots, Y_k).
\]

**Proof:** For the sake of brevity, denote by \( Y \) the vector \( (Y_1, \ldots, Y_k) \) and by \( y \) the vector \( (y_1, \ldots, y_k) \) and let \( \prod_{j=1}^k \mathbb{R}^{p_j} \equiv \mathbb{R}^p \). For \( E \) a Borel set of \( \mathbb{R}^n \),
\[
\int_{Y^{-1}(E)} X d\mathbb{P} = \int_{\mathbb{R}^n \times \mathbb{R}^p} \mathcal{X}_{\mathbb{R}^n \times E}(x, y) x d\lambda_{x, y}
\]
\[
= \int_{E} \int_{\mathbb{R}^n} x d\lambda_{x|y} d\lambda_{y}. \quad (11.3.5)
\]
Consider the function
\[
y \to \int_{\mathbb{R}^n} x d\lambda_{x|y}.
\]
Since \( d\lambda_y \) is a Radon measure having inner and outer regularity, it follows the above function is equal to a Borel function for \( \lambda_y \) a.e. \( y \). This function will be denoted by \( g \). Then from 11.3.5
\[
\int_{Y^{-1}(E)} X d\mathbb{P} = \int_{E} g(y) d\lambda_y = \int_{\mathbb{R}^p} \mathcal{X}_E(y) g(y) d\lambda_y
\]
\[
= \int_{\Omega} \mathcal{X}_E(Y(\omega)) g(Y(\omega)) d\mathbb{P}
\]
\[
= \int_{Y^{-1}(E)} g(Y(\omega)) d\mathbb{P}
\]
and since $Y^{-1}(E)$ is an arbitrary element of $\sigma(Y)$, this shows that since $X$ is $\sigma(Y)$ measurable,

$$X = g(Y) \ P \ a.e.$$  

What about the case where $X$ is not necessarily in $\sigma(Y_1, \cdots, Y_k)$?

**Lemma 11.3.4** There exists a unique function $Z(\omega)$ which satisfies

$$\int_F X(\omega) \, dP = \int_F Z(\omega) \, dP$$

for all $F \in \sigma(Y_1, \cdots, Y_k)$ such that $Z$ is $\sigma(Y_1, \cdots, Y_k)$ measurable. It is denoted by

$$E(X|\sigma(Y_1, \cdots, Y_k))$$

**Proof:** It is like the above. Letting $E$ be a Borel set in $\mathbb{R}^p$,

$$\int_{Y^{-1}(E)} X \, dP = \int_{\mathbb{R}^n \times \mathbb{R}^p} \chi_{\mathbb{R}^n \times E}(x, y) x \, d\lambda_{(x, y)}$$

$$= \int_E \int_{\mathbb{R}^n} x \, d\lambda_{X|y} \, d\lambda_y.$$  

Now let $g(y) \equiv E(X|y_1, \cdots, y_k)$ be a Borel representative of

$$\int_{\mathbb{R}^n} x \, d\lambda_{X|y}$$

It follows $\omega \to g(Y(\omega)) = E(X|Y_1(\omega), \cdots, Y_k(\omega))$ is $\sigma(Y_1, \cdots, Y_k)$ measurable because by definition $\omega \to Y(\omega)$ is $\sigma(Y_1, \cdots, Y_k)$ measurable and a Borel measurable function composed with a measurable one is still measurable. It follows that for all $E$ Borel in $\mathbb{R}^p$,

$$\int_{Y^{-1}(E)} X \, dP = \int_E E(X|y_1, \cdots, y_k) \, d\lambda_Y$$

$$= \int_{Y^{-1}(E)} E(X|Y_1(\omega), \cdots, Y_k(\omega)) \, dP$$

and so $Z(\omega) = E(X|Y_1(\omega), \cdots, Y_k(\omega))$ works because a generic set of $\sigma(Y_1, \cdots, Y_k)$ is $Y^{-1}(E)$ for $E$ a Borel set in $\mathbb{R}^p$. If both $Z, Z_1$ work, then for all $F \in \sigma(Y_1, \cdots, Y_k)$,

$$\int_F (Z - Z_1) \, dP = 0$$

Since $F$ is arbitrary, some routine computations show $Z = Z_1$ a.e.  

**Observation 11.3.5** Note that a.e.

$$E(X|Y_1(\omega), \cdots, Y_k(\omega)) = E(X|\sigma(Y_1, \cdots, Y_k))$$

where the one on the left is the expected value of $X$ given values of $Y_j(\omega)$. This one corresponds to the sort of thing we say in words. The one on the right is an abstract concept which is usually obtained using the Radon Nikodym theorem and its description is given in the lemma. This lemma shows that its meaning is really to take the expected value of $X$ given values for the $Y_k$.

### 11.4 Characteristic Functions For Measures

Recall the characteristic function for a random variable having values in $\mathbb{R}^n$. I will give a review of this to begin with. Then the concept will be generalized to random variables (vectors) which have values in a real separable Banach space.
11.4. CHARACTERISTIC FUNCTIONS FOR MEASURES

Definition 11.4.1 Let $X$ be a random variable. The characteristic function is
\[ \phi_X(t) \equiv E(e^{itX}) \equiv \int_{\Omega} e^{itX(\omega)} dP = \int_{\mathbb{R}^p} e^{itx} d\lambda_X \]
the last equation holding by Proposition [11.4.11] on Page [174].

Recall the following fundamental lemma and definition, Lemma 11.4.3 on Page [124].

Definition 11.4.2 For $T \in \mathbb{G}^*$, define $FT, F^{-1}T \in \mathbb{G}^*$ by
\[ FT(\phi) \equiv T(F\phi), \quad F^{-1}T(\phi) \equiv T(F^{-1}\phi) \]
Lemma 11.4.3 $F$ and $F^{-1}$ are both one to one, onto, and are inverses of each other.

The main result on characteristic functions is the following in Theorem 11.4.4 on Page [174] which is stated here for convenience.

Theorem 11.4.4 Let $X$ and $Y$ be random vectors with values in $\mathbb{R}^p$ and suppose $E(e^{itX}) = E(e^{itY})$ for all $t \in \mathbb{R}^p$. Then $\lambda_X = \lambda_Y$.

I want to do something similar for random variables which have values in a separable real Banach space, $E$ instead of $\mathbb{R}^p$.

Corollary 11.4.5 Let $K$ be a $\pi$ system of subsets of $\Omega$ and suppose two probability measures, $\mu$ and $\nu$ defined on $\sigma(K)$ are equal on $K$. Then $\mu = \nu$.

Proof: This follows from the Lemma 11.4.3 on Page [8]. Let
\[ \mathcal{G} \equiv \{E \in \sigma(K): \mu(E) = \nu(E)\} \]
Then $\mathcal{K} \subseteq \mathcal{G}$, since $\mu$ and $\nu$ are both probability measures, it follows that if $E \in \mathcal{G}$, then so is $E^C$. Since these are measures, if $\{A_i\}$ is a sequence of disjoint sets from $\mathcal{G}$ then
\[ \mu(\bigcup_{i=1}^{\infty} A_i) = \sum_i \mu(A_i) = \sum_i \nu(A_i) = \nu(\bigcup_{i=1}^{\infty} A_i) \]
and so from Lemma 11.4.3, $\mathcal{G} = \sigma(K)$.

Next recall the following fundamental lemma used to prove Pettis’ theorem. It is proved on Page [125] but is stated here for convenience.

Lemma 11.4.6 If $E$ is a separable Banach space with $B'$ the closed unit ball in $E'$, then there exists a sequence $\{f_n\}_{n=1}^{\infty} \equiv D' \subseteq B'$ with the property that for every $x \in E$,
\[ ||x|| = \sup_{f \in D'} |f(x)| \]
Definition 11.4.7 Let $E$ be a separable real Banach space. A cylindrical set is one which is of the form
\[ \{x \in E : x_i^* (x) \in \Gamma_i, i = 1, 2, \cdots, m\} \]
where here $x_i^* \in E'$ and $\Gamma_i$ is a Borel set in $\mathbb{R}$.

It is obvious that $\emptyset$ is a cylindrical set and that the intersection of two cylindrical sets is another cylindrical set. Thus the cylindrical sets form a $\pi$ system. What is the smallest $\sigma$ algebra containing the cylindrical sets? It is the Borel sets of $E$. This is a special case of Lemma 11.4.11. Recall why this was. Letting $\{f_n\}_{n=1}^{\infty} = D'$ be the sequence of Lemma 11.4.6 it follows that
\[ \{x \in E : ||x - a|| \leq \delta\} \]
\[ = \{x \in E : \sup_{f \in D'} |f(x - a)| \leq \delta\} \]
\[ = \{x \in E : \sup_{f \in D'} |f(x) - f(a)| \leq \delta\} \]
\[ = \cap_{n=1}^{\infty} \{x \in E : f_n(x) \in \overline{B(f_n(a), \delta)}\} \]
which yields a countable intersection of cylindrical sets. It follows the smallest $\sigma$ algebra containing the cylindrical sets contains the closed balls and hence the open balls and consequently the open sets and so it contains the Borel sets. However, each cylindrical set is a Borel set and so in fact this $\sigma$ algebra equals $\mathcal{B}(E)$.

From Corollary 11.4.7 it follows that two probability measures which are equal on the cylindrical sets are equal on the Borel sets $\mathcal{B}(E)$.

**Definition 11.4.8** Let $\mu$ be a probability measure on $\mathcal{B}(E)$ where $E$ is a real separable Banach space. Then for $x^* \in E'$,

$$\phi_{\mu}(x^*) \equiv \int_E e^{ix^*(x)} d\mu(x).$$

$\phi_{\mu}$ is called the characteristic function for the measure $\mu$.

Note this is a little different than earlier when the symbol $\phi_X(t)$ was used and $X$ was a random variable. Here the focus is more on the measure than a random variable, $X$ such that $\mathcal{L}(X) = \mu$. It might appear this is a more general concept but in fact this is not the case. You could just consider the separable Banach space or Polish space with the Borel $\sigma$ algebra as your probability space and then consider the identity map as a random variable having the given measure as a distribution measure. Of course a major result is the one which says that the characteristic function determines the measures.

**Theorem 11.4.9** Let $\mu$ and $\nu$ be two probability measures on $\mathcal{B}(E)$ where $E$ is a separable real Banach space. Suppose

$$\phi_{\mu}(x^*) = \phi_{\nu}(x^*)$$

for all $x^* \in E'$. Then $\mu = \nu$.

**Proof:** It suffices to verify that $\mu(A) = \nu(A)$ for all $A \in \mathcal{K}$ where $\mathcal{K}$ is the set of cylindrical sets. Fix $g_n \in (E')^n$. Thus the two measures are equal if for all such $g_n$, $n \in \mathbb{N}$,

$$\mu(g_n^{-1}(B)) = \nu(g_n^{-1}(B))$$

for $B$ a Borel set in $\mathbb{R}^n$. Of course, for such a choice of $g_n \in (E')^n$, there are measures defined on the Borel sets of $\mathbb{R}^n$ $\mu_n$ and $\nu_n$ which are given by

$$\mu_n(B) \equiv \mu(g_n^{-1}(B)) \quad \nu_n(B) \equiv \nu(g_n^{-1}(B))$$

and so it suffices to verify that these two measures are equal. So what are their characteristic functions? Note that $g_n$ is a random variable taking $E$ to $\mathbb{R}^n$ and $\mu_n$, $\nu_n$ are just the probability distribution measures of this random variable. Therefore,

$$\phi_{\mu_n}(t) \equiv \int_{\mathbb{R}^n} e^{it \cdot s} d\mu_n = \int_E e^{it \cdot g_n(x)} d\mu$$

Similarly,

$$\phi_{\nu_n}(t) \equiv \int_{\mathbb{R}^n} e^{it \cdot s} d\nu_n = \int_E e^{it \cdot g_n(x)} d\nu$$

Now $t \cdot g_n \in E'$ and so by assumption, the two ends of the above are equal. Hence $\phi_{\mu_n}(t) = \phi_{\nu_n}(t)$ and so by Theorem 11.4.9, $\mu_n = \nu_n$ which, as shown above, implies $\mu = \nu$. $\blacksquare$

### 11.5 Characteristic Functions And Independence In Banach Space

I will consider the relation between the characteristic function and independence of random variables having values in a Banach space. Recall an earlier proposition which relates independence of random vectors with characteristic functions. It is proved starting on Page 183.

**Proposition 11.5.1** Let $\{X_k\}_{k=1}^n$ be random vectors such that $X_k$ has values in $\mathbb{R}^p$. Then the random vectors are independent if and only if

$$E\left(e^{itP}\right) = \prod_{j=1}^n E\left(e^{it_j \cdot X_j}\right)$$

where $P \equiv \sum_{j=1}^n t_j \cdot X_j$ for $t_j \in \mathbb{R}^p$. 
11.5. CHARACTERISTIC FUNCTIONS AND INDEPENDENCE IN BANACH SPACE

It turns out there is a generalization of the above proposition to the case where the random variables have values in a real separable Banach space. Before proving this recall an earlier theorem which had to do with reducing to the case where the random variables had values in \( \mathbb{R}^n \), Theorem 11.5.2. It is restated here for convenience.

**Theorem 11.5.2** The random variables \( \{X_i\}_{i \in I} \) are independent if and only if

\[
\{i_1, \ldots, i_n\} \subseteq I,
\]

\( m_i, \ldots, m_n \) are positive integers, and \( g_{m_i}, \ldots, g_{m_n} \) are in

\[
(E')^{m_1}, \ldots, (E')^{m_n}
\]

respectively. \( \{g_{m_i} \circ X_{i_j}\}_{j=1}^n \) are independent random vectors having values in

\[
\mathbb{R}^{m_1}, \ldots, \mathbb{R}^{m_n}
\]

respectively.

Now here is the theorem about independence and the characteristic functions.

**Theorem 11.5.3** Let \( \{X_k\}_{k=1}^n \) be random variables such that \( X_k \) has values in \( E_k \), a real separable Banach space. Then the random variables are independent if and only if

\[
E(e^{itP}) = \prod_{j=1}^n E(e^{it_j(X_j)})
\]

where \( P = \sum_{j=1}^n t_j^*(X_j) \) for \( t_j^* \in E_j' \).

**Proof:** If the random variables are independent, then so are the random variables, \( t_j^*(X_j) \) and so the equation follows.

The interesting case is when the equation holds.

It suffices to consider only the case where each \( E_k = E \). This is because you can consider each \( X_j \) to have values in \( \prod_{k=1}^n E_k \) by letting \( X_j \) take its values in the \( j \)th component of the product and 0 in the other components. Can you draw the conclusion the random variables are independent? By Theorem 11.5.2, it suffices to show the random variables \( \{g_{m_k} \circ X_k\}_{k=1}^n \) are independent where \( g_{m_k} = (x_1^*, \ldots, x_n^*) \in (E')^{m_k} \). This happens if whenever \( t_{m_k} \in \mathbb{R}^{m_k} \) and

\[
P = \sum_{k=1}^n t_{m_k} \cdot (g_{m_k} \circ X_k),
\]

it follows

\[
E(e^{itP}) = \prod_{k=1}^n E(e^{it_{m_k}(g_{m_k} \circ X_k)}). \tag{11.5.6}
\]

However, the expression on the right in (11.5.6) equals

\[
\prod_{k=1}^n E(e^{it_{m_k}g_{m_k} \circ X_k})
\]

and \( t_{m_k} \cdot g_{m_k} = \sum_{j=1}^{m_k} t_{j}x_j^* \in E' \). Also the expression on the left equals \( E(e^{i \sum_{k=1}^n t_{m_k}g_{m_k} \circ X_k}) \) Therefore, by assumption, (11.5.6) holds. \( \blacksquare \)

There is an obvious corollary which is useful.

**Corollary 11.5.4** Let \( \{X_k\}_{k=1}^n \) be random variables such that \( X_k \) has values in \( E_k \), a real separable Banach space. Then the random variables are independent if and only if

\[
E(e^{itP}) = \prod_{j=1}^n E(e^{it_j(X_j)})
\]

where \( P = \sum_{j=1}^n t_j^*(X_j) \) for \( t_j^* \in M_j \) where \( M_j \) is a dense subset of \( E_j' \).
Proof: The easy direction follows from Theorem 11.6.4. Suppose then the above equation holds for all $t^*_j \in M_j$. Then let $t^*_j \in E'$ and let $\{t^*_{nj}\}$ be a sequence in $M_j$ such that
\[
\lim_{n \to \infty} t^*_{nj} = t^*_j \text{ in } E'
\]
Then define
\[
P \equiv \sum_{j=1}^n t^*_j X_j, \quad P_n \equiv \sum_{j=1}^n t^*_{nj} X_j.
\]
It follows
\[
E(e^{itP}) = \lim_{n \to \infty} E(e^{itP_n}) = \lim_{n \to \infty} \prod_{j=1}^n E(e^{it^*_{nj}(X_j)}) = \prod_{j=1}^n E(e^{it^*_j(X_j)}) \quad \blacksquare
\]

11.6 Convolution And Sums

Lemma 11.6.3 on Page 155 makes possible a definition of convolution of two probability measures defined on $\mathcal{B}(E)$ where $E$ is a separable Banach space as well as some other interesting theorems which held earlier in the context of locally compact spaces. I will first show a little theorem about density of continuous functions in $L^p(E)$ and then define the convolution of two finite measures. First here is a simple technical lemma.

Lemma 11.6.1 Suppose $K$ is a compact subset of $U$ an open set in $E$ a metric space. Then there exists $\delta > 0$ such that
\[
dist(x, K) + dist(x, U^C) \geq \delta \quad \text{for all } x \in E.
\]
Proof: For each $x \in K$, there exists a ball, $B(x, \delta_x)$ such that $B(x, 3\delta_x) \subseteq U$. Finitely many of these balls cover $K$ because $K$ is compact, say $\{B(x_i, \delta_{x_i})\}_{i=1}^m$. Let
\[
0 < \delta < \min(\delta_{x_i} : i = 1, 2, \ldots, m).
\]
Now pick any $x \in K$. Then $x \in B(x_i, \delta_{x_i})$ for some $x_i$ and so $B(x, \delta) \subseteq B(x_i, 2\delta_{x_i}) \subseteq U$. Therefore, for any $x \in K$, $dist(x, U^C) \geq \delta$. If $x \in B(x_i, 2\delta_{x_i})$ for some $x_i$, it follows $dist(x, U^C) \geq \delta$ because then $B(x, \delta) \subseteq B(x_i, 3\delta_{x_i}) \subseteq U$. If $x \notin B(x_i, 2\delta_{x_i})$ for any of the $x_i$, then $x \notin B(y, \delta)$ for any $y \in K$ because all these sets are contained in some $B(x_i, 2\delta_{x_i})$. Consequently $dist(x, K) \geq \delta$. This proves the lemma.

From this lemma, there is an easy corollary.

Corollary 11.6.2 Suppose $K$ is a compact subset of $U$, an open set in $E$ a metric space. Then there exists a uniformly continuous function $f$ defined on all of $E$, having values in $[0, 1]$ such that $f(x) = 0$ if $x \notin U$ and $f(x) = 1$ if $x \in K$.

Proof: Consider
\[
f(x) = \frac{dist(x, U^C)}{dist(x, U^C) + dist(x, K)}.
\]
Then some algebra yields
\[
|f(x) - f(x')| \leq \frac{1}{\delta} (|dist(x, U^C) - dist(x', U^C)| + |dist(x, K) - dist(x', K)|)
\]
where $\delta$ is the constant of Lemma 11.6.1. Now it is a general fact that
\[
|dist(x, S) - dist(x', S)| \leq d(x, x').
\]
Therefore,
\[
|f(x) - f(x')| \leq \frac{2}{\delta}d(x, x')
\]
and this proves the corollary.

Now suppose \( \mu \) is a finite measure defined on the Borel sets of a separable Banach space, \( E \). It was shown above that \( \mu \) is inner and outer regular. Lemma 10.1.9 on Page 155 shows that \( \mu \) is inner regular in the usual sense with respect to compact sets. This makes possible the following theorem.

**Theorem 11.6.3** Let \( \mu \) be a finite measure on \( \mathcal{B}(E) \) where \( E \) is a separable Banach space and let \( f \in L^p(E;\mu) \). Then for any \( \varepsilon > 0 \), there exists a uniformly continuous, bounded \( g \) defined on \( E \) such that

\[
||f - g||_{L^p(E)} < \varepsilon.
\]

**Proof:** As usual in such situations, it suffices to consider only \( f \geq 0 \). Then by standard considerations there exists a simple measurable function,

\[
s(x) = \sum_{k=1}^m c_k \chi_{A_k}(x)
\]

such that \( ||f - s||_{L^p(E)} < \varepsilon/2 \). Now by regularity of \( \mu \) there exist compact sets, \( K_k \) and open sets, \( V_k \) such that

\[
2 \sum_{k=1}^m |c_k| \mu(V_k \setminus K)^{1/p} < \varepsilon/2
\]

and by Corollary 11.6.2 there exist uniformly continuous functions \( g_k \) having values in \([0,1]\) such that \( g_k = 1 \) on \( K_k \) and 0 on \( V_k^C \). Then consider

\[
g(x) = \sum_{k=1}^m c_k g_k(x).
\]

This function is bounded and uniformly continuous. Furthermore,

\[
||s - g||_{L^p(E)} \leq \left( \int_E \left| \sum_{k=1}^m c_k \chi_{A_k}(x) - \sum_{k=1}^m c_k g_k(x) \right|^p \mu dx \right)^{1/p}
\]

\[
\leq \left( \int_E \left( \sum_{k=1}^m |c_k| |\chi_{A_k}(x) - g_k(x)| \right)^p \mu dx \right)^{1/p}
\]

\[
\leq \sum_{k=1}^m |c_k| \left( \int_{V_k \setminus K_k} |\chi_{A_k}(x) - g_k(x)|^p dx \mu \right)^{1/p}
\]

\[
\leq \sum_{k=1}^m |c_k| \left( \int_{V_k \setminus K_k} 2^p dx \mu \right)^{1/p}
\]

\[
= 2 \sum_{k=1}^m |c_k| \mu(V_k \setminus K)^{1/p} < \varepsilon/2.
\]

Therefore,

\[
||f - g||_{L^p} \leq ||f - s||_{L^p} + ||s - g||_{L^p} < \varepsilon/2 + \varepsilon/2.
\]

This proves the theorem.

**Lemma 11.6.4** Let \( A \in \mathcal{B}(E) \) where \( \mu \) is a finite measure on \( \mathcal{B}(E) \) for \( E \) a separable Banach space. Also let \( x_i \in E \) for \( i = 1, 2, \ldots, m \). Then for \( x \in E^m \),

\[
x \rightarrow \mu \left( A + \sum_{i=1}^m x_i \right), \quad x \rightarrow \mu \left( A - \sum_{i=1}^m x_i \right)
\]

are Borel measurable functions. Furthermore, the above functions are

\[
\mathcal{B}(E) \times \cdots \times \mathcal{B}(E)
\]

measurable where the above denotes the product measurable sets.

**Proof:** First consider the case where \( A = U \), an open set. Let

\[
y \in \left\{ x \in E^m : \mu \left( U + \sum_{i=1}^m x_i \right) > \alpha \right\}
\]

(11.6.7)
Then from Lemma 10.1.9 on Page 155 there exists a compact set, $K \subseteq U + \sum_{i=1}^{m} y_i$ such that $\mu(K) > \alpha$. Then if $y'$ is close enough to $y$, it follows $K \subseteq U + \sum_{i=1}^{m} y'_i$ also. Therefore, for all $y'$ close enough to $y$,

$$\mu \left( U + \sum_{i=1}^{m} y'_i \right) \geq \mu(K) > \alpha.$$ 

In other words the set described in 11.6.7 is an open set and so $y \rightarrow \mu(U + \sum_{i=1}^{m} y_i)$ is Borel measurable whenever $U$ is an open set in $E$.

Define a $\pi$ system, $\mathcal{K}$ to consist of all open sets in $E$. Then define $\mathcal{G}$ as

$$\mathcal{G} = \{ A \in \sigma(\mathcal{K}) = B(E) : y \rightarrow \mu \left( A + \sum_{i=1}^{m} y_i \right) \text{ is Borel measurable} \}.$$ 

I just showed $\mathcal{G} \supseteq \mathcal{K}$. Now suppose $A \in \mathcal{G}$. Then

$$\mu \left( A^c + \sum_{i=1}^{m} y_i \right) = \mu(E) - \mu \left( A + \sum_{i=1}^{m} y_i \right)$$

and so $A^c \in \mathcal{G}$ whenever $A \in \mathcal{G}$. Next suppose $\{ A_i \}$ is a sequence of disjoint sets of $\mathcal{G}$. Then

$$\mu \left( \bigcup_{i=1}^{\infty} A_i + \sum_{j=1}^{m} y_j \right) = \mu \left( \bigcup_{i=1}^{\infty} A_i + \sum_{j=1}^{m} y_j \right) = \sum_{i=1}^{\infty} \mu \left( A_i + \sum_{j=1}^{m} y_j \right)$$

and so $\bigcup_{i=1}^{\infty} A_i \in \mathcal{G}$ because it is the sum of Borel measurable functions. By the lemma on $\pi$ systems, Lemma 11.6.4 on Page 4, it follows $\mathcal{G} = \sigma(\mathcal{K}) = B(E)$. Similarly, $x \rightarrow \mu \left( A - \sum_{j=1}^{m} x_j \right)$ is also Borel measurable whenever $A \in B(E)$. Finally note that

$$B(E) \times \cdots \times B(E)$$

contains the open sets of $E^m$ because the separability of $E$ implies the existence of a countable basis for the topology of $E^m$ consisting of sets of the form

$$\prod_{i=1}^{m} U_i$$

where the $U_i$ come from a countable basis for $E$. Since every open set is the countable union of sets like the above, each being a measurable box, the open sets are contained in

$$B(E) \times \cdots \times B(E)$$

which implies $B(E^m) \subseteq B(E) \times \cdots \times B(E)$ also. This proves the lemma.

With this lemma, it is possible to define the convolution of two finite measures.

**Definition 11.6.5** Let $\mu$ and $\nu$ be two finite measures on $B(E)$, for $E$ a separable Banach space. Then define a new measure, $\mu * \nu$ on $B(E)$ as follows

$$\mu * \nu(A) = \int_{E} \nu(A-x) \, d\mu(x).$$

This is well defined because of Lemma 11.6.4 which says that $x \rightarrow \nu(A-x)$ is Borel measurable.

Here is an interesting theorem about convolutions. However, first here is a little lemma. The following picture is descriptive of the set described in the following lemma.
Lemma 11.6.6  For $A$ a Borel set in $E$, a separable Banach space, define
\[ S_A \equiv \{(x, y) \in E \times E : x + y \in A\} \]
Then $S_A \in \mathcal{B}(E) \times \mathcal{B}(E)$, the $\sigma$ algebra of product measurable sets, the smallest $\sigma$ algebra which contains all the sets of the form $A \times B$ where $A$ and $B$ are Borel.

Proof: Let $\mathcal{K}$ denote the open sets in $E$. Then $\mathcal{K}$ is a $\pi$ system. Let
\[ \mathcal{G} \equiv \{A \in \sigma(\mathcal{K}) = \mathcal{B}(E) : S_A \in \mathcal{B}(E) \times \mathcal{B}(E)\} . \]
Then $\mathcal{K} \subseteq \mathcal{G}$ because if $U \in \mathcal{K}$ then $S_U$ is an open set in $E \times E$ and all open sets are in $\mathcal{B}(E) \times \mathcal{B}(E)$ by a countable basis for the topology of $E \times E$ are sets of the form $B \times C$ where $B$ and $C$ come from a countable basis for $E$. Therefore, $\mathcal{K} \subseteq \mathcal{G}$. Now let $A \in \mathcal{G}$. For $(x, y) \in E \times E$, either $x + y \in A$ or $x + y \notin A$. Hence $E \times E = S_A \cup S_{A^C}$ which shows that if $A \in \mathcal{G}$ then so is $A^C$. Finally if $\{A_i\}$ is a sequence of disjoint sets of $\mathcal{G}$
\[ S_{\bigcup_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} S_{A_i} \]
and this shows that $\mathcal{G}$ is also closed with respect to countable unions of disjoint sets. Therefore, by the lemma on $\pi$ systems, Lemma 11.6.6 on Page 11.6.8 it follows $\mathcal{G} = \sigma(\mathcal{K}) = \mathcal{B}(E)$. This proves the lemma.

Theorem 11.6.7  Let $\mu$, $\nu$, and $\lambda$ be finite measures on $\mathcal{B}(E)$ for $E$ a separable Banach space. Then
\[ \mu \ast \nu = \nu \ast \mu \quad (11.6.8) \]
\[ (\mu \ast \nu) \ast \lambda = \mu \ast (\nu \ast \lambda) \quad (11.6.9) \]
If $\mu$ is the distribution for an $E$ valued random variable, $X$ and if $\nu$ is the distribution for an $E$ valued random variable, $Y$, and $X$ and $Y$ are independent, then $\mu \ast \nu$ is the distribution for the random variable, $X + Y$. Also the characteristic function of a convolution equals the product of the characteristic functions.

Proof: First consider $(11.6.8)$. Letting $A \in \mathcal{B}(E)$, the following computation holds from Fubini’s theorem and Lemma 11.6.6.
\[ \mu \ast \nu (A) \equiv \int_E \nu (A - x) \, d\mu (x) = \int_E \int_E \chi_{S_A}(x, y) \, d\nu (y) \, d\mu (x) \]
\[ = \int_E \int_E \chi_{S_A}(x, y) \, d\mu (x) \, d\nu (y) = \nu \ast \mu (A) . \]
Next consider $(11.6.9)$. Using $(11.6.8)$ whenever convenient,
\[ (\mu \ast \nu) \ast \lambda (A) \equiv \int_E (\mu \ast \nu) (A - x) \, d\lambda (x) \]
\[ = \int_E \int_E \nu (A - x - y) \, d\mu (y) \, d\lambda (x) \]
while
\[ \mu \ast (\nu \ast \lambda) (A) \equiv \int_E (\nu \ast \lambda) (A - y) \, d\mu (y) \]
\[ = \int_E \int_E \nu (A - y - x) \, d\lambda (x) \, d\mu (y) \]
\[ = \int_E \int_E \nu (A - y - x) \, d\mu (y) \, d\lambda (x) . \]
The necessary product measurability comes from Lemma 11.6.3.

Recall

\[(\mu * \nu) (A) \equiv \int_E \nu (A - x) \, d\mu (x).\]

Therefore, if \(s\) is a simple function, \(s (x) = \sum_{k=1}^n c_k X_k (x),\)

\[
\int_E s d(\mu * \nu) = \sum_{k=1}^n c_k \int_E \nu (A_k - x) \, d\mu (x) \\
= \int_E \sum_{k=1}^n c_k \nu (A_k - x) \, d\mu (x) \\
= \int_E \int_E s (x + y) \, d\nu (x) \, d\mu (y)
\]

Approximating with simple functions it follows that whenever \(f\) is bounded and measurable or nonnegative and measurable,

\[
\int_E f d(\mu * \nu) = \int_E \int_E f (x + y) \, d\nu (y) \, d\mu (x)
\]

Therefore, letting \(Z = X + Y\), and \(\lambda\) the distribution of \(Z\), it follows from independence of \(X\) and \(Y\) that for \(t^* \in E'\),

\[
\phi_\lambda (t^*) \equiv E \left( e^{it^* (Z)} \right) = E \left( e^{it^* (X + Y)} \right) = E \left( e^{it^* (X)} \right) E \left( e^{it^* (Y)} \right)
\]

But also, it follows from \(11.6.11\)

\[
\phi_{(\mu * \nu)} (t^*) = \int_E \left( e^{it^* (Z)} \right) d(\mu * \nu) (z) \\
= \int_E \int_E \left( e^{it^* (x+y)} \right) d\nu (y) \, d\mu (x) \\
= \int_E \int_E \left( e^{it^* (x)} \right) \left( e^{it^* (y)} \right) d\nu (y) \, d\mu (x) \\
= \left( \int_E \left( e^{it^* (y)} \right) d\nu (y) \right) \left( \int_E \left( e^{it^* (x)} \right) d\mu (x) \right) \\
= E \left( e^{it^* (X)} \right) E \left( e^{it^* (Y)} \right)
\]

Since \(\phi_\lambda (t^*) = \phi_{(\mu * \nu)} (t^*)\), it follows \(\lambda = \mu * \nu\).

Note the last part of this argument shows the characteristic function of a convolution equals the product of the characteristic functions. This proves the theorem.

11.7 The Convergence Of Sums Of Symmetric Random Variables

It turns out that when random variables have symmetric distributions, some remarkable things can be said about infinite sums of these random variables. Conditions are given here that enable one to conclude the convergence of the sequence of partial sums from the convergence of some subsequence of partial sums.

The following lemma is like an earlier result but does not depend on a function being bounded.

**Definition 11.7.1** Let \(X\) be a random variable. \(\mathcal{L} (X) = \mu\) means \(\lambda_X = \mu\). This is called the law of \(X\). It is the same as saying the distribution measure of \(X\) is \(\mu\).

**Lemma 11.7.2** Let \((\Omega, \mathcal{F}, P)\) be a probability space and let \(X : \Omega \to E\) be a random variable, where \(E\) is a real separable Banach space. Also let \(\mathcal{L} (X) = \mu\), a probability measure defined on \(\mathcal{B} (E)\), the Borel sets of \(E\). Suppose \(h : E \to \mathbb{R}\) is continuous and also suppose \(h \circ X\) is in \(L^1 (\Omega)\). Then

\[
\int_{\Omega} (h \circ X) \, dP = \int_E h (x) \, d\mu.
\]
11.7. THE CONVERGENCE OF SUMS OF SYMMETRIC RANDOM VARIABLES

Proof: Let \( \{a_i\}_{i=1}^{\infty} \) be a countable dense subset of \( \mathbb{R} \). Let \( B_i^n \equiv B \left( a_i, \frac{1}{n} \right) \subseteq \mathbb{R} \) and define Borel sets, \( A_i^n \subseteq E \) as follows:

\[ A_1^n = h^{-1} \left( B_1^n \right), \quad A_{k+1}^n = h^{-1} \left( B_{k+1}^n \right) \setminus \left( \bigcup_{i=1}^{k} A_i^n \right). \]

Thus \( \{A_i^n\}_{i=1}^{\infty} \) are disjoint Borel sets, with \( h(A_i^n) \subseteq B_i^n \). Also let \( b_i^n \) denote the endpoint of \( B \left( a_i, \frac{1}{n} \right) \) which is closer to 0.

\[ h_i^n = \begin{cases} b_i^n \text{ if } h^{-1} \left( B_i^n \right) \neq \emptyset \\ 0 \text{ if } h^{-1} \left( B_i^n \right) = \emptyset \end{cases} \]

Then define

\[ h^n(x) = \sum_{i=1}^{\infty} h_i^n \chi_{A_i^n}(x) \]

Thus \( |h^n(x)| \leq |h(x)| \) for all \( x \in E \) and \( |h^n(x) - h(x)| \leq 1/n \) for all \( x \in E \). Then \( h^n \circ X \) is in \( L^1(\Omega) \) and for all \( n \),

\[ |h^n \circ X(\omega)| \leq |h \circ X(\omega)|. \]

Let

\[ h_k^n(x) = \sum_{i=1}^{k} h_i^n \chi_{A_i^n}(x). \]

Then from the construction in which the \( \{A_i^n\}_{i=1}^{\infty} \) are disjoint,

\[ |h_k^n(X(\omega))| \leq |h(X(\omega))|, \quad |h_k^n(x)| \leq |h(x)| \]

and \( |h_k^n(x)| \) is increasing in \( k \).

\[ \int_{\Omega} |h_k^n(X(\omega))| \, dP = \int_{\Omega} \sum_{i=1}^{k} |h_i^n| \chi_{A_i^n}(X(\omega)) \, dP \]

\[ = \int_{\Omega} \sum_{i=1}^{k} |h_i^n| \chi_{X^{-1}(A_i^n)}(\omega) \, dP \]

\[ = \sum_{i=1}^{k} |h_i^n| P(X^{-1}(A_i^n)) \]

\[ = \sum_{i=1}^{k} |h_i^n| \mu(A_i^n) \]

\[ = \int_{E} |h_k^n(x)| \, d\mu. \]

By the monotone convergence theorem, and letting \( k \to \infty \),

\[ \int_{\Omega} |h^n(X(\omega))| \, dP = \int_{E} |h^n(x)| \, d\mu. \]

Now by the uniform convergence in the construction, you can let \( n \to \infty \) and obtain

\[ \int_{\Omega} |h(X(\omega))| \, dP = \int_{E} |h(x)| \, d\mu. \]

Thus \( h \in L^1(E, \mu) \). It is obviously Borel measurable, being the limit of a sequence of Borel measurable functions.

Now similar reasoning to the above and using the dominated convergence theorem when necessary yields

\[ \int_{\Omega} h^n(X(\omega)) \, dP = \int_{E} h^n(x) \, d\mu \]

Now another application of the dominated convergence theorem yields

\[ \int_{\Omega} h(X(\omega)) \, dP = \int_{E} h(x) \, d\mu. \]

This proves the lemma.

First is a simple definition and lemma about random variables whose distribution is symmetric.
Definition 11.7.3 Let \( X \) be a random variable defined on a probability space, \((\Omega, \mathcal{F}, P)\) having values in a Banach space, \( E \). Then it has a symmetric distribution if whenever \( A \) is a Borel set, \[
P (|X| \in A) = P (|X| \in -A)
\]
In terms of the distribution, \[
\lambda_X = \lambda_{-X}.
\]

It is good to observe that if \( X, Y \) are independent random variables defined on a probability space, \((\Omega, \mathcal{F}, P)\) such that each has symmetric distribution, then \( X + Y \) also has symmetric distribution. Here is why. Let \( A \) be a Borel set in \( E \). Then by Theorem 11.6.7 on Page 193, \[
\lambda_X + \lambda_Y (A) = \int_E \lambda_X (A - z) d\lambda_Y (z)
\]
\[
= \int_E \lambda_{-X} (A - z) d\lambda_{-Y} (z)
\]
\[
= \lambda_{-(X+Y)} (A) = \lambda_{X+Y} (-A)
\]
By induction, it follows that if you have \( n \) independent random variables each having symmetric distribution, then their sum has symmetric distribution.

Here is a simple lemma about random variables having symmetric distributions. It will depend on Lemma 11.6.7 on Page 193.

Lemma 11.7.4 Let \( \mathbf{X} = (X_1, \cdots, X_n) \) and \( Y \) be random variables defined on a probability space, \((\Omega, \mathcal{F}, P)\) such that \( X_i, i = 1, 2, \cdots, n \) and \( Y \) have values in \( E \) a separable Banach space. Thus \( \mathbf{X} \) has values in \( E^n \). Suppose also that \( \{X_1, \cdots, X_n, Y\} \) are independent and that \( Y \) has symmetric distribution. Then if \( A \in \mathcal{B}(E^n) \), it follows \[
P \left( |\mathbf{X}| \in A \cap \left[ \left| \sum_{i=1}^{n} X_i + Y \right| < r \right] \right)
= P \left( |\mathbf{X}| \in A \cap \left[ \left| \sum_{i=1}^{n} X_i - Y \right| < r \right] \right)
\]
You can also change the inequalities in the obvious way, \( < \) to \( \leq \), \( > \) or \( \geq \).

Proof: Denote by \( \lambda_{\mathbf{X}} \) and \( \lambda_Y \) the distribution measures for \( \mathbf{X} \) and \( Y \) respectively. Since the random variables are independent, the distribution for the random variable, \((\mathbf{X}, Y)\) mapping into \( E^{n+1} \) is \( \lambda_{\mathbf{X}} \times \lambda_Y \) where this denotes product measure. Since the Banach space is separable, the Borel sets are contained in the product measurable sets. Then by symmetry of the distribution of \( Y \)
\[
P \left( |\mathbf{X}| \in A \cap \left[ \left| \sum_{i=1}^{n} X_i + Y \right| < r \right] \right)
= \int_{E^n \times E} \chi_A (x) \chi_{B(0,r)} \left( \sum_{i=1}^{n} x_i + y \right) d (\lambda_{\mathbf{X}} \times \lambda_Y) (x, y)
= \int_E \int_{E^n} \chi_A (x) \chi_{B(0,r)} \left( \sum_{i=1}^{n} x_i + y \right) d\lambda_X d\lambda_Y
= \int_{E^n} \int_E \chi_A (x) \chi_{B(0,r)} \left( \sum_{i=1}^{n} x_i + y \right) d\lambda_X d\lambda_{-Y}
= \int_{E^n} \int_{E^n} \chi_A (x) \chi_{B(0,r)} \left( \sum_{i=1}^{n} x_i + y \right) d (\lambda_{\mathbf{X}} \times \lambda_{-Y}) (x, y)
= P \left( |\mathbf{X}| \in A \cap \left[ \left| \sum_{i=1}^{n} X_i + (-Y) \right| < r \right] \right)
\]
This proves the lemma. Other cases are similar.

Now here is a really interesting lemma.
Lemma 11.7.5 Let $E$ be a real separable Banach space. Assume $\xi_1, \ldots, \xi_N$ are independent random variables having values in $E$, a separable Banach space which have symmetric distributions. Also let $S_k = \sum_{i=1}^k \xi_i$. Then for any $r > 0$,

$$P\left(\sup_{k \leq N} ||S_k|| > r\right) \leq 2P\left(||S_N|| > r\right).$$

**Proof:** First of all,

$$P\left(\sup_{k \leq N} ||S_k|| > r\right)$$

$$= P\left(\sup_{k \leq N} ||S_k|| > r \text{ and } ||S_N|| > r\right)$$

$$+ P\left(\sup_{k \leq N-1} ||S_k|| > r \text{ and } ||S_N|| \leq r\right)$$

$$\leq P\left(||S_N|| > r\right) + P\left(\sup_{k \leq N-1} ||S_k|| > r \text{ and } ||S_N|| \leq r\right).$$

(11.7.11)

I need to estimate the second of these terms. Let

$$A_1 \equiv ||S_1|| > r, \ldots, A_k \equiv ||S_k|| > r, \forall j < k.$$

Thus $A_k$ consists of those $\omega$ where $||S_k(\omega)|| > r$ for the first time at $k$. Thus

$$\left[\sup_{k \leq N-1} ||S_k|| > r \text{ and } ||S_N|| \leq r\right] = \bigcup_{j=1}^{N-1} A_j \cap [||S_N|| \leq r]$$

and the sets in the above union are disjoint. Consider $A_j \cap [||S_N|| \leq r]$. For $\omega$ in this set,

$$||S_j(\omega)|| > r, \forall i < j.$$

Since $||S_N(\omega)|| \leq r$ in this set, it follows

$$||S_N(\omega)|| = \left\| S_j(\omega) + \sum_{i=j+1}^N \xi_i(\omega) \right\| \leq r$$

Thus

$$P\left(A_j \cap [||S_N|| \leq r]\right) \quad (11.7.12)$$

$$= P\left(\bigcap_{i=1}^{j-1} [||S_i|| \leq r] \cap [||S_j|| > r] \cap \left\{ \left\| S_j + \sum_{i=j+1}^N \xi_i \right\| \leq r \right\} \right) \quad (11.7.13)$$

Now $\bigcap_{i=1}^{j-1} [||S_i|| \leq r] \cap [||S_j|| > r]$ is of the form

$$[(\xi_1, \ldots, \xi_j) \in A]$$

for some Borel set, $A$. Then letting $Y = \sum_{i=j+1}^N \xi_i$ in Lemma 11.7.2 and $X_i = \xi_i$, Lemma 11.7.2 equals

$$P\left(\bigcap_{i=1}^{j-1} [||S_i|| \leq r] \cap [||S_j|| > r] \cap \left\{ \left\| S_j - \sum_{i=j+1}^N \xi_i \right\| \leq r \right\} \right)$$

$$= P\left(\bigcap_{i=1}^{j-1} [||S_i|| \leq r] \cap [||S_j|| > r] \cap [||S_j - (S_N - S_j)|| \leq r]\right)$$

$$= P\left(\bigcap_{i=1}^{j-1} [||S_i|| \leq r] \cap [||S_j|| > r] \cap [||2S_j - S_N|| \leq r]\right)$$

Now since $||S_j(\omega)|| > r$,

$$[||2S_j - S_N|| \leq r] \subseteq [2||S_j|| - ||S_N|| \leq r] \subseteq [2r - ||S_N|| < r] = [||S_N|| > r]$$
and so, referring to [11.7.12], this has shown

\[ P(A_j \cap \{|S_N| \leq r\}) \]

\[ = P\left(\bigcap_{i=1}^{j-1} \{|S_i| \leq r\} \cap \{|S_j| > r\} \cap \{|2S_j - S_N| \leq r\}\right) \]

\[ \leq P\left(\bigcap_{i=1}^{j-1} \{|S_i| \leq r\} \cap \{|S_j| > r\} \cap \{|S_N| > r\}\right) \]

\[ = P(A_j \cap \{|S_N| > r\}). \]

It follows that

\[ P\left(\sup_{k \leq N-1} \{|S_k| > r\} \cap \{|S_N| \leq r\}\right) = \sum_{i=1}^{N-1} P(A_j \cap \{|S_N| \leq r\}) \]

\[ \leq \sum_{i=1}^{N-1} P(A_j \cap \{|S_N| > r\}) \]

\[ \leq P(\{|S_N| > r\}). \]

and using [11.7.11], this proves the lemma.

This interesting lemma will now be used to prove the following which concludes a sequence of partial sums converges given a subsequence of the sequence of partial sums converges.

**Lemma 11.7.6** Let \( \{\zeta_k\} \) be a sequence of independent random variables having values in a separable real Banach space, \( E \) whose distributions are symmetric. Letting \( S_k \equiv \sum_{i=1}^{k} \zeta_i \), suppose \( \{S_{n_k}\} \) converges a.e. Also suppose that for every \( m > n_k \),

\[ P\left(|S_m - S_{n_k}|_E > 2^{-k}\right) < 2^{-k}. \] (11.7.14)

Then in fact,

\[ S_k(\omega) \rightarrow S(\omega) \text{ a.e.} \] (11.7.15)

and off a set of measure zero, the convergence of \( S_k \) to \( S \) is uniform.

**Proof:** Let \( n_k \leq l \leq m \). Then by Lemma [11.7.5],

\[ P\left(\left|\sup_{n_k < \ell \leq m} |S_\ell - S_{n_k}|\right| > 2^{-k}\right) \leq 2P\left(|S_m - S_{n_k}| > 2^{-k}\right) \]

In using this lemma, you could renumber the \( \zeta_i \) so that the sum

\[ \sum_{j=n_k+1}^{l} \zeta_j \]

corresponds to

\[ \sum_{j=1}^{l-n_k} \xi_j \]

where \( \xi_j = \zeta_{j+n_k} \).

Then using [11.7.6],

\[ P\left(\left|\sup_{n_k < \ell \leq m} |S_\ell - S_{n_k}|\right| > 2^{-k}\right) \leq 2P\left(|S_m - S_{n_k}| > 2^{-k}\right) < 2^{-(k-1)} \]

If \( S_l(\omega) \) fails to converge then \( \omega \) must be in infinitely many of the sets,

\[ \left[\sup_{n_k < \ell} |S_\ell - S_{n_k}| > 2^{-k}\right] \]

each of which has measure no more than \( 2^{-(k-1)} \). Thus \( \omega \) must be in a set of measure zero. This proves the lemma.


11.8 The Multivariate Normal Distribution

Definition 11.8.1 A random vector, $X$, with values in $\mathbb{R}^p$ has a multivariate normal distribution written as

$$X \sim N_p(m, \Sigma)$$

if for all Borel $E \subseteq \mathbb{R}^p$,

$$\lambda_X(E) = \int_{\mathbb{R}^p} \lambda_E(x) \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{1/2}} e^{-\frac{1}{2}(x-m)^*\Sigma^{-1}(x-m)} dx$$

for $\mu$ a given vector and $\Sigma$ a given positive definite symmetric matrix.

Theorem 11.8.2 For $X \sim N_p(m, \Sigma)$, $m = E(X)$ and

$$\Sigma = E((X-m)(X-m)^*)$$

Proof: Let $R$ be an orthogonal transformation such that

$$R\Sigma R^* = D = diag(\sigma_1^2, \ldots, \sigma_p^2).$$

Changing the variable by $x - m = R^*y$,

$$E(X) = \int_{\mathbb{R}^p} xe^{-\frac{1}{2}(x-m)^*\Sigma^{-1}(x-m)} dx \left( \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{1/2}} \right)$$

$$= \int_{\mathbb{R}^p} (R^*y + m) e^{-\frac{1}{2}y^*D^{-1}y} dy \left( \frac{1}{(2\pi)^{p/2} \prod_{i=1}^p \sigma_i} \right)$$

$$= m \int_{\mathbb{R}^p} e^{-\frac{1}{2}y^*D^{-1}y} dy \left( \frac{1}{(2\pi)^{p/2} \prod_{i=1}^p \sigma_i} \right) = m$$

by Fubini’s theorem and the easy to establish formula

$$\frac{1}{\sqrt{2\pi\sigma}} \int_{\mathbb{R}} e^{-\frac{y^2}{2\sigma^2}} dy = 1.$$ 

Next let $M \equiv E((X-m)(X-m)^*)$. Thus, changing the variable as above by $x - m = R^*y$

$$M = \int_{\mathbb{R}^p} (x - m)(x - m)^* e^{-\frac{1}{2}(x-m)^*\Sigma^{-1}(x-m)} dx \left( \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{1/2}} \right)$$

$$= R^* \int_{\mathbb{R}^p} yy^* e^{-\frac{1}{2}y^*D^{-1}y} dy \left( \frac{1}{(2\pi)^{p/2} \prod_{i=1}^p \sigma_i} \right) R$$

Therefore,

$$(RMR^*)_{ij} = \int_{\mathbb{R}^p} y_i y_j e^{-\frac{1}{2}y^*D^{-1}y} dy \left( \frac{1}{(2\pi)^{p/2} \prod_{i=1}^p \sigma_i} \right) = 0,$$

so $RMR^*$ is a diagonal matrix.

$$\Sigma = \int_{\mathbb{R}^p} y_i^2 e^{-\frac{1}{2}y^*D^{-1}y} dy \left( \frac{1}{(2\pi)^{p/2} \prod_{i=1}^p \sigma_i} \right)$$

Using Fubini’s theorem and the easy to establish equations,

$$\frac{1}{\sqrt{2\pi\sigma}} \int_{\mathbb{R}} e^{-\frac{y^2}{2\sigma^2}} dy = 1, \quad \frac{1}{\sqrt{2\pi\sigma}} \int_{\mathbb{R}} y^2 e^{-\frac{y^2}{2\sigma^2}} dy = \sigma^2,$$

it follows $(RMR^*)_{ii} = \sigma_i^2$. Hence $RMR^* = D$ and so $M = R^*DR = \Sigma$. This proves the theorem.
Theorem 11.8.3 Suppose $X_1 \sim N_p(m_1, \Sigma_1)$, $X_2 \sim N_p(m_2, \Sigma_2)$ and the two random vectors are independent. Then
\[
X_1 + X_2 \sim N_p(m_1 + m_2, \Sigma_1 + \Sigma_2).
\] (11.8.16)

Also, if $X \sim N_p(m, \Sigma)$ then $-X \sim N_p(-m, \Sigma)$. Furthermore, if $X \sim N_p(m, \Sigma)$ then
\[
E(e^{it\cdot X}) = e^{it\cdot m} e^{-\frac{1}{2}t^\top \Sigma t}.
\] (11.8.17)

Also if $a$ is a constant and $X \sim N_p(m, \Sigma)$ then $aX \sim N_p(\sigma^2, a^2 \Sigma)$.

**Proof:** Consider $E(e^{it\cdot X})$ for $X \sim N_p(m, \Sigma)$.
\[
E(e^{it\cdot X}) = \frac{1}{(2\pi)^p/2 (\det \Sigma)^{1/2}} \int_{\mathbb{R}^p} e^{it\cdot x} e^{-\frac{1}{2}(x-m)^\top \Sigma^{-1}(x-m)} dx.
\]

Let $R$ be an orthogonal transformation such that
\[
R \Sigma R^\top = D = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2).
\]

Then let $R(x - m) = y$. Then
\[
E(e^{it\cdot X}) = \frac{1}{(2\pi)^p/2 \prod_{i=1}^p \sigma_i} \int_{\mathbb{R}^p} e^{it\cdot (R^\top y + m)} e^{-\frac{1}{2}y^\top D^{-1}y} dx.
\]

Therefore
\[
E(e^{it\cdot X}) = \frac{1}{(2\pi)^p/2 \prod_{i=1}^p \sigma_i} \int_{\mathbb{R}^p} e^{is\cdot (y + Rm)} e^{-\frac{1}{2}y^\top D^{-1}y} dx
\]

where $s = R t$. This equals
\[
e^{it\cdot m} \prod_{i=1}^p \left( \int_{\mathbb{R}} e^{isy_i} e^{-\frac{1}{2}\sigma_i^2 y_i^2} dy_i \right) \frac{1}{\sqrt{2\pi} \sigma_i}
= e^{it\cdot m} \prod_{i=1}^p \left( \int_{\mathbb{R}} e^{isy_i u} e^{-\frac{1}{2}u^2} du \right) \frac{1}{\sqrt{2\pi}}
= e^{it\cdot m} \prod_{i=1}^p e^{-\frac{1}{2}s_i^2 \sigma_i^2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}(u^2 - is_i \sigma_i)\cdot s_i^2} du
= e^{it\cdot m} e^{-\frac{1}{2}\Sigma_{i=1}^p s_i^2 \sigma_i^2} = e^{it\cdot m} e^{-\frac{1}{2}t^\top \Sigma t}
\]

This proves (11.8.18).

Since $X_1$ and $X_2$ are independent, $e^{it\cdot X_1}$ and $e^{it\cdot X_2}$ are also independent. Hence
\[
E(e^{it\cdot X_1} + e^{it\cdot X_2}) = E(e^{it\cdot X_1}) E(e^{it\cdot X_2}).
\]

Thus,
\[
E(e^{it\cdot X_1 + X_2}) = E(e^{it\cdot X_1}) E(e^{it\cdot X_2})
= e^{it\cdot m_1} e^{-\frac{1}{2}t^\top \Sigma_1 t} e^{it\cdot m_2} e^{-\frac{1}{2}t^\top \Sigma_2 t}
= e^{it\cdot (m_1 + m_2)} e^{-\frac{1}{2}t^\top (\Sigma_1 + \Sigma_2) t}
\]

which is the characteristic function of a random vector distributed as
\[
N_p(m_1 + m_2, \Sigma_1 + \Sigma_2).
\]

Now it follows that $X_1 + X_2 \sim N_p(m_1 + m_2, \Sigma_1 + \Sigma_2)$ by Theorem (11.8.17). This proves (11.8.18).

The assertion about $-X$ is also easy to see because
\[
E(e^{it\cdot (-X)}) = E(e^{it\cdot X})
= \frac{1}{(2\pi)^p/2 (\det \Sigma)^{1/2}} \int_{\mathbb{R}^p} e^{i(-t)\cdot x} e^{-\frac{1}{2}(x-m)^\top \Sigma^{-1}(x-m)} dx
= \frac{1}{(2\pi)^p/2 (\det \Sigma)^{1/2}} \int_{\mathbb{R}^p} e^{it\cdot x} e^{-\frac{1}{2}(x+m)^\top \Sigma^{-1}(x+m)} dx
\]
which is the characteristic function of a random variable which is $N(-\mathbf{m}, \Sigma)$. Theorem 11.8.3 again implies $-\mathbf{X} \sim N(-\mathbf{m}, \Sigma)$. Finally consider the last claim. You apply what is known about $\mathbf{X}$ with $\mathbf{t}$ replaced with $a\mathbf{t}$ and then massage things. This gives the characteristic function for $a\mathbf{X}$ is given by

$$E\left(\exp (it \cdot a\mathbf{X})\right) = \exp (it \cdot a\mathbf{m}) \exp \left(-\frac{1}{2} \mathbf{t}^* \mathbf{\Sigma} a^2 \mathbf{t}\right)$$

which is the characteristic function of a normal random vector having mean $a\mathbf{m}$ and covariance $a^2 \mathbf{\Sigma}$. This proves the theorem.

Following [11.8.4] a random vector has a generalized normal distribution if its characteristic function is given as

$$e^{it \cdot \mathbf{m}} e^{-\frac{1}{2} \mathbf{t}^* \mathbf{\Sigma} \mathbf{t}} \quad (11.8.18)$$

where $\mathbf{\Sigma}$ is symmetric and has nonnegative eigenvalues. For a random real valued variable, $\mathbf{m}$ is scalar and so is $\mathbf{\Sigma}$ so the characteristic function of such a generalized normally distributed random variable is

$$e^{it\mu} e^{-\frac{1}{2} t^2 \sigma^2} \quad (11.8.19)$$

These generalized normal distributions do not require $\mathbf{\Sigma}$ to be invertible, only that the eigenvalues be nonnegative. In one dimension this would correspond the characteristic function of a dirac measure having point mass 1 at $\mu$. In higher dimensions, it could be a mixture of such things with more familiar things. I won’t try very hard to distinguish between generalized normal distributions and normal distributions in which the covariance matrix has all positive eigenvalues.

Here are some other interesting results about normal distributions found in [11.8.5]. The next theorem has to do with the question whether a random vector is normally distributed in the above generalized sense.

**Theorem 11.8.4** Let $\mathbf{X} = (X_1, \cdots, X_p)$ where each $X_i$ is a real valued random variable. Then $\mathbf{X}$ is normally distributed in the above generalized sense if and only if every linear combination, $\sum_{j=1}^{p} a_j X_i$ is normally distributed. In this case the mean of $\mathbf{X}$ is

$$\mathbf{m} = (E(X_1), \cdots, E(X_p))$$

and the covariance matrix for $\mathbf{X}$ is

$$\mathbf{\Sigma}_{jk} = E((X_j - m_j)(X_k - m_k)^*) .$$

**Proof:** Suppose first $\mathbf{X}$ is normally distributed. Then its characteristic function is of the form

$$\phi_{\mathbf{X}}(\mathbf{t}) = E(e^{i \mathbf{t} \cdot \mathbf{X}}) = e^{i \mathbf{t} \cdot \mathbf{m}} e^{-\frac{1}{2} \mathbf{t}^* \mathbf{\Sigma} \mathbf{t}} .$$

Then letting $\mathbf{a} = (a_1, \cdots, a_p)$

$$E\left(e^{i \sum_{j=1}^{p} a_j X_i}\right) = E\left(e^{i \mathbf{a} \cdot \mathbf{X}}\right) = e^{i \mathbf{a} \cdot \mathbf{m}} e^{-\frac{1}{2} \mathbf{a}^* \mathbf{\Sigma} \mathbf{a} t^2}$$

which is the characteristic function of a normally distributed random variable with mean $\mathbf{a} \cdot \mathbf{m}$ and variance $\sigma^2 = \mathbf{a}^* \mathbf{\Sigma} \mathbf{a}$. This proves half of the theorem.

Next suppose $\sum_{j=1}^{p} a_j X_j = \mathbf{a} \cdot \mathbf{X}$ is normally distributed with mean $\mu$ and variance $\sigma^2$ so that its characteristic function is given in [11.8.5]. I will now relate $\mu$ and $\sigma^2$ to various quantities involving the $X_j$. Letting $m_j = E(X_j), \mathbf{m} = (m_1, \cdots, m_p)^*$

$$\mu = \sum_{j=1}^{p} a_j E(X_j) = \sum_{j=1}^{p} a_j m_j, \quad \sigma^2 = E\left(\left(\sum_{j=1}^{p} a_j X_j - \sum_{j=1}^{p} a_j m_j\right)^2\right)$$

$$= E\left(\left(\sum_{j=1}^{p} a_j (X_j - m_j)\right)^2\right) = \sum_{j,k} a_j a_k E((X_j - m_j)(X_k - m_k))$$

It follows the mean of the normally distributed random variable, $\mathbf{a} \cdot \mathbf{X}$ is

$$\mu = \sum_j a_j m_j = \mathbf{a} \cdot \mathbf{m}$$
Let $X = (X_1, \ldots, X_p)$, $Y = (Y_1, \ldots, Y_p)$ where each $X_i, Y_i$ is a real valued random variable. Suppose also that for every $a \in \mathbb{R}^p$, $a \cdot X$ and $a \cdot Y$ are both normally distributed with the same mean and variance. Then $X$ and $Y$ are both multivariate normal random vectors with the same mean and variance.

**Proof:** In the proof of Theorem 11.8.4 the proof implies that the characteristic functions of $a \cdot X$ and $a \cdot Y$ are both of the form

$$e^{itm} e^{-\frac{1}{2}t^2 \sigma^2}.$$ 

Then as in the proof of that theorem, it must be the case that

$$m = \sum_{j=1}^{p} a_j m_j$$

where $E(X_i) = m_i = E(Y_i)$ and

$$\sigma^2 = a^* E((X - m)(X - m)^*) a$$

and this last equation must hold for every $a$. Therefore,

$$E((X - m)(X - m)^*) = E((Y - m)(Y - m)^*) \equiv \Sigma$$

and so the characteristic function of both $X$ and $Y$ is $e^{ism} e^{-\frac{1}{2}s^* \Sigma s}$ as in the proof of Theorem 11.8.4. This proves the corollary.

**Theorem 11.8.6** Let $X$ and $Y$ be random vectors having values in $\mathbb{R}^p$ and $\mathbb{R}^q$ respectively. Suppose also that $(X, Y)$ is multivariate normally distributed and

$$E((X - E(X))(Y - E(Y))^*) = 0.$$

Then $X$ and $Y$ are independent random vectors.

**Proof:** Let $Z = (X, Y)$, $m = p + q$. Then by hypothesis, the characteristic function of $Z$ is of the form

$$E(e^{itZ}) = e^{itm} e^{-\frac{1}{2}t^* \Sigma t}$$

where $m = (m_X, m_Y) = E(Z) = E(X, Y)$ and

$$\Sigma = \begin{pmatrix} E((X - E(X))(X - E(X))^*) & 0 \\ 0 & E((Y - E(Y))(Y - E(Y))^*) \end{pmatrix} = \begin{pmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{pmatrix}.$$
Therefore, letting \( t = (u, v) \) where \( u \in \mathbb{R}^p \) and \( v \in \mathbb{R}^q \)

\[
E(e^{it \mathbf{Z}}) = E(e^{i(u, v) \mathbf{X}} v \mathbf{Y}) = E(e^{i(u, v) \mathbf{X} + v \mathbf{Y}})
= e^{i u \mathbf{m}_X e^{-i u \Sigma_X u} e^{i v \mathbf{m}_Y} e^{-i v \Sigma_Y v}}
= E(e^{i u \mathbf{X}}) E(e^{i v \mathbf{Y}}).
\]

(11.8.20)

Where the last equality needs to be justified. When this is done it will follow from Proposition that \( \mathbf{X} \) and \( \mathbf{Y} \) are independent. Thus all that remains is to verify

\[
E(e^{i u \mathbf{X}}) = e^{i u \mathbf{m}_X e^{-i u \Sigma_X u}, E(e^{i v \mathbf{Y}}) = e^{i v \mathbf{m}_Y} e^{-i v \Sigma_Y v}.
\]

However, this follows from (11.8.20). To get the first formula, let \( v = 0 \). To get the second, let \( u = 0 \). This proves the Theorem.

Note that to verify the conclusion of this theorem, it suffices to show

\[
E (X_i - E(X_i)) (Y_j - E(Y_j)) = 0.
\]

Theorem 11.8.7 Suppose \( \mathbf{X} = (X_1, \cdots, X_p) \) is normally distributed with mean \( \mathbf{m} \) and covariance \( \Sigma \). Then if \( X_1 \) is uncorrelated with any of the \( X_i \), meaning

\[
E((X_1 - m_1) (X_j - m_j)) = 0 \quad \text{for} \quad j > 1,
\]

then \( X_1 \) and \( (X_2, \cdots, X_p) \) are both normally distributed and the two random vectors are independent. Here \( m_j \equiv E(X_j) \). More generally, if the covariance matrix is a diagonal matrix, the random variables, \( \{X_1, \cdots, X_p\} \) are independent.

Proof: From Theorem 11.8.6

\[
\Sigma = E((X - \mathbf{m}) (X - \mathbf{m})^*).
\]

Then by assumption,

\[
\Sigma = \left( \begin{array}{cc}
\sigma_1^2 & 0 \\
0 & \Sigma_{p-1}
\end{array} \right).
\]

(11.8.21)

I need to verify that if \( E \in \mathcal{H}_{X_1} (\sigma (X_1)) \) and \( F \in \mathcal{H}_{(X_2, \cdots, X_p)} (\sigma (X_2, \cdots, X_p)) \), then

\[
P(E \cap F) = P(E) P(F).
\]

Let \( E = X_1^{-1} (A) \) and

\[
F = (X_2, \cdots, X_p)^{-1} (B)
\]

where \( A \) and \( B \) are Borel sets in \( \mathbb{R} \) and \( \mathbb{R}^{p-1} \) respectively. Thus I need to verify that

\[
P([(X_1, (X_2, \cdots, X_p)) \in (A, B)]) = 
\]

\[
\mu((X_1, (X_2, \cdots, X_p)) \times (A, B)) = \mu(X_1) \mu(X_2, \cdots, X_p) (B).
\]

(11.8.22)

Using (11.8.4) Fubini's theorem, and definitions,

\[
\mu((X_1, (X_2, \cdots, X_p)) \times (A, B)) = 
\]

\[
\int_{\mathbb{R}^p} X_{A \times B} (x) \frac{1}{(2\pi)^{p/2} \det \Sigma^{1/2}} e^{-\frac{1}{2} (x - \mathbf{m})^* \Sigma^{-1} (x - \mathbf{m})} dx
\]

\[
= \int_{\mathbb{R}} X_A (x_1) \int_{\mathbb{R}^{p-1}} X_B (X_2, \cdots, X_p) \cdot
\]

\[
\frac{1}{(2\pi)^{(p-1)/2} \sqrt{2\pi} (\sigma_1^2)^{1/2} \det (\Sigma_{p-1})^{1/2}} e^{-\frac{1}{2} (x_1 - m_1)^2} \cdot
\]

\[
e^{-\frac{1}{2} (x' - \mathbf{m}')^* \Sigma_{p-1}^{-1} (x' - \mathbf{m}')} dx' dx_1.
\]
where \( \mathbf{x}' = (x_2, \cdots, x_p) \) and \( \mathbf{m}' = (m_2, \cdots, m_p) \). Now this equals

\[
\int_{\mathbb{R}} \mathcal{X}_A(x_1) \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x_1-m_1)^2}{2\sigma_1^2}} \int_B \frac{1}{(2\pi)^{(p-1)/2} \det (\Sigma_{p-1})^{1/2}} e^{-\frac{1}{2} (\mathbf{x}'-\mathbf{m}')\Sigma_{p-1}^{-1} (\mathbf{x}'-\mathbf{m}')} \, d\mathbf{x}' \, dx.
\]

(11.8.23)

In case \( B = \mathbb{R}^{p-1} \), the inside integral equals 1 and

\[
\lambda_{X_1}(A) = \lambda_{(X_1, (X_2, \cdots, X_p))} (A \times \mathbb{R}^{p-1}) = \int_{\mathbb{R}} \mathcal{X}_A(x_1) \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x_1-m_1)^2}{2\sigma_1^2}} \, dx_1
\]

which shows \( X_1 \) is normally distributed as claimed. Similarly, letting \( A = \mathbb{R} \),

\[
\lambda_{(X_2, \cdots, X_p)}(B) = \lambda_{(X_1, (X_2, \cdots, X_p))} (\mathbb{R} \times B) = \int_{\mathbb{R}} \frac{1}{(2\pi)^{(p-1)/2} \det (\Sigma_{p-1})^{1/2}} e^{-\frac{1}{2} (\mathbf{x}'-\mathbf{m}')\Sigma_{p-1}^{-1} (\mathbf{x}'-\mathbf{m}')} \, d\mathbf{x}'
\]

and \((X_2, \cdots, X_p)\) is also normally distributed with mean \( \mathbf{m}' \) and covariance \( \Sigma_{p-1} \). Now from \( \mathbb{R}^p \), \( \Sigma_{p-1} \) follows. In case the covariance matrix is diagonal, the above reasoning extends in an obvious way to prove the random variables, \( \{X_1, \cdots, X_p\} \) are independent.

However, another way to prove this is to use Proposition 11.3.1 on Page 183 and consider the characteristic function. Let \( E(X_j) = m_j \) and

\[
P = \sum_{j=1}^p t_j X_j.
\]

Then since \( \mathbf{X} \) is normally distributed and the covariance is a diagonal,

\[
D = \begin{pmatrix}
\sigma_1^2 & 0 \\
0 & \ddots \\
0 & 0 & \sigma_p^2
\end{pmatrix},
\]

\[
E(e^{it\mathbf{X}}) = E(e^{it \cdot (\mathbf{X}-\mathbf{m})}) = e^{it \cdot \mathbf{m}} e^{-\frac{1}{2} \mathbf{t}^{\prime} \Sigma \mathbf{t}}
\]

\[
= \exp \left( \sum_{j=1}^p it_j m_j - \frac{1}{2} t_j^2 \sigma_j^2 \right)
\]

\[
= \prod_{j=1}^p \exp \left( it_j m_j - \frac{1}{2} t_j^2 \sigma_j^2 \right)
\]

(11.8.25)

Also,

\[
E(e^{i t_j X_j}) = E \left( \exp \left( it_j X_j + \sum_{k \neq j} i0X_k \right) \right)
\]

\[
= \exp \left( it_j m_j - \frac{1}{2} t_j^2 \sigma_j^2 \right)
\]

With \( \mathbb{R}^p \), this shows

\[
E(e^{it \mathbf{X}}) = \prod_{j=1}^p E(e^{it_j X_j})
\]

which shows by Proposition 11.3.1 that the random variables,

\[
\{X_1, \cdots, X_p\}
\]

are independent. This proves the theorem.
11.9 Finding Moments

Let $X$ be a random variable with characteristic function

$$
\phi_X(t) \equiv E(\exp(itX))
$$

Then this can be used to find moments of the random variable assuming they exist. The $k^{th}$ moment is defined as

$$
E(X^k).
$$

This can be done by using the dominated convergence theorem to differentiate the characteristic function with respect to $t$ and then plugging in $t = 0$. For example,

$$
\phi_X'(t) = E(iX \exp(itX))
$$

and now plugging in $t = 0$ you get $iE(X)$. Doing another differentiation you obtain

$$
\phi_X''(t) = E(-X^2 \exp(itX))
$$

and plugging in $t = 0$ you get $-E(X^2)$ and so forth.

An important case is where $X$ is normally distributed with mean 0 and variance $\sigma^2$. In this case, as shown above, the characteristic function is

$$
e^{-\frac{1}{2}t^2\sigma^2}
$$

Also all moments exist when $X$ is normally distributed. So what are these moments?

$$
D_t \left( e^{-\frac{1}{2}t^2\sigma^2} \right) = -t\sigma^2 e^{-\frac{1}{2}t^2\sigma^2}
$$

and plugging in $t = 0$ you find the mean equals 0 as expected.

$$
D_t \left( -\sigma^2 e^{-\frac{1}{2}t^2\sigma^2} \right) = -\sigma^2 e^{-\frac{1}{2}t^2\sigma^2} + t^2\sigma^4 e^{-\frac{1}{2}t^2\sigma^2}
$$

and plugging in $t = 0$ you find the second moment is $\sigma^2$. Then do it again.

$$
D_t \left( -\sigma^2 e^{-\frac{1}{2}t^2\sigma^2} + t^2\sigma^4 e^{-\frac{1}{2}t^2\sigma^2} \right) = 3\sigma^4 te^{-\frac{1}{2}t^2\sigma^2} - t^3\sigma^6 e^{-\frac{1}{2}t^2\sigma^2}
$$

Then $E(X^3) = 0$.

$$
D_t \left( 3\sigma^4 te^{-\frac{1}{2}t^2\sigma^2} - t^3\sigma^6 e^{-\frac{1}{2}t^2\sigma^2} \right) = 3\sigma^4 e^{-\frac{1}{2}t^2\sigma^2} - 6\sigma^6 t^2 e^{-\frac{1}{2}t^2\sigma^2} + t^4\sigma^8 e^{-\frac{1}{2}t^2\sigma^2}
$$

and so $E(X^4) = 3\sigma^4$. By now you can see the pattern. If you continue this way, you find the odd moments are all 0 and

$$
E(X^{2m}) = C_m (\sigma^2)^m. \quad (11.9.26)
$$

This is an important observation.

11.10 The Central Limit Theorem

The central limit theorem is one of the most marvelous theorems in mathematics. It can be proved through the use of characteristic functions. Recall for $x \in \mathbb{R}^p$,

$$
||x||_\infty \equiv \max \{|x_j|, j = 1, \cdots, p\}.
$$

Also recall the definition of the distribution function for a random vector, $X$.

$$
F_X(x) \equiv P(X_j \leq x_j, j = 1, \cdots, p).
$$
Definition 11.10.1 Let \( \{X_n\} \) be random vectors with values in \( \mathbb{R}^p \). Then \( \{\lambda_{X_n}\}_{n=1}^{\infty} \) is called “tight” if for all \( \varepsilon > 0 \) there exists a compact set, \( K_\varepsilon \) such that
\[
\lambda_{X_n}(\{x \notin K_\varepsilon\}) < \varepsilon
\]
for all \( \lambda_{X_n} \). Similarly, if \( \{\mu_n\} \) is a sequence of probability measures defined on the Borel sets of \( \mathbb{R}^p \), then this sequence is “tight” if for each \( \varepsilon > 0 \) there exists a compact set, \( K_\varepsilon \) such that
\[
\mu_n(\{x \notin K_\varepsilon\}) < \varepsilon
\]
for all \( \mu_n \).

Lemma 11.10.2 If \( \{X_n\} \) is a sequence of random vectors with values in \( \mathbb{R}^p \) such that
\[
\lim_{n \to \infty} \phi_{X_n}(t) = \psi(t)
\]
for all \( t \), where \( \psi(0) = 1 \) and \( \psi \) is continuous at \( 0 \), then \( \{\lambda_{X_n}\}_{n=1}^{\infty} \) is tight.

Proof: Let \( e_j \) be the \( j \)th standard unit basis vector.
\[
\frac{1}{u} \int_{-u}^{u} (1 - \phi_{X_n}(te_j)) \, dt = \frac{1}{u} \int_{-u}^{u} (1 - \int_{\mathbb{R}^p} e^{itx_j} \, d\lambda_{X_n}) \, dt = \frac{1}{u} \int_{-u}^{u} \left( \int_{\mathbb{R}^p} (1 - e^{itx_j}) \, d\lambda_{X_n} \right) \, dt
\]
\[
= \left| \left| \int_{\mathbb{R}^p} e^{itx_j} \, d\lambda_{X_n} \right| \right| = \left| \left| \int_{\mathbb{R}^p} e^{itx_j} \, d\lambda_{X_n} \right| \right| = \left| \left| \int_{\mathbb{R}^p} e^{itx_j} \, d\lambda_{X_n} \right| \right|
\]
\[
\geq 2 \int_{|x_j| \geq \frac{u}{2}} \left( 1 - \frac{1}{|u|} \right) \, d\lambda_{X_n} (x)
\]
\[
\geq 2 \int_{|x_j| \geq \frac{u}{2}} \left( 1 - \frac{1}{|u| (2/|u|)} \right) \, d\lambda_{X_n} (x)
\]
\[
= \left| \left| \int_{|x_j| \geq \frac{u}{2}} 1 \, d\lambda_{X_n} (x) \right| \right| = \lambda_{X_n} \left( \left\{ x : |x_j| \geq \frac{2}{u} \right\} \right).
\]
If \( \varepsilon > 0 \) is given, there exists \( r > 0 \) such that if \( u \leq r \),
\[
\frac{1}{u} \int_{-u}^{u} (1 - \psi(te_j)) \, dt < \varepsilon/p
\]
for all \( j = 1, \ldots, p \) and so, by the dominated convergence theorem, the same is true with \( \phi_{X_n} \) in place of \( \psi \) provided \( n \) is large enough, say \( n \geq N(u) \). Thus, if \( u \leq r \) and \( n \geq N(u) \),
\[
\lambda_{X_n} \left( \left\{ x : |x_j| \geq \frac{2}{u} \right\} \right) < \varepsilon/p
\]
for all \( j \in \{1, \ldots, p\} \). It follows that for \( u \leq r \) and \( n \geq N(u) \),
\[
\lambda_{X_n} \left( \left\{ x : \|x\|_\infty \geq \frac{2}{u} \right\} \right) < \varepsilon.
\]
because
\[
\left\{ x : \|x\|_\infty \geq \frac{2}{u} \right\} \subseteq \bigcup_{j=1}^{p} \left\{ x : |x_j| \geq \frac{2}{u} \right\}
\]
This proves the lemma because there are only finitely many measures, \( \lambda_{X_n} \) for \( n < N(u) \) and the compact set can be enlarged finitely many times to obtain a single compact set, \( K_\varepsilon \) such that for all \( n \), \( \lambda_{X_n} (\{x \notin K_\varepsilon\}) \) < \( \varepsilon \). This proves the lemma.
Lemma 11.10.3 If \( \phi_{X_n}(t) \to \phi_X(t) \) for all \( t \), then whenever \( \psi \in \mathcal{S} \),

\[
\lambda_{X_n}(\psi) \equiv \int_{\mathbb{R}^p} \psi(y) d\lambda_{X_n}(y) \to \int_{\mathbb{R}^p} \psi(y) d\lambda_X(y) \equiv \lambda_X(\psi)
\]
as \( n \to \infty \).

**Proof:** Recall that if \( X \) is any random vector, its characteristic function is given by

\[
\phi_X(y) \equiv \int_{\mathbb{R}^p} e^{iyx} d\lambda_X(x).
\]

Also remember the inverse Fourier transform. Letting \( \psi \in \mathcal{S} \), the Schwartz class,

\[
F^{-1}(\lambda_X)(\psi) \equiv \lambda_X\left(F^{-1}\psi \right) = \int_{\mathbb{R}^p} F^{-1}\psi d\lambda_X
\]

\[
= \frac{1}{(2\pi)^{p/2}} \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} e^{iyx} \psi(x) dxd\lambda_X(y)
\]

\[
= \frac{1}{(2\pi)^{p/2}} \int_{\mathbb{R}^p} \psi(x) \int_{\mathbb{R}^p} e^{iyx} d\lambda_X(y) dx
\]

\[
= \frac{1}{(2\pi)^{p/2}} \int_{\mathbb{R}^p} \psi(x) \phi_X(x) dx
\]

and so, considered as elements of \( \mathcal{S}^* \),

\[
F^{-1}(\lambda_X) = \phi_X(\cdot)(2\pi)^{-(p/2)} \in L^\infty.
\]

By the dominated convergence theorem

\[
(2\pi)^{p/2} F^{-1}(\lambda_{X_n})(\psi) \equiv \int_{\mathbb{R}^p} \phi_{X_n}(t) \psi(t) dt
\]

\[
\to \int_{\mathbb{R}^p} \phi_X(t) \psi(t) dt
\]

\[
= (2\pi)^{p/2} F^{-1}(\lambda_X)(\psi)
\]

whenever \( \psi \in \mathcal{S} \). Thus

\[
\lambda_{X_n}(\psi) = FF^{-1}\lambda_{X_n}(\psi) \equiv F^{-1}\lambda_{X_n}(F\psi) \to F^{-1}\lambda_X(F\psi)
\]

\[
= FF^{-1}\lambda_X(\psi) = \lambda_X(\psi).
\]

This proves the lemma.

**Lemma 11.10.4** If \( \phi_{X_n}(t) \to \phi_X(t) \), then if \( \psi \) is any bounded uniformly continuous function,

\[
\lim_{n \to \infty} \int_{\mathbb{R}^p} \psi d\lambda_{X_n} = \int_{\mathbb{R}^p} \psi d\lambda_X.
\]

**Proof:** Let \( \varepsilon > 0 \) be given, let \( \psi \) be a bounded function in \( C^\infty(\mathbb{R}^p) \). Now let \( \eta \in C^\infty_c(Q_r) \) where \( Q_r \equiv [-r, r]^p \) satisfy the additional requirement that \( \eta = 1 \) on \( Q_{r/2} \) and \( \eta(x) \in [0, 1] \) for all \( x \). By Lemma 11.10.2 the set, \( \{\lambda_{X_n}\}_{n=1}^\infty \), is tight and so if \( \varepsilon > 0 \) is given, there exists \( r \) sufficiently large such that for all \( n \),

\[
\int_{[x \notin Q_{r/2}]} |1-\eta| |\psi| d\lambda_{X_n} < \frac{\varepsilon}{3},
\]

and

\[
\int_{[x \notin Q_{r/2}]} |1-\eta| |\psi| d\lambda_X < \frac{\varepsilon}{3}.
\]

Thus,

\[
\left| \int_{\mathbb{R}^p} \psi d\lambda_{X_n} - \int_{\mathbb{R}^p} \psi d\lambda_X \right| \leq \left| \int_{\mathbb{R}^p} \psi d\lambda_{X_n} - \int_{\mathbb{R}^p} \psi \eta d\lambda_{X_n} \right| + \left| \int_{\mathbb{R}^p} \psi \eta d\lambda_{X_n} - \int_{\mathbb{R}^p} \psi \eta d\lambda_X \right| + \left| \int_{\mathbb{R}^p} \psi \eta d\lambda_X \right|
\]
whenever $n$ is large enough by Lemma 11.10.3 because $\psi \eta \in \mathcal{G}$. This establishes the conclusion of the lemma in the case where $\psi$ is also infinitely differentiable. To consider the general case, let $\psi$ only be uniformly continuous and let $\psi_k = \psi * \phi_k$ where $\phi_k$ is a mollifier whose support is in $(-1/k, 1/k)^P$. Then $\psi_k$ converges uniformly to $\psi$ and so the desired conclusion follows for $\psi$ after a routine estimate.

**Definition 11.10.5** Let $\mu$ be a Radon measure on $\mathbb{R}^P$. A Borel set, $A$, is a $\mu$ continuity set if $\mu(\partial A) = 0$ where $\partial A \equiv \overline{A} \setminus \text{interior}(A)$.

The main result is the following continuity theorem. More can be said about the equivalence of various criteria [7].

**Theorem 11.10.6** If $\phi_{X_n}(t) \to \phi_X(t)$ then $\lambda_{X_n}(A) \to \lambda_X(A)$ whenever $A$ is a $\lambda_X$ continuity set.

**Proof:** First suppose $K$ is a closed set and let

$$\psi_k(x) \equiv (1 - k \text{ dist } (x, K))^+.$$ 

Thus, since $K$ is closed $\lim_{k \to \infty} \psi_k(x) = \lambda_K(x)$. Choose $k$ large enough that

$$\int_{\mathbb{R}^P} \psi_k d\lambda_X \leq \lambda_X(K) + \varepsilon.$$ 

Then by Lemma 11.10.3, applied to the bounded uniformly continuous function $\psi_k$,

$$\lim \sup_{n \to \infty} \lambda_{X_n}(K) \leq \lim \sup_{n \to \infty} \int \psi_k d\lambda_{X_n} = \lim_{k \to \infty} \int \psi_k d\lambda_X \leq \lambda_X(K) + \varepsilon.$$ 

Since $\varepsilon$ is arbitrary, this shows

$$\lim \sup_{n \to \infty} \lambda_{X_n}(K) \leq \lambda_X(K)$$

for all $K$ closed.

Next suppose $V$ is open and let

$$\psi_k(x) = 1 - \left(1 - k \text{ dist } (x, V^C)^+\right).$$

Thus $\psi_k(x) \in [0, 1], \psi_k = 1$ if $\text{ dist } (x, V^C) \geq 1/k$, and $\psi_k = 0$ on $V^C$. Since $V$ is open, it follows

$$\lim_{k \to \infty} \psi_k(x) = \lambda_V(x).$$

Choose $k$ large enough that

$$\int \psi_k d\lambda_X \geq \lambda_X(V) - \varepsilon.$$ 

Then by Lemma 11.10.3,

$$\lim \inf_{n \to \infty} \lambda_{X_n}(V) \geq \lim \inf_{n \to \infty} \int \psi_k(x) d\lambda_{X_n} = \int \psi_k(x) d\lambda_X \geq \lambda_X(V) - \varepsilon$$ 

and since $\varepsilon$ is arbitrary,

$$\lim \inf_{n \to \infty} \lambda_{X_n}(V) \geq \lambda_X(V).$$ 

Now let $\lambda_X(\partial A) = 0$ for $A$ a Borel set.

$$\lambda_X(\text{interior}(A)) \leq \lim \inf_{n \to \infty} \lambda_{X_n}(\text{interior}(A)) \leq \lim \inf_{n \to \infty} \lambda_{X_n}(A) \leq \lim \sup_{n \to \infty} \lambda_{X_n}(A) \leq \lim \sup_{n \to \infty} \lambda_{X_n}(\overline{A}) \leq \lambda_X(\overline{A}).$$
11.10. THE CENTRAL LIMIT THEOREM

But \( \lambda_X (\text{interior} (A)) = \lambda_X (\overline{A}) \) by assumption and so \( \lim_{n \to \infty} \lambda_X (A) = \lambda_X (A) \) as claimed. This proves the theorem.

As an application of this theorem the following is a version of the central limit theorem in the situation in which the limit distribution is multivariate normal. It concerns a sequence of random vectors, \( \{X_k\}_{k=1}^{\infty} \), which are identically distributed, have finite mean \( \mathbf{m} \), and satisfy

\[
E \left( |X_k|^2 \right) < \infty. \tag{11.10.27}
\]

**Theorem 11.10.7** Let \( \{X_k\}_{k=1}^{\infty} \) be random vectors satisfying (11.10.27), which are independent and identically distributed with mean \( \mathbf{m} \) and positive definite covariance \( \Sigma \equiv E \left( (X - \mathbf{m})(X - \mathbf{m})^* \right) \). Let

\[
Z_n = \sum_{j=1}^{n} \frac{X_j - \mathbf{m}}{\sqrt{n}}. \tag{11.10.28}
\]

Then, for \( Z \sim N_p (0, \Sigma) \),

\[
\lim_{n \to \infty} F_{Z_n} (x) = F_{Z} (x) \tag{11.10.29}
\]

for all \( x \).

**Proof:** The characteristic function of \( Z_n \) is given by

\[
\phi_{Z_n} (t) = E \left( e^{it \sum_{j=1}^{n} \frac{X_j - \mathbf{m}}{\sqrt{n}}} \right) = \prod_{j=1}^{n} E \left( e^{it \left( \frac{X_j - \mathbf{m}}{\sqrt{n}} \right)} \right).
\]

By Taylor’s theorem applied to real and imaginary parts of \( e^{ix} \), it follows

\[
e^{ix} = 1 + ix - f(x) \frac{x^2}{2}
\]

where \( |f(x)| < 2 \) and

\[
\lim_{x \to 0} f(x) = 1.
\]

Denoting \( X_j \) as \( X \), this implies

\[
e^{it \left( \frac{X - \mathbf{m}}{\sqrt{n}} \right)} = 1 + it \cdot \frac{X - \mathbf{m}}{\sqrt{n}} - f \left( t \cdot \left( \frac{X - \mathbf{m}}{\sqrt{n}} \right) \right) \frac{(t \cdot (X - \mathbf{m}))^2}{2n}
\]

Thus

\[
e^{it \left( \frac{X - \mathbf{m}}{\sqrt{n}} \right)} = 1 + it \cdot \frac{X - \mathbf{m}}{\sqrt{n}} - \frac{(t \cdot (X - \mathbf{m}))^2}{2n} + \left( 1 - f \left( t \cdot \left( \frac{X - \mathbf{m}}{\sqrt{n}} \right) \right) \right) \frac{(t \cdot (X - \mathbf{m}))^2}{2n}.
\]

Thus

\[
\phi_{Z_n} (t) = \prod_{j=1}^{n} \left[ 1 - E \left( \frac{(t \cdot (X - \mathbf{m}))^2}{2n} \right) \right] + E \left( 1 - f \left( t \cdot \left( X - \mathbf{m} \right) \right) \right) \frac{(t \cdot (X - \mathbf{m}))^2}{2n} \right]
\]

\[
= \prod_{j=1}^{n} \left[ 1 - \frac{1}{2n} t^* \Sigma t + \frac{1}{2n} E \left( 1 - f \left( t \cdot \left( X - \mathbf{m} \right) \right) \right) \frac{(t \cdot (X - \mathbf{m}))^2}{2n} \right]. \tag{11.10.30}
\]

(Note \( (t \cdot (X - \mathbf{m}))^2 = t^* (X - \mathbf{m}) (X - \mathbf{m})^* t \).) Now here is a simple inequality for complex numbers whose moduli are no larger than one. I will give a proof of this at the end. It follows easily by induction.

\[
|z_1 \cdots z_n - w_1 \cdots w_n| \leq \sum_{k=1}^{n} |z_k - w_k| \tag{11.10.31}
\]
Also for each $t$, and all $n$ large enough,
\[
\left| \frac{1}{2n} E \left( \left( 1 - f \left( \frac{X - m}{\sqrt{n}} \right) \right) (t \cdot (X - m))^2 \right) \right| < 1.
\]

Applying \[11.10.31\] to \[11.10.30\],
\[
\phi_{Z_n}(t) = \left( \prod_{j=1}^{n} \left( 1 - \frac{1}{2n} t^* \Sigma t \right) \right) + e_n
\]
where
\[
|e_n| \leq \sum_{j=1}^{n} \left| \frac{1}{2n} E \left( \left( 1 - f \left( \frac{X - m}{\sqrt{n}} \right) \right) (t \cdot (X - m))^2 \right) \right|
\]
which converges to 0 as $n \to \infty$ by the Dominated Convergence theorem. Therefore,
\[
\lim_{n \to \infty} \phi_{Z_n}(t) = 0
\]
and so
\[
\lim_{n \to \infty} \phi_{Z_n}(t) = e^{-\frac{1}{2} t^* \Sigma t} = \phi_Z(t)
\]
where $Z \sim N_p(0, \Sigma)$. Therefore, $F_{Z_n}(x) \to F_Z(x)$ for all $x$ because $R_x = \prod_{k=1}^{p} (\infty, x_k)$ is a set of $\lambda_Z$ continuity due to the assumption that $\lambda_Z \ll m_p$ which is implied by $Z \sim N_p(0, \Sigma)$. This proves the theorem.

Here is the proof of the little inequality used above. The inequality is obviously true if $n = 1$. Assume it is true for $n$. Then since all the numbers have absolute value no larger than one,
\[
\prod_{i=1}^{n+1} z_i - \prod_{i=1}^{n+1} w_i \leq \prod_{i=1}^{n+1} z_i - z_{n+1} \prod_{i=1}^{n} w_i + \prod_{i=1}^{n} w_i - w_{n+1} \prod_{i=1}^{n+1} w_i \leq \prod_{i=1}^{n} |z_i - w_i| + |z_{n+1} - w_{n+1}|
\]
by induction.

Suppose $X$ is a random vector with covariance $\Sigma$ and mean $m$, and suppose also that $\Sigma^{-1}$ exists. Consider $\Sigma^{-1/2} (X - m) \equiv Y$. Then $E(Y) = 0$ and
\[
E(YY^*) = E \left( \Sigma^{-1/2} (X - m) (X^* - m) \Sigma^{-1/2} \right)
\]
\[
= \Sigma^{-1/2} E((X - m) (X^* - m)) \Sigma^{-1/2} = I.
\]
Thus $Y$ has zero mean and covariance $I$. This implies the following corollary to Theorem \[11.10.7\].

**Corollary 11.10.8** Let independent identically distributed random variables,
\[
\{X_j\}_{j=1}^{\infty}
\]
have mean $m$ and positive definite covariance $\Sigma$ where $\Sigma^{-1}$ exists. Then if

$$Z_n = \sum_{j=1}^{n} \Sigma^{-(1/2)} \frac{X_j - m}{\sqrt{n}},$$

it follows that for $Z \sim N_p(0, I)$,

$$F_{Z_n}(x) \to F_{Z}(x)$$

for all $x$.

### 11.11 Characteristic Functions Of Probability Measures

Recall one can define the characteristic function of a probability measure. In a sense it is more natural.

**Definition 11.11.1** Let $\mu$ be a probability measure defined on the Borel sets of $\mathbb{R}^n$. Then

$$\phi_{\mu}(t) \equiv \int_{\mathbb{R}^n} e^{it \cdot x} d\mu.$$

Then there is a version of Lemma 11.11.2 whose proof is identical to the proof of that lemma.

**Lemma 11.11.2** If $\{\mu_n\}$ is a sequence of Borel probability measures defined on the Borel sets of $\mathbb{R}^p$ such that

$$\lim_{n \to \infty} \phi_{\mu_n}(t) = \psi(t)$$

for all $t$, where $\psi(0) = 1$ and $\psi$ is continuous at $0$, then $\{\mu_n\}_{n=1}^\infty$ is tight.

As before, there are simple modifications of Lemmas 11.11.3 and 11.11.4. The version of Lemma 11.11.3 being the following.

**Lemma 11.11.3** If $\phi_{\mu_n}(t) \to \phi_{\mu}(t)$ where $\{\mu_n\}$ and $\mu$ are probability measures defined on the Borel sets of $\mathbb{R}^p$, then if $\psi$ is any bounded uniformly continuous function,

$$\lim_{n \to \infty} \int_{\mathbb{R}^p} \psi d\mu_n = \int_{\mathbb{R}^p} \psi d\mu.$$

The next theorem is really important. It gives the existence of a measure based on an assumption that a set of measures is tight.

**Theorem 11.11.4** Let $\Lambda = \{\mu_n\}_{n=1}^\infty$ be a sequence of probability measures defined on the Borel sets of $\mathbb{R}^p$. If $\Lambda$ is tight then there exists a probability measure, $\lambda$, and a subsequence of $\{\mu_n\}_{n=1}^\infty$, still denoted by $\{\mu_n\}_{n=1}^\infty$ such that whenever $\phi$ is a continuous bounded complex valued function defined on $E$,

$$\lim_{n \to \infty} \int_{\mathbb{R}^p} \phi d\mu_n = \int_{\mathbb{R}^p} \phi d\lambda.$$

**Proof:** By tightness, there exists an increasing sequence of compact sets, $\{K_n\}$ such that

$$\mu(K_n) > 1 - \frac{1}{n}$$

for all $\mu \in \Lambda$. Now letting $\mu \in \Lambda$ and $\phi \in C(K_n)$ such that $\|\phi\|_{\infty} \leq 1$, it follows

$$\left| \int_{K_n} \phi d\mu \right| \leq \mu(K_n) \leq 1$$

and so the restrictions of the measures of $\Lambda$ to $K_n$ are contained in the unit ball of $C(K_n)'$. Recall from the Riesz representation theorem, the dual space of $C(K_n)$ is a space of complex Borel measures. Theorem 11.1.2 on Page 270 implies the unit ball of $C(K_n)'$ is weak * sequentially compact. This follows from the observation that $C(K_n)$ is separable which follows easily from the Weierstrass approximation theorem. Thus the unit ball in $C(K_n)'$ is
actually metrizable by Theorem 3.52 on Page 124. Therefore, there exists a subsequence of \( \Lambda, \{ \mu_{1k} \} \) such that their restrictions to \( K_1 \) converge weak \( \ast \) to a measure, \( \lambda_1 \in C(K_1)' \). That is, for every \( \phi \in C(K_1) \),

\[
\lim_{k \to \infty} \int_{K_1} \phi d\mu_{1k} = \int_{K_1} \phi d\lambda_1
\]

By the same reasoning, there exists a further subsequence \( \{ \mu_{2k} \} \) such that the restrictions of these measures to \( K_2 \) converge weak \( \ast \) to a measure \( \lambda_2 \in C(K_2)' \) etc. Continuing this way,

\[
\mu_{11}, \mu_{12}, \mu_{13}, \ldots \to \text{Weak} \ast \text{ in } C(K_1)'
\]

\[
\mu_{21}, \mu_{22}, \mu_{23}, \ldots \to \text{Weak} \ast \text{ in } C(K_2)'
\]

\[
\mu_{31}, \mu_{32}, \mu_{33}, \ldots \to \text{Weak} \ast \text{ in } C(K_3)'
\]

\[
\vdots
\]

Here the \( j^{th} \) sequence is a subsequence of the \( (j - 1)^{th} \). Let \( \lambda_n \) denote the measure in \( C(K_n)' \) to which the sequence \( \{ \mu_{nk} \}_{k=1}^{\infty} \) converges weak \( \ast \). Let \( \{ \mu_n \} \equiv \{ \mu_{nn} \} \), the diagonal sequence. Thus this sequence is ultimately a subsequence of every one of the above sequences and so \( \mu_n \) converges weak \( \ast \) in \( C(K_m)' \) to \( \lambda_n \) for each \( m \).

**Claim:** For \( p > n \), the restriction of \( \lambda_p \) to the Borel sets of \( K_n \) equals \( \lambda_n \).

**Proof of claim:** Let \( H \) be a compact subset of \( K_n \). Then there are sets, \( V_l \) open in \( K_n \) which are decreasing and whose intersection equals \( H \). This follows because this is a metric space. Then let \( H \prec \phi_l \prec V_l \). It follows

\[
\lambda_n (V_l) \geq \int_{K_n} \phi_l d\lambda_n = \lim_{k \to \infty} \int_{K_n} \phi_l d\mu_k
\]

\[
= \lim_{k \to \infty} \int_{K_p} \phi_l d\mu_k = \int_{K_p} \phi_l d\lambda_p \geq \lambda_p (H).
\]

Now considering the ends of this inequality, let \( l \to \infty \) and pass to the limit to conclude

\[
\lambda_n (H) \geq \lambda_p (H).
\]

Similarly,

\[
\lambda_n (H) \leq \int_{K_n} \phi_l d\lambda_n = \lim_{k \to \infty} \int_{K_n} \phi_l d\mu_k
\]

\[
= \lim_{k \to \infty} \int_{K_p} \phi_l d\mu_k = \int_{K_p} \phi_l d\lambda_p \leq \lambda_p (V_l).
\]

Then passing to the limit as \( l \to \infty \), it follows

\[
\lambda_n (H) \leq \lambda_p (H).
\]

Thus the restriction of \( \lambda_p, \lambda_p|_{K_n} \) to the compact sets of \( K_n \) equals \( \lambda_n \). Then by inner regularity it follows the two measures, \( \lambda_p|_{K_n} \) and \( \lambda_n \) are equal on all Borel sets of \( K_n \). Recall that for finite measures on the Borel sets of separable metric spaces, regularity is obtained for free.

It is fairly routine to exploit regularity of the measures to verify that \( \lambda_m (F) \geq 0 \) for all \( F \) a Borel subset of \( K_m \). (Whenever \( \phi \geq 0, \int_{K_m} \phi d\lambda_m \geq 0 \) because \( \int_{K_m} \phi d\mu_k \geq 0 \). Now you can approximate \( X^{F} \) with a suitable nonnegative \( \phi \) using regularity of the measure.) Also, letting \( \phi \equiv 1 \),

\[
1 \geq \lambda_m (K_m) \geq 1 - \frac{1}{m}
\]

(11.11.32)

Define for \( F \) a Borel set,

\[
\lambda (F) \equiv \lim_{n \to \infty} \lambda_n (F \cap K_n).
\]

The limit exists because the sequence on the right is increasing due to the above observation that \( \lambda_n = \lambda_m \) on the Borel subsets of \( K_m \) whenever \( n > m \). Thus for \( n > m \)

\[
\lambda_n (F \cap K_n) \geq \lambda_n (F \cap K_m) = \lambda_m (F \cap K_m).
\]
Now let \( \{F_k\} \) be a sequence of disjoint Borel sets. Then
\[
\lambda(\bigcup_{k=1}^{\infty} F_k) = \lim_{n \to \infty} \lambda_n(\bigcup_{k=1}^{\infty} F_k \cap K_n) = \lim_{n \to \infty} \lambda_n(\bigcup_{k=1}^{\infty} (F_k \cap K_n)) = \lim_{n \to \infty} \sum_{k=1}^{\infty} \lambda_n(F_k \cap K_n) = \sum_{k=1}^{\infty} \lambda(F_k)
\]
the last equation holding by the monotone convergence theorem.

It remains to verify
\[
\lim_{k \to \infty} \int \phi d\mu_k = \int \phi d\lambda
\]
for every \( \phi \) bounded and continuous. This is where tightness is used again. Suppose \( ||\phi||_{\infty} < M \). Then as noted above,
\[
\lambda_n(K_n) = \lambda(K_n)
\]
because for \( p \to n \),\( \lambda_p(K_n) = \lambda_n(K_n) \) and so letting \( p \to \infty \), the above is obtained. Also, from Definition 11.11.5,
\[
\lambda (K_n^C) \leq \lim sup_{p \to \infty} (\lambda_p(K_p) - \lambda_p(K_n)) \leq \lim sup_{p \to \infty} (\lambda_p(K_p) - \lambda_n(K_n)) \leq \lim sup_{p \to \infty} \left( 1 - \left( 1 - \frac{1}{n} \right) \right) = \frac{1}{n}
\]
Consequently,
\[
\left| \int \phi d\mu_k - \int \phi d\lambda \right| \leq \int_{K_n^C} \phi d\mu_k + \int_{K_n} \phi d\mu_k - \left( \int_{K_n} \phi d\lambda + \int_{K_n^C} \phi d\lambda \right) \\
\leq \left| \int_{K_n} \phi d\mu_k - \int_{K_n} \phi d\lambda \right| + \int_{K_n} \phi d\mu_k - \int_{K_n} \phi d\lambda \\
\leq \left| \int_{K_n} \phi d\mu_k - \int_{K_n} \phi d\lambda \right| + \int_{K_n} \phi d\mu_k - \int_{K_n} \phi d\lambda + \int_{K_n^C} \phi d\lambda \\
\leq \left| \int_{K_n} \phi d\mu_k - \int_{K_n} \phi d\lambda \right| + \frac{M}{n} + \frac{M}{n}
\]
First let \( n \) be so large that \( 2M/n < \varepsilon/2 \) and then pick \( k \) large enough that the above expression is less than \( \varepsilon \). This proves the theorem.

Definition 11.11.5 Let \( \mu, \{\mu_n\} \) be probability measures defined on the Borel sets of \( \mathbb{R}^p \) and let the sequence of probability measures, \( \{\mu_n\} \) satisfy
\[
\lim_{n \to \infty} \int \phi d\mu_n = \int \phi d\mu.
\]
for every \( \phi \) a bounded continuous function. Then \( \mu_n \) is said to converge weakly to \( \mu \).

With the above, it is possible to prove the following amazing theorem of Levy.

Theorem 11.11.6 Suppose \( \{\mu_n\} \) is a sequence of probability measures defined on the Borel sets of \( \mathbb{R}^p \) and let \( \{\phi_{\mu_n}\} \) denote the corresponding sequence of characteristic functions. If there exists \( \psi \) which is continuous at 0, \( \psi(0) = 1 \), and for all \( t \),
\[
\phi_{\mu_n}(t) \to \psi(t),
\]
then there exists a probability measure, \( \lambda \) defined on the Borel sets of \( \mathbb{R}^p \) and
\[
\phi_{\lambda}(t) = \psi(t).
\]
That is, \( \psi \) is a characteristic function of a probability measure. Also, \( \{\mu_n\} \) converges weakly to \( \lambda \).
Proof: By Lemma 11.11, \( \{ \mu_n \} \) is tight. Therefore, there exists a subsequence \( \{ \mu_{n_k} \} \) converging weakly to a probability measure, \( \lambda \). In particular,

\[
\phi_\lambda (t) = \int e^{itx} d\lambda (x) = \lim_{n \to \infty} \int e^{itx} d\mu_{n_k} (x)
\]

\[
= \lim_{n \to \infty} \phi_{\mu_{n_k}} (t) = \psi (t)
\]

The last claim follows from this and Lemma 11.11.3. This proves the theorem.

Note how it was only necessary to assume \( \psi (0) = 1 \) and \( \psi \) is continuous at 0 in order to conclude that \( \psi \) is a characteristic function. Thus you find that \( |\psi (t)| \leq 1 \) for free. This helps to see why Prokhorov’s and Levy’s theorems are so amazing.

11.12 Generalized Multivariate Normal

In this section is a further explanation of generalized multivariable normal random variables. Recall that these have characteristic function equal to \( e^{it^\prime \Sigma t} \) where \( \Sigma \geq 0, \Sigma = \Sigma ^* \). The new detail is the case that \( \det \Sigma = 0 \).

Definition 11.12.1 A random vector, \( X \), with values in \( \mathbb{R}^p \) has a multivariate normal distribution written as

\[
X \sim N_p (m, \Sigma)
\]

if for all Borel \( E \subseteq \mathbb{R}^p \),

\[
\lambda_X (E) = \int_{\mathbb{R}^p} \chi_E (x) \frac{1}{(2\pi)^{p/2}} \det \Sigma ^{1/2} e^{-\frac{1}{2} \Sigma ^{-1} (x-m)^\prime (x-m)} dx
\]

for \( \mu \) a given vector and \( \Sigma \) a given positive definite symmetric matrix. Recall also that the characteristic function of this random variable is

\[
E (e^{it\cdot X}) = e^{it\cdot m - \frac{1}{2} t^\prime \Sigma t}
\]

(11.12.33)

So what if \( \det \Sigma = 0? \) Is there a probability measure having characteristic equation

\[
e^{it\cdot m - \frac{1}{2} t^\prime \Sigma t}?
\]

Let \( \Sigma_n \to \Sigma \) in the Frobenius norm, \( \det \Sigma_n > 0 \). That is the \( i^{th} \) components converge. Let \( X_{n} \) be the random variable which is associated with \( m \) and \( \Sigma_n \). Thus for \( \phi \in C_0 (\mathbb{R}^p) \),

\[
|\lambda_{X_n} (\phi)| = \left| \int_{\mathbb{R}^p} \phi (x) \frac{1}{(2\pi)^{p/2}} \det \Sigma_n ^{1/2} e^{-\frac{1}{2} \Sigma_n ^{-1} (x-m)^\prime (x-m)} dx \right| \leq \| \phi \|_{C_0 (\mathbb{R}^p)}
\]

Thus these \( \lambda_{X_n} \) are bounded in the weak * topology of \( C_0 (\mathbb{R}^p)' \) which is the space of signed measures. By the separability of \( C_0 (\mathbb{R}^p) \) and the Banach Alaoglu theorem and the Riesz representation theorem for \( C_0 (\mathbb{R}^p)' \), there is a subsequence still denoted as \( \lambda_{X_n} \) which converges weak * to a finite measure \( \mu \). Is \( \mu \) a probability measure? Is the characteristic function of this measure \( e^{it\cdot m - \frac{1}{2} t^\prime \Sigma t} \)?

Note that \( E (e^{it\cdot X_n}) = e^{it\cdot m - \frac{1}{2} t^\prime \Sigma_n \cdot t} \to e^{it\cdot m - \frac{1}{2} t^\prime \Sigma t} \) and this last function of \( t \) is continuous at 0. Therefore, by Lemma 11.10.2, these measures \( \lambda_{X_n} \) are also tight. Let \( \varepsilon > 0 \) be given. Then there is a compact set \( K \) such that \( \lambda_{X_n} (x \notin K) < \varepsilon \). Now let \( \phi = 1 \) on \( K \) and \( \phi \in C_c (\mathbb{R}^p) \), \( \phi \geq 0 \). Then

\[
(1 - \varepsilon) \leq \int_{\mathbb{R}^p} \phi d\lambda_{X_n} \to \int_{\mathbb{R}^p} \phi d\mu \leq \mu (\mathbb{R}^p)
\]

and so, since \( \varepsilon \) is arbitrary, this shows that \( \mu (\mathbb{R}^p) \geq 1 \). However, \( \mu (\mathbb{R}^p) \leq 1 \) because

\[
\mu (\mathbb{R}^n) \leq \int_{\mathbb{R}^p} \psi d\mu + \varepsilon \leq \int_{\mathbb{R}^p} \psi d\lambda_{X_n} + 2\varepsilon \leq 1 + 2\varepsilon
\]

for suitable \( \psi \in C_c (\mathbb{R}^p) \) having values in \([0, 1]\) and \( n \). Thus \( \mu \) is indeed a probability measure.

Now what of its characteristic function?

\[
e^{it\cdot m - \frac{1}{2} t^\prime \Sigma t} = \lim_{n \to \infty} e^{it\cdot m - \frac{1}{2} t^\prime \Sigma_n \cdot t} = \lim_{n \to \infty} \int_{\mathbb{R}^p} e^{it\cdot x} d\lambda_{X_n} (x)
\]

(11.12.34)
Is this equal to
\[ \int_{\mathbb{R}^p} e^{it \cdot x} d\mu (x) \]

Using tightness again,
\[
\left| \int_{\mathbb{R}^p} e^{it \cdot x} d\mu (x) - \int_{\mathbb{R}^p} e^{it \cdot x} d\lambda_{X_n} (x) \right| \leq \left| \int_{\mathbb{R}^p} e^{it \cdot x} d\mu (x) - \int_{\mathbb{R}^p} \psi e^{it \cdot x} d\mu (x) \right|
\]
\[
+ \left| \int_{\mathbb{R}^p} \psi e^{it \cdot x} d\mu (x) - \int_{\mathbb{R}^p} \psi e^{it \cdot x} d\lambda_{X_n} (x) \right| + \left| \int_{\mathbb{R}^p} \psi e^{it \cdot x} d\lambda_{X_n} (x) - \int_{\mathbb{R}^p} e^{it \cdot x} d\lambda_{X_n} (x) \right|
\]
\[
\leq \varepsilon + \int_{\mathbb{R}^p} \psi e^{it \cdot x} d\mu (x) - \int_{\mathbb{R}^p} \psi e^{it \cdot x} d\lambda_{X_n} (x) + \varepsilon
\]

for a suitable choice of \( \psi \in C_c (\mathbb{R}^p) \) having values in \([0, 1]\). The middle term is less than \( \varepsilon \) if \( n \) large enough. Hence the last limit in \( \lim_{n \to \infty} \int_{\mathbb{R}^p} e^{it \cdot x} d\mu (x) \) as hoped. Letting \( X \) be a random variable having \( \mu \) as its distribution measure, (You could take \( \Omega = \mathbb{R}^p \) and the measurable sets the Borel sets.) what about \( E ((X - m) (X - m)^*) \)? Is it equal to \( \Sigma \) as it is for \( X_n \)? What about the question whether \( X \in L^q (\Omega; \mathbb{R}^p) \) for all \( q > 1 \)? This is clearly true for the case where \( \Sigma^{-1} \) exists, but what of the case where \( \det (\Sigma) = 0 \)?

For simplicity, say \( m = 0 \).

\[
\int_{\Omega} |X|^q dP = \int_{0}^{\infty} P (|X|^q > \lambda) d\lambda = \int_{0}^{\infty} \mu (|X|^q > \lambda) d\lambda
\]
\[
\leq \int_{0}^{\infty} \mu (|X|^q > \lambda) d\lambda \leq \int_{0}^{\infty} \int_{\mathbb{R}^p} (1 - \psi_\lambda) d\mu d\lambda
\]

where \( \psi_\lambda = 1 \) on \( B (0, \frac{1}{2} \lambda^{1/q}) \) is nonnegative, and is in \( C_c (B (0, \lambda^{1/q})) \). Now from the above, \( \mu (\mathbb{R}^p) = \lambda_{X_n} (\mathbb{R}^p) = 1 \) and so the inside integral satisfies
\[
\int_{\mathbb{R}^p} (1 - \psi_\lambda) d\mu = \lim_{n \to \infty} \int_{\mathbb{R}^p} (1 - \psi_\lambda) d\lambda_{X_n}
\]

and each of these integrals on the right is no larger than 1. Hence from Fatou’s lemma,

\[
\int_{\Omega} |X|^q dP \leq \int_{0}^{\infty} \int_{\mathbb{R}^p} (1 - \psi_\lambda) d\mu d\lambda \leq \lim \inf_{n \to \infty} \int_{\mathbb{R}^p} (1 - \psi_\lambda) d\lambda_{X_n} d\lambda
\]

Is this on the right finite? It is dominated by
\[
\lim \inf_{n \to \infty} \int_{0}^{\infty} \lambda_{X_n} \left( |X|^q > \frac{1}{2^n} \lambda \right) d\lambda = \lim \inf_{n \to \infty} 2^q \int_{0}^{\infty} \lambda_{X_n} (|X|^q > \delta) d\delta
\]
\[
= \lim \inf_{n \to \infty} 2^q E (|X_n|^q)
\]

So is a subsequence of \( \{ E (|X_n|^q) \} \) bounded? It equals
\[
\int_{\mathbb{R}^p} \frac{1}{(2\pi)^{p/2} \det (\Sigma_n)^{1/2}} e^{\frac{1}{2} (x - m)^* \Sigma_n^{-1} (x - m)} dx
\]

and for \( q \) an even integer, this moment can be computed using the characteristic function.
\[
e^{-\frac{1}{2} t^* \Sigma_n t} = \int_{\mathbb{R}^p} e^{it \cdot x} d\lambda_{X_n}
\]

Also, it suffices to consider \( E (X_n^q) \). Differentiate both sides. Using the repeated index summation convention,
\[
e^{-\frac{1}{2} t^* \Sigma_n t} (-\Sigma_{nkj} t_j) = \int_{\mathbb{R}^p} x_k e^{it \cdot x} d\mu
\]

Now differentiate again.
\[
e^{-\frac{1}{2} t^* \Sigma_n t} (-\Sigma_{nkj} t_j) (-\Sigma_{nkj} t_j) + (-\Sigma_{nkk}) = -\int_{\mathbb{R}^p} x_k^2 e^{it \cdot x} d\lambda_{X_n}
\]
Next let $t = 0$ to conclude that $E\left( X_{nk}^2 \right) = \Sigma_{nnk}$. Of course you can continue differentiating as long as desired and obtain $E\left( X_{nk}^{2m} \right)$ is equal to some polynomial formula involving $\Sigma_{nnk}$ and these are given to converge to $\Sigma_{kk}$. Therefore, for any $q > 1, \{ E\left( \left| X_n \right|^q \right) \}$ is bounded and so from the above,

$$\int_{\Omega} |X|^q \, dP \leq \lim_{n \to \infty} \inf 2^q E\left( \left| X_n \right|^q \right) < \infty$$

So yes, $X$ is indeed in $L^q(\Omega, \mathbb{R}^p)$ for every $q$. What about the covariance?

From the definition of the characteristic function,

$$e^{-\frac{1}{2}t^* \Sigma t} = \int_{\mathbb{R}^p} e^{it^* x} \, d\mu$$

and so taking the derivative with respect to $t_k$ of both sides,

$$e^{-\frac{1}{2}t^* \Sigma t} (-\Sigma_{kj} t_j) = \int_{\mathbb{R}^p} ix_k e^{it^* x} \, d\mu$$

Now differentiate with respect to $t_l$ on both sides.

$$e^{-\frac{1}{2}t^* \Sigma t} (-\Sigma_{li} t_i) (-\Sigma_{kj} t_j) + e^{-\frac{1}{2}t^* \Sigma t} (-\Sigma_{kl}) = \int_{\mathbb{R}^p} ix_k (ix_l) e^{it^* x} \, d\mu - \int_{\mathbb{R}^p} x_k x_l e^{it^* x} \, d\mu$$

Now let $t = 0$ to obtain

$$\Sigma_{kl} = \int_{\mathbb{R}^p} x_k x_l e^{it^* x} \, d\mu = E\left( X_k X_l \right)$$

If $m \neq 0$, the same kind of argument holds with a little more details. This proves the following theorem.

**Theorem 11.12.2** Let $\Sigma$ be nonnegative and self adjoint $p \times p$ matrix. Then there exists a random variable $X$ whose distribution measure $\lambda_X$ has characteristic function

$$e^{it^* m} e^{-\frac{1}{2}t^* \Sigma t}$$

Also

$$E\left( (X - m)(X - m)^* \right) = \Sigma$$

that is

$$E\left( (X - m)_i (X - m)_j \right) = \Sigma_{ij}$$

This is generalized normally distributed random variable.

There is an interesting corollary to this theorem.

**Corollary 11.12.3** Let $H$ be a real Hilbert space. Then there exist random variables $W(h)$ for $h \in H$ such that each is normally distributed with mean 0 and for every $h, g, (W(h), W(g))$ is normally distributed and

$$E\left( W(h) W(g) \right) = (h, g)_H$$

Furthermore, if $\{e_i\}$ is an orthogonal set of vectors of $H$, then $\{ W(e_i) \}$ are independent random variables. Also for any finite set $\{f_1, f_2, \cdots, f_n\}, (W(f_1), W(f_2), \cdots, W(f_n))$ is normally distributed.

**Proof:** Let $\mu_{h_1, \cdots, h_m}$ be a multivariate normal distribution with covariance $\Sigma_{ij} = (h_i, h_j)$. Thus the characteristic function of this measure is

$$e^{-\frac{1}{2}t^* \Sigma t}$$

Now suppose $\mu_{k_1, \cdots, k_n}$ is another such measure where for simplicity, $\{h_1, \cdots, h_m, k_{m+1}, \cdots, k_n\} = \{k_1, \cdots, k_n\}$. Let $\nu$ be a measure on $\mathcal{B}(\mathbb{R}^m)$ which is given by

$$\nu(E) \equiv \mu_{k_1, \cdots, k_n}(E \times \mathbb{R}^{n-m})$$
Then does it follow that \( \nu = \mu_{h_1 \cdots h_m} \)? If so, then the Kolmogorov consistency condition will hold for these measures \( \mu_{h_1 \cdots h_m} \). To determine whether this is so, take the characteristic function of \( \nu \). Let \( \Sigma_1 \) be the \( n \times n \) matrix which comes from the \( \{ k_1 \cdots k_n \} \) and let \( \Sigma_2 \) be the one which comes from the \( \{ h_1 \cdots h_m \} \).

\[
\int_{\mathbb{R}^m} e^{it \cdot x} d\nu(x) = \int_{\mathbb{R}^{n-m}} \int_{\mathbb{R}^m} e^{i(t,0) \cdot (x,y)} d\mu_{k_1 \cdots k_n}(x,y) = e^{-\frac{1}{2} (t^*,0^*) \Sigma_1 (t,0)} = e^{-\frac{1}{2} t^* \Sigma_2 t}
\]

which is the characteristic function for \( \mu_{h_1 \cdots h_m} \). Therefore, these two measures are the same and the Kolmogorov consistency condition holds. It follows that there exists a measure \( \mu \) defined on the Borel sets of \( \prod_{h \in H} \mathbb{R} \) which extends all of these measures. This argument also shows that if a random vector \( X \) has characteristic function \( e^{-\frac{1}{2} t^* \Sigma t} \), then if \( X_k \) is one of its components, then the characteristic function of \( X_k \) is \( e^{-\frac{1}{2} |h_k|^2 |t|^2} \) so this scalar valued random variable has mean zero and variance \( |h_k|^2 \). Then if \( \omega \in \prod_{h \in H} \mathbb{R} \)

\[
W(h)(\omega) \equiv \pi_h(\omega)
\]

where \( \pi_h \) denotes the projection onto position \( h \) in this product. By the Kolmogorov extension theorem, this is the desired random variable. Its distribution measure is normally distributed with mean 0 and variance \( |h|^2 \). If you have two of them, \( W(g), W(h) \), then this theorem also gives the assertion that \( E(W(h)W(g)) = (h,g)_H \).

Finally consider the claim about independence. Any finite subset of \( \{ W(e_i) \} \) is generalized normal with the covariance matrix being a diagonal. Hence \( E(\exp(i \sum_{k=1}^m t_k W(e_k))) = \prod_{k=1}^m \exp(it_k W(e_k)) \) and so this follows from Proposition 11.3.1.

### 11.13 Positive Definite Functions, Bochner’s Theorem

First here is a nice little lemma about matrices.

**Lemma 11.13.1** Suppose \( M \) is an \( n \times n \) matrix. Suppose also that

\[
\alpha^* M \alpha = 0
\]

for all \( \alpha \in \mathbb{C}^n \). Then \( M = 0 \).

**Proof:** Suppose \( \lambda \) is an eigenvalue for \( M \) and let \( \alpha \) be an associated eigenvector.

\[
0 = \alpha^* M \alpha = \alpha^* \lambda \alpha = \lambda |\alpha|^2
\]

and so all the eigenvalues of \( M \) equal zero. By Schur’s theorem there is a unitary matrix \( U \) such that

\[
M = U \begin{pmatrix} 0 & \ast_1 \\ \ast_{m} & 0 \end{pmatrix} U^*
\]

(11.13.35)

where the matrix in the middle has zeros down the main diagonal and zeros below the main diagonal. Thus

\[
M^* = U \begin{pmatrix} 0 & 0 \\ \ast_2 & 0 \end{pmatrix} U^*
\]

where \( M^* \) has zeros down the main diagonal and zeros above the main diagonal. Also taking the adjoint of the given equation for \( M \), it follows that for all \( \alpha \),

\[
\alpha^* M^* \alpha = 0
\]

Therefore, \( M + M^* \) is Hermitian and has the property that

\[
\alpha^* (M + M^*) \alpha = 0.
\]
Thus $M + M^* = 0$ because it is unitarily similar to a diagonal matrix and the above equation can only hold for all $\alpha$ if $M + M^*$ has all zero eigenvalues which implies the diagonal matrix has zeros down the main diagonal. Therefore, from the formulas for $M, M^*$,

$$0 = U \left( \begin{pmatrix} 0 & \cdots & 0 \\ \cdots & \ddots & \cdots \\ 0 & \cdots & 0 \end{pmatrix} + \begin{pmatrix} 0 & \ast_1 \\ \ast_2 & \ddots \\ \ddots & \ddots & 0 \end{pmatrix} \right) U^*$$

and so the sum of the two matrices in the middle must also equal 0. Hence the entries of the matrix in the middle in 11.13.35 are all equal to zero. Thus $M = 0$ as claimed.

**Definition 11.13.2** A Borel measurable function, $f : \mathbb{R}^n \to \mathbb{C}$ is called positive definite if whenever \( \{t_k\}_{k=1}^p \subseteq \mathbb{R}^n, \alpha \in \mathbb{C}^p \)

$$\sum_{k,j} f(t_j - t_k) \alpha_j \overline{\alpha_k} \geq 0 \quad (11.13.36)$$

The first thing to notice about a positive definite function is the following which implies these functions are automatically bounded.

**Lemma 11.13.3** If $f$ is positive definite then whenever \( \{t_k\}_{k=1}^p \) are $p$ points in $\mathbb{R}^n$, $|f(t_j - t_k)| \leq f(0)$. In particular, for all $t$, $|f(t)| \leq f(0)$.

**Proof:** Let $F$ be the $p \times p$ matrix such that

$$F_{kj} = f(t_j - t_k).$$

Then is of the form

$$\alpha^* F \alpha = (F \alpha, \alpha) \geq 0 \quad (11.13.37)$$

where this is the inner product in $\mathbb{C}^p$. Letting $[\alpha, \beta] \equiv (F \alpha, \beta) = \beta^* F \alpha$, it is obvious that $[\alpha, \beta]$ satisfies

$$[\alpha, \alpha] \geq 0, [a \alpha + b \beta, \gamma] = a [\alpha, \gamma] + b [\beta, \gamma].$$

I claim it also satisfies

$$[\alpha, \beta] = [\beta, \alpha].$$

To verify this last claim, note that since $\alpha^* F \alpha$ is real,

$$\alpha^* F^* \alpha = \alpha^* F \alpha \geq 0$$

and so for all $\alpha \in \mathbb{C}^p$,

$$\alpha^* (F^* - F) \alpha = 0$$

which from Lemma implies $F^* = F$. Hence $F$ is self adjoint and it follows

$$[\alpha, \beta] \equiv \beta^* F \alpha = \beta^* F^* \alpha = \alpha^T F^* \beta = \overline{\alpha^T F \beta} = [\beta, \alpha].$$

Therefore, the Cauchy Schwarz inequality holds for $[\cdot, \cdot]$ and it follows

$$||[\alpha, \beta]|| = |(F \alpha, \beta)| \leq (F \alpha, \alpha)^{1/2} (F \beta, \beta)^{1/2}.$$

Letting $\alpha = e_k$ and $\beta = e_j$, it follows $F_{ss} \geq 0$ for all $s$ and

$$|F_{kj}| \leq F_{kk}^{1/2} F_{jj}^{1/2}$$

which says nothing more than

$$|f(t_j - t_k)| \leq f(0)^{1/2} f(0)^{1/2} = f(0).$$

This proves the lemma.

With this information, here is another useful lemma involving positive definite functions. It is interesting because it looks like the formula which defines what it means for the function to be positive definite.
Lemma 11.13.4 Let \( f \) be a positive definite function as defined above and let \( \mu \) be a finite Borel measure. Then
\[
\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(\mathbf{x} - \mathbf{y}) \, d\mu(\mathbf{x}) \, d\mu(\mathbf{y}) \geq 0. \tag{11.13.38}
\]
If \( \mu \) also has the property that it is symmetric, \( \mu(F) = \mu(-F) \) for all \( F \) Borel, then
\[
\int_{\mathbb{R}^n} f(\mathbf{x}) \, d\mu(\mathbf{x}) \geq 0. \tag{11.13.39}
\]

**Proof:** By definition if \( \{t_j\}_{j=1}^p \subseteq \mathbb{R}^n \), and letting \( \alpha = (1, \ldots, 1)^T \in \mathbb{R}^n \),
\[
\sum_{j,k} f(t_j - t_k) \geq 0.
\]
Therefore, integrating over each of the variables,
\[
0 \leq \sum_{j=1}^p \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(t_j - t_j) \, d\mu(t_j) \, d\mu(t_j) + \sum_{j \neq k} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(t_j - t_k) \, d\mu(t_j) \, d\mu(t_k)
\]
and so
\[
0 \leq f(0) \mu(\mathbb{R}^n)^2 p + p(p - 1) \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(\mathbf{x} - \mathbf{y}) \, d\mu(\mathbf{x}) \, d\mu(\mathbf{y}).
\]
Dividing both sides by \( p(p - 1) \) and letting \( p \to \infty \), it follows
\[
0 \leq \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(\mathbf{x} - \mathbf{y}) \, d\mu(\mathbf{x}) \, d\mu(\mathbf{y})
\]
which shows \( \mu \) is symmetric.

To verify \( \mu \) is a probability measure, use \( \mu(\mathbb{R}^n) = 1 \).

Lemma 11.13.5 Let \( \mu_t \) be the measure defined on \( \mathcal{B}(\mathbb{R}^n) \) by
\[
\mu_t(F) = \int_F \frac{1}{(\sqrt{2\pi t})^n} e^{-\frac{1}{2t}|s|^2} \, ds
\]
for \( t > 0 \). Then \( \mu_t \) is a probability measure.

**Proof:** By Theorem 11.6.7,
\[
\phi_{\mu_t*\mu_t}(s) = \phi_{\mu_t}(s) \phi_{\mu_t}(s) = \left(e^{-\frac{1}{2t}|s|^2}\right)^2 = e^{-\frac{1}{2}(2t)|s|^2} = \phi_{\mu_{2t}}(s).
\]
Each \( \mu_t \) is a probability measure because it is the distribution of a normally distributed random variable of mean 0 and covariance \( tI \).

Now let \( \mu \) be a probability measure on \( \mathcal{B}(\mathbb{R}^n) \).
\[
\phi_{\mu}(t) = \int e^{it \cdot \mathbf{y}} \, d\mu(\mathbf{y})
\]
and so by the dominated convergence theorem, \( \phi_{\mu} \) is continuous and also \( \phi_{\mu}(0) = 1 \). I claim \( \phi_{\mu} \) is also positive definite. Let \( \alpha \in \mathbb{C}^p \) and \( \{t_k\}_{k=1}^p \) a sequence of points of \( \mathbb{R}^n \). Then
\[
\sum_{k,j} \phi_{\mu}(t_k - t_j) \alpha_k \overline{\alpha_j} = \sum_{k,j} \int e^{it_k \cdot \mathbf{y}} \alpha_k e^{-it_j \cdot \mathbf{y}} \overline{\alpha_j} \, d\mu(\mathbf{y})
\]
\[ \int \sum_{k,j} e^{it_k \cdot y} \alpha_k e^{it_j \cdot y} \alpha_j d\mu(y). \]

Now let \( \beta(y) \equiv (e^{it_k \cdot y} \alpha_1, \cdots, e^{it_p \cdot y} \alpha_p)^T \). Then the above equals
\[ \int (1, \cdots, 1) \beta(y) \beta^*(y) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} d\mu \]

The integrand is of the form
\[
\begin{pmatrix} \beta^* \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \\ \beta^* \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \end{pmatrix} \geq 0
\]

because it is just a complex number times its conjugate.

Thus every characteristic function is continuous, equals 1 at 0, and is positive definite. Bochner’s theorem goes the other direction.

To begin with, suppose \( \mu \) is a finite measure on \( \mathcal{B}(\mathbb{R}^n) \). Then for \( \mathcal{S} \) the Schwartz class, \( \mu \) can be considered to be in the space of linear transformations defined on \( \mathcal{S}, \mathcal{S}^* \) as follows.
\[ \mu(f) \equiv \int f d\mu. \]

Recall \( F^{-1}(\mu) \) is defined as
\[ F^{-1}(\mu)(f) \equiv \mu(F^{-1}f) = \int_{\mathbb{R}^n} F^{-1}f d\mu \]
\[ = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{ix \cdot y} f(y) dy d\mu \]
\[ = \int_{\mathbb{R}^n} \left( \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{ix \cdot y} d\mu \right) f(y) dy \]
and so \( F^{-1}(\mu) \) is the bounded continuous function
\[ y \rightarrow \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{ix \cdot y} d\mu. \]

Now the following lemma has the main ideas for Bochner’s theorem.

**Lemma 11.13.6** Suppose \( \psi(t) \) is positive definite, \( t \rightarrow \psi(t) \) is in \( L^1(\mathbb{R}^n, m_n) \) where \( m_n \) is Lebesgue measure, \( \psi(0) = 1 \), and \( \psi \) is continuous at 0. Then there exists a unique probability measure, \( \mu \) defined on the Borel sets of \( \mathbb{R}^n \) such that
\[ \phi_{\mu}(t) = \psi(t). \]

**Proof:** If the conclusion is true, then
\[ \psi(t) = \int_{\mathbb{R}^n} e^{it \cdot x} d\mu(x) = (2\pi)^{n/2} F^{-1}(\mu)(t). \]

Recall that \( \mu \in \mathcal{S}^* \), the algebraic dual of \( \mathcal{S} \). Therefore, in \( \mathcal{S}^* \),
\[ \frac{1}{(2\pi)^{n/2}} F(\psi) = \mu. \]

That is, for all \( f \in \mathcal{S} \),
\[ \int_{\mathbb{R}^n} f(y) d\mu(y) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} F(\psi)(y) f(y) dy \]
\[ = \frac{1}{(2\pi)^{n}} \int_{\mathbb{R}^n} f(y) \left( \int_{\mathbb{R}^n} e^{-iy \cdot x} \psi(x) dx \right) dy. \]  (11.13.40)
I will show
\[ f \rightarrow \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} f(y) \left( \int_{\mathbb{R}^n} e^{-iy \cdot x} \psi(x) \, dx \right) \, dy \]
is a positive linear functional and then it will follow from Lemma 11.13 that \( \mu \) is unique. Thus it is needed to show the inside integral in (11.13.2) is nonnegative. First note that the integrand is a positive definite function of \( x \) for each fixed \( y \). This follows from
\[
\sum_{k,j} e^{-iy \cdot (x_k - x_j)} \psi(x_k - x_j) \alpha_k \alpha_j
\]
\[ = \sum_{k,j} \psi(x_k - x_j) \left( e^{-iy \cdot (x_k) \alpha_k} \right) \left( e^{-iy \cdot (x_j)} \alpha_j \right) \geq 0. \]

Let \( t > 0 \) and
\[ h_{2t}(x) \equiv \frac{1}{(4\pi t)^{n/2}} e^{-\frac{1}{4t}|x|^2}. \]

Then by dominated convergence theorem,
\[
\int_{\mathbb{R}^n} e^{-iy \cdot x} \psi(x) \, dx = \lim_{t \to \infty} \int_{\mathbb{R}^n} e^{-iy \cdot x} \psi(x) \, h_{2t}(x) \, dx
\]

Letting \( d\eta_{2t} = h_{2t}(x) \, dx \), it follows from Lemma 11.13 \( \eta_{2t} = \eta_t * \eta_t \) and since these are symmetric measures, it follows from Lemma 11.13 that the above equals
\[
\lim_{t \to \infty} \int_{\mathbb{R}^n} e^{-iy \cdot x} \psi(x) \, d(\eta_t * \eta_t) \geq 0
\]

Thus the above functional is a positive linear functional and so there exists a unique Radon measure, \( \mu \) satisfying
\[
\int_{\mathbb{R}^n} f(y) \, d\mu(y) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} F(\psi)(y) \, f(y) \, dy
\]
\[
= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} f(y) \left( \int_{\mathbb{R}^n} e^{-iy \cdot x} \psi(x) \, dx \right) \, dy
\]
\[
= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \psi(x) \left( \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(y) \, e^{-iy \cdot x} \, dy \right) \, dx
\]

for all \( f \in C_c(\mathbb{R}^n) \). Thus from the dominated convergence theorem, the above holds for all \( f \in \mathcal{S} \) also. Hence for all \( f \in \mathcal{S} \) and considering \( \mu \) as an element of \( \mathcal{S}^* \),
\[
F^{-1} \mu(Ff) = \mu(f) = \int_{\mathbb{R}^n} f(y) \, d\mu(y)
\]
\[
= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \psi(x) \, F(f)(x) \, dx
\]
\[
= \frac{1}{(2\pi)^{n/2}} F(\psi)(f) = \frac{1}{(2\pi)^{n/2}} \psi(Ff).
\]

It follows that in \( \mathcal{S}^* \),
\[ \psi = (2\pi)^{n/2} F^{-1} \mu \]

Thus
\[ \psi(t) = \int_{\mathbb{R}^n} e^{it \cdot x} \, d\mu \]
in \( L^1 \). Since the right side is continuous and the left is given continuous at \( t = 0 \) and equal to 1 there, it follows
\[
1 = \psi(0) = \int_{\mathbb{R}^n} e^{i0 \cdot x} \, d\mu = \mu(\mathbb{R}^n)
\]
and so \( \mu \) is a probability measure as claimed. This proves the lemma.

The following is Bochner’s theorem.
**Theorem 11.13.7** Let $\psi$ be positive definite, continuous at $0$, and $\psi(0) = 1$. Then there exists a unique Radon probability measure $\mu$ such that $\psi = \phi_\mu$.

**Proof:** If $\psi \in L^1(\mathbb{R}^n, m_n)$, then the result follows from Lemma 11.13.6. By Lemma 11.13.3 $\psi$ is bounded. Consider

$$
\psi_t(x) \equiv \psi(x) \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}t|x|^2}.
$$

Then $\psi_t(0) = 1$, $x \mapsto \psi_t(x)$ is continuous at $0$, and $\psi_t \in L^1(\mathbb{R}^n, m_n)$. Therefore, by Lemma 11.13.6 there exists a unique Radon probability measure $\mu_t$ such that

$$
\psi_t(x) = \int_{\mathbb{R}^n} e^{ix \cdot y} d\mu_t(y) = \phi_{\mu_t}(x).
$$

Now letting $t \to \infty$,

$$
\lim_{t \to \infty} \psi_t(x) = \lim_{t \to \infty} \phi_{\mu_t}(x) = \psi(x).
$$

By Levy’s theorem, Theorem 11.11.6 it follows there exists $\mu$, a probability measure on $\mathcal{B}(\mathbb{R}^n)$ such that $\psi(x) = \phi_\mu(x)$. The measure is unique because the characteristic functions are uniquely determined by the measure. This proves the theorem.
Chapter 12

Discrete Martingales

12.1 Conditional Expectation

From Observation 11.3.5 on Page 186, it was shown that the conditional expectation of a random variable $X$ given some others really is just what the words suggest. Given $\omega \in \Omega$, it results in a value for the “other” random variables and then you essentially take the expectation of $X$ given this information which yields the value of the conditional expectation of $X$ given the other random variables. It was also shown in Lemma 11.3.4 that this gives the same result as finding a $\sigma(X_1, \cdots, X_n)$ measurable function $Z$ such that for all $F \in \sigma(X_1, \cdots, X_n)$,

$$\int_F X dP = \int_F Z dP$$

This was done for a particular type of $\sigma$ algebra but there is no need to be this specialized. The following is the general version of conditional expectation given a $\sigma$ algebra. It makes perfect sense to ask for the conditional expectation given a $\sigma$ algebra and this is what will be done from now on.

**Definition 12.1.1** Let $(\Omega, \mathcal{M}, P)$ be a probability space and let $\mathcal{S} \subseteq \mathcal{F}$ be two $\sigma$ algebras contained in $\mathcal{M}$. Let $f$ be $\mathcal{F}$ measurable and in $L^1(\Omega)$. Then $E(f|\mathcal{S})$, called the conditional expectation of $f$ with respect to $\mathcal{S}$ is defined as follows:

$$E(f|\mathcal{S})$$ is $\mathcal{S}$ measurable

For all $E \in \mathcal{S}$,

$$\int_E E(f|\mathcal{S}) dP = \int_E f dP$$

**Lemma 12.1.2** The above is well defined. Also, if $\mathcal{S} \subseteq \mathcal{F}$ then

$$E(X|\mathcal{S}) = E(E(X|\mathcal{F})|\mathcal{S}).$$

If $Z$ is bounded and measurable in $\mathcal{S}$ then

$$ZE(X|\mathcal{S}) = E(ZX|\mathcal{S}).$$

**Proof:** Let a finite measure on $\mathcal{S}$, $\mu$ be given by

$$\mu(E) = \int_E f dP.$$ 

Then $\mu \ll P$ and so by the Radon Nikodym theorem, there exists a unique $\mathcal{S}$ measurable function, $E(f|\mathcal{S})$ such that

$$\int_E f dP = \mu(E) = \int_E E(f|\mathcal{S}) dP$$

for all $E \in \mathcal{S}$.

Let $F \in \mathcal{S}$. Then

$$\int_F E(E(X|\mathcal{F})|\mathcal{S}) dP = \int_F E(X|\mathcal{F}) dP$$

$$\equiv \int_F E(X|\mathcal{S}) dP$$
and so, by uniqueness, $E(E(X|F)|S) = E(X|S)$. This shows (12.1.1). To establish (12.1.2), note that if $Z = X_F$ where $F \in S$, 
\[
\int X_F E(X|S) \, dP = \int X_F X \, dP = \int E(X_F X|S) \, dP
\]
which shows (12.1.2) in the case where $Z$ is the indicator function of a set in $S$. It follows this also holds for simple functions. Let $\{s_n\}$ be a sequence of simple functions which converges uniformly to $Z$ and let $F \in S$. Then by what was just shown, 
\[
\int_F s_n E(X|S) \, dP = \int_F s_n X \, dP = \int_F E(s_n X|S) \, dP
\]
Now
\[
\left| \int_F E(s_n X|S) \, dP - \int_F E(Z X|S) \, dP \right| 
\leq \int_F |s_n X - Z| X \, dP = \int_F |s_n - Z| |X| \, dP
\]
which converges to 0 by the dominated convergence theorem. Then passing to the limit using the dominated convergence theorem, yields
\[
\int_F Z E(X|S) \, dP = \int_F Z X \, dP \equiv \int_F E(Z X|S) \, dP.
\]
Since this holds for every $F \in S$, this shows (12.1.2). ■

The next major result is a generalization of Jensen’s inequality whose proof depends on the following lemma about convex functions.

**Lemma 12.1.3** Let $\phi$ be a convex real valued function defined on an interval $I$. Then for each $x \in I$, there exists $a_x$ such that for all $t \in I$,
\[
\phi(t) \geq a_x (t - x) + \phi(x).
\]
Also $\phi$ is continuous on $I$.

**Proof:** Let $x \in I$ and let $t > x$. Then by convexity of $\phi$,
\[
\frac{\phi(x + \lambda (t - x)) - \phi(x)}{\lambda (t - x)} \leq \frac{\phi(x) (1 - \lambda) + \lambda \phi(t) - \phi(x)}{\lambda (t - x)}
\]
\[
= \frac{\phi(t) - \phi(x)}{t - x}.
\]
Therefore $t \to \frac{\phi(t) - \phi(x)}{t - x}$ is increasing if $t > x$. If $t < x$
\[
\frac{\phi(x + \lambda (t - x)) - \phi(x)}{\lambda (t - x)} \geq \frac{\phi(x) (1 - \lambda) + \lambda \phi(t) - \phi(x)}{\lambda (t - x)}
\]
\[
= \frac{\phi(t) - \phi(x)}{t - x}
\]
and so $t \to \frac{\phi(t) - \phi(x)}{t - x}$ is increasing for $t \neq x$. Let
\[
a_x \equiv \inf \left\{ \frac{\phi(t) - \phi(x)}{t - x} : t > x \right\}.
\]
Then if $t_1 < x$, and $t > x$,
\[
\frac{\phi(t_1) - \phi(x)}{t_1 - x} \leq a_x \leq \frac{\phi(t) - \phi(x)}{t - x}.
\]
Thus for all $t \in I$,
\[
\phi(t) \geq a_x (t - x) + \phi(x). \tag{12.1.3}
\]

The continuity of $\phi$ follows easily from this and the observation that convexity simply says that the graph of $\phi$ lies below the line segment joining two points on its graph. Thus, we have the following picture which clearly implies continuity. ■
Lemma 12.1.4 Let $I$ be an open interval on $\mathbb{R}$ and let $\phi$ be a convex function defined on $I$. Then there exists a sequence $\{(a_n, b_n)\}$ such that
\[
\phi(t) = \sup \{a_n t + b_n, n = 1, \cdots\}.
\]

Proof: Let $a_x$ be as defined in the above lemma. Let
\[
\psi(x) \equiv \sup \{a_r (x - r) + \phi(r) : r \in \mathbb{Q} \cap I\}.
\]
Thus if $r_1 \in \mathbb{Q}$,
\[
\psi(r_1) \equiv \sup \{a_r (r_1 - r) + \phi(r) : r \in \mathbb{Q} \cap I\} \geq \phi(r_1)
\]
Then $\psi$ is convex on $I$ so $\psi$ is continuous. Therefore,
\[
\psi(t) \geq \phi(t) \geq \sup \{a_r (t - r) + \phi(r), r \in \mathbb{Q} \cap I\} \equiv \psi(t).
\]
Thus $\psi(t) = \phi(t)$. Let $\mathbb{Q} \cap I = \{r_n\}$, $a_n = a_{r_n}$ and $b_n = -a_{r_n}r_n + \phi(r_n)$. ■

Lemma 12.1.5 If $X \leq Y$, then $E(X|S) \leq E(Y|S)$ a.e. Also
\[
X \rightarrow E(X|S)
\]
is linear.

Proof: Let $A \in S$.
\[
\int_A E(X|S) dP \equiv \int_A XdP
\]
\[
\leq \int_Y dP \equiv \int_A E(Y|S) dP.
\]
Hence $E(X|S) \leq E(Y|S)$ a.e. as claimed. It is obvious $X \rightarrow E(X|S)$ is linear.

Theorem 12.1.6 (Jensen’s inequality) Let $X(\omega) \in I$ and let $\phi : I \rightarrow \mathbb{R}$ be convex. Suppose
\[
E(|X|), E(|\phi(X)|) < \infty.
\]
Then
\[
\phi(E(X|S)) \leq E(\phi(X)|S).
\]

Proof: Let $\phi(x) = \sup \{a_n x + b_n\}$. Letting $A \in S$,
\[
\frac{1}{P(A)} \int_A E(X|S) dP = \frac{1}{P(A)} \int_A XdP \in I \text{ a.e.}
\]
whenever $P(A) \neq 0$. Hence $E(X|S)(\omega) \in I$ a.e. and so it makes sense to consider $\phi(E(X|S))$. Now
\[
a_n E(X|S) + b_n = E(a_n X + b_n|S) \leq E(\phi(X)|S).
\]
Thus
\[
\sup \{a_n E(X|S) + b_n\}
\]
\[
= \phi(E(X|S)) \leq E(\phi(X)|S) \text{ a.e. ■}
\]
12.2 Discrete Martingales

Definition 12.2.1 Let \( S_k \) be an increasing sequence of \( \sigma \) algebras which are subsets of \( \mathcal{S} \) and \( X_k \) be a sequence of real-valued random variables with \( E(|X_k|) < \infty \) such that \( X_k \) is \( S_k \) measurable. Then this sequence is called a martingale if
\[
E(X_{k+1}|S_k) = X_k,
\]
a submartingale if
\[
E(X_{k+1}|S_k) \geq X_k,
\]
and a supermartingale if
\[
E(X_{k+1}|S_k) \leq X_k.
\]
Saying that \( X_k \) is \( S_k \) measurable is referred to by saying \( \{X_k\} \) is adapted to \( S_k \).

Note that if \( \{X_k\} \) is a martingale, then \( \{|X_k|\} \) is a submartingale and that if \( \{X_k\} \) is a submartingale and \( \phi \) is convex and increasing, then \( \{\phi(X_k)\} \) is a submartingale.

An upcrossing occurs when a sequence goes from \( a \) up to \( b \). Thus it crosses the interval, \([a, b]\) in the up direction, hence upcrossing. More precisely,

Definition 12.2.2 Let \( \{x_i\}_{i=1}^I \) be any sequence of real numbers, \( I \leq \infty \). Define an increasing sequence of integers \( \{m_k\} \) as follows. \( m_1 \) is the first integer \( \geq 1 \) such that \( x_{m_1} \leq a \), \( m_2 \) is the first integer larger than \( m_1 \) such that \( x_{m_2} \geq b \), \( m_3 \) is the first integer larger than \( m_2 \) such that \( x_{m_3} \leq a \), etc. Then each sequence, \( \{x_{m_1}, \cdots, x_{m_k}\} \), is called an upcrossing of \([a, b]\).

Here is a picture of an upcrossing.

```
  b
     .
  
 a
```

Proposition 12.2.3 Let \( \{X_i\}_{i=1}^n \) be a finite sequence of real random variables defined on \( \Omega \) where \( (\Omega, \mathcal{S}, \mathcal{P}) \) is a probability space. Let \( U_{[a,b]}(\omega) \) denote the number of upcrossings of \( X_i(\omega) \) of the interval \([a, b]\). Then \( U_{[a,b]} \) is a random variable.

**Proof:** Let \( X_0(\omega) \equiv a + 1 \), let \( Y_0(\omega) \equiv 0 \), and let \( Y_k(\omega) \) remain 0 for \( k = 0, \cdots, l \) until \( X_l(\omega) \leq a \). When this happens (if ever), \( Y_{l+1}(\omega) \equiv 1 \). Then let \( Y_i(\omega) \) remain 1 for \( i = l + 1, \cdots, r \) until \( X_r(\omega) \geq b \) when \( Y_{r+1}(\omega) \equiv 0 \). Let \( Y_k(\omega) \) remain 0 for \( k \geq r + 1 \) until \( X_k(\omega) \leq a \) when \( Y_k(\omega) \equiv 1 \) and continue in this way. Thus the upcrossings of \( X_i(\omega) \) are identified as unbroken strings of ones for \( Y_k \) with a zero at each end, with the possible exception of the last string of ones which may be missing the zero at the upper end and may or may not be an upcrossing.

Note also that \( Y_0 \) is measurable because it is identically equal to 0 and that if \( Y_k \) is measurable, then \( Y_{k+1} \) is measurable because the only change in going from \( k \) to \( k + 1 \) is a change from 0 to 1 or from 1 to 0 on a measurable set determined by \( X_k \). In particular,
\[
Y_{k+1}^{-1}(1) = ([Y_k = 1] \cap [X_k < b]) \cup ([Y_k = 0] \cap [X_k \leq a])
\]
This set is in \( \mathcal{S} \) by induction. Of course, \( Y_{k+1}^{-1}(0) \) is just the complement of this set. Thus \( Y_{k+1} \) is \( \mathcal{S} \) measurable since 0, 1 are the only two values possible. Now let
\[
Z_k(\omega) = \begin{cases} 1 & \text{if } Y_k(\omega) = 1 \text{ and } Y_{k+1}(\omega) = 0, \\ 0 & \text{otherwise,} \end{cases}
\]
if \( k < n \) and
\[
Z_n(\omega) = \begin{cases} 1 & \text{if } Y_n(\omega) = 1 \text{ and } X_n(\omega) \geq b, \\ 0 & \text{otherwise.} \end{cases}
\]
Thus \( Z_k(\omega) = 1 \) exactly when an upcrossing has been completed and each \( Z_i \) is a random variable.

\[
U_{[a,b]}(\omega) = \sum_{k=1}^{n} Z_k(\omega)
\]

so \( U_{[a,b]} \) is a random variable as claimed.

The following corollary collects some key observations found in the above construction.

**Corollary 12.2.4** \( U_{[a,b]}(\omega) \leq \) the number of unbroken strings of ones in the sequence, \( \{Y_k(\omega)\} \) there being at most one unbroken string of ones which produces no upcrossing. Also

\[
Y_i(\omega) = \psi_i \left( \{X_j(\omega)\}_{j=1}^{i-1} \right),
\]

where \( \psi_i \) is some function of the past values of \( X_j(\omega) \).

**Lemma 12.2.5** Let \( \phi \) be a convex and increasing function and suppose

\[
\{(X_n, S_n)\}
\]

is a submartingale. Then if \( E(|\phi(X_n)|) < \infty \), it follows

\[
\{(\phi(X_n), S_n)\}
\]

is also a submartingale.

**Proof:** It is given that \( E(X_{n+1}, S_n) \geq X_n \) and so

\[
\phi(X_n) \leq \phi(E(X_{n+1}|S_n)) \leq E(\phi(X_{n+1})|S_n)
\]

by Jensen’s inequality.

The following is called the upcrossing lemma.

**12.2.1 Upcrossings**

**Lemma 12.2.6** (upcrossing lemma) Let \( \{(X_i, S_i)\}_{i=1}^{n} \) be a submartingale and let \( U_{[a,b]}(\omega) \) be the number of upcrossings of \([a, b] \). Then

\[
E(U_{[a,b]}) \leq \frac{E(|X_n|) + |a|}{b - a}.
\]

**Proof:** Let \( \phi(x) = a + (x - a)^+ \) so that \( \phi \) is an increasing convex function always at least as large as \( a \). By Lemma 12.2.4 it follows that \( \{\phi(X_k), S_k\} \) is also a submartingale.

\[
\phi(X_{k+r}) - \phi(X_k) = \sum_{i=k+1}^{k+r} \phi(X_i) - \phi(X_{i-1})
\]

\[
= \sum_{i=k+1}^{k+r} (\phi(X_i) - \phi(X_{i-1})) Y_i + \sum_{i=k+1}^{k+r} (\phi(X_i) - \phi(X_{i-1})) (1 - Y_i).
\]

Observe that \( Y_i \) is \( S_{i-1} \) measurable from its construction in Proposition 12.2.3, \( Y_i \) depending only on \( X_j \) for \( j < i \).

Now let the unbroken strings of ones for \( \{Y_i(\omega)\} \) be

\[
\{k_1, \cdots, k_1 + r_1\}, \{k_2, \cdots, k_2 + r_2\}, \cdots, \{k_m, \cdots, k_m + r_m\}
\]

where \( n = V(\omega) \equiv \) the number of unbroken strings of ones in the sequence \( \{Y_i(\omega)\} \). By Corollary 12.2.4 \( V(\omega) \geq U_{[a,b]}(\omega) \).

\[
\phi(X_n(\omega)) - \phi(X_1(\omega))
\]

\[
= \sum_{k=1}^{n} (\phi(X_k(\omega)) - \phi(X_{k-1}(\omega))) Y_k(\omega)
\]

\[
+ \sum_{k=1}^{n} (\phi(X_k(\omega)) - \phi(X_{k-1}(\omega))) (1 - Y_k(\omega)).
\]
The first sum in the above reduces to summing over the unbroken strings of ones because the terms in which \( Y_i(\omega) = 0 \) contribute nothing. Therefore,
\[
\phi(X_n(\omega)) - \phi(X_1(\omega)) \\
\geq U_{[a,b]}(\omega) (b - a) + 0 + \\
\sum_{k=1}^{n} (\phi(X_k(\omega)) - \phi(X_{k-1}(\omega))) (1 - Y_k(\omega))
\]
where the zero on the right side results from a string of ones which does not produce an upcrossing. It is here that it is important that \( \phi(x) \geq a \). Such a string begins with \( \phi(X_k(\omega)) = a \) and results in an expression of the form \( \phi(X_{k+m}(\omega)) - \phi(X_k(\omega)) \geq 0 \) since \( \phi(X_{k+m}(\omega)) \geq a \). If \( X_k \) had not been replaced with \( \phi(X_k) \), it would have been possible for \( \phi(X_{k+m}(\omega)) \) to be less than \( a \) and the zero in the above could have been a negative number. This would have been inconvenient.

Next take the expected value of both sides in (12.2.6). This results in
\[
E(\phi(X_n) - \phi(X_1)) \geq (b - a) E(U_{[a,b]}) + E\left(\sum_{k=1}^{n} (\phi(X_k) - \phi(X_{k-1}))(1 - Y_k)\right) \\
\geq (b - a) E(U_{[a,b]})
\]
The reason for the last inequality where the term at the end was dropped is
\[
E((\phi(X_k) - \phi(X_{k-1}))(1 - Y_k)) = E(E((\phi(X_k) - \phi(X_{k-1}))(1 - Y_k)|\mathcal{F}_{k-1})) \\
= E((1 - Y_k)E(\phi(X_k)|\mathcal{F}_{k-1}) - (1 - Y_k)E(\phi(X_{k-1})|\mathcal{F}_{k-1})) \\
\geq E((1 - Y_k)(\phi(X_{k-1}) - \phi(X_{k-1}))) = 0.
\]
Recall that \( Y_k \) is \( S_{k-1} \) measurable and that \( (\phi(X_k), S_k) \) is a submartingale.

The reason for this lemma is to prove the amazing submartingale convergence theorem.

### 12.2.2 The Submartingale Convergence Theorem

**Theorem 12.2.7 (submartingale convergence theorem)** Let
\[
\{(X_i, S_i)\}_{i=1}^{\infty}
\]
be a submartingale with \( K \equiv \sup E(|X_n|) < \infty \). Then there exists a random variable, \( X \), such that \( E(|X|) \leq K \) and
\[
\lim_{n \to \infty} X_n(\omega) = X(\omega) \text{ a.e.}
\]

**Proof:** Let \( a, b \in \mathbb{Q} \) and let \( a < b \). Let \( U_{[a,b]}^n(\omega) \) be the number of upcrossings of \( \{X_i(\omega)\}_{i=1}^{n} \). Then let
\[
U_{[a,b]}(\omega) \equiv \lim_{n \to \infty} U_{[a,b]}^n(\omega) = \text{number of upcrossings of } \{X_i\}.
\]
By the upcrossing lemma,
\[
E\left(U_{[a,b]}^n\right) \leq E(|X_n|) + |a| \leq \frac{K + |a|}{b - a}
\]
and so by the monotone convergence theorem,
\[
E(U_{[a,b]}) \leq \frac{K + |a|}{b - a} < \infty
\]
which shows \( U_{[a,b]}(\omega) \) is finite a.e., for all \( \omega \notin S_{[a,b]} \) where \( P(S_{[a,b]}) = 0 \). Define
\[
S \equiv \cup \{S_{[a,b]} : a, b \in \mathbb{Q}, a < b\}.
\]
12.3. Optional Sampling and Stopping Times

Then \( P(S) = 0 \) and if \( \omega \notin S \), \( \{X_k\}_{k=1}^{\infty} \) has only finitely many upcrossings of every interval having rational endpoints. For such \( \omega \) it cannot be the case that

\[
\limsup_{k \to \infty} X_k(\omega) > \liminf_{k \to \infty} X_k(\omega)
\]

because then you could pick rational \( a, b \) such that \([a, b]\) is between the limsup and the liminf and there would be infinitely many upcrossings of \([a, b]\). Thus, for \( \omega \notin S \),

\[
\limsup_{k \to \infty} X_k(\omega) = \liminf_{k \to \infty} X_k(\omega) = \lim_{k \to \infty} X_k(\omega) = X_\infty(\omega).
\]

Letting \( X_\infty(\omega) \equiv 0 \) for \( \omega \in S \), Fatou’s lemma implies

\[
\int \Omega \left| X_\infty \right| dP = \int \Omega \liminf_{n \to \infty} \left| X_n \right| dP \leq \liminf_{n \to \infty} \int \Omega \left| X_n \right| dP \leq K \]

12.2.3 Doob Submartingale Estimate

Another very interesting result about submartingales is the Doob submartingale estimate.

**Theorem 12.2.8** Let \( \{(X_i, S_i)\}_{i=1}^{\infty} \) be a submartingale. Then for \( \lambda > 0 \),

\[
P \left( \max_{1 \leq k \leq n} X_k \geq \lambda \right) \leq \frac{1}{\lambda} \int \Omega X_n^+ dP
\]

**Proof:** Let

\[
A_1 \equiv [X_1 \geq \lambda], A_2 \equiv [X_2 \geq \lambda] \setminus A_1, \\
\cdots, A_k \equiv [X_k \geq \lambda] \setminus (\cup_{i=1}^{k-1} A_i), \cdots
\]

Thus each \( A_k \) is \( S_k \) measurable, the \( A_k \) are disjoint, and their union equals \( \max_{1 \leq k \leq n} X_k \geq \lambda \). Therefore from the definition of a submartingale and Jensen’s inequality,

\[
P \left( \max_{1 \leq k \leq n} X_k \geq \lambda \right) = \sum_{k=1}^{n} P(A_k) \leq \frac{1}{\lambda} \sum_{k=1}^{n} \int_{A_k} X_k dP
\]

\[
\leq \frac{1}{\lambda} \sum_{k=1}^{n} \int_{A_k} E(X_n|S_k) dP
\]

\[
\leq \frac{1}{\lambda} \sum_{k=1}^{n} \int_{A_k} E(X_n^+|S_k) dP
\]

\[
\leq \frac{1}{\lambda} \sum_{k=1}^{n} \int_{A_k} X_n^+ dP \leq \frac{1}{\lambda} \int \Omega X_n^+ dP.
\]

12.3 Optional Sampling and Stopping Times

12.3.1 Stopping Times and Their Properties

First it is necessary to define the notion of a stopping time.

**Definition 12.3.1** Let \((\Omega, \mathcal{F}, P)\) be a probability space and let \( \{\mathcal{F}_n\}_{n=1}^{\infty} \) be an increasing sequence of \( \sigma \) algebras each contained in \( \mathcal{F} \). A stopping time is a measurable function, \( T \) which maps \( \Omega \) to \( \mathbb{N} \),

\[
T^{-1}(A) \in \mathcal{F} \text{ for all } A \in \mathcal{P}(\mathbb{N}),
\]
such that for all \( n \in \mathbb{N} \),
\[
[T \leq n] \in \mathcal{F}_n.
\]

Note this is equivalent to saying
\[
[T = n] \in \mathcal{F}_n
\]
because
\[
[T = n] = [T \leq n] \setminus [T \leq n - 1].
\]

For \( T \) a stopping time define \( \mathcal{F}_T \) as follows.
\[
\mathcal{F}_T \equiv \{ A \in \mathcal{F} : A \cap [T \leq n] \in \mathcal{F}_n \text{ for all } n \in \mathbb{N} \}
\]

These sets in \( \mathcal{F}_T \) are referred to as “prior” to \( T \).

The following lemma is fundamental to understand.

**Lemma 12.3.2** In the situation of Definition 12.3.1, if \( S \leq T \) for two stopping times, \( S \) and \( T \), then \( \mathcal{F}_S \subseteq \mathcal{F}_T \). Also \( \mathcal{F}_T \) is a \( \sigma \) algebra.

**Proof:** Let \( A \in \mathcal{F}_S \). Then this means
\[
A \cap [S \leq n] \in \mathcal{F}_n \text{ for all } n.
\]

Then
\[
A \cap [T \leq n] = \bigcup_{i=1}^n (A \cap [S \leq i]) \cap [T \leq n] \tag{12.3.7}
\]

Suppose \( \omega \) is in the set on the left. Then if \( T(\omega) < n \), it is clearly in the set on the right. If \( T(\omega) = n \), then \( \omega \in [S \leq i] \) for some \( i \leq n \) and it is also in \( [T \leq n] \). Thus the set on the left is contained in the set on the right. Next suppose \( \omega \) is in the set on the right. Then \( \omega \in [T \leq n] \) and it only remains to verify \( \omega \in A \). However, \( \omega \in A \cap [S \leq i] \) for some \( i \) and so \( \omega \in A \) also.

Now from (12.3.7) it follows \( A \cap [T \leq n] \in \mathcal{F}_n \) because
\[
A \cap [S \leq i] \in \mathcal{F}_i \subseteq \mathcal{F}_n
\]
and \( [T \leq n] \in \mathcal{F}_n \) because \( T \) is a stopping time. Since \( n \) is arbitrary, this shows \( \mathcal{F}_S \subseteq \mathcal{F}_T \).

It remains to verify \( \mathcal{F}_T \) is a \( \sigma \) algebra. Suppose \( \{A_i\} \) is a sequence of sets in \( \mathcal{F}_T \). Then I need to show that
\[
\bigcup_{i=1}^\infty A_i \cap [T \leq j] = \bigcup_{i=1}^\infty (A_i \cap [T \leq j])
\]

Now each \( (A_i \cap [T \leq j]) \) is in \( \mathcal{F}_j \) and so the countable union of these sets is also in \( \mathcal{F}_j \). Next suppose \( A \in \mathcal{F}_T \). I need to verify \( A^c \cap [T \leq j] \in \mathcal{F}_j \) for all \( j \). However, \( [T \leq j] \in \mathcal{F}_j \) and \( \Omega \in \mathcal{F}_j \) so \( \Omega \in \mathcal{F}_T \). Thus
\[
\Omega \cap [T \leq j] = (A \cap [T \leq j]) \cup (A^c \cap [T \leq j])
\]
and so
\[
(A^c \cap [T \leq j]) = \Omega \cap [T \leq j] \setminus (A \cap [T \leq j]) \in \mathcal{F}_j.
\]

This proves the lemma.

**Lemma 12.3.3** Let \( T \) be a stopping time and let \( \{X_n\} \) be a sequence of random variables such that \( X_n \) is \( \mathcal{F}_n \) measurable. Then \( X_T(\omega) \equiv X_{T(\omega)}(\omega) \) is also a random variable and it is measurable with respect to \( \mathcal{F}_T \).

**Proof:** I assume the \( X_n \) have values in some topological space and each is measurable because the inverse image of an open set is in \( \mathcal{F}_n \). I need to show \( X_T^{-1}(U) \cap [T \leq n] \in \mathcal{F}_n \) for all \( n \) whenever \( U \) is open.

\[
X_T^{-1}(U) = \bigcup_{i=1}^\infty X_i^{-1}(U) \cap [T = i].
\]

It follows \( X_T^{-1}(U) \in \mathcal{F} \). Furthermore,
\[
X_T^{-1}(U) \cap [T \leq n] = \bigcup_{i=1}^n X_i^{-1}(U) \cap [T = i] \cap [T \leq n] = X_i^{-1}(U) \cap [T = i] \cap [T \leq n]
\]
and so \( X_T \) is \( \mathcal{F}_T \) measurable as claimed. This proves the lemma.
Lemma 12.3.4 Let $S \leq T$ be two stopping times such that $T$ is bounded above and let $\{X_n\}$ be a submartingale (martingale) adapted to the increasing sequence of $\sigma$ algebras, $\{\mathcal{F}_n\}$. Then

$$E(X_T|\mathcal{F}_S) \geq X_S$$

in the case where $\{X_n\}$ is a submartingale and

$$E(X_T|\mathcal{F}_S) = X_S$$

in the case where $\{X_n\}$ is a martingale.

Proof: I will prove the case where $\{X_n\}$ is a submartingale and note the other case will only involve replacing $\geq$ with $=$. First recall that from Lemma 12.3.2, $\mathcal{F}_S \subseteq \mathcal{F}_T$. Also let $m$ be an upper bound for $T$. Then it follows from this that

$$E(|X_T|) = \sum_{i=1}^{m} \int_{\{T=i\}} |X_i| \, dP < \infty$$

with a similar formula holding for $E(|X_S|)$. Thus it makes sense to speak of $E(X_T|\mathcal{F}_S)$.

I need to show that if $B \in \mathcal{F}_S$, so that $B \cap [S \leq n] \in \mathcal{F}_n$ for all $n$, then

$$\int_B X_T \, dP \geq \int_B X_S \, dP. \quad (12.3.8)$$

It suffices to do this for $B$ of the special form

$$B = A \cap \left[ S = i \right]$$

because if this is done, then the result follows from summing over all possible values of $S$. Note that if $B = A \cap \left[ S = m \right]$, then $X_T = X_S = X_m$ and there is nothing to prove in (12.3.8) so it can be assumed $i \leq m - 1$. Then let $B$ be of this form.

$$\int_{A \cap \left[ S = i \right]} X_T \, dP = \sum_{j=i}^{m-1} \int_{A \cap \left[ S = i \cap T = j \right]} X_T \, dP$$

$$= \sum_{j=i}^{m-1} \int_{A \cap \left[ S = i \cap T = j \right]} X_T \, dP + \int_{A \cap \left[ S = i \cap T \geq m \right]} X_m \, dP$$

And so

$$\int_{A \cap \left[ S = i \right]} X_T \, dP = \sum_{j=i}^{m-1} \int_{A \cap \left[ S = i \cap T = j \right]} X_T \, dP + \int_{A \cap \left[ S = i \cap T \geq m \right]} X_m \, dP \quad (12.3.9)$$

$$= \sum_{j=i}^{m-1} \int_{A \cap \left[ S = i \cap T = j \right]} X_T \, dP + \int_{A \cap \left[ S = i \cap T \leq m - 1 \right]^C} X_m \, dP$$

$$\geq \sum_{j=i}^{m-1} \int_{A \cap \left[ S = i \cap T = j \right]} X_T \, dP + \int_{A \cap \left[ S = i \cap T \leq m - 1 \right]^C} X_{m-1} \, dP$$

$$= \sum_{j=i}^{m-1} \int_{A \cap \left[ S = i \cap T = j \right]} X_T \, dP + \int_{A \cap \left[ S = i \cap T > m - 1 \right]} X_{m-1} \, dP$$

provided $m - 1 \geq i$ because $\{X_n\}$ is a submartingale and

$$A \cap \left[ S = i \right] \cap \left[ T \leq m - 1 \right]^C \in \mathcal{F}_{m-1}$$

Now combine the top term of the sum with the term on the right to obtain

$$= \sum_{j=i}^{m-2} \int_{A \cap \left[ S = i \cap T = j \right]} X_T \, dP + \int_{A \cap \left[ S = i \cap T \geq m - 1 \right]} X_{m-1} \, dP$$

which is exactly the same form as (12.3.9) except m is replaced with m − 1. Now repeat this process till you get the following inequality

\[ \int_{A \cap [S=i]} X_T dP \geq \sum_{j=i}^{i+1} \int_{A \cap [S=i \cap T=j]} X_T dP + \int_{A \cap [S=i \cap T \geq i + 2]} X_{i+2} dP \]

The right hand side equals

\[ \sum_{j=i}^{i+1} \int_{A \cap [S=i \cap T=j]} X_T dP + \int_{A \cap [S=i \cap T \leq i + 1]} X_{i+1} dP \]

\[ = \int_{A \cap [S=i \cap T=i]} X_T dP + \int_{A \cap [S=i \cap T \geq i + 1]} X_{i+1} dP \]

\[ = \int_{A \cap [S=i \cap T=i]} X_T dP + \int_{A \cap [S=i \cap T \leq i]} X_{i+1} dP \]

In the case where \( \{X_n\} \) is a martingale, you replace every occurrence of \( \geq \) in the above argument with \( = \). This proves the lemma.

This lemma is called the optional sampling theorem. Another version of this theorem is the case where you have an increasing sequence of stopping times, \( \{T_n\}_{n=1}^{\infty} \). Thus if \( \{X_n\} \) is a sequence of random variables each \( F_n \) measurable, the sequence \( \{X_{T_n}\} \) is also a sequence of random variables such that \( X_{T_n} \) is measurable with respect to \( F_{T_n} \) where \( F_{T_n} \) is an increasing sequence of \( \sigma \) fields. In the case where \( X_n \) is a submartingale (martingale) it is reasonable to ask whether \( \{X_{T_n}\} \) is also a submartingale (martingale). The optional sampling theorem says this is often the case.

**Theorem 12.3.5** Let \( \{T_n\} \) be an increasing bounded sequence of stopping times and let \( \{X_n\} \) be a submartingale (martingale) adapted to the increasing sequence of \( \sigma \) algebras, \( \{F_n\} \). Then \( \{X_{T_n}\} \) is a submartingale (martingale) adapted to the increasing sequence of \( \sigma \) algebras \( \{F_{T_n}\} \).

**Proof:** This follows from Lemma 12.3.4.

**Example 12.3.6** Let \( \{X_n\} \) be a sequence of real random variables such that \( X_n \) is \( F_n \) measurable and let \( A \) be a Borel subset of \( \mathbb{R} \). Let \( T(\omega) \) denote the first time \( X_n(\omega) \) is in \( A \). Then \( T \) is a stopping time. It is called the first hitting time.

To see this is a stopping time,

\[ [T \leq t] = \bigcup_{i=1}^{t} X_i^{-1}(A) \in \mathcal{F}_t. \]
12.4 Optional Stopping Times And Martingales

12.4.1 Stopping Times And Their Properties

The purpose of this section is to consider a special optional sampling theorem for martingales which is superior to the one presented earlier. I have presented a different treatment of the fundamental properties of stopping times also. See Kallenberg [53] for more.

**Definition 12.4.1** Let \((Ω, \mathcal{F}, P)\) be a probability space and let \(\{\mathcal{F}_n\}_{n=1}^{∞}\) be an increasing sequence of \(σ\) algebras each contained in \(\mathcal{F}\). A stopping time is a measurable function, \(τ\) which maps \(Ω\) to \(\mathbb{N}\), 

\[
τ^{-1}(A) ∈ \mathcal{F} \text{ for all } A ∈ \mathcal{P}(\mathbb{N}),
\]

such that for all \(n ∈ \mathbb{N}\),

\[
[τ ≤ n] ∈ \mathcal{F}_n.
\]

Note this is equivalent to saying

\[
[τ = n] ∈ \mathcal{F}_n
\]

because

\[
[τ = n] = [τ ≤ n] \setminus [τ ≤ n − 1].
\]

For \(τ\) a stopping time define \(\mathcal{F}_τ\) as follows.

\[
\mathcal{F}_τ \equiv \{A ∈ \mathcal{F} : A ∩ [τ ≤ n] ∈ \mathcal{F}_n \text{ for all } n ∈ \mathbb{N}\}
\]

These sets in \(\mathcal{F}_τ\) are referred to as “prior” to \(τ\).

First note that for \(τ\) a stopping time, \(\mathcal{F}_τ\) is a \(σ\) algebra. This is in the next proposition.

**Proposition 12.4.2** For \(τ\) a stopping time, \(\mathcal{F}_τ\) is a \(σ\) algebra and if \(Y(k)\) is \(\mathcal{F}_k\) measurable for all \(k\), then

\[
ω → Y(τ(ω))
\]

is \(\mathcal{F}_τ\) measurable.

**Proof:** Let \(A_n ∈ \mathcal{F}_τ\). I need to show \(∪_n A_n ∈ \mathcal{F}_τ\). In other words, I need to show that

\[
∪_n A_n ∩ [τ ≤ k] ∈ \mathcal{F}_k
\]

The left side equals

\[
∪_n (A_n ∩ [τ ≤ k])
\]

which is a countable union of sets of \(\mathcal{F}_k\) and so \(\mathcal{F}_τ\) is closed with respect to countable unions. Next suppose \(A ∈ \mathcal{F}_τ\).

\[
(A^C ∩ [τ ≤ k]) \cup (A ∩ [τ ≤ k]) = Ω ∩ [τ ≤ k]
\]

and \(Ω ∩ [τ ≤ k] ∈ \mathcal{F}_k\). Therefore, so is \(A^C ∩ [τ ≤ k]\). It remains to verify the last claim.

\[
[Y(τ) ≤ a] = ∪_k [τ = k] ∩ [Y(k) ≤ a]
\]

Thus

\[
[Y(τ) ≤ a] ∩ [τ ≤ l] = ∪_k [τ = k] ∩ [Y(k) ≤ a] ∩ [τ ≤ l]
\]

Consider a term in the union. If \(l ≥ k\) the term reduces to \([τ = k] ∩ [Y(k) ≤ a] ∈ \mathcal{F}_k\) while if \(l < k\), this term reduces to \(∅\), also a set of \(\mathcal{F}_k\). Therefore, \(Y(τ)\) must be \(\mathcal{F}_τ\) measurable. This proves the proposition.

The following lemma gives the fundamental properties of stopping times.

**Lemma 12.4.3** In the situation of Definition 12.4.1, let \(σ, τ\) be two stopping times. Then

1. \(τ\) is \(\mathcal{F}_τ\) measurable
2. \(\mathcal{F}_σ ∩ [σ ≤ τ] ≤ \mathcal{F}_σ ∧ τ = \mathcal{F}_σ ∩ \mathcal{F}_τ\)
3. \( \mathcal{F}_\tau = \mathcal{F}_k \) on \([\tau = k]\) for all \(k\). That is if \(A \in \mathcal{F}_k\), then \(A \cap [\tau = k] \in \mathcal{F}_\tau\) and if \(A \in \mathcal{F}_\tau\) then \(A \cap [\tau = k] \in \mathcal{F}_k\). Also if \(A \in \mathcal{F}_\tau\),

\[
\int_{A \cap [\tau = k]} E(Y|\mathcal{F}_\tau) \, dP = \int_{A \cap [\tau = k]} E(Y|\mathcal{F}_k) \, dP
\]

and

\[E(Y|\mathcal{F}_\tau) = E(Y|\mathcal{F}_k) \text{ a.e.}\]
on \([\tau = k]\).

**Proof:** Consider the first claim. I need to show that \([\tau \leq a] \cap [\tau \leq k] \in \mathcal{F}_k\) for every \(k\). However, this is easy if \(a \geq k\) because the left side is then \([\tau \leq k]\) which is given to be in \(\mathcal{F}_k\) since \(\tau\) is a stopping time. If \(a < k\), it is also easy because then the left side is \([\tau \leq a] \in \mathcal{F}_{[a]}\) where \([a]\) is the greatest integer less than or equal to \(a\).

Next consider the second claim. Let \(A \in \mathcal{F}_\sigma\). I want to show first that

\[A \cap [\sigma \leq \tau] \in \mathcal{F}_\tau\]

(12.4.10)

In other words, I want to show

\[A \cap [\sigma \leq \tau] \cap [\tau \leq k] \in \mathcal{F}_k\]

for all \(k\). This will be done if I can show

\[A \cap [\sigma \leq j] \cap [\tau \leq k] \in \mathcal{F}_k\]

for each \(j \leq k\) because

\[\cup_{j \leq k} A \cap [\sigma \leq j] \cap [\tau \leq k] = A \cap [\sigma \leq \tau] \cap [\tau \leq k]\]

However, since \(\sigma \in \mathcal{F}_\sigma\), it follows \(A \cap [\sigma \leq j] \in \mathcal{F}_j \subseteq \mathcal{F}_k\) for each \(j \leq k\) and \([\tau \leq k] \in \mathcal{F}_k\) and so this has shown what I wanted to show, \(A \cap [\sigma \leq \tau] \in \mathcal{F}_\tau\).

Now replace the stopping time, \(\tau\) with the stopping time \(\tau \wedge \sigma\) in what was just shown. Note

\[[\tau \wedge \sigma \leq n] = [\tau \leq n] \cup [\sigma \leq n] \in \mathcal{F}_n\]

so \(\tau \wedge \sigma\) really is a stopping time. This yields

\[A \cap [\sigma \leq \tau \wedge \sigma] \in \mathcal{F}_{\tau \wedge \sigma}\]

However the left side equals \(A \cap [\sigma \leq \tau]\). Thus

\[A \cap [\sigma \leq \tau] \in \mathcal{F}_{\tau \wedge \sigma}\]

This has shown the first part of 2.), \(\mathcal{F}_\sigma \cap [\sigma \leq \tau] \subseteq \mathcal{F}_{\tau \wedge \sigma}\). Now (12.4.10) implies if \(A \in \mathcal{F}_{\sigma \wedge \tau}\),

\[A = A \cap \underbrace{\Omega}_{\text{all of } \Omega} \in \mathcal{F}_\tau\]

and so \(\mathcal{F}_{\sigma \wedge \tau} \subseteq \mathcal{F}_\tau\). Similarly, \(\mathcal{F}_{\sigma \wedge \tau} \subseteq \mathcal{F}_\sigma\) which shows

\[\mathcal{F}_{\sigma \wedge \tau} \subseteq \mathcal{F}_\tau \cap \mathcal{F}_\sigma\]

Next let \(A \in \mathcal{F}_\tau \cap \mathcal{F}_\sigma\). Then is it in \(\mathcal{F}_{\sigma \wedge \tau}\)? Is \(A \cap [\sigma \wedge \tau \leq k] \in \mathcal{F}_k\)? Of course this is so because

\[A \cap [\sigma \wedge \tau \leq k] = A \cap ([\sigma \leq k] \cup [\tau \leq k]) = (A \cap [\sigma \leq k]) \cup (A \cap [\tau \leq k]) \in \mathcal{F}_k\]

since both \(\sigma, \tau\) are stopping times. This proves part 2.).

Now consider part 3.). Note that \([\tau = k]\) is in both \(\mathcal{F}_k\) and \(\mathcal{F}_\tau\). Let \(A \in \mathcal{F}_k\). I need to show

\[\mathcal{F}_\tau \cap [\tau = k] = \mathcal{F}_k \cap [\tau = k]\]

where \(G \cap [\tau = k]\) means all sets of the form \(A \cap [\tau = k]\) where \(A \in G\). Let \(A \in \mathcal{F}_\tau\). Then

\[A \cap [\tau = k] = (A \cap [\tau \leq k]) \setminus (A \cap [\tau \leq k - 1]) \in \mathcal{F}_k\]
Therefore, there exists $B \in \mathcal{F}_k$ such that $B = A \cap [\tau = k]$ and so

$$B \cap [\tau = k] = A \cap [\tau = k]$$

which shows $\mathcal{F}_\tau \cap [\tau = k] \subseteq \mathcal{F}_k \cap [\tau = k]$. Now let $A \in \mathcal{F}_k$ so that

$$A \cap [\tau = k] \in \mathcal{F}_k \cap [\tau = k]$$

Then

$$A \cap [\tau = k] \cap [\tau \leq j] \in \mathcal{F}_j$$

because in case $j < k$, the set on the left is $\emptyset$ and if $j \geq k$ it reduces to $A \cap [\tau = k]$ and both $A$ and $[\tau = k]$ are in $\mathcal{F}_k \subseteq \mathcal{F}_j$. Therefore, the two $\sigma$ algebras of subsets of $[\tau = k]$,

$$\mathcal{F}_\tau \cap [\tau = k], \mathcal{F}_k \cap [\tau = k]$$

are equal. Thus for $A$ in either $\mathcal{F}_\tau$ or $\mathcal{F}_k$, $A \cap [\tau = k]$ is a set of both $\mathcal{F}_\tau$ and $\mathcal{F}_k$ because if $A \in \mathcal{F}_k$, then from the above, there exists $B \in \mathcal{F}_\tau$ such that

$$A \cap [\tau = k] = B \cap [\tau = k] \in \mathcal{F}_\tau$$

with similar reasoning holding if $A \in \mathcal{F}_\tau$. In other words, if $g$ is $\mathcal{F}_\tau$ or $\mathcal{F}_k$ measurable, then the restriction of $g$ to $[\tau = k]$ is measurable with respect to $\mathcal{F}_\tau \cap [\tau = k]$ and $\mathcal{F}_k \cap [\tau = k]$. Let $Y$ be an arbitrary random variable in $L^1(\Omega, \mathcal{F})$. It follows

$$\int_{A \cap [\tau = k]} E(Y|\mathcal{F}_\tau) \, dP = \int_{A \cap [\tau = k]} Y \, dP$$

$$= \int_{A \cap [\tau = k]} E(Y|\mathcal{F}_k) \, dP$$

Since this holds for an arbitrary set in $\mathcal{F}_\tau \cap [\tau = k] = \mathcal{F}_k \cap [\tau = k]$, it follows

$$E(Y|\mathcal{F}_\tau) = E(Y|\mathcal{F}_k) \text{ a.e. on } [\tau = k]$$

This proves the third claim and the Lemma.

With this lemma, here is a major theorem, the optional sampling theorem of Doob. This one is special for martingales.

**Theorem 12.4.4** Let $\{M(k)\}$ be a real valued martingale with respect to the increasing sequence of $\sigma$ algebras, $\{\mathcal{F}_k\}$ and let $\sigma, \tau$ be two stopping times such that $\tau$ is bounded. Then $M(\tau)$ defined as

$$\omega \to M(\tau(\omega))$$

is integrable and

$$M(\sigma \wedge \tau) = E(M(\tau)|\mathcal{F}_\sigma).$$

**Proof:** Consider first the measurability of $\omega \to M(\tau(\omega))$. Say $\tau \leq l$.

$$[M(\tau) \leq a] = \bigcup_{i=1}^l [\tau = i] \cap [M(i) \leq a] \in \mathcal{F}_l$$

so there is at least no problem with measurability. By Proposition 12.4.3, $M(\tau)$ is also $\mathcal{F}_\tau$ measurable.

Next note that since $\tau$ is bounded,

$$\int_{\Omega} |M(\tau(\omega))| \, dP \leq \sum_{i=1}^l \int_{[\tau = i]} |M(i)| \, dP < \infty.$$ 

This proves the first assertion and makes possible the consideration of conditional expectation.

Let $l \geq \tau$ as described above. Then for $k \leq l$, by Lemma 12.4.3,

$$\mathcal{F}_k \cap [\tau = k] = \mathcal{F}_\tau \cap [\tau = k]$$
and so if \( A \in \mathcal{F}_k \cap [\tau = k] \),
\[
\int_A E(M(l) \mid \mathcal{F}_\tau) \, dP = \int_A M(l) \, dP = \int_A E(M(l) \mid \mathcal{F}_k) \, dP = \int_A M(k) \, dP = \int_A M(\tau) \, dP.
\]

Therefore, since \( A \) was arbitrary,
\[
E(M(l) \mid \mathcal{F}_\tau) = M(\tau) \text{ a.e.}
\]
on \([\tau = k]\) for every \( k \leq l \). It follows
\[
E(M(l) \mid \mathcal{F}_\tau) = M(\tau) \text{ a.e.}
\]
since it is true on each \([\tau = k]\) for all \( k \leq l \). Now let
\[
A \in \mathcal{F}_k \cap [\tau = k] = \mathcal{F}_\tau \cap [\tau = k]
\]
Then from the above lemma, \( A \in \mathcal{F}_\tau \cap \mathcal{F}_k \). Therefore,
\[
\int_A E(M(l) \mid \mathcal{F}_\tau) \equiv \int_A M(l) = \int_A E(M(l) \mid \mathcal{F}_k) = \int_A M(k) = \int_A M(\tau)
\]
the last because on \( A, \tau = k \). Thus, since \( A \) arbitrary in the \( \sigma \) algebra \( \mathcal{F}_k \cap [\tau = k] = \mathcal{F}_\tau \cap [\tau = k] \), it must be the case that
\[
E(M(l) \mid \mathcal{F}_\tau) = M(\tau) \text{ a.e. on } [\tau = k]
\]
There are only finitely many values for \( \tau \) and so this shows \( E(M(l) \mid \mathcal{F}_\tau) = M(\tau) \text{ a.e.} \) Consider \( E(M(\tau) \mid \mathcal{F}_\sigma) \) on \([\tau = i] \cap [\sigma = j]\).
\[
E(M(\tau) \mid \mathcal{F}_\sigma) = E(E(M(l) \mid \mathcal{F}_\tau) \mid \mathcal{F}_\sigma) = E(E(M(l) \mid \mathcal{F}_i) \mid \mathcal{F}_j)
\]
Case 1: \( i \geq j \). Then the right side reduces to \( E(M(l) \mid \mathcal{F}_j) = M(j) = M(\tau \land \sigma) \).
Case 2: \( i < j \). Then the right side reduces to \( E(M(l) \mid \mathcal{F}_i) = M(i) = M(\tau \land \sigma) \).
This is a really amazing theorem. Note it says \( M(\sigma \land \tau) = E(M(\tau) \mid \mathcal{F}_\sigma) \). This would not be so surprising if it had said
\[
M(\sigma \land \tau) = E(M(\tau) \mid \mathcal{F}_{\sigma \land \tau}).
\]
What about submartingales? Recall \( \{X(k)\}_{k=0}^\infty \) is a submartingale if
\[
E(X(k+1) \mid \mathcal{F}_k) \geq X(k)
\]
where the \( \mathcal{F}_k \) are an increasing sequence of \( \sigma \) algebras in the usual way. The following is a very interesting result.

**Lemma 12.4.5** Let \( \{X(k)\}_{k=0}^\infty \) be a submartingale adapted to the increasing sequence of \( \sigma \) algebras, \( \{\mathcal{F}_k\} \). Then there exists a unique increasing process \( \{A(k)\}_{k=0}^\infty \) such that \( A(0) = 0 \) and \( A(k+1) \) is \( \mathcal{F}_k \) measurable for all \( k \) and a martingale, \( \{M(k)\}_{k=0}^\infty \) such that
\[
X(k) = A(k) + M(k).
\]
Furthermore, for \( \tau \) a stopping time, \( A(\tau) \) is \( \mathcal{F}_\tau \) measurable.

**Proof:** Define \( \sum_{k=0}^{n-1} \neq 0 \). First consider the uniqueness assertion. Suppose \( A \) is a process which does what is supposed to do.
\[
\sum_{k=0}^{n-1} E(X(k+1) - X(k) \mid \mathcal{F}_k) = \sum_{k=0}^{n-1} E(A(k+1) - A(k) \mid \mathcal{F}_k)
\]
\[
+ \sum_{k=0}^{n-1} E(M(k+1) - M(k) \mid \mathcal{F}_k)
\]
Then since \( \{M(k)\} \) is a martingale,
\[
\sum_{k=0}^{n-1} E(X(k+1) - X(k) \mid \mathcal{F}_k) = \sum_{k=0}^{n-1} A(k+1) - A(k) = A(n)
\]
This shows uniqueness and gives a formula for $A(n)$ assuming it exists. It is only a matter of verifying this does work. Define
\[ A(n) = \sum_{k=0}^{n-1} E (X(k+1) - X(k) | F_k), \quad A(0) = 0. \]
Then $A$ is increasing because from the definition,
\[ A(n+1) - A(n) = E (X(n+1) - X(n) | F_n) \geq 0. \]
Also from the definition above, $A(n)$ is $F_{n-1}$ measurable, consider
\[ \{X(k) - A(k)\}. \]

Why is this a martingale?
\[
E (X(k+1) - A(k+1) | F_k) \\
= E (X(k+1) | F_k) - A(k+1) \\
= E (X(k+1) | F_k) - \sum_{j=0}^{k} E (X(j+1) - X(j) | F_j) \\
= E (X(k+1) | F_k) - E (X(k+1) - X(k) | F_k) \\
- \sum_{j=0}^{k-1} E (X(j+1) - X(j) | F_j) \\
= X(k) - \sum_{j=0}^{k-1} E (X(j+1) - X(j) | F_j) = X(k) - A(k)
\]
Let $M(k) \equiv X(k) - A(k)$. $A(\tau)$ is $F_\tau$ measurable by Proposition 13.7.3.

Note the nonnegative integers could be replaced with any finite set or ordered countable set of numbers with no change in the conclusions of this lemma or the above optional sampling theorem.

Next consider the case of a submartingale.

**Theorem 12.4.6** Let \{X(k)\} be a submartingale with respect to the increasing sequence of $\sigma$ algebras, \{F_k\} and let $\sigma, \tau$ be two stopping times such that $\tau$ is bounded. Then $X(\tau)$ defined as
\[ \omega \to X(\tau(\omega)) \]
is integrable and
\[ X(\sigma \wedge \tau) \leq E (X(\tau) | F_\sigma). \]

**Proof:** The claim about $X(\tau)$ being integrable is the same as in Theorem 13.7.5. If $\tau \leq l$,
\[
E ([X(\tau(\omega))] = \sum_{i=1}^{l} \int_{[\tau=i]} |X(i)| dP < \infty
\]
By Lemma 12.4.6 there is a martingale, \{M(k)\} and an increasing process \{A(k)\} such that $A(k+1)$ is $F_k$ measurable such that
\[ X(k) = M(k) + A(k). \]
Then using Theorem 13.7.5 on the martingale and the fact $A$ is increasing
\[
E (X(\tau) | F_\sigma) = E (M(\tau) + A(\tau) | F_\sigma) = M(\tau \wedge \sigma) + E (A(\tau) | F_\sigma) \\
\geq M(\tau \wedge \sigma) + E (A(\tau \wedge \sigma) | F_\sigma) \\
= M(\tau \wedge \sigma) + A(\tau \wedge \sigma) = X(\tau \wedge \sigma). \]
because in the above, it follows from Lemma 12.4.6 $A(\tau \wedge \sigma)$ is $F_{\tau \wedge \sigma}$ measurable and from Lemma 13.7.4
\[ F_{\tau \wedge \sigma} = F_\tau \cap F_\sigma \subseteq F_\sigma \]
and so
\[ E (A(\tau \wedge \sigma) | F_\sigma) = A(\tau \wedge \sigma). \]
12.5 Submartingale Convergence Theorem

12.5.1 Upcrossings

Let \( \{X(k)\} \) be an adapted stochastic process, \( k = 0, 1, 2, \ldots, M \) adapted to the increasing \( \sigma \) algebras \( \mathcal{F}_k \). Also let [a, b] be an interval. An upcrossing occurs when \( X(k) < a \) and you have \( X(k+1) > b \) while \( X(r) < b \) for all \( r \in [k, k+1) \). In order to understand upcrossings, consider the following:

\[
\tau_0 = \min(\inf\{k : X(k) \leq a\}, M),
\tau_1 = \min(\inf\{k : (X(k \lor \tau_0) - X(\tau_0))_+ \geq b - a\}, M),
\tau_2 = \min(\inf\{k : (X(\tau_1) - X(k \lor \tau_1))_+ \geq b - a\}, M),
\tau_3 = \min(\inf\{k : (X(k \lor \tau_2) - X(\tau_2))_+ \geq b - a\}, M),
\tau_4 = \min(\inf\{k : (X(\tau_3) - X(k \lor \tau_3))_+ \geq b - a\}, M),
\vdots
\]

As usual, \( \inf(\emptyset) \equiv \infty \). Are the above stopping times? If \( \alpha \geq 0 \), and \( \tau \) is a stopping time, is \( k \to (X(\tau) - X(k \lor \tau))_+ \) adapted?

\[
[(X(\tau) - X(k \lor \tau))_+ > \alpha] = [(X(\tau) - X(k))_+ > \alpha] \cap [\tau \leq k]
\]

Now

\[
[(X(\tau) - X(k))_+ > \alpha] \cap [\tau \leq k] = \bigcup_{i=0}^k [(X(i) - X(k))_+ > \alpha] \cap [\tau \leq k] \in \mathcal{F}_k
\]

If \( \alpha < 0 \), then \( [(X(\tau_1) - X(k \lor \tau_1))_+ > \alpha] = \Omega \) and so \( k \to (X(\tau) - X(k \lor \tau))_+ \) is adapted. Similarly \( k \to (X(k \lor \tau) - X(\tau))_+ \) is adapted. Therefore, all those \( \tau_k \) are stopping times.

Now consider the following random variable for odd \( M, 2n + 1 = M \)

\[
U_M^{[a, b]} = \lim_{\varepsilon \to 0} \sum_{k=0}^n \frac{X(\tau_{2k+1}) - X(\tau_{2k})}{\varepsilon + X(\tau_{2k+1}) - X(\tau_{2k})} \leq \frac{1}{b-a} \sum_{k=0}^n X(\tau_{2k+1}) - X(\tau_{2k})
\]

Now suppose \( \{X(k)\} \) is a nonnegative submartingale. Then since \( E(X(2\tau)|\mathcal{F}_{2\tau-1}) \geq X(\tau_{2k-1}) \)

\[
E\left(\sum_{k=1}^n X(\tau_{2k}) - X(\tau_{2k-1})\right) \geq 0
\]

Hence

\[
E\left(U_M^{[a, b]}\right) \leq \frac{1}{b-a} \sum_{k=0}^n E(X(\tau_{2k+1}) - X(\tau_{2k}))
\]

\[
\leq \frac{1}{b-a} \sum_{k=0}^n E(X(\tau_{2k+1}) - X(\tau_{2k})) + \frac{1}{b-a} \sum_{k=1}^n E(X(\tau_{2k}) - X(\tau_{2k-1}))
\]

\[
= \frac{1}{b-a} \sum_{k=0}^n E(X(\tau_k) - X(\tau_{k-1})) \leq \frac{1}{b-a} E(X(M))
\]

Now by the optional sampling theorem \( X(0), X(\tau_k), X(M) \) is a submartingale. Therefore, the above is no larger than

\[
\frac{1}{b-a} E(|X(M)|)
\]

Now note that \( U_M^{[a, b]} \) is at least as large as the number of upcrossings of \( \{X(k)\} \) for \( k \leq M \). This is because every time an upcrossing occurs, it will follow that \( X(\tau_{2k+1}) - X(\tau_{2k}) > 0 \) and so one will occur in the above sum which defines \( U_M^{[a, b]} \). However, this might be larger than the number of upcrossings. The above discussion has proved the following upcrossing lemma.

**Lemma 12.5.1** Let \( \{X(k)\} \) be a nonnegative submartingale. Let

\[
U_M^{[a, b]} = \lim_{\varepsilon \to 0} \sum_{k=0}^n \frac{X(\tau_{2k+1}) - X(\tau_{2k})}{\varepsilon + X(\tau_{2k+1}) - X(\tau_{2k})}, 2n + 1 = M
\]
Then
\[ E \left( T_{M}^{[a,b]} \right) \leq \frac{1}{b-a} E (X (M)) \]

Suppose that there exists a constant \( C \geq E (X (M)) \) for all \( M \). That is, \( \{X (k)\} \) is bounded in \( L^1 (\Omega) \). Then letting \( U^{[a,b]} = \lim_{M \to \infty} U_M^{[a,b]} \),

it follows that
\[ E \left( U^{[a,b]} \right) \leq C \frac{1}{b-a} \]

The second half follows from the first part and the monotone convergence theorem.

Now with this estimate, it is easy to prove the submartingale convergence theorem.

**Theorem 12.5.2** Let \( \{X (k)\} \) be a submartingale which is bounded in \( L^1 (\Omega) \),

\[ \|X (k)\|_{L^1 (\Omega)} \leq C \]

Then there is a set of measure zero \( N \) such that for \( \omega \notin N \), \( \lim_{k \to \infty} X (k) (\omega) \) exists. If \( X (\omega) = \lim_{k \to \infty} X (k) (\omega) \), then \( X \in L^1 (\Omega) \).

**Proof:** Let \( a < b \) and consider the submartingale \( (X (k) - a)_+ \). Let \( U^{[0,b-a]} \) be the random variable of the above lemma which is associated with this submartingale. Thus
\[ E \left( U^{[0,b-a]} \right) \leq \frac{C}{b-a} \]

It follows that \( U^{[0,b-a]} \) is finite for a.e. \( \omega \). As noted above, \( U^{[0,b-a]} \) is an upper bound to the number of upcrossings of \( (X (k) - a)_+ \) and each of these corresponds to an upcrossing of \( [a, b] \) by \( X (k) \). Thus for all \( \omega \notin N_{a,b} \) where \( P (N_{a,b}) = 0 \), it follows that
\[ U^{[0,b-a]} < \infty. \]

If \( \lim_{k \to \infty} X (k) (\omega) \) fails to exist, then there exists \( a < b \) both rational such that
\[ \limsup_{k \to \infty} X (k) > b > a > \liminf_{k \to \infty} X (k) \]

Thus \( \omega \in N_{a,b} \) because there are infinitely many upcrossings of \( [a, b] \). Let \( N = \cup \{N_{a,b} : a, b \in \mathbb{Q} \} \). Then for \( \omega \notin N \), the limit just discussed must exist. Letting \( X (\omega) = \lim_{k \to \infty} X (k) (\omega) \) for \( \omega \notin N \) and letting \( X (\omega) = 0 \) on \( N \), it follows from Fatou’s lemma that \( X \) is in \( L^1 (\Omega) \).

**12.5.2 Maximal Inequalities**

Next I will show that stopping times and the optional stopping theorem can be used to establish maximal inequalities for submartingales very easily.

**Lemma 12.5.3** Let \( \{X (k)\} \) be real valued and adapted to the increasing sequence of \( \sigma \) algebras \( \{\mathcal{F}_k\} \). Let
\[ T (\omega) \equiv \inf \{k : X (k) \geq \lambda \} \]

Then \( T \) is a stopping time. Similarly,
\[ T (\omega) \equiv \inf \{k : X (k) \leq \lambda \} \]

is a stopping time.

**Proof:** Is \( [T \leq p] \in \mathcal{F}_p \) for all \( p \)?
\[ [T = p] = \bigcap_{i=1}^{p-1} [X (i) < \lambda] \cup [X (p) \geq \lambda] \]

Therefore,
\[ [T \leq p] = \bigcup_{i=1}^{p} [T = i] \in \mathcal{F}_p \]
Theorem 12.5.4 Let \( \{X_k\} \) be a real valued submartingale with respect to the \( \sigma \) algebras \( \{\mathcal{F}_k\} \). Then for \( \lambda > 0 \)

\[
\lambda P \left( \max_{1 \leq k \leq n} X_k \geq \lambda \right) \leq E \left( X_n^+ \right),
\]

(12.5.12)

\[
\lambda P \left( \min_{1 \leq k \leq n} X_k \leq -\lambda \right) \leq E \left( |X_n| + |X_1| \right),
\]

(12.5.13)

\[
\lambda P \left( \max_{1 \leq k \leq n} |X_k| \geq \lambda \right) \leq 2E \left( |X_n| + |X_1| \right).
\]

(12.5.14)

**Proof:** Let \( T(\omega) \) be the first time \( X_k(\omega) \) is \( \geq \lambda \) or if this does not happen for \( k \leq n \), then \( T(\omega) \equiv n \). Thus

\[
T(\omega) \equiv \min \{ \min \{ k : X_k(\omega) \geq \lambda \}, n \}
\]

Note

\[
[T > k] = \bigcap_{i=1}^{k} [X_i < \lambda] \in \mathcal{F}_k
\]

and so the complement, \( [T \leq k] \) is also in \( \mathcal{F}_k \), which shows \( T \) is indeed a stopping time.

Then \( 1, T(\omega), n \) are stopping times, \( 1 \leq T(\omega) \leq n \). Therefore, from the optional sampling theorem, Lemma 12.3.4, \( X_1, X_T, X_n \) is a submartingale. It follows

\[
E(X_n) \geq E(X_T) = \int_{\{\max_k X_k \geq \lambda\}} X_T dP + \int_{\{\max_k X_k < \lambda\}} X_T dP
\]

\[
= \int_{\{\max_k X_k \geq \lambda\}} X_T dP + \int_{\{\max_k X_k < \lambda\}} X_n dP
\]

and so, subtracting the last term on the right from both sides,

\[
E(X_n^+) \geq \int_{\{\max_k X_k \geq \lambda\}} X_n dP = \int_{\{\max_k X_k \geq \lambda\}} X_T dP
\]

\[
\geq \lambda P \left( \max_k X_k \geq \lambda \right)
\]

because \( X_T(\omega) \geq \lambda \) on \( \{\max_k X_k \geq \lambda\} \) from the definition of \( T \). This establishes (12.5.13).

Next let \( T(\omega) \) be the first time \( X_k(\omega) \) is \( \leq -\lambda \) or if this does not happen for \( k \leq n \), then \( T(\omega) \equiv n \). Then this is a stopping time by similar reasoning and \( 1 \leq T(\omega) \leq n \) are stopping times and so by the optional stopping theorem, \( X_1, X_T, X_n \) is a submartingale. Therefore, on

\[
\left[ \min_k X_k \leq -\lambda \right], \quad X_T(\omega) \leq -\lambda
\]

and \( E(X_T|\mathcal{F}_1) \geq X_1 \) and so

\[
E(X_1) \leq E (E(X_T|\mathcal{F}_1)) = E(X_T)
\]

which implies

\[
E(X_1) \leq E(X_T) = \int_{\{\min_k X_k \leq -\lambda\}} X_T dP + \int_{\{\min_k X_k > -\lambda\}} X_T dP
\]

\[
= \int_{\{\min_k X_k \leq -\lambda\}} X_T dP + \int_{\{\min_k X_k > -\lambda\}} X_n dP
\]

and so

\[
E(X_1) - \int_{\{\min_k X_k > -\lambda\}} X_n dP \leq \int_{\{\min_k X_k \leq -\lambda\}} X_T dP \leq -\lambda P \left( \min_k X_k \leq -\lambda \right)
\]
which implies
\[
\lambda P \left( \min_k X_k \leq -\lambda \right) \leq \int_{\{X_k > -\lambda\}} X_n dP - E(X_1)
\]
and this proves \( \lambda P \left( \min_k X_k \leq -\lambda \right) \leq \int_{\{X_k > -\lambda\}} X_n dP - E(X_1) \)

The last estimate follows from these. Here is why.

\[
\left[ \max_{1 \leq k \leq n} |X_k| \geq \lambda \right] \subseteq \left[ \max_{1 \leq k \leq n} X_k \geq \lambda \right] \cup \left[ \min_{1 \leq k \leq n} X_k \leq -\lambda \right]
\]
and so
\[
\lambda P \left( \left[ \max_{1 \leq k \leq n} |X_k| \geq \lambda \right] \right) \leq \lambda P \left( \left[ \max_{1 \leq k \leq n} X_k \geq \lambda \right] \cup \left[ \min_{1 \leq k \leq n} X_k \leq -\lambda \right] \right)
\]
\[
\leq \lambda P \left( \left[ \max_{1 \leq k \leq n} X_k \geq \lambda \right] \right) + \lambda P \left( \left[ \min_{1 \leq k \leq n} X_k \leq -\lambda \right] \right)
\]
and this proves the last estimate.

### 12.5.3 The Upcrossing Estimate

A very interesting example of stopping times is next. It has to do with upcrossings. First here is a lemma.

**Lemma 12.5.5** Let \( \{\mathcal{F}_k\} \) be an increasing sequence of \( \sigma \) algebras and let \( \{X(k)\} \) be adapted to this sequence. Suppose that \( X(k) \) has all values in \([a,b]\) and suppose \( \sigma \) is a stopping time with the property that \( X(\sigma) = a \). Let \( \tau(\omega) \) be the first \( k > \sigma \) such that \( X(k) = b \). If no such \( k \) exists, then \( \tau \equiv \infty \). Then \( \tau \) is a stopping time. Also, you can switch \( a, b \) in the above and obtain the same conclusion that \( \tau \) is a stopping time.

**Proof:** Let \( I \) be an interval and consider \( X(k \lor \sigma) \). Is \( k \rightarrow X(k \lor \sigma) \) adapted? Let \( I \) be an interval. Is \( A \equiv X(k \lor \sigma)^{-1}(I) \in \mathcal{F}_k \)?

We know that this set is in \( \mathcal{F}_{k \lor \sigma} \).

\[
A = A \cap [\sigma \leq k] \cup \left( X(k \lor \sigma)^{-1}(I) \cap [\sigma > k] \right)
\]  
(\( \spadesuit \))

Consider the second set in \( \spadesuit \). There are two cases, \( a \in I \) and \( a \notin I \). First suppose \( a \notin I \). Then if \( \omega \in [\sigma > k] \), it follows that \( X(k \lor \sigma) = X(\sigma) = a \). Therefore, in this case, the set on the right in \( \spadesuit \) is empty and the empty set is in \( \mathcal{F}_k \). Next suppose \( a \in I \). Then for \( \omega \in [\sigma > k] \),

\[
X(k \lor \sigma(\omega)) = X(\sigma(\omega)) = a \in I
\]
and so each \( \omega \in [\sigma > k] \) is in the set \( X(k \lor \sigma)^{-1}(I) \) and so, in this case, the set on the right equals

\[
[\sigma > k] \in \mathcal{F}_k
\]

Now consider the first set in \( \spadesuit \).

\[
A \cap [\sigma \leq k] = A \cap [\sigma \lor k \leq k] \in \mathcal{F}_k
\]
by the definition of what it means for the set \( A \) to be in \( \mathcal{F}_{k \lor \sigma} \). The argument proceeds in the same way when you switch \( a, b \).

**Definition 12.5.6** Let \( \{X_k\} \) be a sequence of random variables adapted to the increasing sequence of \( \sigma \) algebras, \( \{\mathcal{F}_k\} \). Let \([a,b]\) be an interval. An upcrossing is a sequence \( X_n(\omega), \cdots, X_{n+p}(\omega) \) such that \( X_n(\omega) \leq a, X_{n+i}(\omega) < b \) for \( i < p \), and \( X_{n+p}(\omega) \geq b \).
Example 12.5.7 Let \( \{F_n\} \) be an increasing sequence of \( \sigma \) algebras contained in \( F \) where \((\Omega, F, P)\) is a probability space and let \( \{X_n\} \) be a sequence of real valued random variables such that \( X_n \) is \( F_n \) measurable. Also let \( a < b \). Define
\[
T_0 \equiv \inf \{ n : X(n) \leq a \}
\]
\[
T_1 \equiv \inf \{ n > T_0 : X(n) \geq b \}
\]
\[
T_2 \equiv \inf \{ n > T_1 : X(n) \leq a \}
\]
\[
\vdots
\]
\[
T_{2k-1} \equiv \inf \{ n > T_{2k-2} : X(n) \geq b \}
\]
\[
T_{2k} \equiv \inf \{ n > T_{2k-1} : X(n) \leq a \}
\]

If \( X_n(\omega) \) is never in the desired interval for any \( n > T_j(\omega) \), then define \( T_{j+1}(\omega) \equiv \infty \). Then this is an increasing sequence of stopping times.

It happens that the above gives an increasing sequence of stopping times.

Lemma 12.5.8 The above example gives an increasing sequence of stopping times.

Proof: You could consider the modified random variables
\[
Y(k) \equiv (X(k) \vee a) \wedge b
\]

Then these new random variables stay in \([a, b]\) and if you replace \( X(n) \) in the above with \( Y(n) \), you get the same sequence of stopping times. Now apply Lemma 12.5.7

Now there is an interesting application of these stopping times to the concept of upcrossings. Let \( \{X_n\} \) be a submartingale such that \( X_n \) is \( F_n \) measurable and let \( a < b \). Assume \( X_0(\omega) \leq a \). The function, \( x \to (x-a)^+ \) is increasing and convex so \( \{(X_n-a)^+\} \) is also a submartingale. Furthermore, \( \{X_n\} \) goes from \( \leq a \) to \( \geq b \) if and only if \( \{(X_n-a)^+\} \) goes from 0 to \( \geq b-a \). That is, a subsequence of the form \( Y_n(\omega), Y_{n+1}(\omega), \ldots, Y_{n+r}(\omega) \) for \( Y \) equal to either \( X \) or \( (X-a)^+ \) starts out below \( a \) (0) and ends up above \( b (b-a) \). Such a sequence is called an upcrossing of \([a, b]\). The idea is to estimate the expected number of upcrossings for \( n \leq N \). For the stopping times defined in Example 12.5.7, let \( T_k \equiv \min (T_k, N) \). Thus \( T_k \), a continuous function of the stopping time, is also a stopping time which is bounded. Moreover, \( T_k \leq T_{k+1} \). Now pick \( n \) such that \( 2n > N \). Then for each \( \omega \in \Omega \)
\[
(X_N(\omega) - a)^+ - (X_0(\omega) - a)^+
\]

must equal the sum of all successive terms of the form
\[
\left( \left( X_{T_{k+1}}^{(r)}(\omega) - a \right)^+ - \left( X_{T_k}^{(r)}(\omega) - a \right)^+ \right)
\]

for \( k = 1, 2, \ldots, 2n \). This is because \( \{T_k^{(r)}(\omega)\} \) is a strictly increasing sequence which starts with 0 due to the assumption \( X_0(\omega) \leq a \) and ends with \( N < 2n \). Therefore,
\[
(X_N - a)^+ - (X_0 - a)^+ = \sum_{k=1}^{2n} \left( X_{T_k}^{(r)} - a \right)^+ - \left( X_{T_{k-1}}^{(r)} - a \right)^+
\]
\[
= \sum_{k=0}^{n-1} \left( X_{T_{2k+1}}^{(r)} - a \right)^+ - \left( X_{T_{2k}}^{(r)} - a \right)^+ + \sum_{k=1}^{n} \left( X_{T_{2k}}^{(r)} - a \right)^+ - \left( X_{T_{2k-1}}^{(r)} - a \right)^+.
\]

Now denote by \( U_{[a, b]}^N \) the number of upcrossings. When \( T_k^{(r)} \) is such that \( k \) is odd, \( \left( X_{T_k}^{(r)} - a \right)^+ \) is above \( b-a \) and when \( k \) is even, it equals 0. Therefore, in the first sum \( X_{T_{2k+1}}^{(r)} - X_{T_{2k}}^{(r)} \geq b-a \) and there are \( U_{[a, b]}^N \) terms which are nonzero in this sum. (Note this might not be \( n \) because many of the terms in the sum could be 0 due to the definition of \( T_k^{(r)} \).) Hence
\[
(X_N - a)^+ - (X_0 - a)^+ = (X_N - a)^+
\]
\[ \geq (b - a) U_{[a,b]}^N + \sum_{k=1}^n \left( (X_{T_{2k}^n} - a)^+ - (X_{T_{2k-1}^n} - a)^+ \right). \]  

Now \( U_{[a,b]}^N \) is a random variable. To see this, let \( Z_k(\omega) = 1 \) if \( T_{2k}^n > T_{2k-1}^n \) and 0 otherwise. Thus \( U_{[a,b]}^N(\omega) = \sum_{k=0}^{n-1} Z_k(\omega) \). Therefore, it makes sense to take the expected value of both sides of (12.5.15). By the optional sampling theorem, \( \left\{ (X_{T_k} - a)^+ \right\} \) is a submartingale and so

\[
E \left( \left( X_{T_{2k}} - a \right)^+ - \left( X_{T_{2k-1}} - a \right)^+ \right) = \int_{\Omega} E \left( \left( X_{T_{2k}} - a \right)^+ | \mathcal{F}_{T_{2k-1}} \right) dP - \int_{\Omega} \left( X_{T_{2k-1}} - a \right)^+ dP \geq 0.
\]

Therefore,

\[ E \left( (X_N - a)^+ \right) \geq (b - a) E \left( U_{[a,b]}^N \right). \]  

This proves most of the following fundamental upcrossing estimate.

**Theorem 12.5.9** Let \( \{X_n\} \) be a real valued submartingale such that \( X_n \) is \( \mathcal{F}_n \) measurable. Then letting \( U_{[a,b]}^N \) denote the upcrossings of \( \{X_n\} \) from \( a \) to \( b \) for \( n \leq N \),

\[ E \left( U_{[a,b]}^N \right) \leq \frac{1}{b-a} E \left( (X_N - a)^+ \right). \]

**Proof:** The estimate (12.5.15) was based on the assumption that \( X_0(\omega) \leq a \). If this is not so, modify \( X_0 \). Change it to \( \min(0, X_0, a) \). Then the inequality holds for the modified submartingale which has at least as many upcrossings. Therefore, the inequality remains. \( \blacksquare \)

Note this theorem holds if the submartingale starts at the index 1 rather than 0. Just adjust the argument.

### 12.6 The Submartingale Convergence Theorem

With this estimate it is now possible to prove the amazing submartingale convergence theorem.

**Theorem 12.6.1** Let \( \{X_n\} \) be a real valued submartingale such that

\[ E \left( |X_n| \right) < M \]

for all \( n \). Then there exists \( X \in L^1(\Omega, \mathcal{F}) \) such that \( X_n(\omega) \) converges to \( X(\omega) \) a.e. \( \omega \) and \( X \in L^1(\Omega) \).

**Proof:** Let \( a < b \) be two rational numbers. From Theorem 12.5.14 it follows that for all \( N \),

\[
\int_{\Omega} U_{[a,b]}^N dP \leq \frac{1}{b-a} E \left( (X_N - a)^+ \right) \leq \frac{1}{b-a} \left( E \left( |X_N| \right) + |a| \right) \leq \frac{M + |a|}{b-a}.
\]

Therefore, letting \( N \to \infty \), it follows that for a.e. \( \omega \), there are only finitely many upcrossings of \([a, b]\). Denote by \( S_{[a,b]} \) the exceptional set. Then letting \( S = \bigcup_{a,b \in \mathbb{Q}} S_{[a,b]} \), it follows that \( P(S) = 0 \) and for \( \omega \notin S \), \( \{X_n(\omega)\} \) is a Cauchy sequence because if

\[
\limsup_{n \to \infty} X_n(\omega) > \liminf_{n \to \infty} X_n(\omega)
\]

then you can pick \( \liminf_{n \to \infty} X_n(\omega) < a < b < \limsup_{n \to \infty} X_n(\omega) \) with \( a, b \) rational and conclude \( \omega \in S_{[a,b]} \).

Let \( X(\omega) = \lim_{n \to \infty} X_n(\omega) \) if \( \omega \notin S \) and let \( X(\omega) = 0 \) if \( \omega \in S \). Then it only remains to verify \( X \in L^1(\Omega) \).

Since \( X \) is the pointwise limit of measurable functions, it follows \( X \) is measurable. By Fatou’s lemma,

\[
\int_{\Omega} |X(\omega)| dP \leq \liminf_{n \to \infty} \int_{\Omega} |X_n(\omega)| dP
\]

Thus \( X \in L^1(\Omega) \). This proves the theorem.

As a simple application, here is an easy proof of a nice theorem about convergence of sums of independent random variables.
Theorem 12.6.2  Let $\{X_k\}$ be a sequence of independent real valued random variables such that $E(|X_k|) < \infty$, $E(X_k) = 0$, and
\[ \sum_{k=1}^{\infty} E(X_k^2) < \infty. \]
Then $\sum_{k=1}^{\infty} X_k$ converges a.e.

Proof: Let $F_n \equiv \sigma(X_1, \cdots, X_n)$. Consider $S_n \equiv \sum_{k=1}^{n} X_k$.
\[ E(S_{n+1}|F_n) = S_n + E(X_{n+1}|F_n). \]
Letting $A \in F_n$ it follows from independence that
\[ \int_A E(X_{n+1}|F_n) \, dP = \int_A X_{n+1} \, dP = \int_\Omega X_{n+1} \, dP = 0 \]
and so $E(X_{n+1}|F_n) = 0$. Therefore, $\{S_n\}$ is a martingale. Now using independence again,
\[ E(|S_n|) \leq E(|S_n^2|) = \sum_{k=1}^{n} E(X_k^2) \leq \sum_{k=1}^{\infty} E(X_k^2) < \infty \]
and so $\{S_n\}$ is an $L^1$ bounded martingale. Therefore, it converges a.e. and this proves the theorem.

Corollary 12.6.3  Let $\{X_k\}$ be a sequence of independent real valued random variables such that $E(|X_k|) < \infty$, $E(X_k) = m_k$, and
\[ \sum_{k=1}^{\infty} E(|X_k - m_k|^2) < \infty. \]
Then $\sum_{k=1}^{\infty} (X_k - m_k)$ converges a.e.

This can be extended to the case where the random variables have values in a separable Hilbert space.

Theorem 12.6.4  Let $\{X_k\}$ be a sequence of independent $H$ valued random variables where $H$ is a real separable Hilbert space such that $E(|X_k|_H) < \infty$, $E(X_k) = 0$, and
\[ \sum_{k=1}^{\infty} E(|X_k|_H^2) < \infty. \]
Then $\sum_{k=1}^{\infty} X_k$ converges a.e.

Proof: Let $\{e_k\}$ be an orthonormal basis for $H$. Then $\{(X_n, e_k)_H\}_{n=1}^{\infty}$ are real valued, independent, and their mean equals 0. Also
\[ \sum_{n=1}^{\infty} E\left(|(X_n, e_k)_H|^2\right) \leq \sum_{n=1}^{\infty} E\left(|X_n|_H^2\right) < \infty \]
and so from Theorem 12.6.2, the series,
\[ \sum_{n=1}^{\infty} (X_n, e_k)_H \]
converges a.e. Therefore, there exists a set of measure zero such that for $\omega$ not in this set, $\sum_n (X_n(\omega), e_k)_H$ converges for each $k$. For $\omega$ not in this exceptional set, define
\[ Y_k(\omega) \equiv \sum_{n=1}^{\infty} (X_n(\omega), e_k)_H \]
Next define
\[ S(\omega) \equiv \sum_{k=1}^{\infty} Y_k(\omega) e_k. \] (12.6.17)

Of course it is not clear this even makes sense. I need to show \( \sum_{k=1}^{\infty} |Y_k(\omega)|^2 < \infty \). Using the independence of the \( X_n \)

\[
E\left(|Y_k|^2\right) = E\left(\left(\sum_{n=1}^{\infty} (X_n, e_k)_H\right)^2\right) = E\left(\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} (X_n, e_k)_H (X_m, e_k)_H\right) \leq \liminf_{N \to \infty} E\left(\left(\sum_{n=1}^{N} \sum_{m=1}^{N} (X_n, e_k)_H (X_m, e_k)_H\right)^2\right) = \liminf_{N \to \infty} E\left(\sum_{n=1}^{N} (X_n, e_k)_H^2\right) = \sum_{n=1}^{\infty} E\left((X_n, e_k)_H^2\right)
\]

Hence from the above,

\[
E\left(\sum_k |Y_k|^2\right) = \sum_k E\left(|Y_k|^2\right) \leq \sum_k \sum_n E\left((X_n, e_k)_H^2\right)
\]

and by the monotone convergence theorem or Fubini’s theorem,

\[
= E\left(\sum_k \sum_n (X_n, e_k)_H^2\right) = E\left(\sum_n \sum_k (X_n, e_k)_H^2\right) = E\left(\sum_n |X_n|_H^2\right) = \sum_n E\left(|X_n|_H^2\right) < \infty \tag{12.6.18}
\]

Therefore, for \( \omega \) off a set of measure zero, and for

\[ Y_k(\omega) \equiv \sum_{n=1}^{\infty} (X_n(\omega), e_k)_H, \]

\[ \sum_k |Y_k(\omega)|^2 < \infty \]

and also for these \( \omega \),

\[ \sum_n \sum_k (X_n(\omega), e_k)_H^2 < \infty. \]

It follows from the estimate (12.6.18) that for \( \omega \) not on a suitable set of measure zero, \( S(\omega) \) defined by (12.6.17)

\[ S(\omega) \equiv \sum_{k=1}^{\infty} Y_k(\omega) e_k \]

makes sense. Thus for these \( \omega \)

\[
S(\omega) = \sum_l \langle S(\omega), e_l \rangle e_l = \sum_l Y_l(\omega) e_l \equiv \sum_l \sum_n (X_n(\omega), e_l)_H e_l = \sum_n \sum_l (X_n(\omega), e_l)_H e_l = \sum_n X_n(\omega).
\]

This proves the theorem.

Now with this theorem, here is a strong law of large numbers.
Theorem 12.6.5 Suppose \( \{X_k\} \) are independent random variables and \( E(|X_k|) < \infty \) for each \( k \) and \( E(X_k) = m_k \). Suppose also
\[
\sum_{j=1}^{\infty} \frac{1}{j^2} E \left( |X_j - m_j|^2 \right) < \infty.
\]
Then
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} (X_j - m_j) = 0 \text{ a.e.}
\]

Proof: Consider the sum
\[
\sum_{j=1}^{\infty} \frac{X_j - m_j}{j}.
\]
This sum converges a.e. because of (12.6.19) and Theorem 12.6.4 applied to the random vectors \( \left\{ \frac{X_j - m_j}{j} \right\} \). Therefore, from Lemma 10.2.24 it follows that for a.e. \( \omega \),
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} (X_j(\omega) - m_j) = 0
\]
This proves the theorem.

The next corollary is often called the strong law of large numbers. It follows immediately from the above theorem.

Corollary 12.6.6 Suppose \( \{X_j\}_{j=1}^{\infty} \) are independent random vectors, \( \lambda_{X_i} = \lambda_{X_j} \) for all \( i, j \) having mean \( m \) and variance equal to
\[
\sigma^2 = \int_{\Omega} |X_j - m|^2 dP < \infty.
\]
Then for a.e. \( \omega \in \Omega \)
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} X_j(\omega) = m
\]

12.7 A Reverse Submartingale Convergence Theorem

Definition 12.7.1 Let \( \{X_n\}_{n=0}^{\infty} \) be a sequence of real random variables such that \( E(|X_n|) < \infty \) for all \( n \) and let \( \{F_n\} \) be a sequence of \( \sigma \) algebras such that \( F_n \supseteq F_{n+1} \) for all \( n \). Then \( \{X_n\} \) is called a reverse submartingale if for all \( n \),
\[
E(X_n|F_{n+1}) \geq X_{n+1}.
\]

Note it is just like a submartingale only the indices are going the other way. Here is an interesting lemma. This lemma gives uniform integrability for a reverse submartingale. First here is a definition.

Definition 12.7.2 Let \( (\Omega, \mathcal{S}, \mu) \) be a measure space in which \( \mu(\Omega) < \infty \). Then \( K \subseteq L^1(\Omega, \mathcal{S}, \mu) \) is said to be equi integrable if
\[
\lim_{\lambda \to \infty} \sup_{f \in K} \int_{\{|f| \geq \lambda\}} |f| d\mu = 0.
\]

If functions defined on a probability space are equi integrable, then one can show easily that they are also uniformly integrable.

Lemma 12.7.3 Suppose \( E(|X_n|) < \infty \) for all \( n \), \( X_n \) is \( F_n \) measurable, \( F_{n+1} \subseteq F_n \) for all \( n \in \mathbb{N} \), and there exist \( X_\infty, F_\infty \) measurable such that \( F_\infty \subseteq F_n \) for all \( n \) and \( X_0, F_0 \) measurable such that \( F_0 \supseteq F_n \) for all \( n \) such that for all \( n \in \{0, 1, \cdots \} \),
\[
E(X_n|F_{n+1}) \geq X_{n+1}, \quad E(X_n|F_\infty) \geq X_\infty,
\]
where \( E(|X_\infty|) < \infty \). Then \( \{X_n : n \in \mathbb{N}\} \) is uniformly integrable.
12.7. A REVERSE SUBMARTINGALE CONVERGENCE THEOREM

Proof:

\[ E(X_{n+1}) \leq E(E(X_n|\mathcal{F}_{n+1})) = E(X_n) \]

Therefore, the sequence \( \{E(X_n)\} \) is a decreasing sequence bounded below by \( E(X_{\infty}) \) so it has a limit. I am going to show the functions are equi integrable. Let \( k \) be large enough that

\[ |E(X_k) - \lim_{m \to \infty} E(X_m)| < \varepsilon \quad (12.7.20) \]

and suppose \( n > k \). Then if \( \lambda > 0 \),

\[
\int_{|X_n| \geq \lambda} |X_n| \, dP = \int_{|X_n| \geq \lambda} X_n dP + \int_{|X_n| < \lambda} (-X_n) \, dP
\]

\[ = \int_{|X_n| \geq \lambda} X_n dP + \int_{\Omega} (-X_n) \, dP - \int_{|X_n| < \lambda} (-X_n) \, dP \]

\[ = \int_{|X_n| \geq \lambda} X_n dP - \int_{\Omega} X_n dP + \int_{|X_n| < \lambda} X_n dP \]

From \( 12.7.20 \),

\[ \leq \int_{|X_n| \geq \lambda} X_n dP - \int_{\Omega} X_k dP + \varepsilon + \int_{|X_n| < \lambda} X_n dP \]

By assumption,

\[ E(X_k|\mathcal{F}_n) \geq X_n \]

and so the above

\[
\leq \int_{|X_n| \geq \lambda} E(X_k|\mathcal{F}_n) \, dP - \int_{\Omega} X_k dP + \varepsilon + \int_{|X_n| < \lambda} E(X_k|\mathcal{F}_n) \, dP
\]

\[ = \int_{|X_n| \geq \lambda} X_k dP - \int_{\Omega} X_k dP + \varepsilon + \int_{|X_n| > \lambda} X_k dP \]

\[ = \int_{|X_n| \geq \lambda} X_k dP - \int_{\Omega} X_k dP + \varepsilon + \int_{|X_n| < \lambda} X_k dP \]

\[ = \int_{|X_n| \geq \lambda} X_k dP + \left( \int_{\Omega} (-X_k) \, dP - \int_{|X_n| > \lambda} (-X_k) \, dP \right) + \varepsilon \]

\[ = \int_{|X_n| \geq \lambda} X_k dP + \int_{|X_n| < \lambda} (-X_k) \, dP + \varepsilon = \int_{|X_n| \geq \lambda} |X_k| \, dP + \varepsilon \]

Applying the maximal inequality for submartingales, Theorem \( 12.7.4 \),

\[ P(\max \{|X_j| : j = n, \ldots, 1\} \geq \lambda) \leq \frac{1}{\lambda} (E(|X_0|) + E(|X_{\infty}|)) \leq \frac{C}{\lambda} \]

and taking sup for all \( n \),

\[ P(\sup \{|X_j| \geq \lambda\}) \leq \frac{C}{\lambda} \]

It follows since the single function, \( X_k \) is equiintegrable that for all \( \lambda \) large enough,

\[ \int_{|X_n| \geq \lambda} |X_n| \, dP \leq 2\varepsilon \]

and since \( \varepsilon \) is arbitrary, this shows \( \{X_n\} \) for \( n > k \) is equiintegrable. Since there are only finitely many \( X_j \) for \( j \leq k \), this shows \( \{X_n\} \) is equiintegrable. Hence \( \{X_n\} \) is uniformly integrable. This proves the lemma.

Now with this lemma and the upcrossing lemma it is easy to prove an important convergence theorem.

**Theorem 12.7.4** Let \( \{X_n, \mathcal{F}_n\}_{n \geq 0} \) be a backwards submartingale as described above and suppose \( \sup_{n \geq 0} E(|X_n|) < \infty \). Then \( \{X_n\} \) converges a.e. and in \( L^1(\Omega) \) to a function, \( X_\infty \).
so is $D \equiv \bigcup S \{\}$

Lemma 12.8.1

Theorem follows

First are some preparatory lemmas. The approach followed here is from Ash [\textcite{ash}]. There is a version of the strong law of large numbers which does not depend on the random variables having finite integrable, it follows $X$ and so $U$ it would follow because if inequality holds, then letting $N$.

Letting $N \rightarrow \infty$, it follows the expected number of upcrossings, $E(U_{[a,b]})$ is bounded. Therefore, there exists a set of measure 0 $N_{ab}$ such that if $\omega \notin N_{ab}, U_{[a,b]}(\omega) < \infty$. Let $N = \cup \{N_{ab} : a, b \in \mathbb{Q}\}$. Then for $\omega \notin N$,

$$\lim\sup_{n \rightarrow \infty} X_n(\omega) = \lim\inf_{n \rightarrow \infty} X_n(\omega)$$

because if inequality holds, then letting

$$\lim\inf_{n \rightarrow \infty} X_n(\omega) < a < b < \lim\sup_{n \rightarrow \infty} X_n(\omega)$$

it would follow $U_{[a,b]}(\omega) = \infty$, contrary to $\omega \notin N_{ab}$.

Let $X_\infty(\omega) \equiv \lim_{n \rightarrow \infty} X_n(\omega)$. Then by Fatou’s lemma,

$$\int \Omega |X_\infty(\omega)| dP \leq \lim\inf_{n \rightarrow \infty} \int \Omega |X_n| dP < \infty,$$

and so $X_\infty$ is in $L^1(\Omega)$. By the Vitali convergence theorem and Lemma 12.8.1 which shows $\{|X_n|\}$ is uniformly integrable, it follows

$$\lim_{n \rightarrow \infty} \int \Omega |X_\infty(\omega) - X_n(\omega)| dP = 0.$$

This proves the theorem.

12.8 Strong Law Of Large Numbers

There is a version of the strong law of large numbers which does not depend on the random variables having finite variance. First are some preparatory lemmas. The approach followed here is from Ash [\textcite{ash}].

Lemma 12.8.1 Let $\{X_n\}$ be a sequence of independent random variables such that $E(|X_k|) < \infty$ for all $k$ and let $S_n = \sum_{k=1}^{n} X_k$. Then for $k \leq n$,

$E(X_k|\sigma(S_n)) = E(X_k|\sigma(S_n,Y)) \ a.e.$

where $Y = (X_{n+1}, X_{n+2}, \cdots) \in \mathbb{R}^N$. Also for $k \leq n$ as above,

$\sigma(S_n,Y) = \sigma(S_n, S_{n+1}, \cdots)$.

Proof: Note that $\mathbb{R}^N$ with the usual product topology has a countable basis. Here it is. Let $\mathcal{B}_N$ denote sets of the form $\prod_{i=1}^{\infty} D_i$ where for $i \leq N, D_i \in \mathcal{B},$ a countable basis for $\mathbb{R}$ and for $i > N, D_i = \mathbb{R}$. Then $\mathcal{B}_N$ is countable and so is $\mathcal{D} := \cup_{N=1}^{\infty} \mathcal{B}_N$. From the definition of the product topology, this is a countable basis for the product topology.

Let $V \in \mathcal{D}$ and $U$ be an open set of $\mathbb{R}$. Then if $A \in (S_n, Y)_{-1}^{-1}(U \times V)$, by independence of the $\{X_n\}$,

$$\int_{(S_n, Y)^{-1}(U \times V)} E(X_k|\sigma(S_n, Y)) dP = \int_{(S_n, Y)^{-1}(U \times V)} X_k dP$$

$$= \int_{\Omega} X_{S_n^{-1}(U)}(\omega) X_{Y^{-1}(V)}(\omega) X_k dP = P(Y^{-1}(V)) \int_{\Omega} X_{S_n^{-1}(U)}(\omega) X_k dP$$

$$= P(Y^{-1}(V)) \int_{S_n^{-1}(U)} E(X_k|\sigma(S_n)) dP.$$ 

Now by independence again, $\{S_n, X_{n+1}, X_{n+2}, \cdots\}$ are independent and so the above equals

$$\int_{S_n^{-1}(U)} X_{Y^{-1}(V)} E(X_k|\sigma(S_n)) dP = \int_{(S_n, Y)^{-1}(U \times V)} E(X_k|\sigma(S_n)) dP.$$
Letting

\[ S = \left\{ A \in B \left( \mathbb{R} \times \mathbb{R}^N \right) : \int_{(S_n, Y)^{-1}(A)} E(X_k|\sigma(S_n)) \, dP \right\} \]

the above has shown this is true for all \( A \) in a countable basis. Therefore, it is true for all \( A \) open in \( \mathbb{R} \times \mathbb{R}^N \). Finally, it is clear that \( S \) is a \( \sigma \) algebra which shows the above holds for all \( A \) Borel in \( \mathbb{R} \times \mathbb{R}^N \). Thus, for all \( B \in \sigma(S_n, Y) \),

\[ \int_B E(X_k|\sigma(S_n)) \, dP = \int_B E(X_k|\sigma(S_n, Y)) \, dP \]

and thus \( E(X_k|\sigma(S_n)) = E(X_k|\sigma(S_n, Y)) \) a.e.

It only remains to prove the last assertion. For \( k > 0 \),

\[ X_{n+k} = S_{n+k} - S_{n+k-1} \]

Thus

\[ \sigma(S_n, Y) = \sigma(S_n, X_{n+1}, \cdots) \]
\[ = \sigma(S_n, (S_{n+1} - S_n), (S_{n+2} - S_{n+1}), \cdots) \]
\[ \subseteq \sigma(S_n, S_{n+1}, \cdots) \]

On the other hand,

\[ \sigma(S_n, S_{n+1}, \cdots) = \sigma(S_n, X_{n+1} + S_n, X_{n+2} + X_{n+1} + S_n, \cdots) \]
\[ \subseteq \sigma(S_n, X_{n+1}, X_{n+2}, \cdots) \]

To see this, note that for an open set, and hence for a Borel set, \( B \),

\[ \left( S_n + \sum_{k=1}^m X_k \right)^{-1} (B) = (S_n, X_{n+1}, \cdots, X_m)^{-1} (B') \]

for some \( B' \in \mathbb{R}^{m+1} \). Thus \( (S_n + \sum_{k=1}^m X_k)^{-1} (B) \) for \( B \) a Borel set is contained in \( \sigma(S_n, X_{n+1}, X_{n+2}, \cdots) \). Similar considerations apply to the other inclusion stated earlier. This proves the lemma.

**Lemma 12.8.2** Let \( \{X_k\} \) be a sequence of independent identically distributed random variables such that \( E(|X_k|) < \infty \). Then letting \( S_n = \sum_{k=1}^n X_k \), it follows that for \( k \leq n \)

\[ E(X_k|\sigma(S_n, S_{n+1}, \cdots)) = E(X_k|\sigma(S_n)) = \frac{S_n}{n} \]

**Proof:** It was shown in Lemma [K.S.I] the first equality holds. It remains to show the second. Letting \( A = S_n^{-1}(B) \) where \( B \) is Borel, it follows there exists \( B' \subseteq \mathbb{R}^n \) a Borel set such that

\[ S_n^{-1}(B) = (X_1, \cdots, X_n)^{-1} (B') \]

Then

\[ \int_A E(X_k|\sigma(S_n)) \, dP = \int_{S_n^{-1}(B)} X_k \, dP \]
\[ = \int_{(X_1, \cdots, X_n)^{-1}(B')} X_k \, dP = \int_{(X_1, \cdots, X_n)^{-1}(B')} x_k \, d\lambda(x_1, \cdots, x_n) \]
\[ = \int \cdots \int \lambda(X_1, \cdots, X_n)^{-1}(B') \, x_1 \, d\lambda x_1 \, d\lambda x_2 \cdots d\lambda x_n \]
\[ = \int \cdots \int \lambda(X_1, \cdots, X_n)^{-1}(B') \, x_1 \, d\lambda x_1 \, d\lambda x_2 \cdots d\lambda x_n \]
\[ \int_A E(X|\sigma(S_n)) \, dP \]

and so since \( A \in \sigma(S_n) \) is arbitrary,
\[ E(X|\sigma(S_n)) = E(X_k|\sigma(S_n)) \]

for each \( k, l \leq n \). Therefore,
\[ S_n = E(S_n|\sigma(S_n)) = \sum_{j=1}^{n} E(X_j|\sigma(S_n)) = nE(X_k|\sigma(S_n)) \text{ a.e.} \]

and so
\[ E(X_k|\sigma(S_n)) = \frac{S_n}{n} \text{ a.e.} \]

as claimed. This proves the lemma.

With this preparation, here is the strong law of large numbers for identically distributed random variables.

**Theorem 12.8.3** Let \( \{X_k\} \) be a sequence of independent identically distributed random variables such that \( E(|X_k|) < \infty \) for all \( k \). Letting \( m = E(X_k) \),
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} X_k(\omega) = m \text{ a.e.} \]

and convergence also takes place in \( L^1(\Omega) \).

**Proof:** Consider the reverse submartingale \( \{E(X_1|\sigma(S_n, S_{n+1}, \ldots))\} \). By Theorem 12.7.4, this converges a.e. and in \( L^1(\Omega) \) to a random variable, \( X_\infty \). However, from Lemma 12.8.2, \( E(X_1|\sigma(S_n, S_{n+1}, \ldots)) = S_n/n \). Therefore, \( S_n/n \) converges a.e. and in \( L^1(\Omega) \) to \( X_\infty \). I need to argue that \( X_\infty \) is constant and also that it equals \( m \). For \( a \in \mathbb{R} \) let
\[ E_a = [X_\infty \geq a] \]

For \( a \) small enough, \( P(E_a) \neq 0 \). Then since \( E_a \) is a tail event for the independent random variables, \( \{X_k\} \) it follows from the Kolmogorov zero one law, Theorem 10.2.18, that \( P(E_a) = 1 \). Let \( b = \sup \{a : P(E_a) = 1\} \). The sets, \( E_a \) are decreasing as \( a \) increases. Let \( \{a_n\} \) be a strictly increasing sequence converging to \( b \). Then
\[ [X_\infty \geq b] = \cap_n [X_\infty \geq a_n] \]

and so
\[ 1 = P(E_b) = \lim_{n \to \infty} P(E_{a_n}) \]

On the other hand, if \( c > b \), then \( P(E_c) < 1 \) and so \( P(E_c) = 0 \). Hence \( P([X = b]) = 1 \). It remains to show \( b = m \). This is easy because by the \( L^1 \) convergence,
\[ b = \int \Omega X_\infty dP = \lim_{n \to \infty} \int \Omega \frac{S_n}{n} dP = \lim_{n \to \infty} m = m. \]

This proves the theorem.
Chapter 13

Stochastic Processes In Banach Space

13.1 Conditional Expectation In Banach Spaces

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X \in L^1(\Omega; \mathbb{R})$. Also let $\mathcal{G} \subseteq \mathcal{F}$ where $\mathcal{G}$ is also a $\sigma$ algebra. Then the usual conditional expectation is defined by

$$\int_A X dP = \int_A E(X|\mathcal{G}) dP$$

where $E(X|\mathcal{G})$ is $\mathcal{G}$ measurable and $A \in \mathcal{G}$ is arbitrary. Recall this is an application of the Radon Nikodym theorem. Also recall $E(X|\mathcal{G})$ is unique up to a set of measure zero.

I want to do something like this here. Denote by $L^1(\Omega; E, \mathcal{G})$ those functions in $L^1(\Omega; E)$ which are measurable with respect to $\mathcal{G}$.

**Theorem 13.1.1** Let $E$ be a separable Banach space and let $X \in L^1(\Omega; E, \mathcal{F})$ where $X$ is measurable with respect to $\mathcal{F}$ and let $\mathcal{G}$ be a $\sigma$ algebra which is contained in $\mathcal{F}$. Then there exists a unique $Z \in L^1(\Omega; E, \mathcal{G})$ such that for all $A \in \mathcal{G}$,

$$\int_A X dP = \int_A Z dP$$

Denoting this $Z$ as $E(X|\mathcal{G})$, it follows

$$\|E(X|\mathcal{G})\| \leq E(\|X\| |\mathcal{G}) .$$

**Proof:** First consider uniqueness. Suppose $Z'$ is another in $L^1(\Omega; E, \mathcal{G})$ which works. Consider a dense subset of $E \{a_n\}_{n=1}^\infty$. Then the balls $\{B(a_n, \|a_n\|/4)\}_{n=1}^\infty$ must cover $E \setminus \{0\}$. Here is why. If $y \neq 0$, pick $a_n \in B\left(y, \frac{\|y\|}{5}\right)$.

Then $\|a_n\| \geq 4\|y\|/5$ and so $\|a_n - y\| < \|y\|/5$. Thus

$$y \in B(a_n, \|y\|/5) \subseteq B\left(a_n, \frac{\|a_n\|}{4}\right)$$

Now suppose $Z$ is $\mathcal{G}$ measurable and

$$\int_A Z dP = 0$$

for all $A \in \mathcal{G}$. The letting $A \equiv Z^{-1}\left(B\left(a_n, \frac{\|a_n\|}{4}\right)\right)$ it follows

$$0 = \int_A Z - a_n + a_n dP$$

251
and so
\[ \|a_n\|P(A) = \left\| \int_A a_n dP \right\| = \left\| \int_A (a_n - Z) dP \right\| \]
\[ \leq \int_{Z^{-1}\left(B\left(a_n, \frac{\|a_n\|}{4}\right)\right)} \|a_n - Z\| dP \leq \frac{\|a_n\|}{4}P(A) \]
which is a contradiction unless \( P(A) = 0 \). Therefore, letting
\[ N \equiv \bigcup_{n=1}^{\infty} Z^{-1}\left(B\left(a_n, \frac{\|a_n\|}{4}\right)\right) = Z^{-1}(E \setminus \{0\}) \]
it follows \( N \) has measure zero and so \( Z = 0 \) a.e. This proves uniqueness because if \( Z, Z' \) both work, then from the above argument, \( Z - Z' = 0 \) a.e.

Next I will show \( Z \) exists. To do this recall Theorem 9.2.5 on Page 135 which is stated below for convenience.

**Theorem 13.1.2** An \( E \) valued function, \( X \), is Bochner integrable if and only if \( X \) is strongly measurable and
\[ \int_{\Omega} \|X(\omega)\| dP < \infty. \] (13.1.1)

In this case there exists a sequence of simple functions \( \{X_n\} \) satisfying
\[ \int_{\Omega} \|X_n(\omega) - X_m(\omega)\| dP \to 0 \text{ as } m, n \to \infty. \] (13.1.2)

\( X_n(\omega) \) converging pointwise to \( X(\omega) \),
\[ \|X_n(\omega)\| \leq 2\|X(\omega)\| \] (13.1.3)
and
\[ \lim_{n \to \infty} \int_{\Omega} \|X(\omega) - X_n(\omega)\| dP = 0. \] (13.1.4)

Now let \( \{X_n\} \) be the simple functions just defined and let
\[ X_n(\omega) = \sum_{k=1}^{m} x_k X_{F_k}(\omega) \]
where \( F_k \in \mathcal{F} \), the \( F_k \) being disjoint. Then define
\[ Z_n = \sum_{k=1}^{m} x_k E(X_{F_k} | \mathcal{G}). \]

Thus, if \( A \in \mathcal{G} \),
\[ \int_{A} Z_n dP = \sum_{k=1}^{m} x_k \int_{A} E(X_{F_k} | \mathcal{G}) dP \]
\[ = \sum_{k=1}^{m} x_k \int_{A} X_{F_k} dP \]
\[ = \sum_{k=1}^{m} x_k P(F_k \cap A) = \int_{A} X_n dP \] (13.1.5)

Then since \( E(X_{F_k} | \mathcal{G}) \geq 0 \),
\[ \|Z_n\| \leq \sum_{k=1}^{m} \|x_k\| E(X_{F_k} | \mathcal{G}) \]
Thus if \( A \in \mathcal{G} \),
\[ E(\|Z_n\| X_A) \leq E \left( \sum_{k=1}^{m} \|x_k\| X_A E(X_{F_k} | \mathcal{G}) \right) = \sum_{k=1}^{m} \|x_k\| \int_{A} E(X_{F_k} | \mathcal{G}) dP \]
\[ = \sum_{k=1}^{m} \|x_k\| \int_{A} X_{F_k} dP = E(X_A \|X_n\|). \] (13.1.6)
13.2. PROPERTIES OF STOCHASTIC PROCESSES

253

Note the use of $\leq$ in the first step in the above. Although the $F_k$ are disjoint, all that is known about $E (X_{F_k} | \mathcal{G})$ is that it is nonnegative. Similarly,

$$E (||Z_n - Z_m||) \leq E (||X_n - X_m||)$$

and this last term converges to 0 as $n, m \to \infty$ by the properties of the $X_n$. Therefore, $\{Z_n\}$ is a Cauchy sequence in $L^1 (\Omega; E; \mathcal{G})$. It follows it converges to $Z$ in $L^1 (\Omega; E; \mathcal{G})$. Then letting $A \in \mathcal{G}$, and using $\mathbf{R}$,

$$\int_A ZdP = \int \mathcal{X}_A ZdP = \lim_{n \to \infty} \int \mathcal{X}_A Z_n dP = \lim_{n \to \infty} \int_A Z_n dP = \lim_{n \to \infty} \int_A X_n dP = \int_A XdP.$$  

Then define $Z \equiv E (X|\mathcal{G})$.

It remains to verify $||E (X|\mathcal{G})|| \equiv ||Z|| \leq E (||X|| | \mathcal{G})$. This follows because, from the above,

$$||Z_n|| \to ||Z||, \; ||X_n|| \to ||X|| \; \text{in} \; L^1 (\Omega)$$

and so if $A \in \mathcal{G}$, then from $\mathbf{R}$,

$$\frac{1}{P(A)} \int_A ||Z_n|| dP \leq \frac{1}{P(A)} \int_A ||X_n|| dP$$

and so, passing to the limit,

$$\frac{1}{P(A)} \int_A ||Z|| dP \leq \frac{1}{P(A)} \int_A ||X|| dP = \frac{1}{P(A)} \int_A E (||X|| | \mathcal{G}) dP$$

Since $A$ is arbitrary, this shows that

$$||E (X|\mathcal{G})|| \equiv ||Z|| \leq E (||X|| | \mathcal{G}). \; \blacksquare$$

In the case where $E$ is reflexive, one could also do this another way. You would define a vector measure on $\mathcal{G}$,

$$\nu (F) \equiv \int_F XdP$$

and then you would use the fact that reflexive separable Banach spaces have the Radon Nikodym property to obtain $Z \in L^1 (\Omega; E; \mathcal{G})$ such that

$$\nu (F) = \int_F XdP = \int_F ZdP.$$  

The function, $Z$ whose existence and uniqueness is guaranteed by Theorem $\mathbf{R}$, is called $E (X|\mathcal{G})$.

13.2 Properties Of Stochastic Processes

Here $E$ will be a separable Banach space and $B (E)$ will be the Borel sets of $E$. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $I$ will be an interval of $\mathbb{R}$. A set of $E$ valued random variables, one for each $t \in I$, $\{X (t) : t \in I\}$ is called a stochastic process. Thus for each $t$, $X (t)$ is a measurable function of $\omega \in \Omega$. Set $X (t, \omega) \equiv X (t) (\omega)$. Functions $t \to X (t, \omega)$ are called trajectories. Thus there is a trajectory for each $\omega \in \Omega$. A stochastic process, $Y$ is called a version or a modification of a stochastic process, $X$ if for all $t \in I$,

$$X (t, \omega) = Y (t, \omega) \; \text{a.e.} \; \omega$$

There are several descriptions of stochastic processes.

1. $X$ is measurable if $X (\cdot, \cdot) : I \times \Omega \to E$ is $B (I) \times \mathcal{F}$ measurable. Note that a stochastic process, $X$ is not necessarily measurable.

2. $X$ is stochastically continuous at $t_0 \in I$ means: for all $\varepsilon > 0$ and $\delta > 0$ there exists $\rho > 0$ such that

$$P (||X (t) - X (t_0)|| \geq \varepsilon) \leq \delta \; \text{whenever} \; |t - t_0| < \rho, \; t \in I.$$  

Note the above condition says that for each $\varepsilon > 0$,

$$\lim_{t \to t_0} P (||X (t) - X (t_0)|| \geq \varepsilon) = 0.$$
3. $X$ is stochastically continuous if it is stochastically continuous at every $t \in I$.

4. $X$ is stochastically uniformly continuous if for every $\varepsilon, \delta > 0$ there exists $\rho > 0$ such that whenever $s, t \in I$ with $|s - t| < \rho$, it follows

$$P (||X(t) - X(s)|| \geq \varepsilon) \leq \delta.$$ 

5. $X$ is mean square continuous at $t_0 \in I$ if

$$\lim_{t \to t_0} E \left(||X(t) - X(t_0)||^2\right) = \lim_{t \to t_0} \int_I ||X(t)(\omega) - X(t_0)(\omega)||^2 dP = 0.$$

6. $X$ is mean square continuous in $I$ if it is mean square continuous at every point of $I$.

7. $X$ is continuous with probability 1 or continuous if $t \to X(t, \omega)$ is continuous for all $\omega$ outside some set of measure 0.

8. $X$ is Hölder continuous if $t \to X(t, \omega)$ is Hölder continuous for a.e. $\omega$.

**Lemma 13.2.1** A stochastically continuous process on $[a, b] \equiv I$ is uniformly stochastically continuous on $[a, b] \equiv I$.

**Proof:** If this is not so, there exists $\varepsilon, \delta > 0$ and points of $I, s_n, t_n$ such that even though

$$|t_n - s_n| < \frac{1}{n},$$

$$P (||X(s_n) - X(t_n)|| \geq \varepsilon) > \delta. \quad (13.2.7)$$

Taking a subsequence, still denoted by $s_n$ and $t_n$ there exists $t \in I$ such that the above hold and

$$\lim_{n \to \infty} s_n = \lim_{n \to \infty} t_n = t.$$

Then

$$P (||X(s_n) - X(t_n)|| \geq \varepsilon) \leq P (||X(s_n) - X(t)|| \geq \varepsilon/2) + P (||X(t) - X(t_n)|| \geq \varepsilon/2).$$

But the sum of the last two terms converges to 0 as $n \to \infty$ by stochastic continuity of $X$ at $t$, violating (13.2.7) for all $n$ large enough. This proves the lemma.

For a stochastically continuous process defined on a closed and bounded interval, there always exists a measurable version. This is significant because then you can do things with product measure and iterated integrals.

**Proposition 13.2.2** Let $X$ be a stochastically continuous process defined on a closed interval, $I \equiv [a, b]$. Then there exists a measurable version of $X$.

**Proof:** By Lemma 13.2.1 $X$ is uniformly stochastically continuous and so there exists a sequence of positive numbers, $\{\rho_n\}$ such that if $|s - t| < \rho_n$, then

$$P \left(||X(t) - X(s)|| \geq \frac{1}{2^n}\right) \leq \frac{1}{2^n}. \quad (13.2.8)$$

Then let $\{t^n_0, t^n_1, \ldots, t^n_m\}$ be a partition of $[a, b]$ in which $|t^n_i - t^n_{i-1}| < \rho_n$. Now define $X_n$ as follows:

$$X_n(t) = \sum_{i=1}^{m_n} X(t^n_{i-1}) X(t^n_{i-1}, t^n_i)(t),$$

$$X_n(b) = X(b).$$

Then $X_n$ is obviously $B(I) \times F$ measurable because it is the sum of functions which are. Consider the set, $A$ on which $\{X_n(t, \omega)\}$ is a Cauchy sequence. This set is of the form

$$A = \cap_{n=1}^{\infty} \cup_{m=1}^{\infty} \cap_{p,q \geq m} \left(||X_p - X_q|| < \frac{1}{n}\right).$$
and so it is a $B(I) \times \mathcal{F}$ measurable set. Now define

$$Y(t, \omega) \equiv \begin{cases} \lim_{n \to \infty} X_n(t, \omega) & \text{if } (t, \omega) \in A \\ 0 & \text{if } (t, \omega) \notin A \end{cases}$$

I claim $Y(t, \omega) = X(t, \omega)$ for a.e. $\omega$. To see this, consider Lemma 13.2.3. From the construction of $X_n$, it follows that for each $t$,

$$P \left( \left| \left| X_n(t) - X(t) \right| \geq \frac{1}{2^n} \right| \right) \leq \frac{1}{2^n}$$

Also, for a fixed $t$, if $X_n(t, \omega)$ fails to converge to $X(t, \omega)$, then $\omega$ must be in infinitely many of the sets,

$$B_n \equiv \left| \left| X_n(t) - X(t) \right| \geq \frac{1}{2^n} \right|$$

which is a set of measure zero by the Borel Cantelli lemma. Recall why this is so.

$$P (\cap_{k=1}^{\infty} \cup_{n=k}^{\infty} B_n) \leq \sum_{n=k}^{\infty} P(B_n) < \frac{1}{2^{k-1}}$$

Therefore, for each $(t, \omega) \in A$ for a.e. $\omega$. Hence $X(t) = Y(t)$ a.e. and so $Y$ is a measurable version of $X$.

**Lemma 13.2.3** Let $D$ be a dense subset of an interval, $I = [0, T]$ and suppose $X : D \to E$ satisfies

$$\left| \left| X(d) - X(d') \right| \right| \leq C |d - d'|^{\gamma}$$

for all $d', d \in D$. Then $X$ extends uniquely to a continuous $Y$ defined on $[0, T]$ such that

$$\left| \left| Y(t) - Y(t') \right| \right| \leq C |t - t'|^{\gamma}.$$  

**Proof:** Let $t \in I$ and let $d_k \to t$ where $d_k \in D$. Then $\{X(d_k)\}$ is a Cauchy sequence because $\left| \left| X(d_k) - X(d_m) \right| \right| \leq C |d_k - d_m|^{\gamma}$. Therefore, $X(d_k)$ converges. The thing it converges to will be called $Y(t)$. Note this is well defined, giving $X(t)$ if $t \in D$. Also, if $d_k \to t$ and $d'_k \to t$, then $\left| \left| X(d_k) - X(d'_k) \right| \right| \leq C |d_k - d'_k|^{\gamma}$ and so $X(d_k)$ and $X(d'_k)$ converge to the same thing. Therefore, it makes sense to define $Y(t) \equiv \lim_{d \to t} X(d)$. It only remains to verify the estimate. But letting $|d - t|$ and $|d' - t|$ be small enough,

$$\left| \left| Y(t) - Y(t') \right| \right| = \left| \left| X(d) - X(d') \right| \right| + \varepsilon \leq C |d' - d| + \varepsilon \leq C |t - t'| + 2\varepsilon.$$

Since $\varepsilon$ is arbitrary, this proves the existence part of the lemma. Uniqueness follows from observing that $Y(t)$ must equal $\lim_{d \to t} X(d)$. This proves the lemma.

### 13.3 Kolmogorov Čentsov Continuity Theorem

**Lemma 13.3.1** Let $r^m_j$ denote $j \left( \frac{T}{2^m} \right)$ where $j \in \{0, 1, \cdots, 2^m \}$. Also let $D_m = \{r^m_j\}_{j=1}^{2^m}$ and $D = \cup_{m=1}^{\infty} D_m$. Suppose $X(t)$ satisfies

$$\left| \left| X(r^m_{j+1}) - X(r^m_j) \right| \right| \leq 2^{-\gamma k}$$

for all $k \geq M$. Then if $d, d' \in D_m$ for $m > n \geq M$ such that $|d - d'| \leq T2^{-n}$, then

$$\left| \left| X(d') - X(d) \right| \right| \leq 2 \sum_{j=n+1}^{m} 2^{-\gamma j}.$$  

Also, there exists a constant $C$ depending on $M$ such that for all $d, d' \in D$,

$$\left| \left| X(d) - X(d') \right| \right| \leq C |d - d'|^{\gamma}.$$
\textbf{Proof:} Suppose \(d' < d\). Suppose first \(m = n + 1\). Then \(d = (k+1)T2^{-(n+1)}\) and \(d' = kT2^{-(n+1)}\). Then from \([13.3.3]\)
\[
\|X(d') - X(d)\| \leq 2^{-\gamma(n+1)} \leq 2 \sum_{j=n+1}^{n+1} 2^{-\gamma j}.
\]
Suppose the claim is true for some \(m > n\) and let \(d, d' \in D_{m+1}\) with \(|d - d'| < T2^{-n}\). If there is no point of \(D_m\) between these, then \(d', d\) are adjacent points either in \(D_m\) or in \(D_{m+1}\). Consequently,
\[
\|X(d') - X(d)\| \leq 2^{-\gamma m} < 2 \sum_{j=n+1}^{m+1} 2^{-\gamma j}.
\]
Assume therefore, there exist points of \(D_m\) between \(d'\) and \(d\). Let \(d' = d'_1 \leq d_1 = d\) where \(d_1, d'_1\) are in \(D_m\) and \(d'_1\) is the smallest element of \(D_m\) which is at least as large as \(d'\) and \(d_1\) is the largest element of \(D_m\) which is no larger than \(d\). Then \(|d' - d'_1| \leq T2^{-(m+1)}\) and \(|d_1 - d| \leq T2^{-(m+1)}\) while all of these points are in \(D_{m+1}\) which contains \(D_m\). Therefore, from \([13.3.3]\) and induction,
\[
\|X(d') - X(d)\| \\
\leq \|X(d') - X(d'_1)\| + \|X(d'_1) - X(d)\| \\
+ \|X(d_1) - X(d)\| \\
\leq 2 \times 2^{-\gamma(m+1)} + 2 \sum_{j=n+1}^{m} 2^{-\gamma j} = 2 \sum_{j=n+1}^{m+1} 2^{-\gamma j} \\
\leq 2 \left( \frac{2^{-\gamma(n+1)} - 1}{1 - 2^{-\gamma}} \right) (T2^{-(n+1)})^\gamma
\] (13.3.10)
It follows the above holds for any \(d, d' \in D\) such that \(|d - d'| \leq T2^{-n}\) because they are both in some \(D_m\) for \(m > n\).

Consider the last claim. Let \(d, d' \in D, |d - d'| \leq T2^{-M}\). Then \(d, d'\) are both in some \(D_m\) for \(m > M\). The number \(|d - d'|\) satisfies
\[
T2^{-(n+1)} < |d - d'| \leq T2^{-n}
\]
for large enough \(n \geq M\). Just pick the first \(n\) such that \(T2^{-(n+1)} < |d - d'|\). Then from \([13.3.10]\),
\[
\|X(d') - X(d)\| \leq \left( \frac{2T^{-\gamma}}{1 - 2^{-\gamma}} \right) (T2^{-(n+1)})^\gamma \\
\leq \left( \frac{2T^{-\gamma}}{1 - 2^{-\gamma}} \right) |d - d'|^\gamma
\]
Now \([0, T]\) is covered by \(2^M\) intervals of length \(T2^{-M}\) and so for any pair \(d, d' \in D,\)
\[
\|X(d) - X(d')\| \leq C |d - d'|^\gamma
\]
where \(C\) is a suitable constant depending on \(2^M\). \(\blacksquare\)

For \(\gamma \leq 1\), you can show, using convexity arguments, that it suffices to have \(C = \left( \frac{2T^{-\gamma}}{1 - 2^{-\gamma}} \right)^{1/\gamma} (2^M)^{1-\gamma}\). Of course the case where \(\gamma > 1\) is not interesting because it would result in \(X\) being a constant.

The following is the amazing Kolmogorov Čentsov continuity theorem\([100]\).

\textbf{Theorem 13.3.2} Suppose \(X\) is a stochastic process on \([0, T]\). Suppose also that there exists a constant, \(C\) and positive numbers \(\alpha, \beta\) such that
\[
E(\|X(t) - X(s)\|^\gamma) \leq C |t - s|^{1+\beta}\] (13.3.11)
Then there exists a stochastic process \(Y\) such that for a.e. \(\omega, t \rightarrow Y(t) (\omega)\) is H"{o}lder continuous with exponent \(\gamma < \frac{\beta}{\alpha}\) and for each \(t, P(\|X(t) - Y(t)\| > 0) = 0\). \((Y\) is a version of \(X\).)

\textbf{Proof:} Let \(n_j\) denote \(j \left( \frac{T}{2^m} \right)\) where \(j \in \{0, 1, \cdots, 2^m\}\). Also let \(D_m = \{n_j\}_{j=1}^{2^m}\) and \(D = \cup_{m=1}^\infty D_m\). Consider the set,
\[
\{\|X(t) - X(s)\| > \delta\}
\]
By \[(13.3.12)\]
\[
P (||X (t) - X (s)|| > \delta) \delta^\alpha \leq \int_{||X(t) - X(s)|| > \delta} ||X(0)||^\alpha dP \leq C|t - s|^{1+\beta}.
\]

Letting \(t = r_j^k, s = r_j^k\) and \(\delta = 2^{-\gamma k}\) where
\[
\gamma \in \left(0, \frac{\beta}{\alpha}\right),
\]
this yields
\[
P (||X (r_{j+1}^k) - X (r_j^k)|| > 2^{-\gamma k}) \leq C2^{\alpha \gamma k} (T 2^{-k})^{1+\beta} = C T^{1+\beta} 2^{k(\alpha \gamma - (1+\beta))}
\]

There are \(2^k\) of these differences and so letting
\[
N_k = \bigcup_{j=1}^{2^k} \{||X (r_{j+1}^k) - X (r_j^k)|| > 2^{-\gamma k}\}
\]
it follows
\[
P (N_k) \leq C2^{\alpha \gamma k} (T 2^{-k})^{1+\beta} 2^k = C2^{k(\alpha \gamma - \beta)} T^{1+\beta}.
\]

Since \(\gamma < \beta/\alpha\),
\[
\sum_{k=1}^{\infty} P (N_k) \leq C T^{1+\beta} \sum_{k=1}^{\infty} 2^{k(\alpha \gamma - \beta)} < \infty
\]
and so by the Borel Cantelli lemma, Lemma \[(13.3.11)\], there exists a set of measure zero \(N\), such that if \(\omega \notin N\), then \(\omega\) is in only finitely many \(N_k\). In other words, for \(\omega \notin N\), there exists \(M (\omega)\) such that if \(k \geq M (\omega)\), then for each \(j\),
\[
||X (r_{j+1}^k) (\omega) - X (r_j^k) (\omega)|| \leq 2^{-\gamma k}.
\]

It follows from Lemma \[(13.3.12)\] that \(t \rightarrow X (t) (\omega)\) is Holder continuous on \(D\) with Holder exponent \(\gamma\). Note the constant is a measurable function of \(\omega\), depending on how many measurable \(N_k\) which contain \(\omega\).

By Lemma \[(13.3.13)\], one can define \(Y (t) (\omega)\) to be the unique function which extends \(d \rightarrow X (d) (\omega)\) off \(D\) for \(\omega \notin N\) and let \(Y (t) (\omega) = 0\) if \(\omega \in N\). Thus by Lemma \[(13.3.14)\], \(t \rightarrow Y (t) (\omega)\) is Holder continuous. Also, \(\omega \rightarrow Y (t) (\omega)\) is measurable because it is the pointwise limit of measurable functions
\[
Y (t) (\omega) = \lim_{d \rightarrow t} X (d) (\omega) \mathcal{X}_{NC} (\omega).
\]

It remains to verify the claim that \(Y (t) (\omega) = X (t) (\omega)\) a.e.
\[
\mathcal{X}_{||Y(t)-X(t)||<\varepsilon} \leq \liminf_{d \rightarrow t} \mathcal{X}_{||X(d)-X(t)||<\varepsilon} (\omega)
\]
because if \(\omega \in N\) both sides are 0 and if \(\omega \in NC\) then the above limit in \[(13.3.15)\] holds and so if \(||Y (t) (\omega) - X (t) (\omega)|| > \varepsilon\), the same is true of \(||X (d) (\omega) - X (t) (\omega)|| > \varepsilon\) whenever \(d\) is close enough to \(t\) and so by Fatou's lemma,
\[
P (||Y (t) - X (t)|| > \varepsilon) = \int \mathcal{X}_{||Y(t)-X(t)||<\varepsilon} (\omega) dP \leq \liminf_{d \rightarrow t} \int \mathcal{X}_{||X(d)-X(t)||<\varepsilon} (\omega) dP \leq \liminf_{d \rightarrow t} \int \mathcal{X}_{||X(d)-X(t)||<\varepsilon} (\omega) dP \leq \liminf_{d \rightarrow t} \epsilon^{-\alpha} \int \mathcal{X}_{||X(d)-X(t)||<\varepsilon} dP \leq \liminf_{d \rightarrow t} \epsilon^{-\alpha} \int \mathcal{X}_{||X(d)-X(t)||<\varepsilon} dP = 0.
\]
Therefore,

\[
P \left( \left| Y(t) - X(t) \right| > 0 \right) \\
= P \left( \bigcup_{k=1}^{\infty} \left[ \left| Y(t) - X(t) \right| > \frac{1}{k} \right] \right) \\
\leq \sum_{k=1}^{\infty} P \left( \left[ \left| Y(t) - X(t) \right| > \frac{1}{k} \right] \right) = 0. \blacksquare
\]

A few observations are interesting. In the proof, the following inequality was obtained.

\[
||X(d')(\omega) - X(d)(\omega)|| \leq \frac{2}{T^{\gamma}(1-2^{\gamma})} \left( T^{2-(n+1)} \right)^{\gamma} \\
\leq \frac{2}{T^{\gamma}(1-2^{\gamma})} (|d-d'|)^{\gamma}
\]

which was so for any \(d', d \in D\) with \(|d' - d| < T^{2-(M(\omega)+1)}\). Thus the Holder continuous version of \(X\) will satisfy

\[
||Y(t)(\omega) - Y(s)(\omega)|| \leq \frac{2}{T^{\gamma}(1-2^{\gamma})} (|t-s|)^{\gamma}
\]

provided \(|t-s| < T^{2-(M(\omega)+1)}\). Does this translate into an inequality of the form

\[
||Y(t)(\omega) - Y(s)(\omega)|| \leq \frac{2}{T^{\gamma}(1-2^{\gamma})} (|t-s|)^{\gamma}
\]

for any pair of points \(t, s \in [0, T]\)? It seems it does not for any \(\gamma < 1\) although it does yield

\[
||Y(t)(\omega) - Y(s)(\omega)|| \leq C (|t-s|)^{\gamma}
\]

where \(C\) depends on the number of intervals having length less than \(T^{2-(M(\omega)+1)}\) which it takes to cover \([0, T]\). First note that if \(\gamma > 1\), then the Holder continuity will imply \(t \rightarrow Y(t)(\omega)\) is a constant. Therefore, the only case of interest is \(\gamma < 1\). Let \(s, t\) be any pair of points and let \(s = x_0 < \cdots < x_n = t\) where \(|x_i - x_{i-1}| < T^{2-(M(\omega)+1)}\). Then

\[
||Y(t)(\omega) - Y(s)(\omega)|| \leq \sum_{i=1}^{n} ||Y(x_i)(\omega) - Y(x_{i-1})(\omega)|| \\
\leq \frac{2}{T^{\gamma}(1-2^{\gamma})} \sum_{i=1}^{n} (|x_i - x_{i-1}|)^{\gamma}
\]

How does this compare to

\[
\left( \sum_{i=1}^{n} |x_i - x_{i-1}| \right)^{\gamma} = |t-s|^{\gamma}?
\]

This last expression is smaller than the right side of (13.3.15) for any \(\gamma < 1\). Thus for \(\gamma < 1\), the constant in the conclusion of the theorem depends on both \(T\) and \(\omega \notin N\).

In the case where \(\alpha \geq 1\), here is another proof of this theorem. It is based on the one in the book by Stroock [SI].

**Theorem 13.3.3** Suppose \(X\) is a stochastic process on \([0, T]\) having values in the Banach space \(E\). Suppose also that there exists a constant, \(C\) and positive numbers \(\alpha, \beta, \gamma \geq 1\), such that

\[
E \left( \left| X(t) - X(s) \right|^{\alpha} \right) \leq C |t-s|^{1+\beta}
\]

Then there exists a stochastic process \(Y\) such that for a.e. \(\omega, t \rightarrow Y(t)(\omega)\) is Holder continuous with exponent \(\gamma < \frac{\beta}{\alpha}\) and for each \(t, P \left( \left| X(t) - Y(t) \right| > 0 \right) = 0. (Y\) is a version of \(X\).) Also

\[
E \left( \sup_{0 \leq s < t \leq T} \frac{||Y(t) - Y(s)||}{(t-s)^{\gamma}} \right) \leq C
\]

where \(C\) depends on \(\alpha, \beta, T, \gamma\).
Proof: The proof considers piecewise linear approximations of $X$ which are automatically continuous. These are shown to converge to $Y$ in $L^\infty (\Omega; C ([0, T], E))$ so it follows that $Y$ must be continuous for a.e. $\omega$. Finally, it is shown that $Y$ is a version of $X$ and is Hölder continuous. In the proof, I will use $C$ to denote a constant which depends on the quantities $\gamma, \alpha, \beta, T$. Let $\{t^n_k\}_{k=0}^{2^n}$ be a uniform partition of the interval $[0, T]$ so that $t^n_{k+1} - t^n_k = T2^{-n}$. Now let

$$M_n = \max_{k \leq 2^n} \| X (t^n_k) - X (t^n_{k-1}) \|$$

Then it follows that

$$M_n \leq \sum_{k=1}^{2^n} \| X (t^n_k) - X (t^n_{k-1}) \|^\alpha$$

and so

$$E (M_n) \leq \sum_{k=1}^{2^n} C (T2^{-n})^{1+\beta} = C2^n 2^{-n(1+\beta)} = C2^{-n\beta} \quad (13.3.17)$$

Next denote by $X_n$ the piecewise linear function which results from the values of $X$ at the points $t^n_k$. Consider the following picture which illustrates a part of the graphs of $X_n$ and $X_{n+1}$.

Then

$$\max_{t \in [0, T]} \| X_{n+1} (t) - X_n (t) \| \leq \max_{1 \leq k \leq 2^n+1} \left\| X_2 (t^{n+1}_{2k-1}) - \frac{X (t^n_k) + X (t^n_{k-1})}{2} \right\|$$

$$\leq \max_{k \leq 2^n+1} \left( \frac{1}{2} \| X (t^{n+1}_{2k-1}) - X (t^n_{2k}) \| + \frac{1}{2} \| X (t^{n+1}_{2k}) - X (t^{n+1}_{2k-2}) \| \right) \leq M_{n+1}$$

Denote by $\| \cdot \|_\infty$ the usual norm in $C ([0, T], E)$,

$$\max_{t \in [0, T]} \| Z (t) \| \equiv \| Z \|_\infty,$$

Then from what was just established,

$$E (\| X_{n+1} - X_n \|_\infty^\alpha) = \int_\Omega \| X_{n+1} - X_n \|_\infty^\alpha dP \leq E (M_{n+1}) = C2^{-n\beta}$$

which shows that

$$\| X_{n+1} - X_n \|_{L^\alpha (\Omega; C ([0, T], E))} = \left( \int_\Omega \| X_{n+1} - X_n \|_\infty^\alpha dP \right)^{1/\alpha} \leq C (2^{(\beta/\alpha)})^{-n}$$

Also, for $m > n$, it follows from the assumption that $\alpha \geq 1$,

$$\| X_m - X_n \|_{L^\alpha (\Omega; C ([0, T], E))} \leq \sum_{k=n}^{\infty} C (2^{(\beta/\alpha)})^{-k} \leq C \frac{(2^{(\beta/\alpha)})^{-n}}{1 - 2^{-(\beta/\alpha)}} = C (2^{(\beta/\alpha)})^{-n} \quad (13.3.18)$$

Thus $\{X_n\}$ is a Cauchy sequence in $L^\alpha (\Omega; C ([0, T], E))$ and so it converges to some $Y$ in this space, a subsequence converging pointwise. Then from Fatou’s lemma,

$$\| Y - X_n \|_{L^\alpha (\Omega; C ([0, T], E))} \leq C (2^{(\beta/\alpha)})^{-n} \quad (13.3.19)$$
Also, for a.e. \( \omega; t \to Y(t) \) is in \( C([0,T], E) \). It remains to verify that \( Y(t) = X(t) \) a.e.

From the construction, it follows that for any \( n \) and \( m \geq n \)
\[
Y(t^n_k) = X_m(t^n_k) = X(t^n_k)
\]

Thus
\[
\|Y(t) - X(t)\| \leq \|Y(t) - Y(t^n_k)\| + \|Y(t^n_k) - X(t)\| = \|Y(t) - Y(t^n_k)\| + \|X(t^n_k) - X(t)\|
\]

Now from the hypotheses of the theorem,
\[
P\left(\|X(t^n_k) - X(t)\|^{\alpha} > \varepsilon\right) \leq \frac{1}{\varepsilon} E\left(\|X(t^n_k) - X(t)\|^{\alpha}\right) \leq \frac{C}{\varepsilon} |t^n_k - t|^{1+\beta}
\]

Thus, there exists a sequence of mesh points \( \{s_n\} \) converging to \( t \) such that
\[
P\left(\|X(s_n) - X(t)\|^{\alpha} > 2^{-n}\right) \leq 2^{-n}
\]

Then by the Borel Cantelli lemma, there is a set of measure zero \( N \) such that for \( \omega \notin N \),
\[
\|X(s_n) - X(t)\|^{\alpha} \leq 2^{-n}
\]

for all \( n \) large enough. Then
\[
\|Y(t) - X(t)\| \leq \|Y(t) - Y(s_n)\| + \|X(s_n) - X(t)\|
\]

which shows that, by continuity of \( Y \), for \( \omega \) not in an exceptional set of measure zero, \( \|Y(t) - X(t)\| = 0 \).

It remains to verify the assertion about Holder continuity of \( Y \). Let \( 0 \leq s < t \leq T \). Then for some \( n \),
\[
2^{-(n+1)T} \leq t - s \leq 2^{-n}T
\]

Thus
\[
\|Y(t) - Y(s)\| \leq \|Y(t) - X_n(t)\| + \|X_n(t) - X_n(s)\| + \|X_n(s) - Y(s)\|
\]
\[
\leq 2 \sup_{\tau \in [0,T]} \|Y(\tau) - X_n(\tau)\| + \|X_n(t) - X_n(s)\|
\]

(13.3.21)

Now
\[
\frac{\|X_n(t) - X_n(s)\|}{t - s} \leq \frac{\|X_n(t) - X_n(s)\|}{2^{-(n+1)T}}
\]

From [13.3.21] a picture like the following must hold.

\[
\begin{array}{cccc}
& t_{k-1} & s & t_k & t_{k+1} \\
1_{n+1} & & & & \\
k-1 & & & & \\
k & & & & \\
& & & & \\
& & & & \\
& & & & \\
\end{array}
\]

Therefore, from the above,
\[
\frac{\|X_n(t) - X_n(s)\|}{t - s} \leq \frac{\|X(t_{k-1}^{n+1}) - X(t_k^{n+1})\|}{2^{-(n+1)T}} + \|X(t_k^{n+1}) - X(t_{k+1}^{n+1})\|
\]
\[
\leq C2^n M_{n+1}
\]

It follows from [13.3.21],
\[
\|Y(t) - Y(s)\| \leq 2 \|Y - X_n\| + C2^n M_{n+1} (t - s)
\]

Next, letting \( \gamma < \beta/\alpha \), and using [13.3.21],
\[
\frac{\|Y(t) - Y(s)\|}{(t - s)^\gamma} \leq 2 \left( T^{-1} 2^{n+1}\right)^\gamma \|Y - X_n\| + C2^n (2^{-n})^{1-\gamma} M_{n+1}
\]
\[
= C2^n \gamma \left( \|Y - X_n\| + M_{n+1}\right)
\]
The above holds for any \( s, t \) satisfying \( 2^{-n}T \leq s < t < T \). Then
\[
\sup_{0 \leq s < t \leq T} \left( \frac{\| Y(t) - Y(s) \|}{(t-s)^\gamma} \right)^\alpha \leq C 2^{n\gamma} (\| Y - X_n \|_\infty + M_{n+1})
\]
Denote by \( P_n \) the ordered pairs \( s, t \) satisfying the above condition that
\[
0 \leq s < t \leq T, |t-s| \in \left[ 2^{-(n+1)}T, 2^{-n}T \right]
\]
Thus for a.e. \( \omega \), and for all \( n \),
\[
\left( \sup_{(s,t) \in P_n} \frac{\| Y(t) - Y(s) \|}{(t-s)^\gamma} \right)^\alpha \leq C \sum_{k=0}^{\infty} 2^{k\alpha\gamma} (\| Y - X_k \|_\infty + M_{k+1})
\]
Note that \( n \) is arbitrary. Hence
\[
\sup_{0 \leq s < t \leq T} \left( \frac{\| Y(t) - Y(s) \|}{(t-s)^\gamma} \right)^\alpha \leq \sum_{k=0}^{\infty} 2^{k\alpha\gamma} (\| Y - X_k \|_\infty + M_{k+1})
\]
By continuity of \( Y \), the result on the left is unchanged if the ordered pairs are restricted to lie in \( \mathbb{Q} \cap [0, T] \times \mathbb{Q} \cap [0, T] \), a countable set. Thus the left side is measurable. It follows from [13.3.20] and [13.3.17] which say
\[
\| Y - X_k \|_{L^\infty([0, T] \times E)} \leq C \left( 2^{(\beta/\alpha)} \right)^{-k}, \quad E(M_k^\alpha) \leq C 2^{-k\beta}
\]
that
\[
E \left( \sup_{0 \leq s < t \leq T} \left( \frac{\| Y(t) - Y(s) \|}{(t-s)^\gamma} \right)^\alpha \right) \leq \sum_{k=0}^{\infty} C 2^{k\alpha\gamma} 2^{-\beta k} \equiv C < \infty
\]
because \( \alpha \gamma - \beta < 0 \). By continuity of \( Y \), there are no measurability concerns in taking the above expectation. Note that this implies, since \( \alpha \geq 1 \),
\[
E \left( \sup_{0 \leq s < t \leq T} \left( \frac{\| Y(t) - Y(s) \|}{(t-s)^\gamma} \right)^\alpha \right) \leq C \left( E \left( \sup_{0 \leq s < t \leq T} \left( \frac{\| Y(t) - Y(s) \|}{(t-s)^\gamma} \right)^\alpha \right) \right)^{1/\alpha} \leq C^{1/\alpha} \equiv C
\]
Now
\[
P \left( \sup_{0 \leq s < t \leq T} \left( \frac{\| Y(t) - Y(s) \|}{(t-s)^\gamma} \right)^\alpha > 2^{\alpha k} \right) \leq \frac{1}{2^{\alpha k}} C
\]
and so there exists a set of measure zero \( N \) such that for \( \omega \notin N \),
\[
\sup_{0 \leq s < t \leq T} \left( \frac{\| Y(t) - Y(s) \|}{(t-s)^\gamma} \right)^\alpha \leq 2^{\alpha k}
\]
for all \( k \) large enough. Pick such a \( k \), depending on \( \omega \notin N \). Then for any \( s, t \),
\[
\frac{\| Y(t) - Y(s) \|}{(t-s)^\gamma} \leq 2^k
\]
and so, this has shown that for \( \omega \notin N \),
\[
\| Y(t) - Y(s) \| \leq C(\omega)(t-s)^\gamma
\]
Note that if \( X(t) \) is known to be continuous off a set of measure zero, then the piecewise linear approximations converge to \( X(t) \) in \( C([0,T],E) \) off this set of measure zero. Therefore, it must be that off a set of measure zero, \( Y(t) = X(t) \) and so in fact \( X(t) \) is Holder continuous off a set of measure zero and the condition on expectation also must hold, that is

\[
E \left( \sup_{0 \leq s < t \leq T} \frac{\|X(t) - X(s)\|}{(t-s)^\gamma} \right) \leq C.
\]

### 13.4 Filtrations

Instead of having a sequence of \( \sigma \) algebras, one can consider an increasing collection of \( \sigma \) algebras indexed by \( t \in \mathbb{R} \). This is called a filtration.

**Definition 13.4.1** Let \( X \) be a stochastic process defined on an interval, \( I = [0,T] \) or \( [0,\infty) \). Suppose the probability space, \( (\Omega, F, P) \) has an increasing family of \( \sigma \) algebras, \( \{F_t\} \). This is called a filtration. If for arbitrary \( t \in I \) the random variable \( X(t) \) is \( F_t \) measurable, then \( X \) is said to be adapted to the filtration \( \{F_t\} \). Denote by \( F_t^+ \) the intersection of all \( F_s \) for \( s > t \). The filtration is normal if

1. \( F_0 \) contains all \( A \in F \) such that \( P(A) = 0 \)
2. \( F_t = F_t^+ \) for all \( t \in I \).

\( X \) is called progressively measurable if for every \( t \in I \), the mapping

\[
(s, \omega) \in [0,t] \times \Omega, \quad (s, \omega) \to X(s, \omega)
\]

is \( B([0,t]) \times F_t \) measurable.

Thus \( X \) is progressively measurable means

\[
(s, \omega) \to X_{[0,t]}(s)X(s, \omega)
\]

is \( B([0,t]) \times F_t \) measurable. As an example of a normal filtration, here is an example.

**Example 13.4.2** For example, you could have a stochastic process, \( X(t) \) and you could define

\[
G_t = \sigma (X(s) : s \leq t),
\]

the completion of the smallest \( \sigma \) algebra such that each \( X(s) \) is measurable for all \( s \leq t \). This gives an example of a filtration to which \( X(t) \) is adapted which satisfies 1. More generally, suppose \( X(t) \) is adapted to a filtration, \( G_t \). Define

\[
F_t = \bigcap_{s > t} G_s
\]

Then

\[
F_t^+ = \bigcap_{s > t} F_s = \bigcap_{s > t} \bigcap_{r > s} G_r = \bigcap_{s > t} G_s = F_t.
\]

and each \( X(t) \) is measurable with respect to \( F_t \). Thus there is no harm in assuming a stochastic process adapted to a filtration can be modified so the filtration is normal. Also note that \( F_t \) defined this way will be complete so if \( A \in F_t \) has \( P(A) = 0 \) and if \( B \subseteq A \), then \( B \in F_t \) also. This is because this relation between the sets and the probability of \( A \) being zero, holds for this pair of sets when considered as elements of each \( G_s \) for \( s > t \). Hence \( B \in G_s \) for each \( s > t \) and is therefore one of the sets in \( F_t \).

What is the description of a progressively measurable set?

It means that for \( Q \) progressively measurable, \( Q \cap [0,t] \times \Omega \) as shown in the above picture is \( B([0,t]) \times F_t \) measurable. It is like saying a little more descriptively that the function is progressively product measurable.

I shall generally assume the filtration is normal.
13.4. FILTRATIONS

Observation 13.4.3 If $X$ is progressively measurable, then it is adapted. Furthermore the progressively measurable sets, those $E \cap [0,T] \times \Omega$ for which $\mathcal{X}_E$ is progressively measurable form a $\sigma$ algebra.

To see why this is, consider $X$ progressively measurable and fix $t$. Then $(s, \omega) \to X(s, \omega)$ for $(s, \omega) \in [0, t] \times \Omega$ is given to be $\mathcal{B}([0, t]) \times \mathcal{F}_t$ measurable, the ordinary product measure and so fixing any $s \in [0, t]$, it follows the resulting function of $\omega$ is $\mathcal{F}_t$ measurable. In particular, this is true upon fixing $s = t$. Thus $\omega \to X(t, \omega)$ is $\mathcal{F}_t$ measurable and so $X(t)$ is adapted.

A set $E \subseteq [0, T] \times \Omega$ is progressively measurable means that $\mathcal{X}_E$ is progressively measurable. That is $\mathcal{X}_E$ restricted to $[0, t] \times \Omega$ is $\mathcal{B}([0, t]) \times \mathcal{F}_t$ measurable. In other words, $E$ is progressively measurable if

$$E \cap ([0, t] \times \Omega) \in \mathcal{B}([0, t]) \times \mathcal{F}_t.$$

If $E_i$ is progressively measurable, does it follow that $E = \bigcup_{i=1}^{\infty} E_i$ is also progressively measurable? Yes.

$$E \cap ([0, t] \times \Omega) = \bigcup_{i=1}^{\infty} E_i \cap ([0, t] \times \Omega) \in \mathcal{B}([0, t]) \times \mathcal{F}_t$$

because each set in the union is in $\mathcal{B}([0, t]) \times \mathcal{F}_t$. If $E$ is progressively measurable, is $E^C$?

$${E^C \cap ([0, t] \times \Omega) \cup \left( E \cap ([0, t] \times \Omega) \right) = \bigcup_{i=1}^{\infty} E^C_i \cap ([0, t] \times \Omega)}$$

and so $E^C \cap ([0, t] \times \Omega) \in \mathcal{B}([0, t]) \times \mathcal{F}_t$. Thus the progressively measurable sets are a $\sigma$ algebra.

Another observation of interest is in the following lemma.

Lemma 13.4.4 Suppose $Q$ is in $\mathcal{B}([0, a]) \times \mathcal{F}_r$. Then if $b \geq a$ and $t \geq r$, then $Q$ is also in $\mathcal{B}([0, b]) \times \mathcal{F}_t$.

Proof: Consider a measurable rectangle $A \times B$ where $A \in \mathcal{B}([0, a])$ and $B \in \mathcal{F}_r$. Is it true that $A \times B \in \mathcal{B}([0, b]) \times \mathcal{F}_t$? This reduces to the question whether $A \in \mathcal{B}([0, b])$. If $A$ is an interval, it is clear that $A \in \mathcal{B}([0, b])$. Consider the $\pi$ system of intervals and let $\mathcal{G}$ denote those Borel sets $A \in \mathcal{B}([0, a])$ such that $A \in \mathcal{B}([0, b])$. If $A \in \mathcal{G}$, then $[0, b] \setminus A \in \mathcal{B}([0, b])$ by assumption (the difference of Borel sets is surely Borel). However, this set equals

$$([0, a] \setminus A) \cup \{a, b\}$$

and so

$$[0, b] = ([0, a] \setminus A) \cup \{a, b\} \cup A$$

The set on the left is in $\mathcal{B}([0, b])$ and the sets on the right are disjoint and two of them are also in $\mathcal{B}([0, b])$. Therefore, the third, $([0, a] \setminus A)$ is in $\mathcal{B}([0, b])$. It is obvious that $\mathcal{G}$ is closed with respect to countable disjoint unions. Therefore, by Lemma 13.4.3, Dynkin’s lemma, $\mathcal{G} \supseteq \sigma$ (Intervals) = $\mathcal{B}([0, a])$.

Therefore, such a measurable rectangle $A \times B$ where $A \in \mathcal{B}([0, a])$ and $B \in \mathcal{F}_r$ is in $\mathcal{B}([0, b]) \times \mathcal{F}_t$ and in fact it is a measurable rectangle in $\mathcal{B}([0, b]) \times \mathcal{F}_t$. Now let $\mathcal{K}$ denote all these measurable rectangles $A \times B$ where $A \in \mathcal{B}([0, a])$ and $B \in \mathcal{F}_r$. Let $\mathcal{G}$ (new $\mathcal{G}$) denote those sets $Q$ of $\mathcal{B}([0, a]) \times \mathcal{F}_r$ which are in $\mathcal{B}([0, b]) \times \mathcal{F}_t$. Then if $Q \in \mathcal{G}$,

$$Q \cup ([0, a] \times \Omega \setminus Q) \cup \{a, b\} \times \Omega = [a, b] \times \Omega$$

Then the sets are disjoint and all but $[0, a] \times \Omega \setminus Q$ are in $\mathcal{B}([0, b]) \times \mathcal{F}_t$. Therefore, this one is also in $\mathcal{B}([0, b]) \times \mathcal{F}_t$. If $Q_i \in \mathcal{G}$ and the $Q_i$ are disjoint, then $\bigcup_i Q_i$ is also in $\mathcal{B}([0, b]) \times \mathcal{F}_t$ and so $\mathcal{G}$ is closed with respect to countable disjoint unions and complements. Hence $\mathcal{G} \supseteq \sigma(\mathcal{K}) = \mathcal{B}([0, a]) \times \mathcal{F}_r$ which shows

$$\mathcal{B}([0, a]) \times \mathcal{F}_r \subseteq \mathcal{B}([0, b]) \times \mathcal{F}_t$$

A significant observation is the following which states that the integral of a progressively measurable function is progressively measurable.

Proposition 13.4.5 Suppose $X : [0, T] \times \Omega \to E$ where $E$ is a separable Banach space. Also suppose that $X(\cdot, \omega) \in L^1([0, T], E)$ for each $\omega$. Here $\mathcal{F}_t$ is a filtration and with respect to this filtration, $X$ is progressively measurable. Then

$$(t, \omega) \to \int_0^t X(s, \omega) \, ds$$

is also progressively measurable.
Proof: Suppose \( Q \in [0, T] \times \Omega \) is progressively measurable. This means for each \( t \),
\[
Q \cap [0, t] \times \Omega \in \mathcal{B}([0, t]) \times \mathcal{F}_t
\]
What about \( (s, \omega) \in [0, t] \times \Omega, (s, \omega) \rightarrow \int_0^s X_Q dr? \)
Is that function on the right \( \mathcal{B}([0, t]) \times \mathcal{F}_t \) measurable? We know that \( Q \cap [0, s] \times \Omega \) is \( \mathcal{B}([0, s]) \times \mathcal{F}_s \) measurable and hence \( \mathcal{B}([0, t]) \times \mathcal{F}_t \) measurable. When you integrate a product measurable function, you do get one which is product measurable. Therefore, this function must be \( \mathcal{B}([0, t]) \times \mathcal{F}_t \) measurable. This shows that \( (t, \omega) \rightarrow \int_0^t X_Q (s, \omega) ds \) is progressively measurable. Here is a claim which was just used.

Claim: If \( Q \in \mathcal{B}([0, t]) \times \mathcal{F}_t \) measurable, then \( (s, \omega) \rightarrow \int_0^s X_Q dr \) is also \( \mathcal{B}([0, t]) \times \mathcal{F}_t \) measurable.

Proof of claim: First consider \( A \times B \) where \( A \in \mathcal{B}([0, t]) \) and \( B \in \mathcal{F}_t \). Then
\[
\int_0^s X_{A \times B} dr = \int_0^s X_A X_B dr = X_B (\omega) \int_0^s X_A (s) dr
\]
This is clearly \( \mathcal{B}([0, t]) \times \mathcal{F}_t \) measurable. It is the product of a continuous function of \( s \) with the indicator function of a set in \( \mathcal{F}_t \). Now let
\[
\mathcal{G} = \left\{ Q \in \mathcal{B}([0, t]) \times \mathcal{F}_t : (s, \omega) \rightarrow \int_0^s X_Q (r, \omega) dr \text{ is } \mathcal{B}([0, t]) \times \mathcal{F}_t \text{ measurable} \right\}
\]
Then it was just shown that \( \mathcal{G} \) contains the measurable rectangles. It is also clear that \( \mathcal{G} \) is closed with respect to countable disjoint unions and complements. Therefore, \( \mathcal{G} \supseteq \sigma (\mathcal{K}_t) = \mathcal{B}([0, t]) \times \mathcal{F}_t \) where \( \mathcal{K}_t \) denotes the measurable rectangles \( A \times B \) where \( B \in \mathcal{F}_t \) and \( A \in \mathcal{B}([0, t]) = \mathcal{B}([0, T]) \cap [0, t] \). This proves the claim.

Thus if \( Q \) is progressively measurable, it follows that \( (s, \omega) \rightarrow \int_0^s X_Q (r, \omega) dr \equiv f (s, \omega) \) is progressively measurable because for \( (s, \omega) \in [0, t] \times \Omega, (s, \omega) \rightarrow f (s, \omega) \) is \( \mathcal{B}([0, t]) \times \mathcal{F}_t \) measurable. This is what was to be proved in this special case.

Now consider the conclusion of the proposition. By considering the positive and negative parts of \( \phi (X) \) for \( \phi \in E' \), and using Pettis theorem, it suffices to consider the case where \( X \geq 0 \). Then there exists an increasing sequence of progressively measurable simple functions \( \{X_n\} \) converging pointwise to \( X \). From what was just shown,

\[
(t, \omega) \rightarrow \int_0^t X_n ds
\]
is progressively measurable. Hence, by the monotone convergence theorem, \( (t, \omega) \rightarrow \int_0^t X ds \) is also progressively measurable.

What else can you do to something which is progressively measurable and obtain something which is progressively measurable? It turns out that shifts in time can preserve progressive measurability. Let \( \mathcal{F}_t \) be a filtration on \([0, T]\) and extend the filtration to be equal to \( \mathcal{F}_0 \) and \( \mathcal{F}_T \) for \( t < 0 \) and \( t > T \), respectively. Recall the following definition of progressively measurable sets.

Definition 13.4.6 Denote by \( \mathcal{P} \) those sets \( Q \) in \( \mathcal{F}_T \times \mathcal{B}([0, T]) \) such that for \( t \in [-\infty, T] \)
\[
\Omega \times (-\infty, t] \cap Q \in \mathcal{F}_t \times \mathcal{B}((-\infty, t]).
\]

Lemma 13.4.7 Define \( Q + h \) as
\[
Q + h \equiv \{(t + h, \omega) : (t, \omega) \in Q\}.
\]
Then if \( Q \in \mathcal{P} \), it follows that \( Q + h \in \mathcal{P} \).

Proof: This is most easily seen through the use of the following diagram. In this diagram, \( Q \) is in \( \mathcal{P} \) so it is progressively measurable.
13.4. FILTRATIONS

By definition, $S$ in the picture is $\mathcal{B}((-\infty, t-h]) \times \mathcal{F}_{t-h}$ measurable. Hence $S + h \equiv Q + h \cap \Omega \times (-\infty, t]$ is $\mathcal{B}((-\infty, t]) \times \mathcal{F}_{t-h}$ measurable. To see this, note that if $B \times A \in \mathcal{B}((-\infty, t-h]) \times \mathcal{F}_{t-h}$, then translating it by $h$ gives a set in $\mathcal{B}((-\infty, t]) \times \mathcal{F}_{t-h}$. Then if $\mathcal{G}$ consists of sets $S$ in $\mathcal{B}((-\infty, t-h]) \times \mathcal{F}_{t-h}$ for which $S + h$ is in $\mathcal{B}((-\infty, t]) \times \mathcal{F}_{t-h}$, $\mathcal{G}$ is closed with respect to countable disjoint unions and complements. Thus, $\mathcal{G}$ equals $\mathcal{B}((-\infty, t-h]) \times \mathcal{F}_{t-h}$. In particular, it contains the set $S$ just described. □

Now for $h > 0$,

$$\tau_h f(t) \equiv \begin{cases} f(t-h) & \text{if } t \geq h, \\ 0 & \text{if } t < h. \end{cases}$$

Lemma 13.4.8 Let $Q \in \mathcal{P}$. Then $\tau_h X_Q$ is $\mathcal{P}$ measurable.

Proof: If $\tau_h X_Q(t,\omega) = 1$, then you need to have $(t-h,\omega) \in Q$ and so $(t,\omega) \in Q + h$. Thus

$$\tau_h X_Q = X_{Q+h},$$

which is $\mathcal{P}$ measurable since $Q \in \mathcal{P}$. In general,

$$\tau_h X_Q = X_{[h,T] \times \Omega} X_{Q+h},$$

which is $\mathcal{P}$ measurable. □

This lemma implies the following.

Lemma 13.4.9 Let $f(t,\omega)$ have values in a separable Banach space and suppose $f$ is $\mathcal{P}$ measurable. Then $\tau_h f$ is $\mathcal{P}$ measurable.

Proof: Taking values in a separable Banach space and being $\mathcal{P}$ measurable, $f$ is the pointwise limit of $\mathcal{P}$ measurable simple functions. If $s_n$ is one of these, then from the above lemmas, $\tau_h s_n$ is $\mathcal{P}$ measurable. Then, letting $n \to \infty$, it follows that $\tau_h f$ is $\mathcal{P}$ measurable. □

The following is similar to Proposition 13.2.2. It shows that under pretty weak conditions, an adapted process has a progressively measurable adapted version.

Proposition 13.4.10 Let $X$ be a stochastically continuous adapted process for a normal filtration defined on a closed interval, $I \equiv [0,T]$. Then $X$ has a progressively measurable adapted version.

Proof: By Lemma 13.2.1 $X$ is uniformly stochastically continuous and so there exists a sequence of positive numbers, $\{\rho_n\}$ such that if $|s-t| < \rho_n$, then

$$P \left( \left| X(t) - X(s) \right| \geq \frac{1}{2^n} \right) \leq \frac{1}{2^n}. \quad (13.4.22)$$

Then let $\{t^n_0, t^n_1, \ldots, t^n_m\}$ be a partition of $[0,T]$ in which $|t^n_i - t^n_{i-1}| < \rho_n$. Now define $X_n$ as follows:

$$X_n(t)(\omega) \equiv \sum_{i=1}^{m_n} X(t^n_{i-1})(\omega) X(t^n_i, t^n_{i-1})(t)$$

$$X_n(T) \equiv X(T).$$

Then $(s,\omega) \to X_n(s,\omega)$ for $(s,\omega) \in [0,t] \times \Omega$ is obviously $B([0,t]) \times \mathcal{F}_t$ measurable. Consider the set, $A$ on which $\{X_n(t,\omega)\}$ is a Cauchy sequence. This set is of the form

$$A = \cap_{n=1}^{\infty} \cup_{m=1}^{\infty} \cap_{p,q \geq m} \left| X_p - X_q \right| < \frac{1}{n}$$

and so it is a $B(I) \times \mathcal{F}$ measurable set and $A \cap [0,t] \times \Omega$ is $B([0,t]) \times \mathcal{F}_t$ measurable for each $t \leq T$ because each $X_q$ in the above has the property that its restriction to $[0,t] \times \Omega$ is $B([0,t]) \times \mathcal{F}_t$ measurable. Now define

$$Y(t,\omega) \equiv \begin{cases} \lim_{n \to \infty} X_n(t,\omega) & \text{if } (t,\omega) \in A \\ 0 & \text{if } (t,\omega) \notin A \end{cases}$$

I claim that for each $t$, $Y(t,\omega) = X(t,\omega)$ for a.e. $\omega$. To see this, consider Proposition 13.2.2. From the construction of $X_n$, it follows that for each $t$,

$$P \left( \left| X_n(t) - X(t) \right| \geq \frac{1}{2^n} \right) \leq \frac{1}{2^n}.$$
Also, for a fixed \( t \), if \( X_n(t, \omega) \) fails to converge to \( X(t, \omega) \), then \( \omega \) must be in infinitely many of the sets,

\[
B_n \equiv \left\{ \|X_n(t) - X(t)\| \geq \frac{1}{2^n} \right\}
\]

which is a set of measure zero by the Borel Cantelli lemma. Recall why this is so.

\[
P(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} B_n) \leq \sum_{n=k}^{\infty} P(B_n) < \frac{1}{2^{k-1}}
\]

Therefore, for each \( t, (t, \omega) \in A \) for a.e. \( \omega \). Hence \( X(t) = Y(t) \) a.e. and so \( Y \) is a measurable version of \( X \). \( Y \) is adapted because the filtration is normal and hence \( F_t \) contains all sets of measure zero. Therefore, \( Y(t) \) differs from \( X(t) \) on a set which is \( F_t \) measurable. ■

There is a more specialized situation in which the measurability of a stochastic process automatically implies it is adapted. Furthermore, this can be defined easily in terms of a \( \pi \) system of sets.

**Definition 13.4.11** Let \( F_t \) be a filtration on \((\Omega, F, P)\) and denote by \( P_\infty \) the smallest \( \sigma \) algebra of sets of \([0, \infty) \times \Omega\) containing the sets

\[
(s, t] \times F, F \in F_s \quad \{ 0 \} \times F, F \in F_0.
\]

This is called the predictable \( \sigma \) algebra. and the sets in this \( \sigma \) algebra are called the predictable sets. Denote by \( P_T \) the intersection of \( P_\infty \) to \([0, T] \times \Omega\). A stochastic process \( X \) which maps either \([0, T] \times \Omega \) or \([0, \infty) \times \Omega \) to \( E \), a separable real Banach space is called predictable if for every Borel set \( A \in B(E) \), it follows \( X^{-1}(A) \in P_T \) or \( P_\infty \).

This is a lot like product measure except one of the \( \sigma \) algebras is changing.

**Proposition 13.4.12** Let \( F_t \) be a filtration as above and let \( X \) be a predictable stochastic process. Then \( X \) is \( F_t \) adapted.

**Proof:** Let \( s_0 > 0 \) and define

\[
\mathcal{G}_{s_0} \equiv \{ S \in P_\infty : S_{s_0} \in F_{s_0} \}
\]

where

\[
S_{s_0} \equiv \{ \omega \in \Omega : (s_0, \omega) \in S \}.
\]

It is clear \( \mathcal{G}_{s_0} \) is a \( \sigma \) algebra. The next step is to show \( \mathcal{G}_{s_0} \) contains the sets

\[
(s, t] \times F, F \in F_s
\]

and

\[
\{ 0 \} \times F, F \in F_0.
\]

It is clear \( \{ 0 \} \times F \) is contained in \( \mathcal{G}_{s_0} \) because \( \{ 0 \} \times F \} \in F_{s_0} \). Similarly, if \( s \geq s_0 \) or if \( s, t < s_0 \) then \( ((s, t] \times F) \in F_{s_0} \). The only case left is for \( s < s_0 \) and \( t \geq s_0 \). In this case, letting \( A_s \in F_s \), \((s, t] \times A_s \} = A_s \in F_s \subseteq F_{s_0} \). Therefore, \( \mathcal{G}_{s_0} \) contains all the sets of the form given in (13.4.23) and (13.4.24) and so since \( P_\infty \) is the smallest \( \sigma \) algebra containing these sets, it follows \( P_\infty = \mathcal{G}_{s_0} \). The case where \( s_0 = 0 \) is entirely similar but shorter. Therefore, if \( X \) is predictable, letting \( A \in B(E) \), \( X^{-1}(A) \) is in \( P_\infty \) or \( P_T \) and so

\[
(X^{-1}(A))_s \equiv \{ \omega \in \Omega : X(s, \omega) \in A \} = X(s)^{-1}(A) \in F_s
\]

showing \( X(t) \) is \( F_t \) adapted. This proves the proposition.

Another way to see this is to recall the progressively measurable functions are adapted. Then show the predictable sets are progressively measurable.
Proposition 13.4.13 Let $\mathcal{P}$ denote the predictable $\sigma$ algebra and let $\mathcal{R}$ denote the progressively measurable $\sigma$ algebra. Then $\mathcal{P} \subseteq \mathcal{R}$.

Proof: Let $\mathcal{G}$ denote those sets of $\mathcal{P}$ such that they are also in $\mathcal{R}$. Then $\mathcal{G}$ clearly contains the $\pi$ system of sets $\{0\} \times A, A \in \mathcal{F}_0$, and $(s, t] \times A, A \in \mathcal{F}_s$. Furthermore, $\mathcal{G}$ is closed with respect to countable disjoint unions and complements. It follows $\mathcal{G}$ contains the $\sigma$ algebra generated by this $\pi$ systems which is $\mathcal{P}$. This proves the proposition.

Proposition 13.4.14 Let $X(t)$ be a stochastic process having values in $E$ a complete metric space and let it be $\mathcal{F}_t$ adapted and left continuous. Then it is predictable. Also, if $X(t)$ is stochastically continuous and adapted on $[0, T]$, then it has a predictable version.

Proof: Define $I_{m,k} \equiv (\{k - 1\} 2^{-m}T, k2^{-m}T]$ if $k \geq 1$ and $I_{m,0} = \{0\}$ if $k = 1$. Then define

$$X_m(t) \equiv \sum_{k=1}^{2^m} X(T(k - 1)2^{-m}) \mathcal{A}_{(k - 1)2^{-m}T, k2^{-m}T]}(t) + X(0) \mathcal{A}_{[0,0]}(t)$$

Here the sum means that $X_m(t)$ has value $X(T(k - 1)2^{-m})$ on the interval $((k - 1)2^{-m}T, k2^{-m}T]$. Thus $X_m$ is predictable because each term in the sum is. Thus

$$X_m^{-1}(U) = \bigcup_{k=1}^{2^m} \{X(T(k - 1)2^{-m}) \mathcal{A}_{(k - 1)2^{-m}T, k2^{-m}T]}^{-1}(U) = \bigcup_{k=1}^{2^m} ((k - 1)2^{-m}T, k2^{-m}T] \times \{X(T(k - 1)2^{-m})\}^{-1}(U),$$

a finite union of predictable sets. Since $X$ is left continuous,

$$X(t, \omega) = \lim_{m \to \infty} X_m(t, \omega)$$

and so $X$ is predictable.

Next consider the other claim. Since $X$ is stochastically continuous on $[0, T]$, it is uniformly stochastically continuous on this interval by Lemma 13.2.1. Therefore, there exists a sequence of partitions of $[0, T]$, the $m^{th}$ being

$$0 = t_{m,0} < t_{m,1} < \cdots < t_{m,m(m)} = T$$

such that for $X_m$ defined as above, then for each $t$

$$P\left([d\left(X_m(t), X(t)\right) \geq 2^{-m}]\right) \leq 2^{-m} \quad (13.4.25)$$

Then as above, $X_m$ is predictable. Let $A$ denote those points of $\mathcal{P}_T$ at which $X_m(t, \omega)$ converges. Thus $A$ is a predictable set because it is just the set where $X_m(t, \omega)$ is a Cauchy sequence. Now define the predictable function $Y$

$$Y(t, \omega) = \begin{cases} \lim_{m \to \infty} X_m(t, \omega) & \text{if } (t, \omega) \in A \\ 0 & \text{if } (t, \omega) \notin A \end{cases}$$

From 13.2.1 it follows from the Borel Cantelli lemma that for fixed $t$, the set of $\omega$ which are in infinitely many of the sets,

$$[d\left(X_m(t), X(t)\right) \geq 2^{-m}]$$

has measure zero. Therefore, for each $t$, there exists a set of measure zero, $N(t)$ such that for $\omega \notin N(t)$ and all $m$ large enough

$$d\left(X_m(t, \omega), X(t, \omega)\right) < 2^{-m}$$

Hence for $\omega \notin N(t)$, $(t, \omega) \in A$ and so $X_m(t, \omega) \rightarrow Y(t, \omega)$ which shows

$$d\left(Y(t, \omega), X(t, \omega)\right) = 0 \text{ if } \omega \notin N(t).$$

The predictable version of $X(t)$ is $Y(t)$.
Here is a summary of what has been shown above.

adapted and left continuous
⇒ predictable
⇒ progressively measurable
⇒ adapted

Also
stochastically continuous and adapted ⇒ progressively measurable version

13.5 Martingales

Definition 13.5.1 Let $X$ be a stochastic process defined on an interval, $I$ with values in a separable Banach space, $E$. It is called integrable if $E(||X(t)||) < \infty$ for each $t \in I$. Also let $\mathcal{F}_t$ be a filtration. An integrable and adapted stochastic process $X$ is called a martingale if for $s \leq t$

$$E(X(t)|\mathcal{F}_s) = X(s) \ P \text{ a.e. } \omega.$$ 

Recalling the definition of conditional expectation this says that for $F \in \mathcal{F}_s$

$$\int_F X(t) \, dP = \int_F E(X(t)|\mathcal{F}_s) \, dP = \int_F X(s) \, dP$$

for all $F \in \mathcal{F}_s$. A real valued stochastic process is called a submartingale if whenever $s \leq t$,

$$E(X(t)|\mathcal{F}_s) \geq X(s) \ \text{a.e.}$$

and a supermartingale if

$$E(X(t)|\mathcal{F}_s) \leq X(s) \ \text{a.e.}$$

Example 13.5.2 Let $\mathcal{F}_t$ be a filtration and let $Z$ be in $L^1(\Omega, \mathcal{F}_T, P)$. Then let $X(t) = E(Z|\mathcal{F}_t)$.

This works because for $s < t$, $E(X(t)|\mathcal{F}_s) = E(E(Z|\mathcal{F}_t)|\mathcal{F}_s) = E(Z|\mathcal{F}_s) = X(s)$.

Proposition 13.5.3 The following statements hold for a stochastic process defined on $[0, T] \times \Omega$ having values in a real separable Banach space, $E$.

1. If $X(t)$ is a martingale then $||X(t)|| , t \in [0, T]$ is a submartingale.

2. If $g$ is an increasing convex function from $[0, \infty)$ to $[0, \infty)$ and $E(g(||X(t)||)) < \infty$ for all $t \in [0, T]$ then $g(||X(t)||) , t \in [0, T]$ is a submartingale.

Proof: Let $s \leq t$ by Theorem [13.1.1] on Page 251

$$E(||X(t)|||\mathcal{F}_s) \geq ||E(X(t)|\mathcal{F}_s)|| = ||X(s)||$$

which shows $||X||$ is a submartingale as claimed.

Consider the second claim. Recall Jensen’s inequality for submartingales, Theorem [12.1.6] on Page 225. From the first part

$$||X(s)|| \leq E(||X(t)|||\mathcal{F}_s) \ \text{a.e.}$$

and so from Jensen’s inequality,

$$g(||X(s)||) \leq g(E(||X(t)|||\mathcal{F}_s)) \leq E(g(||X(t)||)|\mathcal{F}_s) \ \text{a.e.},$$

showing that $g(||X(t)||)$ is also a submartingale. ■
13.6 Some Maximal Estimates

Martingales and submartingales have some very interesting maximal estimates. I will present some of these here. The proofs are fairly general and do not require the filtration to be normal.

Lemma 13.6.1 Let \( \{F_t\} \) be a filtration and let \( \{X(t)\} \) be a nonnegative valued submartingale for \( t \in [S,T] \). Then for \( \lambda > 0 \) and any \( p \geq 1 \), if \( A_t \) is a \( F_t \) measurable subset of \( [X(t) \geq \lambda] \), then

\[
P(A_t) \leq \frac{1}{\lambda^p} \int_{A_t} X(T)^p \, dP.
\]

Proof: From Jensen’s inequality,

\[
\lambda^p P(A_t) \leq \int_{A_t} X(t)^p \, dP \leq \int_{A_t} E(X(T)|F_t)^p \, dP
\]

\[
\leq \int_{A_t} E(X(T)^p | F_t) \, dP = \int_{A_t} X(T)^p \, dP
\]

and this proves the lemma.

The following theorem is the main result.

Theorem 13.6.2 Let \( \{F_t\} \) be a filtration and let \( \{X(t)\} \) be a nonnegative valued right continuous \(^2\) submartingale for \( t \in [S,T] \). Then for all \( \lambda > 0 \) and \( p \geq 1 \), for

\[
X^* \equiv \sup_{t \in [S,T]} X(t),
\]

\[
P([X^* \geq \lambda]) \leq \frac{1}{\lambda^p} \int_{\Omega} X_{[X^* \geq \lambda]} X(T)^p \, dP
\]

In the case that \( p > 1 \), it is also true that

\[
E((X^*)^p) \leq \left( \frac{p}{p-1} \right) E((X(T)^p)^{1/p} (E((X^*)^p))^{1/p'}
\]

Also there are no measurability issues related to the above \( \sup_{t \in [S,T]} X(t) \equiv X^* \)

Proof: Let \( S \leq t_0^m < t_1^m < \cdots < t_m^m = T \) where \( t_j^m - t_{j-1}^m = (T - S)2^{-m} \). First consider \( m = 1 \).

\[
A_{t_0^1} = \{ \omega \in \Omega : X(t_0^1) \geq \lambda \}, \quad A_{t_1^1} = \{ \omega \in \Omega : X(t_1^1) \geq \lambda \} \setminus A_{t_0^1}
\]

\[
A_{t_2^1} = \{ \omega \in \Omega : X(t_2^1) \geq \lambda \} \setminus (A_{t_0^1} \cup A_{t_1^1})
\]

Do this type of construction for \( m = 2, 3, 4, \cdots \) yielding disjoint sets, \( \{A_{t_j^m}\}_{j=0}^{2^m} \) whose union equals

\[
\cup_{t \in D_m} [X(t) \geq \lambda]
\]

where \( D_m = \{t_j^m\}_{j=0}^{2^m} \). Thus \( D_m \subseteq D_{m+1} \). Then also, \( D \equiv \cup_{m=1}^{\infty} D_m \) is dense and countable. From Lemma 13.6.1,

\[
P(\cup_{t \in D_m} [X(t) \geq \lambda]) = P(\sup_{t \in D_m} X(t) \geq \lambda) = \sum_{j=0}^{2^m} P(A_{t_j^m})
\]

\[
\leq \frac{1}{\lambda^p} \sum_{j=0}^{2^m} \int_{A_{t_j^m}} X_{[sup_{t \in D_m} X(t) \geq \lambda]} X(T)^p \, dP
\]

\[
\leq \frac{1}{\lambda^p} \int_{\Omega} X_{[sup_{t \in D} X(t) \geq \lambda]} X(T)^p \, dP \]

\( ^1 t \rightarrow M(t)(\omega) \) is continuous from the right for a.e. \( \omega \).
Let \( m \to \infty \) in the above to obtain
\[
P(\bigcup_{t \in D} [X(t) \geq \lambda]) = P\left( \sup_{t \in D} X(t) \geq \lambda \right) \leq \frac{1}{\lambda^p} \int_{\Omega} X[\sup_{t \in D} X(t) \geq \lambda] X(T)^p \, dP. \tag{13.6.26}
\]

From now on, assume that for a.e. \( \omega \in \Omega \), \( t \to X(t)(\omega) \) is right continuous. Then with this assumption, the following claim holds.
\[
\sup_{t \in [S,T]} X(t) \equiv X^* = \sup_{t \in D} X(t)
\]
which verifies that \( X^* \) is measurable. Then from 13.6.27,
\[
P([X^* \geq \lambda]) = P\left( \sup_{t \in D} X(t) \geq \lambda \right) \leq \frac{1}{\lambda^p} \int_{\Omega} X[\sup_{t \in D} X(t) \geq \lambda] X(T)^p \, dP
\]
\[
= \frac{1}{\lambda^p} \int_{\Omega} X[\sup_{t \in D} X(t) \geq \lambda] X(T)^p \, dP.
\]

Now consider the other inequality. Using the distribution function technique and the above estimate obtained in the first part,
\[
E((X^*)^p) = \int_0^\infty p\alpha^{p-1} P([X^* \geq \alpha]) \, d\alpha \leq \int_0^\infty p\alpha^{p-1} P([X^* \geq \alpha]) \, d\alpha
\]
\[
\leq \int_0^\infty p\alpha^{p-1} \frac{1}{\alpha} \int_{\Omega} X[\sup_{t \in D} X(t) \geq \alpha] X(T) \, dP \, d\alpha
\]
\[
= p \int_0^{X^*} \alpha^{p-2} \, d\alpha \int_{\Omega} X(T) \, dP
\]
\[
= \frac{p}{p-1} \int_{\Omega} (X^*)^{p-1} X(T) \, dP
\]
\[
\leq \frac{p}{p-1} \left( \int_{\Omega} (X(T))^p \right)^{1/p'} \left( \int_{\Omega} X(T)^p \right)^{1/p'}
\]
\[
= \frac{p}{p-1} E(X(T)^p)^{1/p} E((X^*)^{p})^{1/p'}. \tag{13.6.27}
\]

Of course it would be nice to divide both sides by \( E((X^*)^p)^{1/p} \) but we don’t know that this is finite. One can use a stopped submartingale which will have \( X(t) \) bounded, divide, and then let the stopping time increase to \( \infty \). However, this is discussed later.

With Theorem 13.6.2 here is an important maximal estimate for martingales having values in \( E \), a real separable Banach space.

**Theorem 13.6.3** Let \( X(t) \) for \( t \in I = [0,T] \) be an \( E \) valued right continuous martingale with respect to a filtration, \( \mathcal{F}_t \). Then for \( p \geq 1 \),
\[
P\left( \left[ \sup_{t \in I} \|X(t)\| \geq \lambda \right] \right) \leq \frac{1}{\lambda^p} E\left( \|X(T)\|^p \right). \tag{13.6.27}
\]

If \( p > 1 \),
\[
E\left( \left[ \sup_{t \in [S,T]} \|X(t)\| \right]^p \right) \leq \left( \frac{p}{p-1} \right) E\left( \|X(T)\|^{p} \right)^{1/p} E\left( \left[ \sup_{t \in [S,T]} \|X(t)\| \right] \right)^{p'} \tag{13.6.28}
\]

**Proof:** By Proposition 13.5.2, \( \|X(t)\|, t \in I \) is a submartingale and so from Theorem 13.6.2, it follows 13.6.27 and 13.6.28 hold. \( \blacksquare \)
13.7 Optional Sampling Theorems

13.7.1 Stopping Times And Their Properties

The optional sampling theorem is very useful in the study of martingales and submartingales as will be shown.

First it is necessary to define the notion of a stopping time.

**Definition 13.7.1** Let \((\Omega, \mathcal{F}, P)\) be a probability space and let \(\{\mathcal{F}_n\}_{n=1}^{\infty}\) be an increasing sequence of \(\sigma\) algebras each contained in \(\mathcal{F}\), called a discrete filtration. A stopping time is a measurable function, \(\tau\) which maps \(\Omega\) to \(\mathbb{N}\), \(\tau^{-1}(A) \in \mathcal{F}\) for all \(A \in \mathcal{P}(\mathbb{N})\),

such that for all \(n \in \mathbb{N}\),

\([\tau \leq n] \in \mathcal{F}_n\).

Note this is equivalent to saying

\([\tau = n] \in \mathcal{F}_n\)

because

\([\tau = n] = [\tau \leq n] \setminus [\tau \leq n - 1]\).

For \(\tau\) a stopping time define \(\mathcal{F}_\tau\) as follows.

\(\mathcal{F}_\tau \equiv \{A \in \mathcal{F}: A \cap [\tau \leq n] \in \mathcal{F}_n\) for all \(n \in \mathbb{N}\}\)

These sets in \(\mathcal{F}_\tau\) are referred to as “prior” to \(\tau\). Note that this condition is equivalent to saying that

\(\mathcal{F}_\tau \equiv \{A \in \mathcal{F}: A \cap [\tau = n] \in \mathcal{F}_n\) for all \(n \in \mathbb{N}\}\)

The most important example of a stopping time is the first hitting time.

**Example 13.7.2** The first hitting time of an adapted process \(X(n)\) of a Borel set \(G\) is a stopping time. This is defined as

\(\tau \equiv \min\{k: X(k) \in G\}\)

To see this, note that

\([\tau = n] = \cap_{k<n}[X(k) \in G^c] \cap [X(n) \in G] \in \mathcal{F}_n\).

**Proposition 13.7.3** For \(\tau\) a stopping time, \(\mathcal{F}_\tau\) is a \(\sigma\) algebra and if \(Y(k)\) is \(\mathcal{F}_k\) measurable for all \(k\), \(Y(k)\) having values in a separable Banach space \(E\), then

\(\omega \rightarrow Y(\tau(\omega))\)

is \(\mathcal{F}_\tau\) measurable.

**Proof:** Let \(A_n \in \mathcal{F}_\tau\). I need to show \(\cup_n A_n \in \mathcal{F}_\tau\). In other words, I need to show that

\(\cup_n A_n \cap [\tau \leq k] \in \mathcal{F}_k\)

The left side equals

\(\cup_n (A_n \cap [\tau \leq k])\)

which is a countable union of sets of \(\mathcal{F}_k\) and so \(\mathcal{F}_\tau\) is closed with respect to countable unions. Next suppose \(A \in \mathcal{F}_\tau\).

\(A^c \cap [\tau \leq k] \cup (A \cap [\tau \leq k]) = \Omega \cap [\tau \leq k]\)

and \(\Omega \cap [\tau \leq k] \in \mathcal{F}_k\). Therefore, so is \(A^c \cap [\tau \leq k]\).

It remains to verify the last claim. Let \(B\) be an open set in \(E\). Is

\([Y(\tau) \in B] \in \mathcal{F}_\tau\) ?

Is

\([Y(\tau) \in B] \cap [\tau \leq k] \in \mathcal{F}_k\) for all \(k\)?

This equals

\(\cup_{i=1}^k [Y(\tau) \in B] \cap [\tau = i] = \cup_{i=1}^k [Y(i) \in B] \cap [\tau = i] \in \mathcal{F}_k\)

Therefore, \(Y(\tau)\) must be \(\mathcal{F}_\tau\) measurable. 

The following lemma contains the fundamental properties of stopping times for discrete filtrations.
Lemma 13.7.4 In the situation of Definition [13.7.1], let $\sigma, \tau$ be two stopping times. Then

1. $\tau$ is $F_\tau$ measurable
2. $F_\sigma \cap [\sigma \leq \tau] \subseteq F_{\sigma \land \tau} = F_\sigma \cap F_\tau$
3. $F_\tau = F_k$ on $[\tau = k]$ for all $k$. That is if $A \in F_k$, then $A \cap [\tau = k] \in F_\tau$ and if $A \in F_\tau$, then $A \cap [\tau = k] \in F_k$. In other words, the two $\sigma$ algebras 
   \[ [\tau = k] \cap F_\tau, [\tau = k] \cap F_k \]
   are equal. Letting $G$ denote this $\sigma$ algebra, if $g$ is either $F_\tau$ or $F_k$ measurable then its restriction to $[\tau = k]$ is $G$ measurable. Also if $A \in F_\tau$, and $Y \in L^1(\Omega; E)$,
   \[ \int_{A \cap [\tau = k]} E(Y|F_\tau) \, dP = \int_{A \cap [\tau = k]} E(Y|F_k) \, dP \]
   and
   \[ E(Y|F_\tau) = E(Y|F_k) \text{ a.e.} \]
   on $[\tau = k]$.

Proof: Consider the first claim. $[\tau \leq \lambda] \cap [\tau \leq m] = [\tau \leq \lambda \wedge m] \in F_{\lambda \wedge m} \subseteq F_m$ and so $\tau$ is $F_\tau$ measurable. Here $[\lambda]$ is the greatest integer less than or equal to $\lambda$. Next note that $\sigma \wedge \tau$ is a stopping time because
   \[ [\sigma \wedge \tau \leq k] = [\sigma \leq k] \cup [\tau \leq k] \in F_k \]
   Next consider the second claim. Let $A \in F_\sigma$. I want to show
   \[ A \cap [\sigma \leq \tau] \in F_{\sigma \land \tau} \]
   (13.7.29)
   In other words, I want to show
   \[ A \cap [\sigma \leq \tau] \cap [\tau \wedge \sigma \leq k] \in F_k \]
   (13.7.30)
   for all $k$. However, the set on the left equals
   \[ A \cap [\sigma \leq \tau] \cap [\sigma \leq k] \]
   \[ = \bigcup_{j=1}^k A \cap [\sigma = j] \cap [\tau \geq j] \cap [\sigma \leq k] \]
   \[ = \bigcup_{j=1}^k A \cap [\sigma = j] \cap [\tau \leq j-1] \cap [\sigma \leq k] \cap [\tau \leq k] \]
   \[ = \bigcup_{j=1}^k A \cap [\sigma \wedge \tau = i] \cap [\tau = k] \cap A \cap [\sigma = i] \cap [\tau = k] \]
   and so this is in $F_k$. Thus $A \cap [\tau \leq k] \subseteq F_k$ being the finite union of sets which are. Similarly $A \cap [\sigma \leq k] \subseteq F_k$ for all $k$ and so $A \subseteq F_\tau \cap F_\sigma$.
   Next let $A \in F_\tau \cap F_\sigma$. Then is it in $F_{\sigma \land \tau}$? Is $A \cap [\sigma \wedge \tau \leq k] \subseteq F_k$? Of course this is so because
   \[ A \cap [\sigma \wedge \tau \leq k] = A \cap ([\sigma \leq k] \cup [\tau \leq k]) \]
   \[ = (A \cap [\sigma \leq k]) \cup (A \cap [\tau \leq k]) \subseteq F_k \]
   since both $\sigma, \tau$ are stopping times. This proves part 2.)
Now consider part 3). Note that \([\tau = k]\) is in both \(\mathcal{F}_k\) and \(\mathcal{F}_\tau\). First consider the claim it is in \(\mathcal{F}_\tau\).

\[ \{\tau = k\} \cap \{\tau \leq l\} = \emptyset \text{ if } l < k \]

which is in \(\mathcal{F}_l\). If \(l \geq k\), it reduces to \([\tau = k] \in \mathcal{F}_k \subseteq \mathcal{F}_l\) so it is in \(\mathcal{F}_\tau\). \([\tau = k]\) is obviously in \(\mathcal{F}_k\).

I need to show

\[ \mathcal{F}_\tau \cap [\tau = k] = \mathcal{F}_k \cap [\tau = k] \]

where \(\mathcal{H} \cap [\tau = k]\) means all sets of the form \(A \cap [\tau = k]\) where \(A \in \mathcal{H}\). Let \(A \in \mathcal{F}_\tau\). Then

\[ A \cap [\tau = k] = (A \cap [\tau \leq k]) \setminus (A \cap [\tau \leq k - 1]) \in \mathcal{F}_k \]

Therefore, there exists \(B \in \mathcal{F}_k\) such that \(B = A \cap [\tau = k]\) and so

\[ B \cap [\tau = k] = A \cap [\tau = k] \]

which shows \(\mathcal{F}_\tau \cap [\tau = k] \subseteq \mathcal{F}_k \cap [\tau = k]\). Now let \(A \in \mathcal{F}_k\) so that

\[ A \cap [\tau = k] \in \mathcal{F}_k \cap [\tau = k] \]

Then

\[ A \cap [\tau = k] \cap [\tau \leq j] \in \mathcal{F}_j \]

because in case \(j < k\), the set on the left is \(\emptyset\) and if \(j \geq k\) it reduces to \(A \cap [\tau = k]\) and both \(A\) and \([\tau = k]\) are in \(\mathcal{F}_k \subseteq \mathcal{F}_j\). Thus \(A \cap [\tau = k] = B \in \mathcal{F}_\tau\) and so

\[ A \cap [\tau = k] = B \cap [\tau = k] \in \mathcal{F}_\tau \cap [\tau = k]. \]

Therefore, the two \(\sigma\) algebras of subsets of \([\tau = k]\),

\[ \mathcal{F}_\tau \cap [\tau = k], \mathcal{F}_k \cap [\tau = k] \]

are equal. Thus for \(A\) in either \(\mathcal{F}_\tau\) or \(\mathcal{F}_k\), \(A \cap [\tau = k]\) is a set of both \(\mathcal{F}_\tau\) and \(\mathcal{F}_k\) because if \(A \in \mathcal{F}_k\), then from the above, there exists \(B \in \mathcal{F}_\tau\) such that

\[ A \cap [\tau = k] = B \cap [\tau = k] \in \mathcal{F}_\tau \]

with similar reasoning holding if \(A \in \mathcal{F}_\tau\). In other words, if \(g\) is \(\mathcal{F}_\tau\) or \(\mathcal{F}_k\) measurable, then the restriction of \(g\) to \([\tau = k]\) is measurable with respect to \(\mathcal{F}_\tau \cap [\tau = k]\) and \(\mathcal{F}_k \cap [\tau = k]\). Let \(Y\) be an arbitrary random variable in \(L^1(\Omega, \mathcal{F})\). It follows, since \(A \cap [\tau = k]\) is in both \(\mathcal{F}_\tau\) and \(\mathcal{F}_k\),

\[ \int_{A \cap [\tau = k]} E(Y|\mathcal{F}_\tau) \, dP = \int_{A \cap [\tau = k]} Y \, dP = \int_{A \cap [\tau = k]} E(Y|\mathcal{F}_k) \, dP \]

Since this holds for an arbitrary set in \(\mathcal{F}_\tau \cap [\tau = k] = \mathcal{F}_k \cap [\tau = k]\), it follows

\[ E(Y|\mathcal{F}_\tau) = E(Y|\mathcal{F}_k) \text{ a.e. on } [\tau = k] \]

The assertion that

\[ E(Y|\mathcal{F}_\tau) = E(Y|\mathcal{F}_k) \text{ a.e.} \]

on \([\tau = k]\) and that a function \(g\) which is \(\mathcal{F}_\tau\) or \(\mathcal{F}_k\) measurable, when restricted to \([\tau = k]\) is \(\mathcal{G}\) measurable for

\[ \mathcal{G} = [\tau = k] \cap \mathcal{F}_\tau = [\tau = k] \cap \mathcal{F}_k \]

is the main result in the above lemma and this fact leads to the amazing Doob optional sampling theorem below. Also note that if \(Y(k)\) is any process defined on the positive integers \(k\), then by definition, \(Y(k)(\omega) = Y(\tau(\omega))(\omega)\) on the set \([\tau = k]\) because \(\tau\) is constant on this set.
13.7.2 Doob Optional Sampling Theorem

With this lemma, here is a major theorem, the optional sampling theorem of Doob. This one is for martingales having values in a Banach space. To begin with, consider the case of a martingale defined on a countable set.

Theorem 13.7.5 Let \( \{M(k)\} \) be a martingale having values in \( E \) a separable real Banach space with respect to the increasing sequence of \( \sigma \) algebras, \( \{\mathcal{F}_k\} \) and let \( \sigma, \tau \) be two stopping times such that \( \tau \) is bounded. Then \( M(\tau) \) defined as

\[
\omega \to M(\tau(\omega))
\]

is integrable and

\[
M(\sigma \wedge \tau) = E(M(\tau) | \mathcal{F}_\sigma).
\]

Proof: By Proposition 13.7.2 \( M(\tau) \) is \( \mathcal{F}_\tau \) measurable.

Next note that since \( \tau \) is bounded by some \( l \),

\[
\int_\Omega ||M(\tau(\omega))|| dP \leq \sum_{i=1}^l \int_{[\tau=i]} ||M(i)|| dP < \infty.
\]

This proves the first assertion and makes possible the consideration of conditional expectation.

Let \( l \geq \tau \) as described above. Then for \( k \leq l \), by Lemma 13.7.4,

\[
\mathcal{F}_k \cap [\tau = k] = \mathcal{F}_\tau \cap [\tau = k] = \mathcal{G}
\]

implying that if \( g \) is either \( \mathcal{F}_k \) measurable or \( \mathcal{F}_\tau \) measurable, then its restriction to \( [\tau = k] \) is \( \mathcal{G} \) measurable and so if \( A \in \mathcal{F}_k \cap [\tau = k] = \mathcal{F}_\tau \cap [\tau = k] \),

\[
\int_A E(M(l) | \mathcal{F}_\tau) dP = \int_A M(l) dP
\]

\[
= \int_A E(M(l) | \mathcal{F}_k) dP
\]

\[
= \int_A M(k) dP
\]

\[
= \int_A M(\tau) dP \text{ (on } A, \tau = k\text{)}
\]

Therefore, since \( A \) was arbitrary,

\[
E(M(l) | \mathcal{F}_\tau) = M(\tau) \text{ a.e.}
\]

on \( [\tau = k] \) for every \( k \leq l \). It follows

\[
E(M(l) | \mathcal{F}_\tau) = M(\tau) \text{ a.e.}
\]

(13.7.31)

since it is true on each \( [\tau = k] \) for all \( k \leq l \).

Now consider \( E(M(\tau) | \mathcal{F}_\sigma) \) on the set \( [\sigma = i] \cap [\tau = j] \). By Lemma 13.7.4 on this set,

\[
E(M(\tau) | \mathcal{F}_\sigma) = E(M(\tau) | \mathcal{F}_i) = E(E(M(l) | \mathcal{F}_\tau) | \mathcal{F}_i) = E(E(M(l) | \mathcal{F}_j) | \mathcal{F}_i)
\]

If \( j \leq i \), this reduces to

\[
E(M(l) | \mathcal{F}_j) = M(j) = M(\sigma \wedge \tau).
\]

If \( j > i \), this reduces to

\[
E(M(l) | \mathcal{F}_i) = M(i) = M(\sigma \wedge \tau)
\]

and since this exhausts all possibilities for values of \( \sigma \) and \( \tau \), it follows

\[
E(M(\tau) | \mathcal{F}_\sigma) = M(\sigma \wedge \tau) \text{ a.e.} \]

You can also give a version of the above to submartingales. This requires the following very interesting decomposition of a submartingale into the sum of an increasing stochastic process and a martingale.

Theorem 13.7.6 Let \( \{X_n\} \) be a submartingale. Then there exists a unique stochastic process, \( \{A_n\} \) and martingale, \( \{M_n\} \) such that

\[
X_n = A_n + M_n
\]
1. \( A_n(\omega) \leq A_{n+1}(\omega), \quad A_1(\omega) = 0, \)
2. \( A_n \) is \( \mathcal{F}_{n-1} \) adapted for all \( n \geq 1 \) where \( \mathcal{F}_0 \equiv \mathcal{F}_1. \)

and also \( X_n = M_n + A_n. \)

**Proof:** Let \( A_1 = 0 \) and define
\[
A_n = \sum_{k=2}^{n} E(X_k - X_{k-1}|\mathcal{F}_{k-1}).
\]

It follows \( A_n \) is \( \mathcal{F}_{n-1} \) measurable. Since \( \{X_k\} \) is a submartingale, \( A_n \) is increasing because
\[
A_{n+1} - A_n = E(X_{n+1} - X_n|\mathcal{F}_n) \geq 0 \quad (13.7.32)
\]

It is a submartingale because
\[
E(A_n|\mathcal{F}_{n-1}) = E\left(\sum_{k=2}^{n} E(X_k - X_{k-1}|\mathcal{F}_{k-1})|\mathcal{F}_{n-1}\right)
= \sum_{k=2}^{n} E(X_k - X_{k-1}|\mathcal{F}_{k-1}) \equiv A_n \geq A_{n-1}
\]

Now let \( M_n \) be defined by
\[
X_n = M_n + A_n.
\]

Then from \(13.7.32\)
\[
E(M_{n+1}|\mathcal{F}_n) = E(X_{n+1}|\mathcal{F}_n) - E(A_{n+1}|\mathcal{F}_n)
= E(X_{n+1}|\mathcal{F}_n) - E(A_{n+1} - A_n|\mathcal{F}_n) - A_n
= E(X_{n+1}|\mathcal{F}_n) - E(X_{n+1} - X_n|\mathcal{F}_n) - A_n - A_n
= E(X_n|\mathcal{F}_n) - A_n
= X_n - A_n \equiv M_n
\]

This proves the existence part.

It remains to verify uniqueness. Suppose then that
\[
X_n = M_n + A_n = M'_n + A'_n
\]
where \( \{A_n\} \) and \( \{A'_n\} \) both satisfy the conditions of the theorem and \( \{M_n\} \) and \( \{M'_n\} \) are both martingales. Then
\[
M_n - M'_n = A'_n - A_n
\]
and so, since \( A'_n - A_n \) is \( \mathcal{F}_{n-1} \) measurable and \( \{M_n - M'_n\} \) is a martingale,
\[
M_{n-1} - M'_{n-1} = E(M_n - M'_n|\mathcal{F}_{n-1})
= E(A'_n - A_n|\mathcal{F}_{n-1})
= A'_n - A_n = M_n - M'_n.
\]

Continuing this way shows \( M_n - M'_n \) is a constant. However, since \( A'_1 - A_1 = 0 = M_1 - M'_1 \), it follows \( M_n = M'_n \)
and this proves uniqueness. \( \blacksquare \)

Now here is a version of the optional sampling theorem for submartingales.

**Theorem 13.7.7** Let \( \{X(k)\} \) be a real valued submartingale with respect to the increasing sequence of \( \sigma \) algebras, \( \{\mathcal{F}_k\} \) and let \( \sigma \leq \tau \) be two stopping times such that \( \tau \) is bounded. Then \( M(\tau) \) defined as
\[
\omega \rightarrow X(\tau(\omega))
\]
is integrable and
\[
X(\sigma) \leq E(X(\tau)|\mathcal{F}_\sigma).
\]

Without assuming \( \sigma \leq \tau \), one can write
\[
X(\sigma \wedge \tau) \leq E(X(\tau)|\mathcal{F}_\sigma)
\]
Theorem 13.8.2 above to write
\[ X(n) = M(n) + A(n) \]
where \( M \) is a martingale and \( A \) is increasing with \( A(n) \) being \( F_{n-1} \) measurable and \( A(0) = 0 \) as discussed in Theorem 13.7.5. Then
\[ E(X(\tau)|F_\sigma) = E(M(\tau) + A(\tau)|F_\sigma) \]
Now since \( A \) is increasing, you can use the optional sampling theorem for martingales, Theorem 13.7.6 to conclude that, since \( F_{\sigma\wedge\tau} \subseteq F_\sigma \) and \( A(\sigma \wedge \tau) \) is \( F_{\sigma\wedge\tau} \) measurable,
\[ \geq E(M(\tau) + A(\sigma \wedge \tau)|F_\sigma) = E(M(\tau)|F_\sigma) + A(\sigma \wedge \tau) = M(\sigma \wedge \tau) + A(\sigma \wedge \tau) = X(\sigma \wedge \tau). \]

In summary, the main results for stopping times for discrete filtrations are the following definitions and theorems.

\[ |\tau \leq m| \in F_m \]
\[ A \in F_\tau \text{ means } A \cap [\tau \leq m] \in F_m \text{ for any } m \]
\[ X \text{ adapted implies } X(\tau) \text{ is } F_\tau \text{ measurable} \]
\[ F_{\sigma\wedge\tau} = F_\sigma \cap F_\tau \]
\[ [\tau = k] \cap F_k = [\tau = k] \cap F_\tau \]
This last theorem implies the following amazing result. From these fundamental properties, we obtain the optional sampling theorem for martingales and submartingales.
\[ E(Y|F_\tau) = E(Y|F_k) \text{ a.e. on } [\tau = k] \]

13.8 Doob Optional Sampling Continuous Case

13.8.1 Stopping Times

With continuous processes, the discrete filtration is replaced by a normal filtration. Also we tend to feature right continuous or continuous processes. As in the case of discrete martingales, there is something called a stopping time.

**Definition 13.8.1** Let \((\Omega, F, P)\) be a probability space and let \( F_t \) be a filtration. A measurable function, \( \tau : \Omega \to [0, \infty] \) is called a stopping time if
\[ |\tau \leq t| \in F_t \]
for all \( t \geq 0 \). Associated with a stopping time is the \( \sigma \) algebra, \( F_\tau \) defined by
\[ F_\tau \equiv \{ A \in F : A \cap [\tau \leq t] \in F_t \text{ for all } t \}. \]
These sets are also called those “prior” to \( \tau \).

Note that \( F_\tau \) is obviously closed with respect to countable unions. If \( A \in F_\tau \), then
\[ A^C \cap [\tau \leq t] = [\tau \leq t] \setminus (A \cap [\tau \leq t]) \in F_t \]
Thus \( F_\tau \) is a \( \sigma \) algebra.

**Proposition 13.8.2** Let \( B \) be an open subset of topological space \( E \) and let \( X(t) \) be a right continuous \( F_t \) adapted stochastic process such that \( F_t \) is normal. Then define
\[ \tau(\omega) \equiv \inf \{ t > 0 : X(t)(\omega) \in B \}. \]
This is called the first hitting time. Then \( \tau \) is a stopping time. If \( X(t) \) is continuous and adapted to \( F_t \), a normal filtration, then if \( H \) is a nonempty closed set such that \( H = \cap_{n=1}^\infty B_n \) for \( B_n \) open, \( B_n \supseteq B_{n+1} \),
\[ \tau(\omega) \equiv \inf \{ t > 0 : X(t)(\omega) \in H \} \]
is also a stopping time.
13.8. DOOB OPTIONAL SAMPLING CONTINUOUS CASE

**Proof:** Consider the first claim. \( \omega \in [\tau = a] \) implies that for each \( n \in \mathbb{N} \), there exists \( t \in [a, a + \frac{1}{n}] \) such that \( X(t) \in B \). Also for \( t < a \), you would need \( X(t) \notin B \). By right continuity, this is the same as saying that \( X(d) \notin B \) for all rational \( d < a \). (If \( t < a \), you could let \( d_n \downarrow t \) where \( X(d_n) \in B^c \), a closed set. Then it follows that \( X(t) \) is also in the closed set \( B^c \).) Thus, aside from a set of measure zero, for each \( m \in \mathbb{N} \),

\[
[\tau = a] = \left( \bigcap_{n=m}^{\infty} \cup_{t \in [a, a + \frac{1}{n}]} [X(t) \in B] \right) \cap \left( \bigcap_{t \in [0, a]} [X(t) \in B^c] \right)
\]

Since \( X(t) \) is right continuous, this is the same as

\[
\left( \bigcap_{n=m}^{\infty} \cup_{d \in Q \cap [a, a + \frac{1}{n}]} [X(d) \in B] \right) \cap \left( \bigcap_{d \in Q \cap [0, a]} [X(d) \in B^c] \right) \in \mathcal{F}_{a + \frac{1}{m}}
\]

Thus, since the filtration is normal,

\[
[\tau = a] \in \bigcap_{m=1}^{\infty} \mathcal{F}_{a + \frac{1}{m}} = \mathcal{F}_{a+} = \mathcal{F}_a
\]

Now what of \( [\tau < a] \)? This is equivalent to saying that \( X(t) \in B \) for some \( t < a \). Since \( X \) is right continuous, this is the same as saying that \( X(t) \in B \) for some \( t \in \mathbb{Q}, t < a \). Thus

\[
[\tau < a] = \cup_{d \in Q, d < a} [X(d) \in B] \in \mathcal{F}_a
\]

It follows that \( [\tau \leq a] = [\tau < a] \cup [\tau = a] \in \mathcal{F}_a \).

Now consider the claim involving the additional assumption that \( X(t) \) is continuous and it is desired to hit a closed set \( H = \bigcap_{n=1}^{\infty} B_n \) where \( B_n \) is open, \( B_n \supseteq B_{n+1} \). (Note that if the topological space is a metric space, this is always possible so this is not a big restriction.) Then let \( \tau_n \) be the first hitting time of \( B_n \) by \( X(t) \). Then it can be shown that

\[
[\tau \leq a] = \cap_n [\tau_n \leq a] \in \mathcal{F}_a
\]

To show this, first note that \( \omega \in [\tau \leq a] \) if and only if there exists \( t \leq a \) such that \( X(t) (\omega) \in H \). This follows from continuity and the fact that \( H \) is closed. Thus \( \tau_n \leq a \) for all \( n \) because for some \( t \leq a \), \( X(t) \in H \subseteq B_n \) for all \( n \). Next suppose \( \omega \in [\tau_n \leq a] \) for all \( n \). Then for \( \delta_n \downarrow 0 \), there exists \( t_n \in [0, a + \delta_n] \) such that \( X(t_n) (\omega) \in B_n \). It follows there is a subsequence, still denoted by \( t_n \) such that \( t_n \rightarrow t \in [0, a] \). By continuity of \( X \), it must be the case that \( X(t) (\omega) \in H \) and so \( \omega \in [\tau \leq a] \). This shows the above formula. Now by the first part, each \( [\tau_n \leq a] \in \mathcal{F}_a \) and so \( [\tau \leq a] \in \mathcal{F}_a \) also.

Another useful result for real valued stochastic process is the following.

**Proposition 13.8.3** Let \( X(t) \) be a real valued stochastic process which is \( \mathcal{F}_t \) adapted for a normal filtration \( \mathcal{F}_t \), with the property that off a set of measure zero \( t \rightarrow X(t) \) is lower semicontinuous. Then

\[
\tau = \inf \{ t : X(t) > a \}
\]

is a stopping time.

**Proof:** As above

\[
[\tau = a] = \left( \bigcap_{n=m}^{\infty} \cup_{t \in [a, a + \frac{1}{n}]} [X(t) > a] \right) \cap \left( \bigcap_{t \in [0, a]} [X(t) \leq a] \right)
\]

Now

\[
\cap_{t \in [0, a]} [X(t) \leq a] \subseteq \cap_{t \in [0, a], t \in Q} [X(t) \leq a]
\]

If \( \omega \) is in the right side, then for arbitrary \( t < a \), let \( t_n \downarrow t \) where \( t_n \in \mathbb{Q} \) and \( t < a \). Then \( X(t) \leq \lim \inf_{n \rightarrow \infty} X(t_n) \leq a \) and so \( \omega \) is in the left side also. Thus

\[
\bigcap_{t \in [0, a]} [X(t) \leq a] = \bigcap_{t \in (0, a), t \in \mathbb{Q}} [X(t) \leq a]
\]

\[
\bigcup_{t \in [a, a + \frac{1}{n}], t \in \mathbb{Q}} [X(t) > a] \supseteq \bigcup_{t \in [a, a + \frac{1}{n}], t \in \mathbb{Q}} [X(t) > a]
\]

If \( \omega \) is in the left side, then for some \( t \) in the given interval, \( X(t) > a \). If for all \( s \in [a, a + \frac{1}{n}] \cap \mathbb{Q} \) you have \( X(s) \leq a \), then you could take \( s_n \rightarrow t \) where \( X(s_n) \leq a \) and conclude that \( X(t) \leq a \) also by lower semicontinuity. Thus there is some rational \( s \) where \( X(s) > a \) and so the two sides are equal. Hence,

\[
[\tau = a] = \left( \bigcap_{n=m}^{\infty} \cup_{t \in [a, a + \frac{1}{n}], t \in \mathbb{Q}} [X(t) > a] \right) \cap \left( \bigcap_{t \in [0, a], t \in \mathbb{Q}} [X(t) \leq a] \right)
\]
The first set on the right is in $F_{a+(1/m)}$ and so is the next set on the right. Hence $[\tau = a] \in \cap_m F_{a+(1/m)} = F_a$. What of $[\tau < a]$? This equals $\cup_{t \in [0,a]} [X(t) > a] = \cup_{t \in [0,a] \cap Q} [X(t) > a] \in F_a$, the equality following from lower semicontinuity. Thus $[\tau = a] = [\tau = a] \cup [\tau < a] \in F_a$. ■

Thus there do exist stopping times, the first hitting time above being an example. When dealing with continuous stopping times on a normal filtration, one uses the following discrete stopping times

$$\tau_n = \sum_{k=1}^{\infty} \mathcal{N}_{[\tau \in (t^n_k, t^n_{k+1}])} t^n_{k+1}$$

where here $|t^n_k - t^n_{k+1}| = r_n$ for all $k$ where $r_n \to 0$. Then here is an important lemma.

**Lemma 13.8.4** $\tau_n$ is a stopping time ($[\tau_n \leq t] \in F_t$.) Also $F_\tau \subseteq F_{\tau_n}$ and for each $\omega$, $\tau_n (\omega) \downarrow \tau (\omega)$.

**Proof:** Say $t \in (t^n_{k-1}, t^n_k)$. Then $[\tau_n \leq t] = [\tau \leq t^n_{k-1}]$ if $t < t^n_k$ and it equals $[\tau \leq t^n_k]$ if $t = t^n_k$. Either way $[\tau_n \leq t] \in F_t$ so it is a stopping time. Also from the definition, it follows that $\tau_n \geq \tau$ and $[\tau_n (\omega) - \tau (\omega)] \leq r_n$ which is given to converge to 0. Now suppose $A \in F_\tau$ and say $t \in (t^n_{k-1}, t^n_k)$ as above. Then

$$A \cap [\tau_n \leq t] = A \cap [\tau \leq t^n_{k-1}] \in F_{t^n_{k-1}} \subseteq F_t$$

and

$$A \cap [\tau_n \leq t] = A \cap [\tau \leq t^n_k] \in F_{t^n_k} = F_t$$

Thus $F_\tau \subseteq F_{\tau_n}$ as claimed. ■

Next is the claim that if $X(t)$ is adapted to $F_t$, then $X(\tau)$ is adapted to $F_\tau$ just like the discrete case.

**Proposition 13.8.5** Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\sigma \leq \tau$ be two stopping times with respect to a filtration, $F_t$. Then $F_\sigma \subseteq F_\tau$. If $X(t)$ is a right continuous stochastic process adapted to a normal filtration $F_t$ and $\tau$ is a stopping time, $\omega \rightarrow X(\tau (\omega))$ is $F_\tau$ measurable.

**Proof:** Let $A \in F_\sigma$. Then $A \cap [\sigma \leq t] \in F_t$ for all $t \geq 0$. Since $\sigma \leq \tau$,

$$A \cap [\tau \leq t] = A \cap [\sigma \leq t] \cap [\sigma \leq t] \in F_t$$

Thus $A \in F_\tau$ and so $F_\sigma \subseteq F_\tau$.

Consider the following approximation of $\tau$ in which $t^n_k = k2^{-n}$.

$$\tau_n = \sum_{k=1}^{\infty} \mathcal{N}_{[\tau \in (t^n_k, t^n_{k+1})]} t^n_{k+1}$$

Thus $\tau_n \downarrow \tau$. Consider for $U$ an open set, $X(\tau_n)^{-1} (U) \cap [\tau_n < t]$. Say $t \in (t^n_k, t^n_{k+1})$. Then from the above definition of $\tau_n$,

$$[\tau_n < t] = [\tau \leq t^n_k] \in F_{t^n_k} \subseteq F_t$$

It follows that

$$X(\tau_n)^{-1} (U) \cap [\tau_n < t] = \bigcup_{j=1}^{k} X(\tau_j)^{-1} (U) \cap [\tau_n = t^n_j]$$

and so this set is in $F_{t^n_j} \subseteq F_t$. The reason $[\tau_n = t^n_j] \in F_{t^n_j}$ is that it equals $[\tau \in (t^n_{j-1}, t^n_j)] \in F_{t^n_j}$ by assumption that $\tau$ is a stopping time.

By right continuity of $X$, it follows that

$$X(\tau)^{-1} (U) \cap [\tau < t] = \cup_{m=1}^{\infty} \cap_{n \geq m} X(\tau_n)^{-1} (U) \cap [\tau_n < t] \in F_t$$

It follows that for every $m$,

$$X(\tau)^{-1} (U) \cap [\tau \leq t] = \cap_{n=m}^{\infty} X(\tau)^{-1} (U) \cap \left[ \tau < t + \frac{1}{m} \right] \in F_{t + \frac{1}{m}}$$

Since the filtration is normal, it follows that

$$X(\tau)^{-1} (U) \cap [\tau \leq t] \in F_{\tau+} = F_t.$$ ■

Now consider an increasing family of stopping times, $\tau (t)$ ($\omega \rightarrow \tau (t) (\omega)$). It turns out this is a submartingale.
Example 13.8.6 Let \( \{ \tau(t) \} \) be an increasing family of stopping times. Then \( \tau(t) \) is adapted to the \( \sigma \) algebras \( \mathcal{F}_{\tau(t)} \) and \( \{ \tau(t) \} \) is a submartingale adapted to these \( \sigma \) algebras.

First I need to show that a stopping time, \( \tau \) is \( \mathcal{F}_\tau \) measurable. Consider \( [\tau \leq s] \). Is this in \( \mathcal{F}_\tau \)? Is \( [\tau \leq s] \cap [\tau \leq r] \in \mathcal{F}_\tau \) for each \( r \)? This is obviously so if \( s \leq r \) because the intersection reduces to \( [\tau \leq s] \in \mathcal{F}_s \subseteq \mathcal{F}_r \). On the other hand, if \( s > r \) then the intersection reduces to \( [\tau \leq r] \in \mathcal{F}_r \) and so it is clear that \( \tau \) is \( \mathcal{F}_\tau \) measurable. It remains to verify it is a submartingale.

Let \( s < t \) and let \( A \in \mathcal{F}_{\tau(s)} \)

\[
\int_A E(\tau(t) | \mathcal{F}_{\tau(s)}) \, dP \equiv \int_A \tau(t) \, dP \geq \int_A \tau(s) \, dP
\]

and this shows \( E(\tau(t) | \mathcal{F}_{\tau(s)}) \geq \tau(s) \).

Now here is an important example. First note that for \( \tau \) a stopping time, \( \sigma(\tau) \) is a submartingale adapted to the filtration \( \mathcal{F}_\tau \). Therefore, \( \tau \) is \( \mathcal{F}_\tau \) measurable.

Here is a lemma which is the main idea for the proofs of the optional sampling theorem for the continuous case.

**Lemma 13.8.9** Let \( \tau \) be a stopping time and let \( X \) be continuous and adapted to the filtration \( \mathcal{F}_\tau \). Then for \( a > 0 \), define \( \sigma \) as

\[
\sigma(\omega) = \inf \{ t > \tau(\omega) : ||X(t)(\omega) - X(\tau(\omega))|| = a \}
\]

Then \( \sigma \) is also a stopping time.

To see this is so, let

\[
Y(t)(\omega) = ||X(t \vee \tau)(\omega) - X(\tau(\omega))||
\]

Then \( Y(t) \) is \( \mathcal{F}_{t \vee \tau} \) measurable. It is desired to show that \( Y \) is \( \mathcal{F}_\tau \) adapted. Hence if \( U \) is open in \( \mathbb{R} \), then

\[
Y(t)^{-1}(U) = \left( Y(t)^{-1}(U) \cap [\tau \leq t] \right) \cup \left( Y(t)^{-1}(U) \cap [\tau > t] \right)
\]

The second set in the above union on the right equals either \( \emptyset \) or \( [\tau > t] \) depending on whether \( 0 \in U \). If \( \tau > t \), then \( Y(t) = 0 \) and so the second set equals \( [\tau > t] \) if \( 0 \in U \). If \( 0 \notin U \), then the second set equals \( \emptyset \). Thus the second set above is in \( \mathcal{F}_\tau \). It is necessary to show the first set is also in \( \mathcal{F}_\tau \). The first set equals

\[
Y(t)^{-1}(U) \cap [\tau \leq t] = Y(t)^{-1}(U) \cap [\tau \vee t \leq t]
\]

because \( [\tau \vee t \leq t] = [\tau \leq t] \). However, \( Y(t)^{-1}(U) \in \mathcal{F}_{t \vee \tau} \) and so the set on the right in the above is in \( \mathcal{F}_\tau \). Therefore, \( Y(t) \) is adapted. Then \( \alpha \) is just the first hitting time for \( Y(t) \) to equal the closed set \( a \). Therefore, \( \sigma \) is a stopping time by Proposition 13.8.3A.

### 13.8.2 The Optional Sampling Theorem Continuous Case

Next I want a version of the Doob optional sampling theorem which applies to martingales defined on \([0, L], L \leq \infty \). First recall Theorem 13.8.3A part of which is stated as the following lemma.

**Lemma 13.8.8** Let \( f \in L^1(\Omega; E, \mathcal{F}) \) where \( E \) is a separable Banach space. Then if \( \mathcal{G} \) is a \( \sigma \) algebra \( \mathcal{G} \subseteq \mathcal{F} \),

\[
||E(f|\mathcal{G})|| \leq E(||f|| |\mathcal{G})
\]

Here is a lemma which is the main idea for the proofs of the optional sampling theorem for the continuous case.

**Lemma 13.8.9** Let \( X(t) \) be a right continuous nonnegative submartingale such that the filtration \( \{ \mathcal{F}_t \} \) is normal. Recall this includes

\[
\mathcal{F}_t = \bigcap_{s > t} \mathcal{F}_s.
\]

Also let \( \tau \) be a stopping time with values in \([0, T] \). Let \( \mathcal{P}_n = \{ t^n_k \}_{k=1}^{m+n+1} \) be a sequence of partitions of \([0, T] \) which have the property that

\[
\mathcal{P}_n \subseteq \mathcal{P}_{n+1}, \quad \lim_{n \to \infty} ||\mathcal{P}_n|| = 0,
\]

where

\[
||\mathcal{P}_n|| \equiv \sup \{ ||t^n_k - t^n_{k+1}|| : k = 1, 2, \ldots, m_n \}
\]
Then let

\[ \tau_n(\omega) \equiv \sum_{k=0}^{m_n} t_{k+1}^{n} \mathcal{X}_{\tau-1((t_{k}^{n}, t_{k+1}^{n}])}(\omega) \]

It follows that \( \tau_n \) is a stopping time and also the functions \( |X(\tau_n)| \) are uniformly integrable. Furthermore, \( |X(\tau)| \) is integrable.

**Proof:** First of all, say \( t \in (t_{k}^{n}, t_{k+1}^{n}] \). If \( t < t_{k+1}^{n} \), then

\[ [\tau_n \leq t] = [\tau \leq t_{k}^{n}] \in \mathcal{F}_{t_{k}^{n}} \subseteq \mathcal{F}_t \]

and if \( t = t_{k+1}^{n} \), then

\[ [\tau_n \leq t_{k+1}^{n}] = [\tau \leq t_{k+1}^{n}] \in \mathcal{F}_{t_{k+1}^{n}} = \mathcal{F}_t \]

and so \( \tau_n \) is a stopping time. It follows from Proposition 13.8.1 that \( X(\tau_n) \) is \( \mathcal{F}_{\tau_n} \) measurable.

Now from Lemma 13.8.9 or Theorem 13.7.7, \( X(0), X(\tau_n), X(T) \) is a submartingale. Then

\[ \int_{[X(\tau_n) \geq \lambda]} X(\tau_n) \, dP \leq \int_{[X(\tau_n) \geq \lambda]} E(X(T) | \mathcal{F}_{\tau_n}) \, dP = \int_{\Omega} E(X|X(\tau_n) \geq \lambda)X(T) | \mathcal{F}_{\tau_n}) \, dP = \int_{[X(\tau_n) \geq \lambda]} X(T) \, dP \]

From maximal estimates, for example Theorem 12.2.8,

\[ P(|X(\tau_n) \geq \lambda|) \leq \frac{1}{\lambda} \int_{\Omega} X(T) \, dP = \frac{1}{\lambda} \int_{\Omega} X(T) \, dP \]

and now it follows from the above that the random variables \( X(\tau_n) \) are equiintegrable. Recall this means that

\[ \lim_{\lambda \to \infty} \sup_n \int_{[X(\tau_n) \geq \lambda]} X(\tau_n) \, dP = 0 \]

Hence they are uniformly integrable.

To verify that \( |X(\tau)| \) is integrable, note that by right continuity, \( X(\tau_n) \to X(\tau) \) pointwise. Apply the Vitali convergence theorem to obtain

\[ \int_{\Omega} |X(\tau)| \, dP = \lim_{n \to \infty} \int_{\Omega} |X(\tau_n)| \, dP \leq \int_{\Omega} X(T) \, dP < \infty. \]

In fact, you do not need to assume \( X \) is nonnegative.

**Lemma 13.8.10** Let \( X(t) \) be a right continuous submartingale such that the filtration \( \{ \mathcal{F}_t \} \) is normal. Recall this includes

\[ \mathcal{F}_t = \cap_{s > t} \mathcal{F}_s. \]

Also let \( \tau \) be a stopping time with values in \([0, T]\). Let \( \mathcal{P}_n = \{ t_{k}^{n} \}_{k=1}^{m_{n}+1} \) be a sequence of partitions of \([0, T]\) which have the property that

\[ \mathcal{P}_n \subseteq \mathcal{P}_{n+1}, \quad \lim_{n \to \infty} ||\mathcal{P}_n|| = 0, \]

where

\[ ||\mathcal{P}_n|| \equiv \sup \{|t_{k}^{n} - t_{k+1}^{n}| : k = 1, 2, \ldots, m_{n}\} \]

Then let

\[ \tau_n(\omega) \equiv \sum_{k=0}^{m_n} t_{k+1}^{n} \mathcal{X}_{\tau-1((t_{k}^{n}, t_{k+1}^{n}])}(\omega) \]

It follows that \( \tau_n \) is a stopping time and also the functions \( |X(\tau_n)| \) are uniformly integrable. Furthermore, \( |X(\tau)| \) is integrable.
13.8. DOOB OPTIONAL SAMPLING CONTINUOUS CASE

**Proof:** It was shown above that \( \tau_n \) is a stopping time. Also, \( t^n_k \to X (t^n_k) \) is a discrete submartingale. Then by Theorem [3.7.4] there is a martingale \( t^n_k \to M (t^n_k) \) and an increasing submartingale \( t^n_k \to A (t^n_k) \) such that \( A \geq 0 \) and is increasing

\[
X (t^n_k) = M (t^n_k) + A (t^n_k)
\]

You define \( A (t^n_0) \equiv 0 \) and for \( n \geq 1 \),

\[
A (t^n_n) = \sum_{k=1}^n E \left( X (t^n_k) - X (t^n_{k-1}) \mid \mathcal{F}_{t^n_{k-1}} \right)
\]

and repeat the arguments in that theorem. You know that \( A (0), A (\tau_n), A (T) \) is a submartingale by the optional sampling theorem for discrete submartingales given earlier, Theorem [3.7.4], and so

\[
P (A (\tau_n) > \lambda) \leq \frac{1}{\lambda} \int_{[A (\tau_n) > \lambda]} A (\tau_n) \, dP \leq \frac{1}{\lambda} \int_{[A (\tau_n) > \lambda]} A (T) \, dP \leq \frac{\|A (T)\|_{L^1}}{\lambda}
\]

It also follows from the definition of \( A \) that

\[
\|A (T)\|_{L^1} = \int \Omega X (T) - X (0) \, dP < \infty
\]

Hence

\[
\lim_{\lambda \to \infty} \int \Omega [A (\tau_n) > \lambda] A (\tau_n) \, dP \leq \lim_{\lambda \to \infty} \int \Omega [A (\tau_n) > \lambda] A (T) \, dP = 0
\]

Because \( P (A (\tau_n) > \lambda) \to 0 \) and a single function in \( L^1 \) is uniformly integrable. Thus these functions \( A (\tau_n) \) are equi-integrable. Hence they are uniformly integrable. Now \( t^n_k \to |M (t^n_k)| \) is also a nonnegative submartingale. Thus

\[
|M (0)|, |M (\tau_n)|, |M (T)|
\]

is a submartingale by the optional sampling theorem for discrete submartingales given earlier. Therefore,

\[
P (|M (\tau_n)| > \lambda) \leq \frac{1}{\lambda} \int |M (\tau_n)| \, dP \leq \frac{1}{\lambda} \int |M (T)| \, dP \leq \frac{\|M (T)\|_{L^1}}{\lambda}
\]

Of course \( \|M (T)\|_{L^1} \) is finite because it is dominated by

\[
\int \Omega A (T) + |X (T)| \, dP < \infty
\]

Hence

\[
\lim_{\lambda \to \infty} \sup_n \int |M (\tau_n)| \, dP \leq \lim_{\lambda \to \infty} \sup_n \int |M (T)| \, dP = 0
\]

because a single function in \( L^1 \) is uniformly integrable and the above estimate shows that \( P (|M (\tau_n)| > \lambda) \to 0 \) uniformly in \( n \). Thus, in fact \( X (\tau_n) \) must be uniformly integrable since it is the sum of two which are. \( \blacksquare \)

**Theorem 13.8.11** Let \( \{M (t)\} \) be a right continuous martingale having values in \( E \) a separable real Banach space with respect to the increasing sequence of \( \sigma \) algebras, \( \{\mathcal{F}_t\} \) which is assumed to be a normal filtration satisfying,

\[
\mathcal{F}_t = \cap_{s \geq t} \mathcal{F}_s,
\]

for \( t \in [0, L], L \leq \infty \) and let \( \sigma, \tau \) be two stopping times with \( \tau \) bounded. Then \( M (\tau) \) defined as

\[
\omega \to M (\tau (\omega))
\]

is integrable and

\[
M (\sigma \wedge \tau) = E (M (\tau) \mid \mathcal{F}_\sigma).
\]
Proof: Since $M(t)$ is a martingale, $\|M(t)\|$ is a submartingale. Let

$$\tau_n(\omega) \equiv \sum_{k=0}^{\infty} 2^{-n(k+1)} T X_{\tau^{-1}(k2^{-n}T,(k+1)T2^{-n})}(\omega).$$

By Lemma 13.8.10, $\tau_n$ is a stopping time and the functions $\|M(\tau_n)\|$ are uniformly integrable. Also $\|M(\tau)\|$ is integrable. Similarly $\|M(\tau_n \wedge \sigma_n)\|$ are uniformly integrable where $\sigma_n$ is defined similarly to $\tau_n$.

Consider the main claim now. Letting $\sigma, \tau$ be stopping times with $\tau$ bounded, it follows that for $\sigma_n$ and $\tau_n$ as above, it follows from Theorem 13.8.12

$$M(\sigma \wedge \tau_n) = E(M(\tau_n) \mid \mathcal{F}_{\sigma_n})$$

Thus, taking $A \in \mathcal{F}_\sigma$ and recalling $\sigma \leq \sigma_n$ so that by Proposition 13.7.5, $\mathcal{F}_\sigma \subseteq \mathcal{F}_{\sigma_n}$,

$$\int_A M(\sigma \wedge \tau_n) dP = \int_A E(M(\tau_n) \mid \mathcal{F}_{\sigma_n}) dP = \int_A M(\tau_n) dP.$$

Now passing to a limit as $n \to \infty$, the Vitali convergence theorem, Theorem 13.8.10 on Page 282 and the right continuity of $M$ implies one can pass to the limit in the above and conclude

$$\int_A M(\sigma \wedge \tau) dP = \int_A M(\tau) dP.$$

By Proposition 13.8.10, $M(\sigma \wedge \tau)$ is $\mathcal{F}_{\sigma \wedge \tau} \subseteq \mathcal{F}_\sigma$ measurable showing

$$E(M(\tau) \mid \mathcal{F}_\sigma) = M(\sigma \wedge \tau).$$

A similar theorem is available for submartingales defined on $[0, L], L \leq \infty$.

**Theorem 13.8.12** Let $\{X(t)\}$ be a right continuous submartingale with respect to the increasing sequence of $\sigma$ algebras, $\{\mathcal{F}_t\}$ which is assumed to be a normal filtration,

$$\mathcal{F}_t = \cap_{s \geq t} \mathcal{F}_s,$$

for $t \in [0, L], L \leq \infty$ and let $\sigma, \tau$ be two stopping times with $\tau$ bounded. Then $X(\tau)$ defined as

$$\omega \to X(\tau(\omega))$$

is integrable and

$$X(\sigma \wedge \tau) \leq E(X(\tau) \mid \mathcal{F}_\sigma).$$

**Proof:** Let

$$\tau_n(\omega) \equiv \sum_{k=0}^{\infty} 2^{-n(k+1)} T X_{\tau^{-1}(k2^{-n}T,(k+1)T2^{-n})}(\omega).$$

Then by Lemma 13.8.10, $\tau_n$ is a stopping time, the functions $|X(\tau_n)|$ are uniformly integrable, and $|X(\tau)|$ is also integrable. For $\sigma_n$ defined similarly to $\tau_n$, it also follows $|X(\tau_n \wedge \sigma_n)|$ are uniformly integrable.

Let $A \in \mathcal{F}_\sigma$. Since $\sigma \leq \sigma_n$, it follows that $\mathcal{F}_\sigma \subseteq \mathcal{F}_{\sigma_n}$. By the discrete optional sampling theorem for submartingales, Theorem 13.8.4,

$$X(\sigma_n \wedge \tau_n) \leq E(X(\tau_n) \mid \mathcal{F}_{\sigma_n})$$

and so

$$\int_A X(\sigma_n \wedge \tau_n) dP \leq \int_A E(X(\tau_n) \mid \mathcal{F}_{\sigma_n}) dP = \int_A X(\tau_n) dP$$

and now taking $\lim_{n \to \infty}$ of both sides and using the Vitali convergence theorem along with the right continuity of $X$, it follows

$$\int_A X(\sigma \wedge \tau) dP \leq \int_A X(\tau) dP \equiv \int_A E(X(\tau) \mid \mathcal{F}_\sigma) dP$$

By Proposition 13.8.10, $\mathcal{F}_{\sigma \wedge \tau} \subseteq \mathcal{F}_\sigma$, and so since $A \in \mathcal{F}_\sigma$ was arbitrary,

$$E(X(\tau) \mid \mathcal{F}_\sigma) \geq X(\sigma \wedge \tau) \text{ a.e.}$$

Note that a function defined on a countable ordered set such as the integers or equally spaced points is right continuous.

Here is an interesting lemma.
Lemma 13.8.13 Suppose $E(|X_n|) < \infty$ for all $n$, $X_n$ is $\mathcal{F}_n$ measurable, $\mathcal{F}_{n+1} \subseteq \mathcal{F}_n$ for all $n \in \mathbb{N}$, and there exist $X_\infty \mathcal{F}_\infty$ measurable such that $\mathcal{F}_\infty \subseteq \mathcal{F}_n$ for all $n$ and $X_0 \mathcal{F}_0$ measurable such that $\mathcal{F}_0 \supseteq \mathcal{F}_n$ for all $n$ such that for all $n \in \{0, 1, \cdots \}$,

$$E(X_n|\mathcal{F}_n) \geq X_{n+1}, \quad E(X_n|\mathcal{F}_\infty) \geq X_\infty.$$ 

Then $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable.

Proof:

$$E(X_{n+1}) \leq E(E(X_n|\mathcal{F}_{n+1})) = E(X_n)$$

Therefore, the sequence $E(X_n)$ is a decreasing sequence bounded below by $E(X_\infty)$ so it has a limit. Let $k$ be large enough that

$$\left|E(X_k) - \lim_{m \to \infty} E(X_m)\right| < \varepsilon$$

(13.8.33)

and suppose $n > k$. Then if $\lambda > 0$,

$$\int_{\{|X_n| \geq \lambda\}} |X_n| \, dP = \int_{\{|X_n| \geq \lambda\}} X_n \, dP + \int_{\{|X_n| \leq -\lambda\}} (-X_n) \, dP$$

$$= \int_{\{|X_n| \geq \lambda\}} X_n \, dP + \int_{\Omega} (-X_n) \, dP - \int_{\{|X_n| < \lambda\}} (-X_n) \, dP$$

$$= \int_{\{|X_n| \geq \lambda\}} X_n \, dP - \int_{\Omega} X_n \, dP + \int_{\{|X_n| < \lambda\}} X_n \, dP$$

From (13.8.33),

$$\leq \int_{\{|X_n| \geq \lambda\}} X_n \, dP - \int_{\Omega} X_k \, dP + \varepsilon + \int_{\{|X_n| < \lambda\}} X_n \, dP$$

By assumption,

$$E(X_k|\mathcal{F}_n) \geq X_n$$

and so

$$\leq \int_{\{|X_n| \geq \lambda\}} E(X_k|\mathcal{F}_n) \, dP - \int_{\Omega} X_k \, dP + \varepsilon + \int_{\{|X_n| < \lambda\}} E(X_k|\mathcal{F}_n) \, dP$$

$$= \int_{\{|X_n| \geq \lambda\}} X_k \, dP - \int_{\Omega} X_k \, dP + \varepsilon + \int_{\{|X_n| < \lambda\}} X_k \, dP$$

$$= \int_{\{|X_n| \geq \lambda\}} X_k \, dP - \int_{\Omega} X_k \, dP + \varepsilon + \int_{\{|X_n| < \lambda\}} X_k \, dP$$

$$= \int_{\{|X_n| \geq \lambda\}} X_k \, dP + \left(\int_{\Omega} (-X_k) \, dP - \int_{\{|X_n| < \lambda\}} (-X_k) \, dP\right) + \varepsilon$$

$$= \int_{\{|X_n| \geq \lambda\}} X_k \, dP + \int_{\{|X_n| \leq -\lambda\}} (-X_n) \, dP + \varepsilon = \int_{\{|X_n| \geq \lambda\}} |X_k| \, dP + \varepsilon$$

Applying the maximal inequality for submartingales, Theorem 12.8.2,

$$P(\max \{|X_j| : j = n, \cdots, 1\} \geq \lambda) \leq \frac{1}{\lambda}(E(|X_0|) + E(|X_\infty|)) \leq \frac{C}{\lambda}$$

and taking sup for all $n$,

$$P(\sup \{|X_j|\} \geq \lambda) \leq \frac{C}{\lambda}$$

It follows that for all $\lambda$ large enough,

$$\int_{\{|X_n| \geq \lambda\}} |X_n| \, dP \leq 2\varepsilon$$

and since $\varepsilon$ is arbitrary, this shows $\{X_n\}$ for $n > k$ is equiintegrable. Since there are only finitely many $X_j$ for $j \leq k$, this shows $\{X_n\}$ is equiintegrable. Hence $\{X_n\}$ is uniformly integrable. ■
13.9 Some Maximal Inequalities

As in the case of discrete martingales and submartingales, there are maximal inequalities available.

**Lemma 13.9.1** Let \( X \) be right continuous and adapted such that the given filtration is complete in the sense that \( F_0 \) contains all sets \( A \) of \( \mathcal{F} \) such that \( P(A) = 0 \). Then there exists a set of measure zero \( N \) and a \( \mathcal{F} \times \mathcal{B}(\mathbb{R}) \) measurable function \( Y \) such that if \( \omega \notin N \), then \( Y(t)(\omega) = X(t)(\omega) \). Also, if \( f \) is \( \mathcal{F} \) measurable and nonnegative then \( (\lambda, \omega) \to X_{[\lambda>\lambda]} \) is \( \mathcal{F} \times \mathcal{B}(\mathbb{R}) \) measurable.

**Proof:** Let \( \{t^n_0, t^n_1, \cdots, t^n_m\} \) be a partition of \( [0, T] \) in which \( |t^n_i - t^n_{i-1}| < \rho_n \) where \( \rho_n \to 0 \). Now define \( X_n \) as follows:

\[
X_n(t)(\omega) \equiv \sum_{i=1}^{m_n} X(t^n_i)(\omega) X_{[t^n_{i-1}, t^n_i]}(t)
\]

\[
X_n(0) \equiv X(0).
\]

then each \( X_n \) is obviously product measurable because it is the sum of functions which are. By right continuity, \( X_n \) converges pointwise to \( X \) for \( \omega \notin N \) where \( N \) is a set of measure zero and so if \( Y(t)(\omega) \equiv X(t)(\omega) \) for all \( \omega \notin N \) and \( Y(t)(\omega) = 0 \) for all \( \omega \in N \), this is the desired product measurable function.

To see the last claim, let \( s \) be a nonnegative simple function, \( s(\omega) = \sum_{k=1}^{n} c_k X_{E_k}(\omega) \) where the \( c_k \) are strictly increasing in \( k \). Also let \( F_k = \bigcup_{i=k} E_i \). Then

\[
X_{[s>\lambda]} = \sum_{k=1}^{n} X_{[c_{k-1}, c_k]}(\lambda) X_{E_k}(\omega)
\]

which is clearly product measurable. For arbitrary \( f \geq 0 \) and measurable, there is an increasing sequence of simple functions \( s_n \) converging pointwise to \( f \). Therefore,

\[
\lim_{n \to \infty} X_{[s_n>\lambda]} = X_{[f>\lambda]}
\]

and so \( X_{[f>\lambda]} \) is product measurable. \( \blacksquare \)

**Definition 13.9.2** Let \( X(t) \) be a right continuous submartingale for \( t \in I \) and let \( \{\tau_n\} \) be a sequence of stopping times such that \( \lim_{n \to \infty} \tau_n = \infty \). Then \( X_{\tau_n} \) is called the stopped submartingale and it is defined by

\[
X_{\tau_n}(t) \equiv X(t \wedge \tau_n).
\]

**Proposition 13.9.3** The stopped submartingale just defined is a submartingale.

**Proof:** By the optional sampling theorem for submartingales, Theorem [13.8.12], it follows that for \( s < t \),

\[
E(X_{\tau_n}(t) | F_s) \equiv E(X(t \wedge \tau_n) | F_s) \geq X(t \wedge \tau_n \wedge s)
\]

\[
= X(\tau_n \wedge s) \equiv X_{\tau_n}(s) \ . \ \blacksquare
\]

**Theorem 13.9.4** Let \( \{X(t)\} \) be a right continuous nonnegative submartingale adapted to the normal filtration \( F_t \) for \( t \in [0, T] \). Let \( p \geq 1 \). Define

\[
X^*(t) \equiv \sup \{X(s) : 0 < s < t\}, \ X^*(0) \equiv 0.
\]

Then for \( \lambda > 0 \), if \( X(t) \) is in \( L^1(\Omega) \) for each \( t \),

\[
P(\{X^*(T) > \lambda\}) \leq \frac{1}{\lambda^p} \int X_{[X^*(T)>\lambda]} X(T)^p \ dP
\]

(13.9.34)

If \( X(t) \) is continuous, the above inequality holds without this assumption. In case \( p > 1 \), and \( X(t) \) continuous, then for each \( t \leq T \),

\[
\left( \int_{\Omega} |X^*(t)|^p \ dP \right)^{1/p} \leq \frac{p}{p-1} \left( \int_{\Omega} X(T)^p \ dP \right)^{1/p}
\]

(13.9.35)
13.9. SOME MAXIMAL INEQUALITIES

Proof: The first inequality follows from Theorem [K71]. However, it can also be obtained a different way using stopping times.

Define the stopping time

$$\tau = \inf \{ t > 0 : X(t) > \lambda \} \land T.$$  

(The infimum over an empty set will equal $\infty$.) This is a stopping time by [K71] because it is just a continuous function of the first hitting time of an open set. Also from the definition of $X^*$ in which the supremum is taken over an open interval,

$$[\tau < t] = [X^*(t) > \lambda]$$

Note this also shows $X^*(t)$ is $\mathcal{F}_t$ measurable. Then it follows that $X^p(t)$ is also a submartingale since $r^p$ is increasing and convex. By the optional sampling theorem, $X(0)^p$, $X(\tau)^p$, $X(T)^p$ is a submartingale. Also $[\tau < T] \in \mathcal{F}_\tau$ and so

$$\int_{[\tau < T]} X(\tau)^p dP \leq \int_{[\tau < T]} E(X(T)^p | \mathcal{F}_\tau) dP = \int_{[\tau < T]} X(T)^p dP$$

By right continuity, on $[\tau < T]$, $X(\tau) \geq \lambda$. Therefore,

$$\lambda^p P([X^*(T) > \lambda]) = \lambda^p P([\tau < T]) \leq \int_{[\tau < T]} X(\tau)^p dP \leq \int_{[X^*(T) > \lambda]} X(T)^p dP$$

Next suppose $X(t)$ is continuous and let $\{\tau_n\}$ be a localizing sequence,

$$\tau_n = \inf \{ t : X(t) > n \}.$$  

Then by continuity, $X^\tau_n$ is bounded because $X(\tau_n \land t) \leq n$, and so from what was just shown,

$$\lambda^p P([X^{\tau_n}(T) > \lambda]) \leq \int_{[X^{\tau_n}(T) > \lambda]} (X^{\tau_n})(T)^p dP$$

Then $(X^{\tau_n})(T)$ is increasing as $\tau_n \to \infty$ so the result follows from the monotone convergence theorem. This proves the first part.

Let $X^{\tau_n}$ be as just defined. Thus it is a bounded submartingale. To save on notation, the $X$ in the following argument is really $X^{\tau_n}$. This is done so that all the integrals are finite. If $p > 1$, then from the first part using the case of $p = 1$,

$$\int_{\Omega} |X^*(t)|^p dP \leq \int_{\Omega} |X^*(T)|^p dP = \int_0^\infty p \lambda^{p-1} E([X^*(T) > \lambda]) d\lambda$$

$$\leq \int_0^\infty \lambda^{p-1} \frac{1}{\lambda} \int_{X^*(T) > \lambda} X(T) dP d\lambda$$

$$= \int_0^\infty X(T) \int_0^{X^*(T)} \lambda^{p-1} d\lambda dP$$

$$= \int_0^\infty X(T) \frac{X^*(T)^{p-1}}{p-1} dP$$

$$\leq \frac{p}{p-1} \left( \int_{\Omega} X^*(T)^{p} dP \right)^{1/p} \left( \int_{\Omega} X(T)^p dP \right)^{1/p}$$

Now divide both sides by $\left( \int_{\Omega} X(T)^p dP \right)^{1/p'}$. Substituting $X^{\tau_n}$ for $X$,

$$\left( \int_{\Omega} |X^{\tau_n}(t)|^p dP \right)^{1/p} \leq \left( \int_{\Omega} X^{\tau_n}(T)^p dP \right)^{1/p} \leq \frac{p}{p-1} \left( \int_{\Omega} X^{\tau_n}(T)^p dP \right)^{1/p}$$

Now let $n \to \infty$ and use the monotone convergence theorem to obtain the inequality of the theorem. This establishes [K71]. The use of Fubini’s theorem follows from Lemma [K71].

Here is another sort of maximal inequality in which $X(t)$ is not assumed nonnegative.
Theorem 13.9.5 Let \( \{X(t)\} \) be a right continuous submartingale adapted to the normal filtration \( \mathcal{F}_t \) for \( t \in [0, T] \) and \( X^\ast(t) \) defined as in Theorem 13.9.2.

\[
X^\ast(t) \equiv \sup \{X(s) : 0 < s < t\}, \quad X^\ast(0) \equiv 0,
\]

\[
P(\{X^\ast(T) > \lambda\}) \leq \frac{1}{\lambda} E(|X(T)|) \tag{13.9.36}
\]

For \( t > 0 \), let

\[
X_*(t) = \inf \{X(s) : s < t\}.
\]

Then

\[
P(\{X_* (T) < -\lambda\}) \leq \frac{1}{\lambda} E(|X(T)| + |X(0)|) \tag{13.9.37}
\]

Also

\[
P(\{\sup \{|X(s)| : s < T\} > \lambda\}) \leq \frac{2}{\lambda} E(|X(T)| + |X(0)|) \tag{13.9.38}
\]

**Proof:** The function \( f(r) = r^+ \equiv \frac{1}{2} (|r| + r) \) is convex and increasing. Therefore, \( X^+(t) \) is also a submartingale but this one is nonnegative. Also

\[
[X^\ast(T) > \lambda] = \left( (X^+)^\ast(T) > \lambda \right)
\]

and so from Theorem 13.9.2,

\[
P(\{X^\ast(T) > \lambda\}) = P(\{(X^+)^\ast(T) > \lambda\}) \leq \frac{1}{\lambda} E(X^+(T)) \leq \frac{1}{\lambda} E(|X(T)|).
\]

Next let

\[
\tau = \min (\inf \{t : X(t) < -\lambda\}, T)
\]

then as before, \( X(0), X(\tau), X(T) \) is a submartingale and so

\[
\int_{\tau < T} X(\tau) dP + \int_{\tau = T} X(\tau) dP = \int_\Omega X(\tau) dP \geq \int_\Omega X(0) dP
\]

Now for \( \omega \in [\tau < T], X(t)(\omega) < -\lambda \) for some \( t < T \) and so by right continuity, \( X(\tau)(\omega) \leq -\lambda \). Therefore,

\[
-\lambda \int_{\tau < T} dP \geq -\int_{\tau = T} X(T) dP + \int_\Omega X(0) dP
\]

If \( X_*(T) < -\lambda \), then from the definition given above, there exists \( t < T \) such that \( X(t) < -\lambda \) and so \( \tau < T \). If \( \tau < T \), then by definition, there exists \( t < T \) such that \( X(t) < -\lambda \) and so \( X_*(T) < -\lambda \). Hence \( [\tau < T] = [X_*(T) < -\lambda] \).

It follows that

\[
P(\{X_*(T) < -\lambda\}) = P(\{\tau < T\}) \leq \frac{1}{\lambda} \int_{\tau < T} X(T) dP - \frac{1}{\lambda} \int_\Omega X(0) dP \leq \frac{1}{\lambda} E(|X(T)| + |X(0)|)
\]

and this proves 13.9.2.

Finally, combining the above two inequalities,

\[
P(\{\sup \{|X(s)| : s < T\} > \lambda\}) = P(\{X_*(T) < -\lambda\} + P(\{X^\ast(T) > \lambda\}) \leq \frac{2}{\lambda} E(|X(T)| + |X(0)|).
\]

\[ \blacksquare \]
13.10 Continuous Submartingale Convergence Theorem

In this section, \( \{X(t)\} \) will be a right continuous submartingale.

The next task is to consider an upcrossing estimate as was done before for discrete submartingales. Let \( a < b \) and define

\[
\tau_0 \equiv \min \{ \inf \{ t > 0 : X(t) < a \}, M \}, \\
\tau_1 \equiv \min \{ \inf \{ t > 0 : (X(t \vee \tau_0) - X(\tau_0))_+ > b - a \}, M \}, \\
\tau_2 \equiv \min \{ \inf \{ t > 0 : (X(\tau_1) - X(t \vee \tau_1))_+ > b - a \}, M \}, \\
\tau_3 \equiv \min \{ \inf \{ t > 0 : (X(t \vee \tau_2) - X(\tau_2))_+ > b - a \}, M \}, \\
\tau_4 \equiv \min \{ \inf \{ t > 0 : (X(\tau_3) - X(t \vee \tau_3))_+ > b - a \}, M \}, \\
\vdots
\]

**Lemma 13.10.1** The above \( \tau_i \) are stopping times for \( t \in [0, M] \).

**Proof:** It is obvious that \( \tau_0 \) is a stopping time because it is the minimum of \( M \) and the first hitting time of an open set by a right continuous adapted process. Consider a stopping time \( \eta \leq M \) and let

\[
\sigma \equiv \inf \{ t > 0 : (X(t \vee \eta) - X(\eta))_+ > b - a \}
\]

I claim that \( t \to X(t \vee \eta) - X(\eta) \) is adapted to \( \mathcal{F}_t \). Suppose \( \alpha \geq 0 \) and consider

\[
[(X(t \vee \eta) - X(\eta))_+ > \alpha]
\]

The above set equals

\[
([X(t \vee \eta) - X(\eta))_+ > \alpha] \cap [\eta \leq t]) \cap ([X(t \vee \eta) - X(\eta))_+ > \alpha] \cap [\eta > t])
\]

Consider the second of the above two sets. Since \( \alpha \geq 0 \), this set is \( \emptyset \). This is because for \( \eta > t \), \( X(t \vee \eta) - X(\eta) = 0 \). Now consider the first. It equals

\[
[(X(t \vee \eta) - X(\eta))_+ > \alpha] \cap [\eta \vee t \leq t],
\]

a set of \( \mathcal{F}_{t \vee \eta} \) intersected with \( [\eta \vee t \leq t] \) and so it is in \( \mathcal{F}_t \).

If \( \alpha < 0 \), then \( \{\mathcal{F}_{t \vee \eta}\} \) reduces to \( \Omega \), also in \( \mathcal{F}_t \). Therefore, by Proposition 13.10.2, \( \sigma \) is a stopping time because it is the first hitting time of an open set of a right continuous adapted process. It follows that \( \sigma \land M \) is also a stopping time. Similarly \( t \to X(\eta) - X(t \vee \eta) \) is adapted and

\[
\sigma \equiv \inf \{ t > 0 : (X(\eta) - X(t \vee \eta))_+ > b - a \}
\]

is also a stopping time from the same reasoning. It follows that the \( \tau_i \) defined above are all stopping times. \( \blacksquare \)

Note that in the above, if \( \eta = M \), then \( \sigma = M \) also. Thus in the definition of the \( \tau_i \), if any \( \tau_i = M \), it follows that also \( \tau_{i+1} = M \) and so there is no change in the stopping times. Also note that these stopping times \( \tau_i \) are increasing as \( i \) increases.

Let

\[
U_{[a,b]}^{\eta,M} \equiv \lim_{\varepsilon \to 0} \sum_{k=0}^{n} \frac{X(\tau_{2k+1}) - X(\tau_{2k})}{\varepsilon + X(\tau_{2k+1}) - X(\tau_{2k})} \\
\leq \frac{1}{b-a} \sum_{k=0}^{n} X(\tau_{2k+1}) - X(\tau_{2k})
\]

(13.10.40)

Note that if an upcrossing occurs after \( \tau_{2k} \) on \( [0, M] \), then \( \tau_{2k+1} > \tau_{2k} \) because there exists \( t \) such that

\[
(X(t \vee \tau_{2k}) - X(\tau_{2k}))_+ > b - a
\]

However, you could have \( \tau_{2k+1} > \tau_{2k} \) without an upcrossing occuring. This happens when \( \tau_{2k} < M \) and \( \tau_{2k+1} = M \) which may mean that \( X(t) \) never again climbs above \( b \).
Then $U_{[a,b]}^{nM}$ is clearly a random variable which is at least as large as the number of upcrossings occurring for $t \leq M$ using only $2n+1$ of the stopping times. From the optional sampling theorem,

\[
E(X(\tau_{2k})) - E(X(\tau_{2k-1})) = \int_{\Omega} X(\tau_{2k}) - X(\tau_{2k-1}) \, dP
\]

\[
= \int_{\Omega} E(X(\tau_{2k}) | \mathcal{F}_{\tau_{2k-1}}) - X(\tau_{2k-1}) \, dP
\]

\[
\geq \int_{\Omega} X(\tau_{2k-1}) - X(\tau_{2k-1}) \, dP = 0
\]

Note that from right continuity, $X(\tau_{2k}) \leq a$ while $X(\tau_{2k-1}) \geq b$ so the above may seem surprising. However, the two stopping times can both equal $M$ so this is actually possible. For example, it could happen that $X(t) > a$ for all $t \in [0, M]$.

Next, take the expectation of both sides of

\[
E\left(U_{[a,b]}^{nM}\right) \leq \frac{1}{b-a} \sum_{k=0}^{n} E(X(\tau_{2k+1})) - E(X(\tau_{2k}))
\]

\[
\leq \frac{1}{b-a} \sum_{k=0}^{n} E(X(\tau_{2k+1})) - E(X(\tau_{2k})) + \frac{1}{b-a} \sum_{k=1}^{n} E(X(\tau_{2k})) - E(X(\tau_{2k-1}))
\]

\[
= \frac{1}{b-a} \left(E(X(\tau_1)) - E(X(\tau_0))\right) + \frac{1}{b-a} \sum_{k=1}^{n} E(X(\tau_{2k+1})) - E(X(\tau_{2k})))
\]

\[
\leq \frac{1}{b-a} (E(X(\tau_{2n+1})) - E(X(\tau_0)))
\]

\[
\leq \frac{1}{b-a} (E(X(M)) - E(X(0)))
\]

which does not depend on $n$. The last inequality follows because $0 \leq \tau_{2n+1} \leq M$ and $X(t)$ is a submartingale. Let $n \to \infty$ to obtain

\[
E\left(U_{[a,b]}^{M}\right) \leq \frac{1}{b-a} (E(X(M)) - E(X(0)))
\]

where $U_{[a,b]}^{M}$ is an upper bound to the number of upcrossings of $\{X(t)\}$ on $[0, M]$. This proves the following interesting upcrossing estimate.

**Lemma 13.10.2** Let $\{X(t)\}$ be a right continuous submartingale adapted to a normal filtration $\mathcal{F}_t$ for $t \in [0, M]$. Then if $U_{[a,b]}^{M}$ is defined as the above upper bound to the number of upcrossings of $\{X(t)\}$ for $t \in [0, M]$, then this is a random variable and

\[
E\left(U_{[a,b]}^{M}\right) \leq \frac{1}{b-a} (E(X(M)) - E(X(0)))
\]

With this it is easy to prove a continuous submartingale convergence theorem.

**Theorem 13.10.3** Let $\{X(t)\}$ be a right continuous submartingale adapted to a normal filtration such that

\[
\sup_t \{E(|X(t)|)\} = C < \infty.
\]

Then there exists $X_\infty \in L^1(\Omega)$ such that

\[
\lim_{t \to \infty} X(t)(\omega) = X_\infty(\omega) \quad a.e. \ \omega.
\]

**Proof:** Let $U_{[a,b]}$ be defined by

\[
U_{[a,b]} = \lim_{M \to \infty} U_{[a,b]}^{M}.
\]

Thus the random variable $U_{[a,b]}$ is an upper bound for the number of upcrossings. From Lemma 13.10.2 and the assumption of this theorem, there exists a constant $C$ independent of $M$ such that

\[
E\left(U_{[a,b]}^{M}\right) \leq \frac{2C}{b-a}.
\]
Letting \( M \to \infty \), it follows from monotone convergence theorem that
\[
E \left( U_{[a,b]} \right) \leq \frac{2C}{b-a}
\]
also. Therefore, there exists a set of measure 0 \( N_{ab} \) such that if \( \omega \notin N_{ab} \), then \( U_{[a,b]}(\omega) < \infty \). That is, there are only finitely many upcrossings. Now let
\[ N = \cup \{ N_{ab} : a, b \in \mathbb{Q} \}. \]

It follows that for \( \omega \notin N \), it cannot happen that
\[
\limsup_{t \to \infty} X(t)(\omega) - \liminf_{t \to \infty} X(t)(\omega) > 0
\]
because if this expression is positive, there would be arbitrarily large values of \( t \) where \( X(t)(\omega) > b \) and arbitrarily large values of \( t \) where \( X(t)(\omega) < a \) where \( a, b \) are rational numbers chosen such that
\[
\limsup_{t \to \infty} X(t)(\omega) > b > a > \liminf_{t \to \infty} X(t)(\omega)
\]
Thus there would be infinitely many upcrossings which is not allowed for \( \omega \notin N \). Therefore, the limit \( \lim_{t \to \infty} X(t)(\omega) \) exists for a.e. \( \omega \). Let \( X_\infty(\omega) \) equal this limit for \( \omega \notin N \) and let \( X_\infty(\omega) = 0 \) for \( \omega \in N \). Then \( X_\infty \) is measurable and by Fatou’s lemma,
\[
\int_{\Omega} |X_\infty(\omega)| \, dP \leq \liminf_{n \to \infty} \int_{\Omega} |X(n)(\omega)| \, dP < C. \]

Now here is an interesting result due to Doob.

**Theorem 13.10.4** Let \( \{ M(t) \} \) be a right continuous real martingale adapted to the normal filtration \( F_t \). Then the following are equivalent.

1. The random variables \( M(t) \) are equiintegrable.
2. There exists \( M(\infty) \in L^1(\Omega) \) such that \( \lim_{t \to \infty} \| M(\infty) - M(t) \|_{L^1(\Omega)} = 0 \).

In this case, \( M(t) = E(M(\infty) | F_t) \) and convergence also takes place pointwise.

**Proof:** Suppose the equiintegrable condition. Then there exists \( \lambda \) large enough that for all \( t \),
\[
\int_{\|M(t)\|<\lambda} |M(t)| \, dt < 1.
\]
It follows that for all \( t \),
\[
\int_{\Omega} |M(t)| \, dP = \int_{\|M(t)\|<\lambda} |M(t)| \, dP + \int_{\|M(t)\|\geq\lambda} |M(t)| \, dP \leq 1 + \lambda.
\]
Since the martingale is bounded in \( L^1 \), by Theorem 13.10.3 there exists \( M(\infty) \in L^1(\Omega) \) such that \( \lim_{t \to \infty} M(t)(\omega) = M(\infty)(\omega) \) pointwise a.e. By the assumption \( \{ M(t) \} \) are equiintegrable, it follows these functions are uniformly integrable. Letting \( \delta > 0 \) be such that if \( P(E) < \delta \), then
\[
\int_E |M(t)| \, dP < \frac{\varepsilon}{5},
\]
and \( t_n \to \infty \), Egoroff’s theorem implies that there exists a set \( E \) of measure less than \( \delta \) such that on \( E^C \), the convergence of the \( M(t_n) \) is uniform. Thus
\[
\int_{\Omega} |M(t) - M(t_n)| \, dP = \int_E |M(t) - M(t_n)| \, dP + \int_{E^C} |M(t) - M(t_n)| \, dP \leq \frac{2\varepsilon}{5} + \int_{E^C} |M(t) - M(t_n)| \, dP < \varepsilon
\]
whenever \(m, n\) are large enough. Therefore, the sequence \(\{M(t_n)\}\) is Cauchy in \(L^1(\Omega)\) which implies it converges to something in \(L^1(\Omega)\) which must equal \(M(\infty)\) a.e.

Next suppose there is a function \(M(\infty)\) to which \(M(t)\) converges in \(L^1(\Omega)\). Then for \(t\) fixed and \(A \in \mathcal{F}_t\), then as \(s \to \infty, s > t\)

\[
\int_A M(t)\,dP = \int_A E(M(s)\,|\mathcal{F}_t)\,dP = \int_A M(s)\,dP \\
\to \int_A M(\infty)\,dP = \int_A E(M(\infty)\,|\mathcal{F}_t)
\]

which shows \(E(M(\infty)\,|\mathcal{F}_t) = M(t)\) a.e. since \(A \in \mathcal{F}_t\) is arbitrary. By Lemma \([3.5.5]\),

\[
\int_{\{M(t) \geq \lambda\}} |M(t)|\,dP = \int_{\{M(t) \geq \lambda\}} |E(M(\infty)\,|\mathcal{F}_t)|\,dP \\
\leq \int_{\{M(t) \geq \lambda\}} E(\{|M(\infty)|\,|\mathcal{F}_t\})\,dP \\
= \int_{\{M(t) \geq \lambda\}} |M(\infty)|\,dP
\]  

(13.10.41)

Now from this,

\[
\lambda P(\{|M(t)| \geq \lambda\}) \leq \int_{\{M(t) \geq \lambda\}} |M(t)|\,dP \leq \int_{\Omega} |E(M(\infty)\,|\mathcal{F}_t)|\,dP \\
\leq \int_{\Omega} E(\{|M(\infty)|\,|\mathcal{F}_t\})\,dP = \int_{\Omega} |M(\infty)|\,dP
\]

and so

\[
P(\{|M(t)| \geq \lambda\}) \leq \frac{C}{\lambda}
\]

From \([3.10.11]\), this shows \(\{M(t)\}\) is uniformly integrable because this is true of the single function \(|M(\infty)|\). By the submartingale convergence theorem, the convergence to \(M(\infty)\) also takes place pointwise. ■

### 13.11 Hitting This Before That

Let \(\{M(t)\}\) be a real valued martingale for \(t \in [0, T]\) where \(T \leq \infty\) and \(M(0) = 0\). In case \(T = \infty\), assume the conditions of Theorem \([3.10.11]\) are satisfied. Thus there exists \(M(\infty)\) and the \(M(t)\) are equiintegrable. With the Doob optional sampling theorem it is possible to estimate the probability that \(M(t)\) hits \(a\) before it hits \(b\) where \(a < 0 < b\). There is no loss of generality in assuming \(T = \infty\) since if it is less than \(\infty\), you could just let \(M(t) \equiv M(T)\) for all \(t > T\). In this case, the equiintegrability of the \(M(t)\) follows because for \(t < T\),

\[
\int_{\{M(t) > \lambda\}} |M(t)|\,dP = \int_{\{M(t) > \lambda\}} |E(M(T)\,|\mathcal{F}_t)|\,dP \\
\leq \int_{\{M(t) > \lambda\}} |M(T)|\,dP
\]

and from Theorem \([3.10.3]\),

\[
P(\{|M(t)| > \lambda\}) \leq P(\{|M^*(t)| > \lambda\}) \leq \frac{1}{\lambda} \int_{\Omega} |M(T)|\,dP.
\]

**Definition 13.11.1** Let \(M\) be a process adapted to the filtration \(\mathcal{F}_t\) and let \(\tau\) be a stopping time. Then \(M^\tau\), called the stopped process is defined by

\[
M^\tau(t) \equiv M(\tau \wedge t).
\]

With this definition, here is a simple lemma.

**Lemma 13.11.2** Let \(M\) be a right continuous martingale adapted to the normal filtration \(\mathcal{F}_t\) and let \(\tau\) be a stopping time. Then \(M^\tau\) is also a martingale adapted to the filtration \(\mathcal{F}_t\).
Proof: Let \( s < t \). By the Doob optional sampling theorem,
\[
E(\mathcal{M}^\tau(t) | \mathcal{F}_s) = E(M(\tau \wedge t) | \mathcal{F}_s) = M(\tau \wedge t \wedge s) = \mathcal{M}^\tau(s).
\]

**Theorem 13.11.3** Let \( \{M(t)\} \) be a continuous real valued martingale adapted to the normal filtration \( \mathcal{F}_t \) and let
\[
M^* = \sup \{|M(t)| : t \geq 0\}
\]
and \( M(0) = 0 \). Letting
\[
\tau_x = \inf \{t > 0 : M(t) = x\}
\]
Then if \( a < 0 < b \) the following inequalities hold.
\[
(b - a)P(\{\tau_b \leq \tau_a\}) \geq -aP(\{M^* > 0\}) \geq (b - a)P(\{\tau_b < \tau_a\})
\]
and
\[
(b - a)P(\{\tau_a < \tau_b\}) \leq bP(\{M^* > 0\}) \leq (b - a)P(\{\tau_a \leq \tau_b\}).
\]
In words, \( P(\{\tau_b \leq \tau_a\}) \) is the probability that \( M(t) \) hits \( b \) no later than when it hits \( a \). (Note that if \( \tau_a = \infty = \tau_b \) then you would have \( \tau_a = \tau_b \).

**Proof:** For \( x \in \mathbb{R} \), define
\[
\tau_x = \inf \{t \in \mathbb{R} \text{ such that } M(t) = x\}
\]
with the usual convention that \( \inf(\emptyset) = \infty \). Let \( a < 0 < b \) and let
\[
\tau = \tau_a \wedge \tau_b
\]
Then the following claim will be important.

**Claim:** \( E(M(\tau)) = 0 \).

**Proof of the claim:** Let \( t > 0 \). Then by the Doob optional sampling theorem,
\[
E(M(\tau \wedge t)) = E(E(M(t) | \mathcal{F}_\tau)) = E(M(t))
\]
\[
= E(E(M(t) | \mathcal{F}_0)) = E(M(0)) = 0.
\]

Observe the martingale \( M^* \) must be bounded because it is stopped when \( M(t) \) equals either \( a \) or \( b \). There are two cases according to whether \( \tau = \infty \). If \( \tau = \infty \), then \( M(t) \) never hits \( a \) or \( b \) so \( M(t) \) has values between \( a \) and \( b \). In this case \( M^*(t) = M(t) \in [a,b] \). On the other hand, you could have \( \tau < \infty \). Then in this case \( M^*(t) \) is eventually equal to either \( a \) or \( b \) depending on which it hits first. In either case, the martingale \( M^* \) is bounded and by the martingale convergence theorem, Theorem 13.11.3, there exists \( M^*(\infty) \) such that
\[
\lim_{t \to \infty} M^*(t)(\omega) = M^*(\infty)(\omega) = M(\tau)(\omega)
\]
and since \( M^*(t) \) are bounded, the dominated convergence theorem implies
\[
E(M(\tau)) = \lim_{t \to \infty} E(M(\tau \wedge t)) = 0.
\]
This proves the claim.

Recall
\[
M^*(\omega) = \sup \{|M(t)(\omega)| : t \in [0,\infty]\}.
\]
Also note that \( [\tau_a = \tau_b] = [\tau = \infty] \). Now from the claim,
\[
0 = E(M(\tau)) = \int_{[\tau_a < \tau_b]} M(\tau) dP + \int_{[\tau_b < \tau_a]} M(\tau) dP + \int_{[\tau_a = \tau_b]} M(\infty) dP
\]
\[
+ \int_{[\tau_a = \tau_b]} [M^* = 0] M(\infty) dP
\]
\[
= E(M(\tau)) = aP([\tau_a < \tau_b])
\]

Thus, \( \int_{[\tau_a < \tau_b]} M(\tau) dP \) is either equal to \( a \) or \( b \) depending on whether \( \tau_a < \tau_b \) or \( \tau_b < \tau_a \).
Consider this last term. By the definition, $[\tau_a = \tau_b]$ corresponds to $M(t)$ never hitting either $a$ or $b$. Since $M(0) = 0$, this can only happen if $M(t)$ has values in $[a, b]$. Therefore, this last term satisfies

$$aP ([\tau_a = \tau_b] \cap [M^* > 0])$$

$$\leq \int_{[\tau_a = \tau_b] \cap [M^* > 0]} M(\infty) \, dP$$

It follows

$$0 \leq bP ([\tau_a = \tau_b] \cap [M^* > 0]) + aP ([\tau_a < \tau_b] \cup [\tau_a = \tau_b] \cap [M^* > 0]) + aP ([\tau_a < \tau_b] \cup [\tau_a = \tau_b] \cap [M^* > 0])$$

Note that $[\tau_b < \tau_a], [\tau_a < \tau_b] \subseteq [M^* > 0]$ and so

$$[\tau_b < \tau_a] \cup [\tau_a < \tau_b] \cup ([\tau_a = \tau_b] \cap [M^* > 0]) = [M^* > 0]$$

The following diagram may help in keeping track of the various substitutions.

\[
\begin{array}{c|c|c}
\tau_a < \tau_b & \tau_b < \tau_a & \tau_b = \tau_a \cap [M^* > 0] \\
\end{array}
\]

**Left side of (13.11.45)**

From (13.11.45), this yields on substituting for $P ([\tau_a < \tau_b])$

$$0 \geq aP ([\tau_a = \tau_b] \cap [M^* > 0]) = aP ([\tau_a = \tau_b] \cap [M^* > 0]) - bP ([\tau_a < \tau_b])$$

Hence,

$$0 \geq aP ([\tau_a = \tau_b] \cap [M^* > 0]) - aP ([\tau_a = \tau_b] \cap [M^* > 0]) - aP ([\tau_a < \tau_b] \cap [M^* > 0])$$

and so

$$0 \geq -aP ([\tau_a < \tau_b] \cap [M^* > 0]) + aP ([M^* > 0]) + bP ([\tau_b < \tau_a])$$

which implies, since $[\tau_a > \tau_n] \subseteq [M^* > 0]$,

$$0 \geq -aP ([\tau_a > \tau_b] \cap [M^* > 0]) + aP ([M^* > 0]) + bP ([\tau_b < \tau_a])$$

Recall that $a < 0$. Then the above implies

$$-aP ([M^* > 0]) \geq (b - a) P ([\tau_b < \tau_a])$$

Next use (13.11.45) to substitute for $P ([\tau_b < \tau_a])$. The next line is the bottom half of (13.11.44) and then substitution is made.

$$0 \geq aP ([\tau_a = \tau_b] \cap [M^* > 0]) = aP ([\tau_a = \tau_b] \cap [M^* > 0]) + bP ([\tau_b < \tau_a])$$

again $[\tau_a < \tau_b] \subseteq [M^* > 0]$ and so the above equals

$$= aP ([\tau_a \leq \tau_b] \cap [M^* > 0]) + bP ([M^* > 0]) - P ([\tau_a \leq \tau_b] \cap [M^* > 0])$$

and so

$$(b - a) P ([\tau_a \leq \tau_b]) \geq bP ([M^* > 0])$$
Right side of \( \text{Proposition 13.12.2} \)

From \( \text{Definition 13.12.1} \), used to substitute for \( P(\{\tau_a < \tau_b\}) \) this yields
\[
0 \leq bP(\{\tau_a = \tau_b\} \cap [M^* > 0]) + aP(\{\tau_a < \tau_b\}) + bP(\{\tau_b < \tau_a\})
\]
\[
= bP(\{\tau_a = \tau_b\} \cap [M^* > 0]) + aP([M^* > 0]) - P([\tau_a \geq \tau_b] \cap [M^* > 0])
+ bP(\{\tau_b < \tau_a\})
\]
\[
= bP(\{\tau_a \geq \tau_b\} \cap [M^* > 0]) + aP([M^* > 0]) - P([\tau_a \geq \tau_b] \cap [M^* > 0])
\]
and so
\[
(b - a) P(\{\tau_a \geq \tau_b\}) \geq -aP([M^* > 0]) \tag{13.11.51}
\]
Next use \( \text{Proposition 13.11.48} \) to substitute for the term \( P(\{\tau_b < \tau_a\}) \) and write
\[
0 \leq bP(\{\tau_a = \tau_b\} \cap [M^* > 0]) + aP(\{\tau_a < \tau_b\}) + bP(\{\tau_b < \tau_a\})
\]
\[
= bP(\{\tau_a = \tau_b\} \cap [M^* > 0]) + aP([\tau_a < \tau_b])
+ bP([M^* > 0]) - P([\tau_a \leq \tau_b] \cap [M^* > 0])
\]
\[
= aP([\tau_a < \tau_b]) + bP([M^* > 0]) - bP([\tau_a < \tau_b] \cap [M^* > 0])
\]
\[
= aP([\tau_a < \tau_b]) + bP([M^* > 0]) - bP([\tau_a < \tau_b])
\]
and so
\[
(b - a) P([\tau_a < \tau_b]) \leq bP([M^* > 0]) \tag{13.11.52}
\]
Now the boxed in formulas in \( \text{Proposition 13.11.47} \) yield the conclusion of the theorem. This proves the theorem.

Note \( P(\{\tau_a < \tau_b\}) \) means \( M(t) \) hits \( a \) before it hits \( b \) with other occurrences of similar expressions being defined similarly.

13.12 The Space \( \mathcal{M}^p_T(E) \)

Here \( p \geq 1 \).

Definition 13.12.1 Let \( M \) be an \( E \) valued martingale. Then \( M \in \mathcal{M}^p_T(E) \) if \( t \to M(t)(\omega) \) is continuous for a.e. \( \omega \) and
\[
E \left( \sup_{t \in [0,T]} ||M(t)||^p \right) < \infty
\]
Here \( E \) is a separable Banach space.

Proposition 13.12.2 Define a norm on \( \mathcal{M}^p_T(E) \) by
\[
||M||_{\mathcal{M}^p_T(E)} \equiv E \left( \sup_{t \in [0,T]} ||M(t)||^p \right)^{1/p}
\]
Then with this norm, \( \mathcal{M}^p_T(E) \) is a Banach space.

Proof: First it is good to observe that \( \sup_{t \in [0,T]} ||M(t)||^p \) is measurable. This follows because of the continuity of \( t \to M(t) \). Let \( D \) be a dense countable set in \( [0,T] \). Then by continuity,
\[
\sup_{t \in [0,T]} ||M(t)||^p = \sup_{t \in D} ||M(t)||^p
\]
and the expression on the right is measurable because \( D \) is countable.
Next it is necessary to show this is a norm. It is clear that

\[ ||M||_{\mathcal{M}^p_T(E)} \geq 0 \]

and equals 0 only if

\[ 0 = E \left( \sup_{t \in [0,T]} ||M(t)||^p \right) \]

which requires \( M(t) = 0 \) for all \( t \) and \( \omega \) off a set of measure zero so that \( M = 0 \). It is also clear that

\[ ||\alpha M||_{\mathcal{M}^p_T(E)} = |\alpha| ||M||_{\mathcal{M}^p_T(E)}. \]

It remains to check the triangle inequality. Let \( M, N \in \mathcal{M}^p_T(E) \).

\[ ||M + N||_{\mathcal{M}^p_T(E)} \equiv E \left( \sup_{t \in [0,T]} ||M(t) + N(t)||^p \right)^{1/p} \]

\[ \leq E \left( \sup_{t \in [0,T]} (||M(t)|| + ||N(t)||)^p \right)^{1/p} \]

\[ \leq \left( \int_{\Omega} \left( \sup_{t \in [0,T]} ||M(t)|| + \sup_{t \in [0,T]} ||N(t)|| \right)^p dP \right)^{1/p} \]

\[ \equiv \left( \int_{\Omega} ||M(t)||^p dP \right)^{1/p} + \left( \int_{\Omega} ||N(t)||^p dP \right)^{1/p} \]

\[ \equiv ||M||_{\mathcal{M}^p_T(E)} + ||N||_{\mathcal{M}^p_T(E)} \]

Next consider the claim that \( \mathcal{M}^p_T(E) \) is a Banach space. Let \( \{M_n\} \) be a Cauchy sequence. Then

\[ E \left( \sup_{t \in [0,T]} ||M_n(t) - M_m(t)||^p \right) \to 0 \quad (13.12.53) \]

as \( m, n \to \infty \). From continuity,

\[ \sup_{t \in [0,T]} ||M_n(t) - M_m(t)|| = \sup_{t \in [0,T]} ||M_n(t) - M_m(t)|| \]

Then from theorem [13.6.3] or [13.9.4],

\[ P \left( \sup_{t \in [0,T]} ||M_n(t) - M_m(t)|| > \lambda \right) \leq \frac{1}{\lambda^p} E \left( ||M_n(T) - M_m(T)||^p \right) \]

Therefore, one can extract a subsequence \( \{M_{n_k}\} \) such that

\[ P \left( \sup_{t \in [0,T]} ||M_{n_k}(t) - M_{n_{k+1}}(t)|| > 2^{-k} \right) \leq 2^{-k}. \]

By the Borel Cantelli lemma, it follows \( \{M_{n_k}(t)(\omega)\} \) converges uniformly on \([0,T]\) for a.e. \( \omega \). Denote by \( M(t)(\omega) \) the thing to which it converges, a continuous process because of the uniform convergence. Also, because it is the pointwise limit off a set of measure zero, \( \omega \to M(t)(\omega) \) is \( \mathcal{F}_t \) measurable. Also, from [13.6.3] and Fatou’s lemma

\[ \int_{\Omega} \sup_{t \in [0,T]} ||M_n(t) - M(t)||^p dP \]

\[ \leq \lim_{k \to \infty} \inf \int_{\Omega} \sup_{t \in [0,T]} ||M_n(t) - M_{n_k}(t)||^p dP \leq \varepsilon \]
whenever $n$ is large enough, this from the assumption that $\{M_n\}$ is Cauchy. Thus

$$\lim_{n \to \infty} E \left( \sup_{t \in [0,T]} ||M_n(t) - M(t)||^p \right) = 0$$

and so for each $t, M_n(t) \to M(t)$ in $L^p(\Omega)$. This also shows that for large, $n$

$$E \left( \sup_{t \in [0,T]} ||M(t)||^p \right) \leq E \left( \sup_{t \in [0,T]} (||M(t) - M_n(t)|| + ||M_n(t)||^p) \right) \leq 2^{p-1} E \left( \sup_{t \in [0,T]} ||M(t) - M_n(t)||^p + \sup_{t \in [0,T]} (||M_n(t)||^p) \right) < \infty$$

It only remains to verify $M$ is a martingale. Let $s \leq t$ and let $B \in \mathcal{F}_s$. For each $s$, $M_n(s) \to M(s)$ in $L^p(\Omega)$. Then from the above, $\omega \to M(s)(\omega)$ is $\mathcal{F}_s$ measurable. Then it follows that

$$\int_B M(s) dP = \lim_{n \to \infty} \int_B M_n(s) dP = \lim_{n \to \infty} \int_B E(M_n(t)|\mathcal{F}_s) dP = \lim_{n \to \infty} \int_B M(t) dP$$

and so by definition, $E(M(t)|\mathcal{F}_s) = M(s)$ which shows $M$ is a martingale. ■

**Proposition 13.12.3** The functions $M(t)$ for each $M \in \mathcal{M}^p_T(E)$ are equi integrable.

**Proof:** This follows because

$$\int_{\{||M(t)|| \geq \lambda\}} ||M(t)||^p dP \leq \int_{\sup_{t \in [0,T]} ||M(t)|| \geq \lambda} \left( \sup_{t \in [0,T]} ||M(t)||^p \right) dP \quad (13.12.54)$$

which converges to 0 due to the definition of $\mathcal{M}^p_T(E)$ which requires that

$$\sup_{t \in [0,T]} ||M(t)||^p \in L^1(\Omega, \mathcal{F}, P).$$

Since the sets $\left[ \sup_{t \in [0,T]} ||M(t)|| \geq \lambda \right]$ decrease to $\emptyset$ as $\lambda \to \infty$, the dominated convergence theorem implies the integral on the right in (13.12.54) converges to 0. ■
Chapter 14

The Quadratic Variation Of A Martingale

14.1 How To Recognize A Martingale

The main ideas are most easily understood in the special case where it is assumed the martingale is bounded. Then one can extend to more general situations.

Let \( \{ M(t) \} \) be a continuous martingale having values in a separable Hilbert space. The idea is to consider the submartingale \( \{ ||M(t)||^2 \} \) and write it as the sum of a martingale and a submartingale. An important part of the argument is the following lemma which gives a checkable criterion for a stochastic process to be a martingale.

**Lemma 14.1.1** Let \( \{ X(t) \} \) be a stochastic process adapted to the filtration \( \{ F_t \} \) for \( t \geq 0 \). Then it is a martingale for the given filtration if for every stopping time, \( \sigma \) it follows

\[
E(X(t)) = E(X(\sigma)) .
\]

In fact, it suffices to check this on stopping times which have two values.

**Proof:** Let \( s < t \) and \( A \in F_s \). Define a stopping time

\[
\sigma(\omega) \equiv s \chi_A(\omega) + t \chi_{A^c}(\omega)
\]

This is a stopping time because \([\sigma \leq l] = \Omega \) if \( l \geq t \). Also \([\sigma \leq l] = A \in F_s \) if \( l \in [s, t) \) and \([\sigma \leq l] = \emptyset \) if \( l < s \). Then by assumption,

\[
\int_A X(t) \, dP + \int_{A^c} X(t) \, dP = \int_A X(\sigma) \, dP = \int_A X(s) \, dP + \int_{A^c} X(t) \, dP
\]

Therefore,

\[
\int_A X(t) \, dP = \int_A X(s) \, dP
\]

and since \( X(s) \) is \( F_s \) measurable, it follows \( E(X(t) | F_s) = X(s) \) a.e. and this shows \( \{ X(t) \} \) is a martingale. \( \square \)

Note that if \( t \in [0, T] \), it suffices to check the expectation condition for stopping times which have two values no larger than \( T \).

The following lemma will be useful.

**Lemma 14.1.2** Suppose \( X_n \to X \) in \( L^1(\Omega, F, P; E) \) where \( E \) is a separable Banach space. Then letting \( G \) be a \( \sigma \) algebra contained in \( F \),

\[
E(X_n | G) \to E(X | G)
\]

in \( L^1(\Omega) \).
CHAPTER 14. THE QUADRATIC VARIATION OF A MARTINGALE

Proof: This follows from the definitions and Theorem 13.1.1 on Page 251.

\[
\int_{\Omega} \| E(X|\mathcal{G}) - E(X_n|\mathcal{G}) \| \, dP = \int_{\Omega} \| E(X_n - X|\mathcal{G}) \| \, dP \\
\leq \int_{\Omega} E(\|X_n - X\| |\mathcal{G}) \, dP \\
= \int_{\Omega} \|X_n - X\| \, dP \]

Corollary 14.1.3 Let \( X, Y \) be in \( L^2(\Omega, \mathcal{F}, P; H) \) where \( H \) is a separable Hilbert space and let \( X \) be \( \mathcal{G} \) measurable where \( \mathcal{G} \subseteq \mathcal{F} \). Then

\[ E((X,Y)|\mathcal{G}) = (X,E(Y|\mathcal{G})) \text{ a.e.} \]

Proof: First let \( X = aX_B \) where \( B \in \mathcal{G} \). Then for \( A \in \mathcal{G} \),

\[
\int_{A} E((aX_B,Y)|\mathcal{G}) \, dP = \int_{A} X_B E((a,Y)|\mathcal{G}) \, dP = \int_{A} (a, \int_{A} Y \, dP) \\
= (a, \int_{A \cap B} Y \, dP) \\
= (a, \int_{A} X_B E(Y|\mathcal{G}) \, dP) = (a, \int_{A \cap B} Y \, dP) \\
\]

It follows that the formula holds for \( X \) simple.

Therefore, letting \( X_n \) be a sequence of \( \mathcal{G} \) measurable simple functions converging pointwise to \( X \) and also in \( L^2(\Omega) \),

\[ E((X_n,Y)|\mathcal{G}) = (X_n,E(Y|\mathcal{G})) \]

Now the desired formula holds from Lemma 14.1.2. ■

The following is related to something called a martingale transform. It is a lot like what will happen later with the Ito integral.

Proposition 14.1.4 Let \( \{\tau_k\} \) be an increasing sequence of stopping times for the normal filtration \( \{\mathcal{F}_t\} \) such that

\[ \lim_{k \to \infty} \tau_k = \infty, \quad \tau_0 = 0. \]

Also let \( \xi_k \) be \( \mathcal{F}_{\tau_k} \) measurable with values in \( H \), a separable Hilbert space and let \( M(t) \) be a right continuous martingale adapted to the normal filtration \( \mathcal{F}_t \) which has the property that \( M(t) \in L^2(\Omega; H) \) for all \( t, M(0) = 0 \). Then if \( |\xi_k| \leq C \),

\[
E \left( \left( \sum_{k \geq 0} (\xi_k, (M(\tau_{k+1} \land t) - M(\tau_k \land t))) \right)^2 \right) \\
\leq C^2 E \left( \|M(t)\|^2 \right) \tag{14.1.1}
\]

Proof: First of all, the sum converges because eventually \( \tau_k \land t = t \). Therefore, for large enough \( k \), \( M(\tau_{k+1} \land t) - M(\tau_k \land t) \equiv \Delta M_k = 0 \). Consider first the finite sum, \( k \leq q \).

\[
E \left( \sum_{k=0}^{q} (\xi_k, \Delta M_k) \right)^2 \tag{14.1.2}
\]

When the sum is multiplied out, you get mixed terms. Consider one of these mixed terms, \( j < k \)

\[ E \left( (\xi_k, \Delta M_k) (\xi_j, \Delta M_j) \right) \]
Using Corollary 14.1.2 and Doob’s optional sampling theorem, Theorem 14.3.3, this equals
\[
E \left( E \left( (\xi_k, \Delta M_k) (\xi_j, \Delta M_j) | F_{\tau_k} \right) \right) = E \left( (\xi_j, \Delta M_j) E \left( (\xi_k, \Delta M_k) | F_{\tau_k} \right) \right)
\]
\[
= E \left( (\xi_j, \Delta M_j) E \left( M (\tau_{k+1} \wedge t) - M (\tau_k \wedge t) | F_{\tau_k} \right) \right) = E \left( (\xi_j, \Delta M_j) (\xi_k, 0) \right) = 0
\]
Note that in using the optional sampling theorem, the stopping time \( \tau_{k+1} \wedge t \) is bounded.

Therefore, the only terms which survive in \( \sum_{k=0}^{q} E(\xi_k, \Delta M_k)^2 \) are the non mixed terms and so this expression reduces to
\[
\sum_{k=0}^{q} E(\xi_k, \Delta M_k)^2 \leq C^2 \sum_{k=0}^{q} E \left( ||\Delta M_k||^2 \right)
\]
\[
= C^2 \sum_{k=0}^{q} E \left( ||M (\tau_{k+1} \wedge t) - M (\tau_k \wedge t)||^2 \right)
\]
\[
= C^2 \sum_{k=0}^{q} E \left( ||M (\tau_{k+1} \wedge t)||^2 \right) + E \left( ||M (\tau_k \wedge t)||^2 \right)
\]
\[
- 2E \left( (M (\tau_{k} \wedge t), M (\tau_{k+1} \wedge t)) \right)
\]
(14.1.3)

Consider the term \( E \left( (M (\tau_{k} \wedge t), M (\tau_{k+1} \wedge t)) \right) \). By Doob’s optional sampling theorem for martingales and Corollary 14.1.2 again, this equals
\[
E \left( E \left( (M (\tau_{k} \wedge t), M (\tau_{k+1} \wedge t)) | F_{\tau_k} \right) \right)
\]
\[
= E \left( (M (\tau_{k} \wedge t), E \left( (M (\tau_{k+1} \wedge t) | F_{\tau_k} \right) \right) \right)
\]
\[
= E \left( (M (\tau_{k} \wedge t), M (\tau_{k+1} \wedge t \wedge \tau_k)) \right)
\]
\[
= E \left( ||M (\tau_k \wedge t)||^2 \right)
\]
It follows \( \sum_{k=0}^{q} E \left( ||M (\tau_{k+1} \wedge t)||^2 \right) \) equals
\[
C^2 \sum_{k=0}^{q} E \left( ||M (\tau_{k+1} \wedge t)||^2 \right) - E \left( ||M (\tau_k \wedge t)||^2 \right) \leq C^2 E \left( ||M (t)||^2 \right)
\]
Then from Fatou’s lemma,
\[
E \left( \left( \sum_{k=0}^{q} (\xi_k, (M (\tau_{k+1} \wedge t) - M (\tau_k \wedge t))) \right)^2 \right) \leq
\]
\[
\lim_{q \to \infty} E \left( \left( \sum_{k=0}^{q} (\xi_k, (M (\tau_{k+1} \wedge t) - M (\tau_k \wedge t))) \right)^2 \right)
\]
\[
\leq C^2 E \left( ||M (t)||^2 \right)
\]
Now here is an interesting lemma which will be used to prove uniqueness in the main result.

**Lemma 14.1.5** Let \( \mathcal{F} \) be a normal filtration and let \( A (t), B (t) \) be adapted to \( \mathcal{F} \), continuous, and increasing with \( A (0) = B (0) = 0 \) and suppose \( A (t) - B (t) \) is a martingale. Then \( A (t) - B (t) = 0 \) for all \( t \).

**Proof:** I shall show \( A (l) = B (l) \) where \( l \) is arbitrary. Let \( M (t) \) be the name of the martingale. Define a stopping time
\[
\tau \equiv \inf \{ t > 0 : |M (t)| > C \} \wedge l \wedge \inf \{ t > 0 : A (t) > C \}
\]
\[
\wedge \inf \{ t > 0 : B (t) > C \}
\]
where \( \inf (\emptyset) \equiv \infty \) and denote the stopped martingale
\[
M^\tau (t) \equiv M (t \wedge \tau).
\]
Then I claim this is also a martingale with respect to the filtration $\mathcal{F}_t$ because by Doob’s optional sampling theorem for martingales, if $s < t$,

$$E \left( M^\tau (t) \mid \mathcal{F}_s \right) \equiv E \left( M (\tau \wedge t) \mid \mathcal{F}_s \right) = M (\tau \wedge s) = M^\tau (s)$$

Note the bounded stopping time is $\tau \wedge t$ and the other one is $\sigma = s$ in this theorem. Then $M^\tau$ is a continuous martingale which is also uniformly bounded. It equals $A^\tau - B^\tau$. The stopping time ensures $A^\tau$ and $B^\tau$ are uniformly bounded by $C$. Thus all of $|A^\tau (t)|, B^\tau (t), A^\tau (l)$ are bounded by $C$ on $[0,l]$. Now let $\mathcal{P}_n \equiv \{t_k\}_{k=1}^n$ be a uniform partition of $[0,l]$ and let $M^\tau (\mathcal{P}_n)$ denote

$$M^\tau (\mathcal{P}_n) \equiv \max \left\{ |M^\tau (t_{i+1}) - M^\tau (t_i)| \right\}_{i=1}^n .$$

Then

$$E \left( M^\tau (l)^2 \right) = E \left( \left( \sum_{k=0}^{n-1} M^\tau (t_{k+1}) - M^\tau (t_k) \right)^2 \right)$$

Now consider a mixed term in the sum where $j < k$.

$$E \left( (M^\tau (t_{k+1}) - M^\tau (t_k)) (M^\tau (t_{j+1}) - M^\tau (t_j)) \right)$$

$$= E \left( E \left( (M^\tau (t_{k+1}) - M^\tau (t_k)) (M^\tau (t_{j+1}) - M^\tau (t_j)) \mid \mathcal{F}_k \right) \right)$$

$$= E \left( (M^\tau (t_{j+1}) - M^\tau (t_j)) E \left( (M^\tau (t_{k+1}) - M^\tau (t_k)) \mid \mathcal{F}_k \right) \right)$$

$$= E \left( (M^\tau (t_{j+1}) - M^\tau (t_j)) (M^\tau (t_{k+1}) - M^\tau (t_k)) \right) = 0$$

It follows

$$E \left( M^\tau (l)^2 \right) = E \left( \sum_{k=0}^{n-1} (M^\tau (t_{k+1}) - M^\tau (t_k))^2 \right)$$

$$\leq E \left( \sum_{k=0}^{n-1} M^\tau (\mathcal{P}_n) \left| M^\tau (t_{k+1}) - M^\tau (t_k) \right| \right)$$

$$\leq E \left( \sum_{k=0}^{n-1} M^\tau (\mathcal{P}_n) \left( |A^\tau (t_{k+1}) - A^\tau (t_k)| + |B^\tau (t_{k+1}) - B^\tau (t_k)| \right) \right)$$

$$\leq E \left( M^\tau (\mathcal{P}_n) \sum_{k=0}^{n-1} \left( |A^\tau (t_{k+1}) - A^\tau (t_k)| + |B^\tau (t_{k+1}) - B^\tau (t_k)| \right) \right)$$

$$\leq E \left( M^\tau (\mathcal{P}_n) 2C \right)$$

the last step holding because $A$ and $B$ are increasing. Now letting $n \to \infty$, the right side converges to 0 by the dominated convergence theorem and the observation that for a.e. $\omega$,

$$\lim_{n \to \infty} M^\tau (\mathcal{P}_n) (\omega) = 0$$

because of continuity of $M$. Thus for $\tau = \tau_C$ given above,

$$M (l \wedge \tau_C) = 0 \text{ a.e.}$$

Now let $C \in \mathbb{N}$ and let $N_C$ be the exceptional set off which $M (l \wedge \tau_C) = 0$. Then letting $N_l$ denote the union of all these exceptional sets for $C \in \mathbb{N}$, it is also a set of measure zero and for $\omega$ not in this set, $M (l \wedge \tau_C) = 0$ for all $C$. Since the martingale is continuous, it follows for each such $\omega$, eventually $\tau_C > l$ and so $M (l) = 0$. Thus for $\omega \notin N_l$,

$$M (l) (\omega) = 0$$

Now let $N = \cup_{l \in \mathbb{Q} \cap [0,\infty)} N_l$. Then for $\omega \notin N$, $M (l) (\omega) = 0$ for all $l \in \mathbb{Q} \cap [0,\infty)$ and so by continuity, this is true for all positive $l$. ■

Note this shows a continuous martingale is not of bounded variation unless it is a constant.
14.2 The Quadratic Variation

This section is on the quadratic variation of a martingale. Actually, you can also consider the quadratic variation of a local martingale which is more general. Therefore, this concept is defined first. We will generally assume \( M(0) = 0 \) since there is no real loss of generality in doing so. One can simply subtract \( M(0) \) otherwise.

**Definition 14.2.1** Let \( \{M(t)\} \) be adapted to the normal filtration \( \mathcal{F}_t \) for \( t > 0 \). Then \( \{M(t)\} \) is a local martingale (submartingale) if there exist stopping times \( \tau_n \) increasing to infinity such that for each \( n \), the process \( M^\tau_n(t) \equiv M(t \wedge \tau_n) \) is a martingale (submartingale) with respect to the given filtration. The sequence of stopping times \( \tau_n \) is called a localizing sequence. The martingale \( M^\tau_n \) is called the stopped martingale. Exactly the same convention applies to a localized submartingale.

**Proposition 14.2.2** If \( M(t) \) is a continuous local martingale (submartingale) for a normal filtration as above, \( M(0) = 0 \), then there exists a localizing sequence \( \tau_n \) such that for each \( n \) the stopped martingale(submartingale) \( M^\tau_n \) is uniformly bounded. Also if \( M \) is a martingale, then \( M^\tau \) is also a martingale (submartingale). If \( \tau_n \) is an increasing sequence of stopping times such that \( \lim_{n \to \infty} \tau_n = \infty \), and for each \( \tau_n \) and real valued stopping time \( \delta \), there exists a function \( X \) of \( \tau_n \wedge \delta \) such that \( X(\tau_n \wedge \delta) \) is \( \mathcal{F}_{\tau_n \wedge \delta} \) measurable, then \( \lim_{n \to \infty} X(\tau_n \wedge \delta) \equiv X(\delta) \) exists for each \( \omega \) and \( X(\delta) \) is \( \mathcal{F}_\delta \) measurable.

**Proof:** First consider the claim about \( M^\tau \) being a martingale (submartingale) when \( M \) is. By optional sampling theorem,

\[
E(M^\tau(t) | \mathcal{F}_s) = E(M(\tau \wedge t) | \mathcal{F}_s) = M(\tau \wedge t \wedge s) = M^\tau(s).
\]

The case where \( M \) is a submartingale is similar.

Next suppose \( \sigma_n \) is a localizing sequence for the local martingale (submartingale) \( M \). Then define

\[
\eta_n \equiv \inf \{ t > 0 : ||M(t)|| > n \}.
\]

Therefore, by continuity of \( M, ||M(\eta_n)|| \leq n \). Now consider \( \tau_n \equiv \eta_n \wedge \sigma_n \). This is an increasing sequence of stopping times. By continuity of \( M \), it must be the case that \( \eta_n \to \infty \). Hence \( \sigma_n \wedge \eta_n \to \infty \).

Finally, consider the last claim. Pick \( \omega \). Then \( X(\tau_n(\omega) \wedge \delta(\omega))(\omega) \) is eventually constant as \( n \to \infty \) because for all \( n \) large enough, \( \tau_n(\omega) > \delta(\omega) \) and so this sequence of functions converges pointwise. That which it converges to, denoted by \( X(\delta) \), is \( \mathcal{F}_\delta \) measurable because each function \( \omega \to X(\tau_n(\omega) \wedge \delta(\omega))(\omega) \) is \( \mathcal{F}_{\delta \wedge \tau_n} \subseteq \mathcal{F}_{\delta} \) measurable.

One can also give a generalization of Lemma [14.1.9] to conclude a local martingale must be constant or else they must fail to be of bounded variation.

**Corollary 14.2.3** Let \( \mathcal{F}_t \) be a normal filtration and let \( A(t), B(t) \) be adapted to \( \mathcal{F}_t \), continuous, and increasing with \( A(0) = B(0) = 0 \) and suppose \( A(t) - B(t) \equiv M(t) \) is a local martingale. Then \( M(t) = A(t) - B(t) = 0 \) a.e. for all \( t \).

**Proof:** Let \( \{\tau_n\} \) be a localizing sequence for \( M \). For given \( n \), consider the martingale,

\[
M^\tau_n(t) = A^\tau_n(t) - B^\tau_n(t)
\]

Then from Lemma [14.1.9], it follows \( M^\tau_n(t) = 0 \) for all \( t \) for all \( \omega \notin N_n \), a set of measure 0. Let \( N = \cup_n N_n \). Then for \( \omega \notin N \), \( M(\tau_n(\omega) \wedge t)(\omega) = 0 \). Let \( n \to \infty \) to conclude that \( M(t)(\omega) = 0 \). Therefore, \( M(t)(\omega) = 0 \) for all \( t \).

Recall Example [14.3.5] on Page 270. For convenience, here is a version of what it says.

**Lemma 14.2.4** Let \( X(t) \) be continuous and adapted to a normal filtration \( \mathcal{F}_t \) and let \( \eta \) be a stopping time. Then if \( K \) is a closed set with \( 0 \notin K \),

\[
\tau \equiv \inf \{ t > \eta : X(t) \in K \}
\]

is also a stopping time.

**Proof:** First consider \( Y(t) = X(t \vee \eta) - X(\eta) \). I claim that \( Y(t) \) is adapted to \( \mathcal{F}_t \). Consider \( U \) and open set and \( [Y(t) \in U] \). Is it in \( \mathcal{F}_t \)? We know it is in \( \mathcal{F}_{t \vee \eta} \). It equals

\[
([Y(t) \in U] \cap [\eta \leq t]) \cup ([Y(t) \in U] \cap [\eta > t])
\]

Consider the second of these sets. It equals

\[
([X(\eta) - X(\eta) \in U] \cap [\eta > t])
\]
If \( 0 \in U \), then it reduces to \( [\eta > t] \in \mathcal{F}_t \). If \( 0 \notin U \), then it reduces to \( \emptyset \) still in \( \mathcal{F}_t \). Next consider the first set. It equals
\[
[X(\tau \wedge \eta) - X(\eta) \in U] \cap [\eta \leq t] = [X(\tau \wedge \eta) - X(\eta) \in U] \cap [\tau \wedge \eta \leq t] \in \mathcal{F}_t
\]
from the definition of \( \mathcal{F}_{\tau \wedge \eta} \). (You know that \( [X(\tau \wedge \eta) - X(\eta) \in U] \in \mathcal{F}_{\tau \wedge \eta} \) and so when this is intersected with \( [\tau \wedge \eta \leq t] \) one obtains a set in \( \mathcal{F}_t \). This is what it means to be in \( \mathcal{F}_{\tau \wedge \eta} \).) Now \( \tau \) is just the first hitting time of \( Y(t) \) of the closed set. ■

**Proposition 14.2.5** Let \( M(t) \) be a continuous local martingale for \( t \in [0,T] \) having values in \( H \) a separable Hilbert space adapted to the normal filtration \( \{\mathcal{F}_t\} \) such that \( M(0) = 0 \). Then there exists a unique continuous, increasing, nonnegative, local submartingale \( [M](t) \) called the quadratic variation such that
\[
||M(t)||^2 - [M](t)
\]
is a real local martingale and \( [M](0) = 0 \). Here \( t \in [0,T] \). If \( \delta \) is any stopping time
\[
[M^\delta] = [M] ^\delta
\]

**Proof:** First it is necessary to define some stopping times. Define stopping times \( \tau_0^\delta \equiv \eta_0^\delta \equiv 0. \)
\[
\eta_{k+1}^n \equiv \inf \{ s > \eta_k^n : ||M(s) - M(\eta_k^n)|| = 2^{-n} \}, \\
\tau_k^n \equiv \eta_k^n \wedge T
\]
where \( \inf \emptyset \equiv \infty \). These are stopping times by Example 13.8.7 on Page 263. See also Lemma 14.2.3. Then for \( t > 0 \) and \( \delta \) any stopping time, and fixed \( \omega \), for some \( k, \)
\[
t \wedge \delta \in I_k(\omega), I_0(\omega) \equiv [\tau_0^\delta(\omega), \tau_1^\delta(\omega)], I_k(\omega) \equiv [\tau_k^\delta(\omega), \tau_{k+1}^\delta(\omega)] some \ k
\]
Here is why. The sequence \( \{\tau_k^\delta(\omega)\}_{k=1}^\infty \) eventually equals \( T \) for all \( n \) sufficiently large. This is because if it did not, it would converge, being bounded above by \( T \) and then by continuity of \( M \), \( \{M(\tau_k^n(\omega))\}_{k=1}^\infty \) would be a Cauchy sequence contrary to the requirement that
\[
||M(\tau_{k+1}^n(\omega)) - M(\tau_k^\delta(\omega))|| = ||M(\eta_{k+1}^n(\omega)) - M(\eta_k^n(\omega))|| = 2^{-n}.
\]
Note that if \( \delta \) is any stopping time, then
\[
||M(t \wedge \delta \wedge \tau_{k+1}^n) - M(t \wedge \delta \wedge \tau_k^n)|| = ||M^\delta(t \wedge \tau_{k+1}^n) - M^\delta(t \wedge \tau_k^n)|| \leq 2^{-n}
\]
You can see this is the case by considering the cases, \( t \wedge \delta \geq \tau_{k+1}^n, t \wedge \delta \in [\tau_k^n, \tau_{k+1}^n], \) and \( t \wedge \delta < \tau_k^n. \) It is only this approximation property and the fact that the \( \tau_k^n \) partition \([0,T]\) which is important in the following argument.

Now let \( \alpha_n \) be a localizing sequence such that \( M^{\alpha_n} \) is bounded as in Proposition 14.2.3. Thus \( M^{\alpha_n}(t) \in L^2(\Omega) \) and this is all that is needed. In what follows, let \( \delta \) be a stopping time and denote \( M^{\alpha_n + \delta} \) by \( M \) to save notation. Thus \( M \) will be uniformly bounded and from the definition of the stopping times \( \tau_k^n \), for \( t \in [0,T] \),
\[
M(t) \equiv \sum_{k \geq 0} M(t \wedge \tau_{k+1}^n) - M(t \wedge \tau_k^n),
\]
and the terms of the series are eventually 0, as soon as \( \eta_k^n = \infty \). Therefore,
\[
||M(t)||^2 = \left| \sum_{k \geq 0} M(t \wedge \tau_{k+1}^n) - M(t \wedge \tau_k^n) \right|^2
\]
Then this equals
\[
= \sum_{k \geq 0} ||M(t \wedge \tau_{k+1}^n) - M(t \wedge \tau_k^n)||^2
\]
Consider one of these mixed terms for \( j \neq k \):

\[
+ \sum_{j \neq k} ((M(t \land \tau_{k+1}^n) - M(t \land \tau_k^n)) - (M(t \land \tau_{j+1}^n) - M(t \land \tau_j^n)))
\] (14.2.5)

Consider the second sum. It equals

\[
2 \sum_{k \geq 0} \sum_{j=0}^{k-1} ((M(t \land \tau_{k+1}^n) - M(t \land \tau_k^n)) - (M(t \land \tau_{j+1}^n) - M(t \land \tau_j^n)))
\]

\[
= 2 \sum_{k \geq 0} \left( (M(t \land \tau_{k+1}^n) - M(t \land \tau_k^n)) \sum_{j=0}^{k-1} (M(t \land \tau_{j+1}^n) - M(t \land \tau_j^n)) \right)
\]

\[
= 2 \sum_{k \geq 0} ((M(t \land \tau_{k+1}^n) - M(t \land \tau_k^n)) M(t \land \tau_k^n))
\]

This last sum equals \( P_n(t) \) defined as

\[
2 \sum_{k \geq 0} ((M(\tau_k^n), (M(t \land \tau_{k+1}^n) - M(t \land \tau_k^n))) \equiv P_n(t)
\] (14.2.6)

This is because in the \( k \)th term, if \( t \geq \tau_k^n \), then it reduces to

\[
(M(\tau_k^n), (M(t \land \tau_{k+1}^n) - M(t \land \tau_k^n))
\]

while if \( t < \tau_k^n \), then the term reduces to 0 which is also the same as

\[
(M(\tau_k^n), (M(t \land \tau_{k+1}^n) - M(t \land \tau_k^n)))
\]

This is a finite sum because eventually, for large enough \( k \), \( \tau_k^n = T \). However, the number of nonzero terms depends on \( \omega \). This is not a good thing. However, a little more can be said. In fact, the sum also converges in \( L^2(\Omega) \). Say \( ||M(t, \omega)|| \leq C \).

\[
E \left( \sum_{k \geq p} ((M(\tau_k^n), (M(t \land \tau_{k+1}^n) - M(t \land \tau_k^n)))^2 \right)
\]

\[
= \sum_{k \geq p} E \left( ((M(\tau_k^n), (M(t \land \tau_{k+1}^n) - M(t \land \tau_k^n)))^2 \right) + \text{mixed terms}
\] (14.2.7)

Consider one of these mixed terms for \( j < k \).

\[
E \left( ((M(\tau_j^n), \Delta_j, (M(t \land \tau_{j+1}^n) - M(t \land \tau_j^n))) \cdot ((M(\tau_k^n), \Delta_k, (M(t \land \tau_{k+1}^n) - M(t \land \tau_k^n)))
\]

Then it equals

\[
E \left( E((M(\tau_j^n), \Delta_j), (M(\tau_k^n), \Delta_k) | F_{\tau_k}))
\]

\[
= E((M(\tau_j^n), \Delta_j) E((M(\tau_k^n), \Delta_k) | F_{\tau_k}))
\]

\[
= E((M(\tau_j^n), \Delta_j), (M(\tau_k^n), E(\Delta_k | F_{\tau_k}))) = 0
\]

Now since the mixed terms equal 0, it follows from \[**14.2.7**\], that expression is dominated by

\[
C^2 \sum_{k \geq p} E \left( ||M(t \land \tau_{k+1}^n) - M(t \land \tau_k^n)||^2 \right)
\]
Using a similar manipulation to what was just done to show the mixed terms equal 0, this equals
\[
C^2 \sum_{k=p}^{q} E \left( \left\| M \left( t \wedge \tau_{k+1}^n \right) \right\|^2 \right) - E \left( \left\| M \left( t \wedge \tau_k^n \right) \right\|^2 \right)
\]
\[
\leq C^2 E \left( \left\| M \left( t \wedge \tau_{q+1}^n \right) \right\|^2 - \left\| M \left( t \wedge \tau_p^n \right) \right\|^2 \right)
\]
The integrand converges to 0 as \( p, q \to \infty \) and the uniform bound on \( M \) allows a use of the dominated convergence theorem. Thus the partial sums of the series of \( 14.2.6 \) converge in \( L^2(\Omega) \) as claimed.

By adding in the values of \( \left\{ \tau_{k+1}^n \right\} P_n(t) \) can be written in the form
\[
2 \sum_{k \geq 0} (M \left( \tau_{k+1}^n \right), (M \left( t \wedge \tau_{k+1}^n \right) - M \left( t \wedge \tau_k^n \right)))
\]
where \( \tau_{k+1}^n \) has some repeats. From the construction,
\[
\left\| M \left( \tau_{k+1}^n \right) - M \left( \tau_k^n \right) \right\| \leq 2^{-(n+1)}
\]
Thus
\[
P_n(t) - P_{n+1}(t) = 2 \sum_{k \geq 0} (M \left( \tau_{k+1}^n \right) - M \left( \tau_k^n \right), (M \left( t \wedge \tau_{k+1}^n \right) - M \left( t \wedge \tau_k^n \right)))
\]
and so from Proposition \( 14.2.6 \) applied to \( \xi_k \equiv M \left( \tau_{k+1}^n \right) - M \left( \tau_k^n \right) \),
\[
E \left( \left\| P_n(t) - P_{n+1}(t) \right\|^2 \right) \leq 2^{-2n} E \left( \left\| M(t) \right\|^2 \right).
\]
(14.2.8)

Now \( t \to P_n(t) \) is continuous because it is a finite sum of continuous functions. It is also the case that \( \left\{ P_n(t) \right\} \) is a martingale. To see this use Lemma \( 14.2.6 \). Let \( \sigma \) be a stopping time having two values. Then using Corollary \( 14.2.6 \) and the Doob optional sampling theorem, Theorem \( 14.3.11 \)
\[
E \left( \sum_{k=0}^{q} (M \left( \tau_k^n \right), (M \left( \sigma \wedge \tau_{k+1}^n \right) - M \left( \sigma \wedge \tau_k^n \right))) \right)
\]
\[
= \sum_{k=0}^{q} E \left( (M \left( \tau_k^n \right), (M \left( \sigma \wedge \tau_{k+1}^n \right) - M \left( \sigma \wedge \tau_k^n \right))) \right)
\]
\[
= \sum_{k=0}^{q} E \left( (E \left( M \left( \tau_k^n \right), (M \left( \sigma \wedge \tau_{k+1}^n \right) - M \left( \sigma \wedge \tau_k^n \right)) | \mathcal{F}_{\tau_k^n} \right)) \right)
\]
\[
= \sum_{k=0}^{q} E \left( (M \left( \tau_k^n \right), E \left( M \left( \sigma \wedge \tau_{k+1}^n \right) - M \left( \sigma \wedge \tau_k^n \right) \right) | \mathcal{F}_{\tau_k^n} \right))
\]
\[
= \sum_{k=0}^{q} E \left( (M \left( \tau_k^n \right), E \left( M \left( \sigma \wedge \tau_{k+1}^n \wedge \tau_k^n \right) - M \left( \sigma \wedge \tau_k^n \right) \right)) \right) = 0
\]
Note the Doob theorem applies because \( \sigma \wedge \tau_{k+1}^n \) is a bounded stopping time due to the fact \( \sigma \) has only two values. Similarly
\[
E \left( \sum_{k=0}^{q} (M \left( \tau_k^n \right), (M \left( t \wedge \tau_{k+1}^n \right) - M \left( t \wedge \tau_k^n \right))) \right)
\]
\[
= \sum_{k=0}^{q} E \left( (M \left( \tau_k^n \right), (M \left( t \wedge \tau_{k+1}^n \right) - M \left( t \wedge \tau_k^n \right))) \right)
\]
\[
= \sum_{k=0}^{q} E \left( (E \left( M \left( \tau_k^n \right), (M \left( t \wedge \tau_{k+1}^n \right) - M \left( t \wedge \tau_k^n \right)) | \mathcal{F}_{\tau_k^n} \right)) \right)
\]
\[
= \sum_{k=0}^{q} E \left( (M \left( \tau_k^n \right), E \left( M \left( t \wedge \tau_{k+1}^n \right) - M \left( t \wedge \tau_k^n \right) \right) | \mathcal{F}_{\tau_k^n} \right))
\]
\[
= \sum_{k=0}^{q} E \left( (M \left( \tau_k^n \right), E \left( M \left( t \wedge \tau_{k+1}^n \wedge \tau_k^n \right) - M \left( t \wedge \tau_k^n \right) \right)) \right) = 0
\]
14.2. THE QUADRATIC VARIATION

It follows each partial sum for $P_n(t)$ is a martingale. As shown above, these partial sums converge in $L^2(\Omega)$ and so it follows that $P_n(t)$ is also a martingale. Note the Doob theorem applies because $t \wedge \tau^n_{k+1}$ is a bounded stopping time.

I want to argue that $P_n$ is a Cauchy sequence in $\mathcal{M}^2_T(\mathbb{R})$. By Theorem 14.2.2 and continuity of $P_n$

$$E\left(\sup_{t \leq T} |P_n(t) - P_{n+1}(t)|^2 \right)^{1/2} \leq 2E\left(|P_n(T) - P_{n+1}(T)|^2\right)^{1/2}$$

By 14.2.2,

$$\leq 2^{-n}E\left(||M(T)||^2\right)^{1/2}$$

which shows $\{P_n\}$ is indeed a Cauchy sequence in $\mathcal{M}^2_T(\mathbb{R})$.

Therefore, by Proposition 13.9.4, there exists $\{N(t)\} \in \mathcal{M}^2_T(\mathbb{R})$ such that $P_n \to N$ in $\mathcal{M}^2_T(H)$. That is

$$\lim_{n \to \infty} E\left(\sup_{t \in [0,T]} |P_n(t) - N(t)|^2\right)^{1/2} = 0.$$ 

Since $\{N(t)\} \in \mathcal{M}^2_T(\mathbb{R})$, it is a continuous martingale and $N(t) \in L^2(\Omega)$, and $N(0) = 0$ because this is true of each $P_n(0)$. From the above 14.2.9

$$||M(t)||^2 = Q_n(t) + P_n(t)$$

where

$$Q_n(t) = \sum_{k \geq 0} ||M(t \wedge \tau^n_{k+1}) - M(t \wedge \tau^n_k)||^2$$

and $P_n(t)$ is a martingale. Then from 14.2.2, $Q_n(t)$ is a submartingale and converges for each $t$ to something, denoted as $[M](t)$ in $L^1(\Omega)$ uniformly in $t \in [0,T]$. This is because $P_n(t)$ converges uniformly on $[0,T]$ to $N(t)$ in $L^2(\Omega)$ and $||M(t)||^2$ does not depend on $n$. Then also $[M]$ is a submartingale which equals 0 at 0 because this is true of $Q_n$ and because if $A \in \mathcal{F}_s$ where $s < t$,

$$\int_A E([M](t)|\mathcal{F}_s) \ dP = \int_A [M](t) \ dP = \lim_{n \to \infty} \int_A \left(||M(t)||^2 - P_n(t)\right) \ dP$$

$$= \lim_{n \to \infty} \int_A E\left(||M(t)||^2 - P_n(t)\right) \ dP \geq \lim_{n \to \infty} \int_A ||M(s)||^2 - P_n(s) \ dP$$

$$= \lim_{n \to \infty} \int_A Q_n(s) \ dP = \int_A [M](s) \ dP.$$

Note that $Q_n(t)$ is increasing because as $t$ increases, the definition allows for the possibility of more nonzero terms in the sum. Therefore, $[M](t)$ is also increasing in $t$. The function $t \to [M](t)$ is continuous because $||M(t)||^2 = [M](t) + N(t)$ and $t \to N(t)$ is continuous as is $t \to ||M(t)||^2$. That is, off a set of measure zero, these are both continuous functions of $t$ and so the same is true of $[M]$.

Now put back in $M^\alpha \wedge \delta$ in place of $M$. From the above, this has shown

$$||M^\alpha \wedge \delta(t)||^2 = [M^\alpha \wedge \delta](t) = N_p(t)$$

where $N_p$ is a martingale and

$$[M^\alpha \wedge \delta](t) = \lim_{n \to \infty} \sum_{k \geq 0} ||M^\alpha \wedge \delta(t \wedge \tau^n_{k+1}) - M^\alpha \wedge \delta(t \wedge \tau^n_k)||^2$$

$$= \lim_{n \to \infty} \sum_{k \geq 0} ||M(t \wedge \tau^n_{k+1} \wedge \alpha \wedge \delta) - M(t \wedge \tau^n_k \wedge \alpha \wedge \delta)||^2 \text{ in } L^1(\Omega),$$

the convergence being uniform on $[0,T]$. The above formula shows that $[M^\alpha \wedge \delta](t)$ is a $\mathcal{F}_{t \wedge \delta \wedge \alpha_p}$ measurable random variable which depends on $t \wedge \delta \wedge \alpha_p$. (Note that $t \wedge \delta$ is a real valued stopping time even if $\delta = \infty$.) Therefore, by Proposition 14.2.2, there exists a random variable, denoted as $[M^\delta](t)$ which is the pointwise limit as $p \to \infty$ of
these random variables which is \( \mathcal{F}_t \) measurable because, for a given \( \omega \), when \( \alpha_p \) becomes larger than \( t \), the sum in (14.2.10) loses its dependence on \( p \). Thus from pointwise convergence in (14.2.10),

\[
[M^\delta] (t) \equiv \lim_{n \to \infty} \sum_{k \geq 0} ||M (t \land \delta \land \tau^\alpha_{k+1}) - M (t \land \delta \land \tau^\alpha_k)||^2
\]

In case \( \delta = \infty \), the above gives an \( \mathcal{F}_t \) measurable random variable denoted by \([M] (t)\) such that

\[
[M] (t) \equiv \lim_{n \to \infty} \sum_{k \geq 0} ||M (t \land \tau^\alpha_{k+1}) - M (t \land \tau^\alpha_k)||^2
\]

Now stopping with the stopping time \( \delta \), this shows that

\[
[M^\delta] (t) \equiv \lim_{n \to \infty} \sum_{k \geq 0} ||M (t \land \delta \land \tau^\alpha_{k+1}) - M (t \land \delta \land \tau^\alpha_k)||^2 = [M]^\delta (t)
\]

That is, the quadratic variation of the stopped local martingale makes sense a.e. and equals the stopped quadratic variation of the local martingale.

This has now shown that

\[
||M^{\alpha_n} (t)||^2 - [M]^{\alpha_n} (t) = ||M^{\alpha_n} (t)||^2 - [M]^{\alpha_n} (t) = N_n (t), \quad N_n (t) \text{ a martingale}
\]

and both of the random variables on the left converge pointwise as \( n \to \infty \) to a function which is \( \mathcal{F}_t \) measurable. Hence so does \( N_n (t) \). Of course \( N_n (t) \) is likewise a function of \( \alpha_n \land t \) and so by Proposition 14.2.10 again, it converges pointwise to a \( \mathcal{F}_t \) measurable function called \( N (t) \) and \( N (t) \) is a continuous local martingale.

It remains to consider the claim about the uniqueness. Suppose then there are two which work, \([M] \), and \([M]_1\). Then \([M] - [M]_1\) equals a local martingale \( G \) which is 0 when \( t = 0 \). Thus the uniqueness assertion follows from Corollary 14.2.6.

Here is a corollary which tells how to manipulate stopping times. It is contained in the above proposition, but it is worth emphasizing it from a different point of view.

**Corollary 14.2.6** In the situation of Proposition 14.2.4 let \( \tau \) be a stopping time. Then

\[
[M^\tau] = [M]^\tau.
\]

**Proof:**

\[
[M]^\tau (t) + N_1 (t) = \left(||M||^2\right)^\tau (t) = ||M^\tau||^2 (t) = [M]^\tau (t) + N_2 (t)
\]

where \( N_i \) is a local martingale. Therefore, \([M]^\tau (t) - [M^\tau] (t) = N_2 (t) - N_1 (t)\), a local martingale. Therefore, by Corollary 14.2.6, this shows \([M]^\tau (t) - [M^\tau] (t) = 0.\]

### 14.3 The Covariation

**Definition 14.3.1** The covariation of two continuous \( H \) valued local martingales for \( H \) a separable Hilbert space \( M, N, M (0) = 0 = N (0) \), is defined as follows.

\[
[M, N] \equiv \frac{1}{4} ([M + N] - [M - N])
\]

**Lemma 14.3.2** The following hold for the covariation.

\[
[M, N] = [M, M]
\]

\[
[M, N] = \text{local martingale} + \frac{1}{4} \left(||M + N||^2 - ||M - N||^2\right)
\]

\[
= (M, N) + \text{local martingale}.
\]
14.3. THE COVARIATION

Proof: From the definition of covariation,

\[ [M] = ||M||^2 - N_1 \]

\[ [M,M] = \frac{1}{4} ([M + M] - [M - M]) = \frac{1}{4} \left( ||M + M||^2 - N_2 \right) \]

\[ = ||M||^2 - \frac{1}{4} N_2 \]

where \( N_1 \) is a local martingale. Thus \([M] - [M,M]\) is equal to the difference of two increasing continuous adapted processes and it also equals a local martingale. By Corollary 14.2.3, this process must equal 0. Now consider the second claim.

\[ [M,N] = \frac{1}{4} ([M + N] - [M - N]) = \frac{1}{4} \left( ||M + N||^2 - ||M - N||^2 + N \right) \]

\[ = (M,N) + \frac{1}{4} N \]

where \( N \) is a local martingale. ■

Corollary 14.3.3 Let \( M, N \) be two continuous local martingales, \( M(0) = N(0) = 0 \), as in Proposition 14.2.5. Then \([M,N]\) is of bounded variation and

\[ (M,N)_H - [M,N] \]

is a local martingale. Also for \( \tau \) a stopping time,

\[ [M,N]_\tau = [M^\tau, N^\tau] = [M^\tau, N] = [M, N^\tau]. \]

In addition to this,

\[ [M - M^\tau] = [M] - [M^\tau] \leq [M] \]

and also

\[ (M,N) \rightarrow [M,N] \]

is bilinear and symmetric.

Proof: Since \([M,N]\) is the difference of increasing functions, it is of bounded variation.

\[ (M,N)_H - [M,N] = \left( \frac{(M,N)_H}{[M,N]} \right) - \frac{1}{4} ([M + N] - [M - N]) \]

which equals a local martingale from the definition of \([M + N]\) and \([M - N]\). It remains to verify the claim about the stopping time. Using Corollary 14.3.3

\[ [M,N]_\tau = \frac{1}{4} ([M + N] - [M - N])^\tau \]

\[ = \frac{1}{4} ([M + N]^\tau - [M - N]^\tau) \]

\[ = \frac{1}{4} ([M^\tau + N^\tau] - [M^\tau - N^\tau]) \equiv [M^\tau, N^\tau]. \]

The really interesting part is the next equality. This will involve Corollary 14.3.3

\[ [M,N]_\tau - [M^\tau, N] = [M^\tau, N^\tau] - [M^\tau, N] \]

\[ = \frac{1}{4} ([M^\tau + N^\tau] - [M^\tau - N^\tau]) - \frac{1}{4} ([M^\tau + N] - [M^\tau - N]) \]

\[ = \frac{1}{4} ([M^\tau + N^\tau] + [M^\tau - N]) - \frac{1}{4} ([M^\tau + N] + [M^\tau - N^\tau]), \quad (14.3.11) \]
the difference of two increasing adapted processes. Also, this equals

local martingale \( -(M^\tau, N) + (M^\tau, N^\tau) \)

**Claim:** \((M^\tau, N) - (M^\tau, N^\tau) = (M^\tau, N - N^\tau)\) is a local martingale. Let \(\sigma_n\) be a localizing sequence for both \(M\) and \(M\). Such a localizing sequence is of the form \(\tau_n^M \wedge \tau_n^N\) where these are localizing sequences for the indicated local submartingale. Then obviously,

\[ (- (M^\tau, N) + (M^\tau, N^\tau))^{\sigma_n} = -(M^{\sigma_n \wedge \tau}, N^{\sigma_n}) + (M^{\sigma_n \wedge \tau}, N^{\sigma_n \wedge \tau}) \]

where \(N^{\sigma_n}\) and \(M^{\sigma_n}\) are martingales. To save notation, denote these by \(M\) and \(N\) respectively. Now use Lemma 14.1.1. Let \(\sigma\) be a stopping time with two values.

**E**((\(M^\tau (\sigma), N (\sigma) - N^\tau (\sigma)\))) = \(E((M^\tau (\sigma), N (\sigma) - N^\tau (\sigma)) | \mathcal{F}_\tau)\)

Now \(M^\tau (\sigma)\) is \(M (\sigma \wedge \tau)\) which is \(\mathcal{F}_\tau\) measurable and so by the Doob optional sampling theorem,

\[ = E(M^\tau (\sigma), E(N (\sigma) - N^\tau (\sigma)) | \mathcal{F}_\tau) \]

while

\[ = E(M^\tau (\sigma), N (\sigma \wedge \tau) - N (\tau \wedge \sigma)) = 0 \]

Since \(M^\tau (t)\) is \(\mathcal{F}_\tau\) measurable,

\[ = E((M^\tau (t), E(N (t) - N^\tau (t)) | \mathcal{F}_\tau)) \]

\[ = E((M^\tau (t), E(N (t \wedge \tau) - N (t \wedge \tau))) | \mathcal{F}_\tau)) = 0 \]

This shows the claim is true.

Now from [14.1.1] and Corollary 14.1.3,

\[ [M, N]^\tau - [M^\tau, N] = 0. \]

Similarly

\[ [M, N]^\tau - [M, N^\tau] = 0 \]

Now consider the next claim that \([M - M^\tau] = [M] - [M^\tau]\). From the definition, it follows

\[ [M - M^\tau] - ([M] + [M^\tau] - 2[M, M^\tau]) \]

\[ = ||M - M^\tau||^2 - \left(||M||^2 + ||M^\tau||^2 - 2(M, M^\tau)\right) \]

+ local martingale.

By the first part of the corollary which ensures \([M, M^\tau]\) is of bounded variation, the left side is the difference of two increasing adapted processes and so by Corollary 14.2.3 again, the left side equals 0. Thus from the above,

\[ [M - M^\tau] = [M] + [M^\tau] - 2[M, M^\tau] \]

\[ = [M] + [M^\tau] - 2[M^\tau, M^\tau] \]

\[ = [M] + [M^\tau] - 2[M^\tau] \]

\[ = [M] - [M^\tau] \leq [M] \]

Finally consider the claim that \([M, N]\) is bilinear. From the definition, letting \(M_1, M_2, N\) be \(H\) valued local martingales,

\[ (aM_1 + bM_2, N)_{H} = [aM_1 + bM_2, N] + \text{local martingale} \]

\[ a(M_1, N) + b(M_2, N)_{H} = a[M_1, N] + b[M_2, N] + \text{local martingale} \]

Hence

\[ [aM_1 + bM_2, N] - (a[M_1, N] + b[M_2, N]) = \text{local martingale.} \]

The left side can be written as the difference of two increasing functions thanks to \([M, N]\) of bounded variation and so by Lemma 14.1.3 it equals 0. \([M, N]\) is obviously symmetric from the definition. \(\blacksquare\)
14.4 The Burkholder Davis Gundy Inequality

Define

\[ M^* (\omega) \equiv \sup \{ ||M (t) (\omega)|| : t \in [0, T] \} . \]

The Burkholder Davis Gundy inequality is an amazing inequality which involves \( M^* \) and \([M] (T)\).

Before presenting this, here is the good lambda inequality, Theorem 14.4.1 on Page 14.3 listed here for convenience.

**Theorem 14.4.1** Let \((\Omega, \mathcal{F}, \mu)\) be a finite measure space and let \(F\) be a continuous increasing function defined on \([0, \infty)\) such that \(F (0) = 0\). Suppose also that for all \(\alpha > 1\), there exists a constant \(C_\alpha\) such that for all \(x \in [0, \infty)\),

\[ F (ax) \leq C_\alpha F (x) . \]

Also suppose \(f, g\) are nonnegative measurable functions and there exists \(\beta > 1, 0 < r \leq 1\), such that for all \(\lambda > 0\) and \(1 > \delta > 0\,

\[ \mu ([f > \beta \lambda] \cap [g \leq r \delta \lambda]) \leq \phi (\delta) \mu ([f > \lambda]) \]

where \(\lim_{\delta \to 0^+} \phi (\delta) = 0\) and \(\phi\) is increasing. Under these conditions, there exists a constant \(C\) depending only on \(\beta, \phi, r\) such that

\[ \int_{\Omega} F (f (\omega)) d\mu (\omega) \leq C \int_{\Omega} F (g (\omega)) d\mu (\omega) . \]

The proof of this important inequality also will depend on the hitting this before that theorem which is listed next for convenience.

**Theorem 14.4.2** Let \(\{M (t)\}\) be a continuous real valued martingale adapted to the normal filtration \(\mathcal{F}_t\), and let

\[ M^* \equiv \sup \{ ||M (t)| : t \geq 0 \} \]

and \(M (0) = 0\). Letting

\[ \tau_x \equiv \inf \{ t > 0 : M (t) = x \} \]

Then if \(a < 0 < b\) the following inequalities hold.

\[ (b - a) P ([\tau_b \leq \tau_a]) \geq -a P ([M^* > 0]) \geq (b - a) P ([\tau_b < \tau_a]) \]

and

\[ (b - a) P ([\tau_a < \tau_b]) \leq b P ([M^* > 0]) \leq (b - a) P ([\tau_a \leq \tau_b]) . \]

In words, \(P ([\tau_b \leq \tau_a])\) is the probability that \(M (t)\) hits \(b\) no later than when it hits \(a\). (Note that if \(\tau_a = \infty = \tau_b\) then you would have \(\tau_a = \tau_b\).)

Then the Burkholder Davis Gundy inequality is as follows. Generalizations will be presented later.

**Theorem 14.4.3** Let \(\{M (t)\}\) be a continuous \(H\) valued martingale which is uniformly bounded, \(M (0) = 0\), where \(H\) is a separable Hilbert space and \(t \in [0, T]\). Then if \(F\) is a function of the sort described in the good lambda inequality above, there are constants, \(C\) and \(c\) independent of such martingales \(M\) such that

\[ c \int_{\Omega} F \left( ([M] (T))^{1/2} \right) dP \leq \int_{\Omega} F (M^*) dP \leq C \int_{\Omega} F \left( ([M] (T))^{1/2} \right) dP \]

where

\[ M^* (\omega) \equiv \sup \{ ||M (t) (\omega)|| : t \in [0, T] \} . \]

**Proof:** Using Corollary 14.3.3, let

\[ N (t) \equiv ||M (t) - M^* (t)||^2 - [M - M^*] (t) \]

where

\[ \tau \equiv \inf \{ t \in [0, T] : ||M (t)|| > \lambda \} \]

Thus \(N\) is a martingale and \(N (0) = 0\). In fact \(N (t) = 0\) as long as \(t \leq \tau\). As usual \(\inf (\emptyset) \equiv \infty\). Note

\[ [\tau < \infty] = [M^* > \lambda] \supset [N^* > 0] . \]
This is because to say \( \tau < \infty \) is to say there exists \( t < T \) such that \(|M(t)|| > \lambda \) which is the same as saying \( M^* > \lambda \). Thus the first two sets are equal. If \( \tau = \infty \), then from the formula for \( N(t) \) above, \( N(t) = 0 \) for all \( t \in [0, T] \) and so it can’t happen that \( N^* > 0 \). Thus the third set is contained in \( \{ \tau < \infty \} \) as claimed.

Let \( \beta > 2 \) and let \( \delta \in (0, 1) \). Then

\[
\beta - 1 > 1 > \delta > 0
\]

Consider the following which is set up to use the good lambda inequality.

\[
S_r = [M^* > \beta \lambda] \cap \left( ([M](T))^1/2 \leq r \delta \lambda \right)
\]

where \( 0 < r < 1 \). It is shown that \( S_r \) corresponds to hitting “this before that” and there is an estimate for this which involves \( P ([N^* > 0]) \) which is bounded above by \( P ([M^* > \lambda]) \) as discussed above. This will satisfy the hypotheses of the good lambda inequality.

**Claim:** For \( \omega \in S_r \), \( N(t) \) hits \( \lambda^2 \left( 1 - \delta^2 \right) \).

**Proof of claim:** For \( \omega \in S_r \), there exists a \( t < T \) such that \( |M(t)| > \beta \lambda \) and so using Corollary 14.3.3

\[
N(t) \geq ||M(t)|| - ||M^*(t)||^2 - |M - M^*(t)| \geq |\beta \lambda - \lambda^2 - [M(t)]
\]

\[
\geq (\beta - 1)^2 \lambda^2 - \delta^2 \lambda^2
\]

which shows that \( N(t) \) hits \( (\beta - 1)^2 \lambda^2 - \delta^2 \lambda^2 \) for \( \omega \in S_r \). By the intermediate value theorem, it also hits \( \lambda^2 \left( 1 - \delta^2 \right) \).

This proves the claim.

**Claim:** \( N(t)(\omega) \) never hits \( -\delta^2 \lambda^2 \) for \( \omega \in S_r \).

**Proof of claim:** Suppose \( t \) is the first time \( N(t) \) reaches \( -\delta^2 \lambda^2 \). Then \( t > \tau \) and so

\[
N(t) = -\delta^2 \lambda^2 \geq ||M(t)|| - \lambda^2 - [M(t)] + [M^*(t)]
\]

\[
\geq -r^2 \lambda^2 \delta^2,
\]

a contradiction since \( r < 1 \). This proves the claim.

Therefore, for all \( \omega \in S_r \), \( N(t)(\omega) \) reaches \( \lambda^2 \left( 1 - \delta^2 \right) \) before it reaches \( -\delta^2 \lambda^2 \). It follows

\[
P(S_r) \leq P(N(t) \text{ reaches } \lambda^2 \left( 1 - \delta^2 \right) \text{ before } -\delta^2 \lambda^2)
\]

and because of Theorem 13.11.3 this is no larger than

\[
P([N^* > 0]) \frac{\delta^2 \lambda^2}{\lambda^2 (1 - \delta^2) - (-\delta^2 \lambda^2)} = P([N^* > 0]) \delta^2 \leq \delta^2 P([M^* > \lambda]).
\]

Thus

\[
P \left( [M^* > \beta \lambda] \cap \left( ([M](T))^1/2 \leq r \delta \lambda \right) \right) \leq P([M^* > \lambda]) \delta^2
\]

By the good lambda inequality,

\[
\int_{\Omega} F(M^*)dP \leq C \int_{\Omega} F \left( ([M](T))^1/2 \right) dP
\]

which is one half the inequality.

Now consider the other half. This time define the stopping time \( \tau \) by

\[
\tau = \inf \left\{ t \in [0, T] : (|M(t)|)^1/2 > \lambda \right\}
\]

and let

\[
S_r = \left( ([M](T))^1/2 > \beta \lambda \right) \cap [2M^* \leq r \delta \lambda].
\]

Then there exists \( t < T \) such that \( |M(t)| > \beta^2 \lambda^2 \). This time, let

\[
N(t) = |M(t)| - |M^*(t)| - ||M(t) - M^*(t)||^2
\]

This is still a martingale since by Corollary 14.3.3

\[
[M](t) - [M^*(t)] = [M - M^*(t)](t)
\]
Claim: \( N(t)(\omega) \) hits \( \lambda^2 \left( 1 - \delta^2 \right) \) for some \( t < T \) for \( \omega \in S_T \).

Proof of claim: Fix such a \( \omega \in S_T \). Let \( t < T \) be such that \( [M](t) > \beta^2 \lambda^2 \). Then \( t > \tau \) and so for that \( \omega \),

\[
N(t) \geq \beta^2 \lambda^2 - \lambda^2 - \|M(t) - M(\tau)\|^2 \\
\geq (\beta - 1)^2 \lambda^2 - (\|M(t)\| + \|M(\tau)\|)^2 \\
\geq (\beta - 1)^2 \lambda^2 - \tau^2 \delta^2 \lambda^2 \geq \lambda^2 - \delta^2 \lambda^2
\]

By the intermediate value theorem, it hits \( \lambda^2 \left( 1 - \delta^2 \right) \). This proves the claim.

Claim: \( N(t)(\omega) \) never hits \(-\delta^2 \lambda^2 \) for \( \omega \in S_T \).

Proof of claim: By Corollary 14.3.3 if it did at \( t \), then \( t > \tau \) because \( N(t) = 0 \) for \( t \leq \tau \), and so

\[
0 \leq [M](t) - [M]^+(t) = [M](t) - [M](\tau)^2 - \delta^2 \lambda^2 \\
\leq (\|M(t)\| + \|M(\tau)\|)^2 - \delta^2 \lambda^2 \leq r^2 \delta^2 \lambda^2 - \delta^2 \lambda^2 < 0,
\]

a contradiction. This proves the claim.

It follows that for each \( \tau \in (0, 1) \),

\[
P(S_\tau) \leq P(N(t) \text{ hits } \lambda^2 \left( 1 - \delta^2 \right) \text{ before } -\delta^2 \lambda^2)
\]

By Theorem 14.4.3 this is no longer than

\[
P(\{N^* > 0\}) \frac{\delta^2 \lambda^2}{\lambda^2 \left( 1 - \delta^2 \right) + \delta^2 \lambda^2} = P(\{N^* > 0\}) \delta^2
\]

\[
\leq P(\{\tau < \infty\}) \delta^2 = P\left(\left(\|M(T)\|^{1/2} > \lambda\right)\right) \delta^2
\]

Now by the good lambda inequality, there is a constant \( k \) independent of \( M \) such that

\[
\int F\left(\|M(T)\|^{1/2}\right) dP \leq k \int F(2M^*) dP \leq kC_2 \int F(M^*) dP
\]

by the assumptions about \( F \). Therefore, combining this result with the first part,

\[
(kC_2)^{-1} \int F\left(\|M(T)\|^{1/2}\right) dP \leq \int F(M^*) dP \leq C \int F\left(\|M(T)\|^{1/2}\right) dP
\]

Of course, everything holds for local martingales in place of martingales.

**Theorem 14.4.4** Let \( \{M(t)\} \) be a continuous \( H \) valued local martingale, \( M(0) = 0 \), where \( H \) is a separable Hilbert space and \( t \in [0, T] \). Then if \( F \) is a function of the sort described in the good lambda inequality, that is,

\[
F(0) = 0, \ F \text{ continuous, } F \text{ increasing,} \quad F(\alpha x) \leq c_\alpha F(x),
\]

there are constants, \( C \) and \( c \) independent of such local martingales \( M \) such that

\[
c \int F\left(\|M(T)\|^{1/2}\right) dP \leq \int F(M^*) dP \leq C \int F\left(\|M(T)\|^{1/2}\right) dP
\]

where

\[
M^*(\omega) \equiv \sup \{\|M(t)(\omega)\| : t \in [0, T]\}.
\]

**Proof:** Let \( \{\tau_n\} \) be an increasing localizing sequence for \( M \) such that \( M^{\tau_n} \) is uniformly bounded. Such a localizing sequence exists from Proposition 14.3.3. Then from Theorem 14.4.3 there exist constants \( c, C \) independent of \( \tau_n \) such that

\[
c \int F\left(\|M^{\tau_n}(T)\|^{1/2}\right) dP \leq \int F(M^{\tau_n}) dP \leq C \int F\left(\|M^{\tau_n}(T)\|^{1/2}\right) dP
\]
By Corollary 14.3.3, this implies

\[ c \int_{\Omega} F \left( (|M|^{\tau_n}) (T)^{1/2} \right) \, dP \leq \int_{\Omega} F \left( (M^*) \right) \, dP \leq C \int_{\Omega} F \left( (|M|^{\tau_n}) (T)^{1/2} \right) \, dP \]

and now note that \((|M|^{\tau_n}) (T)^{1/2}\) and \((M^*)\) increase in \(n\) to \([M] (T)^{1/2}\) and \(M^*\) respectively. Then the result follows from the monotone convergence theorem. 

Here is a corollary 14.3.3.

**Corollary 14.4.5** Let \(\{M(t)\}\) be a continuous \(H\) valued local martingale and let \(\varepsilon, \delta \in (0, \infty)\). Then there is a constant \(C\), independent of \(\varepsilon, \delta\) such that

\[ P \left( \left| \sup_{t \in [0,T]} |M(t)| \right| \geq \varepsilon \right) \leq \frac{C}{\varepsilon} E \left( [M]^{1/2} (T) \right) + P \left( [M]^{1/2} (T) > \delta \right) \]

**Proof:** Let the stopping time \(\tau\) be defined by

\[ \tau \equiv \inf \left\{ t > 0 : [M]^{1/2} (t) > \delta \right\} \]

Then

\[ P \left( [M^*] \geq \varepsilon \right) = P \left( [M^*] \geq \varepsilon \cap [\tau = \infty] \right) + P \left( [M^*] \geq \varepsilon \cap [\tau < \infty] \right) \]

On the set where \(\tau = \infty\), \(M^* = M\) and so

\[ P \left( [M^*] \geq \varepsilon \right) \leq \frac{1}{\varepsilon} \int_{\Omega} (M^*)^* \, dP + P \left( [M^*] \geq \varepsilon \cap [M]^{1/2} (T) > \delta \right) \]

By Theorem 14.1.2 and Corollary 14.3.3,

\[ \leq \frac{C}{\varepsilon} \int_{\Omega} [M]^* \, dP + P \left( [M^*] \geq \varepsilon \cap [M]^{1/2} (T) > \delta \right) \]

By Corollary 14.4.5.

\[ = \frac{C}{\varepsilon} \int_{\Omega} ([M]^*) \, dP + P \left( [M^*] \geq \varepsilon \cap [M]^{1/2} (T) > \delta \right) \]

\[ \leq \frac{C}{\varepsilon} \int_{\Omega} [M]^{1/2} (T) \, dP + P \left( [M^*] \geq \varepsilon \cap [M]^{1/2} (T) > \delta \right) \]

\[ \leq \frac{C}{\varepsilon} \int_{\Omega} [M]^{1/2} (T) \, dP + P \left( [M]^{1/2} (T) > \delta \right) \]

**14.5 \(M_1^2 (H)\) Is A Hilbert Space**

The Burkholder Davis Gundy inequality along with the properties of the covariation implies the following amazing proposition.

**Proposition 14.5.1** The space \(M_1^2 (H)\) is a Hilbert space. Here \(H\) is a separable Hilbert space.

**Proof:** We already know from Proposition 14.1.2 that this space is a Banach space. It is only necessary to exhibit an equivalent norm which makes it a Hilbert space. However, you can let \(F(\lambda) = \lambda^2\) in the Burkholder Davis Gundy theorem and obtain for \(M \in M_1^2 (H)\), the two norms

\[ \left( \int_{\Omega} [M] (T) \, dP \right)^{1/2} = \left( \int_{\Omega} [M, M] (T) \, dP \right)^{1/2} \]

and

\[ \left( \int_{\Omega} (M^*)^2 \, dP \right)^{1/2} \]
are equivalent. The first comes from an inner product since from Corollary [14.3.3] \([\cdot, \cdot]\) is bilinear and symmetric and nonnegative. If \([M, M](T) = [M](T) = 0\) in \(L^1(\Omega)\), then from the Burkholder Davis Gundy inequality, \(M^* = 0\) in \(L^2(\Omega)\) and so \(M = 0\). Hence

\[
\int_\Omega [M, N](T) \, dP
\]

is an inner product which yields the equivalent norm.

**Example 14.5.2** An example of a real martingale is the Wiener process, \(W(t)\). It has the property that whenever \(t_1 < t_2 < \cdots < t_n\), the increments \(\{W(t_i) - W(t_{i-1})\}\) are independent and whenever \(s < t\), \(W(t) - W(s)\) is normally distributed with mean 0 and variance \((t - s)\). For the Wiener process, we let

\[
\mathcal{F}_t \equiv \cap_{u>t} (W(s) - W(r) : r < s \leq u)
\]

and it is with respect to this normal filtration that \(W\) is a continuous martingale. What is the quadratic variation of such a process?

The quadratic variation of the Wiener process is just \(t\). This is because if \(A \in \mathcal{F}_s, s < t\),

\[
E \left( X_A \left( |W(t)|^2 - t \right) \right) = E \left( X_A \left( |W(t) - W(s)|^2 + |W(s)|^2 + 2(W(s), W(t) - W(s)) - (t - s + s) \right) \right)
\]

Now

\[
E(\mathcal{X}_A(2(W(s), W(t) - W(s)))) = P(A)E(2W(s))E(W(t) - W(s)) = 0
\]

by the independence of the increments. Thus the above reduces to

\[
E \left( X_A \left( |W(t) - W(s)|^2 + |W(s)|^2 - (t - s + s) \right) \right)
\]

\[
= E \left( X_A \left( |W(t) - W(s)|^2 - (t - s) \right) \right) + E \left( X_A \left( |W(s)|^2 - s \right) \right)
\]

\[
= P(A)E \left( |W(t) - W(s)|^2 - (t - s) \right) + E \left( X_A \left( |W(s)|^2 - s \right) \right)
\]

\[
= E \left( X_A \left( |W(s)|^2 - s \right) \right)
\]

and so \(E(|W(t)|^2 - t|\mathcal{F}_s) = |W(s)|^2 - s\) showing that \(t \to |W(t)|^2 - t\) is a martingale. Hence, by uniqueness, \(|W| \equiv t\).

### 14.6 The Quadratic Variation And Stochastic Integration

Let \(\mathcal{F}_t\) be a normal filtration and let \(\{M(t)\}\) be a continuous local martingale adapted to \(\mathcal{F}_t\) having values in \(U\) a separable real Hilbert space.

**Definition 14.6.1** Let \(\mathcal{F}_t\) be a normal filtration and let

\[
f(t) \equiv \sum_{k=0}^{n-1} f_k \mathcal{X}_{(t_k, t_{k+1})} (t)
\]

where \(\{t_k\}_{k=0}^n\) is a partition of \([0, T]\) and each \(f_k\) is \(\mathcal{F}_{t_k}\) measurable, \(f_k M^* \in L^2(\Omega)\) where

\[
M^* (\omega) \equiv \sup_{t \in [0, T]} ||M(t)(\omega)||
\]

Such a function is called an elementary function. Also let \(\{M(t)\}\) be a local martingale adapted to \(\mathcal{F}_t\) which has values in a separable real Hilbert space, \(U\) such that \(M(0) = 0\). For such an elementary real valued function define

\[
\int_0^t f dM \equiv \sum_{k=0}^{n-1} f_k (M(t \wedge t_{k+1}) - M(t \wedge t_k))
\]
Then with this definition, here is a wonderful lemma.

**Lemma 14.6.2** For \( f \) an elementary function as above, \( \{ \int_0^t f \, dM \} \) is a continuous local martingale and

\[
E \left( \left\| \int_0^t f \, dM \right\|_U^2 \right) = \int_0^t f(s)^2 \, d[M](s) \, dP.
\]

(14.6.13)

If \( N \) is another continuous local martingale adapted to \( \mathcal{F}_t \) and both \( f, g \) are elementary functions such that for each \( k \),

\[
f_k M^*, g_k N^* \in L^2(\Omega),
\]

then

\[
E \left( \left( \int_0^t f \, dM, \int_0^t g \, dN \right)_U \right) = \int_0^t f \, g d[M, N]
\]

(14.6.14)

and both sides make sense.

**Proof:** Let \( \{ \tau_l \} \) be a localizing sequence for \( M \) such that \( M^{\tau_l} \) is a bounded martingale. Then from the definition, for each \( \omega \)

\[
\int_0^t f \, dM = \lim_{l \to \infty} \int_0^t f \, dM^{\tau_l} = \lim_{l \to \infty} \left( \int_0^t f \, dM \right)^{\tau_l}
\]

and it is clear that \( \{ \int_0^t f \, dM^{\tau_l} \} \) is a martingale because it is just the sum of some martingales. Thus \( \{ \tau_l \} \) is a localizing sequence for \( \int_0^t f \, dM \). It is also clear \( \int_0^t f \, dM \) is continuous because it is a finite sum of continuous random variables.

Next consider the formula which is really a version of the Ito isometry. There is no loss of generality in assuming the mesh points are the same for the two elementary functions because if not, one can simply add in points to make this happen. It suffices to consider the special case. To begin with, let \( \{ \tau_l \} \) be a localizing sequence which makes both \( M^{\tau_l} \) and \( N^{\tau_l} \) into bounded martingales. Consider the stopped process.

\[
E \left( \left( \int_0^t f \, dM^{\tau_l}, \int_0^t g \, dN^{\tau_l} \right)_U \right)
\]

\[
= E \left( \left( \sum_{k=0}^{n-1} f_k (M^{\tau_l}(t \wedge t_{k+1}) - M^{\tau_l}(t \wedge t_k)) + \sum_{k=0}^{n-1} g_k (N^{\tau_l}(t \wedge t_{k+1}) - N^{\tau_l}(t \wedge t_k)) \right) \right)
\]

To save on notation, write \( M^{\tau_l}(t \wedge t_{k+1}) - M^{\tau_l}(t \wedge t_k) \equiv \Delta M_k(t) \), similar for \( \Delta N_k \). Thus

\[
\Delta M_k = M^{\tau_l \wedge t_{k+1}} - M^{\tau_l \wedge t_k},
\]

similar for \( \Delta N_k \). Then the above equals

\[
E \left( \sum_{k=0}^{n-1} f_k \Delta M_k, \sum_{k=0}^{n-1} g_k \Delta N_k \right) = E \left( \sum_{k,j} f_k g_j (\Delta M_k, \Delta N_j) \right)
\]

Now consider one of the mixed terms with \( j < k \).

\[
E \left( (f_k \Delta M_k, g_j \Delta N_j) \right) = E \left( (f_k \Delta M_k, g_j \Delta N_j) \mid \mathcal{F}_{t_j} \right)
\]

\[
= E \left( g_j \Delta N_j, f_k E \left( \Delta M_k \mid \mathcal{F}_{t_j} \right) \right) = 0
\]

since \( E \left( \Delta M_k \mid \mathcal{F}_{t_k} \right) = E \left( (M^{\tau_l}(t \wedge t_{k+1}) - M^{\tau_l}(t \wedge t_k)) \mid \mathcal{F}_{t_k} \right) = 0 \) by the Doob optional sampling theorem. Thus

\[
E \left( \left( \int_0^t f \, dM^{\tau_l}, \int_0^t g \, dN^{\tau_l} \right)_U \right) = \int_0^t f \, g \, d[M, N]
\]

(14.6.15)
Letting \( l \to \infty \), this reduces to

\[
\int_{\Omega} \int_0^t fg \, d[M+N] - d[M-N]
\]

Now consider the left side of (14.6.16)

\[
E \left( \int_0^t f dM^\tau, \int_0^t gdN^\tau \right)_U
\]

Therefore, its expectation also equals 0. Consequently the above reduces to

\[
\sum_{k=0}^{n-1} E (f_k g_k (\Delta M_k, \Delta N_k))
\]

where \( \mathcal{N}_k \) is a martingale such that \( \mathcal{N}_k (t) = 0 \) for all \( t \leq t_k \). This is because the martingale \((N^\tau)_{t_{k+1}} - (N^\tau)_{t_k} = \Delta N_k\) equals 0 for such \( t \); and so \( E (\mathcal{N}_k (t)) = 0 \). Thus \( f_k g_k \mathcal{N}_k \) is a martingale which equals zero when \( t = 0 \). Therefore, its expectation also equals 0. Consequently the above reduces to

\[
\sum_{k=0}^{n-1} E (f_k g_k (\Delta M_k, \Delta N_k)).
\]

At this point, recall the definition of the covariation. The above equals

\[
\frac{1}{4} \sum_{k=0}^{n-1} E (f_k g_k ([\Delta M_k + \Delta N_k] - [\Delta M_k - \Delta N_k]))
\]

Rewriting this yields

\[
\frac{1}{4} \sum_{k=0}^{n-1} E \left( f_k g_k \left( (M^\tau)_{t_{k+1}} + (N^\tau)_{t_{k+1}} - (M^\tau)_{t_k} + (N^\tau)_{t_k} \right) \right)
\]

\[
- \left( (M^\tau)_{t_{k+1}} - (N^\tau)_{t_{k+1}} - (M^\tau)_{t_k} + (N^\tau)_{t_k} \right)
\]

To save on notation, denote

\[
(M^\tau)_{t_{k+1}} + (N^\tau)_{t_{k+1}} - (M^\tau)_{t_k} + (N^\tau)_{t_k} \equiv \Delta_k (M^\tau + N^\tau)
\]

\[
(M^\tau)_{t_{k+1}} - (N^\tau)_{t_{k+1}} - (M^\tau)_{t_k} - (N^\tau)_{t_k} \equiv \Delta_k (M^\tau - N^\tau)
\]

Thus the above equals

\[
\frac{1}{4} \sum_{k=0}^{n-1} E \left( f_k g_k (\Delta_k (M^\tau + N^\tau) - \Delta_k (M^\tau - N^\tau)) \right)
\]

Now from Corollary [14.6.10],

\[
\sum_{k=0}^{n-1} E (f_k g_k ([\Delta_k (M + N)]^\tau - [\Delta_k (M - N)]^\tau))
\]

Letting \( l \to \infty \), this reduces to

\[
\frac{1}{4} \sum_{k=0}^{n-1} E (f_k g_k ([\Delta_k (M + N)]^\tau - [\Delta_k (M - N)]^\tau))
\]

Now consider the left side of (14.6.16)

\[
\int_{\Omega} \sum_{k,j} f_k g_j \left( (M^\tau (t \land t_{k+1}) - M^\tau (t \land t_k)) \right.
\]

\[
\left. (N^\tau (t \land t_{j+1}) - N^\tau (t \land t_j)) \right) dP
\]
Then for each $\omega$, the integrand converges as $l \to \infty$ to
\[
\sum_{k,j} f_k g_j ((M (t \wedge t_{k+1}) - M (t \wedge t_k)), (N (t \wedge t_{j+1}) - N (t \wedge t_j)))
\]
But also you can do a sloppy estimate which will allow the use of the dominated convergence theorem.
\[
\left\| \sum_{k,j} f_k g_j (M^\tau_l (t \wedge t_{k+1}) - M^\tau_l (t \wedge t_k)), (N^\tau_l (t \wedge t_{j+1}) - N^\tau_l (t \wedge t_j)) \right\|
\leq \sum_{k,j} |f_k| |g_j| 4M^* N^* \in L^1 (\Omega)
\]
by assumption that the functions are elementary functions. Thus the left side of (14.6.13) converges as $l \to \infty$ to
\[
\int_{\Omega} \left( \int_0^t f dM, \int_0^t g dN \right) _U dP
\]
Note for each $\omega$, the inside integral in (14.6.13) is just a Stieltjes integral taken with respect to the increasing integrating function $[M]$.

Of course, with this estimate it is obvious how to extend the integral to a larger class of functions.

**Definition 14.6.3** Let $\nu (\omega)$ denote the Radon measure representing the functional
\[
\Lambda (\omega) (g) \equiv \int_0^T g d [M] (t) (\omega)
\]
$t \to [M] (t) (\omega)$ is a continuous increasing function and $\nu (\omega)$ is the measure representing the Stieltjes integral, one for each $\omega$.) Then let $G_M$ denote functions $f (s, \omega)$ which are the limit of such elementary functions in the space $L^2 (\Omega; L^2 ([0, T], \nu (\cdot)))$, the norm of such functions being
\[
\|f\|_G^2 \equiv \int_{\Omega} \int_0^t f(s)^2 d [M] (s) dP
\]
For $f \in G$ just defined,
\[
\int_0^t f dM \equiv \lim_{n \to \infty} \int_0^t f_n dM
\]
where $\{f_n\}$ is a sequence of elementary functions converging to $f$ in
\[
L^2 (\Omega; L^2 ([0, T], \nu (\cdot)))
\]
Now here is an interesting lemma.

**Lemma 14.6.4** Let $M, N$ be continuous local martingales, $M (0) = N (0) = 0$ having values in a separable Hilbert space, $U$. Then
\[
[M + N]^{1/2} \leq \left( [M]^{1/2} + [N]^{1/2} \right)
\]
(14.6.17)
\[
[M + N] \leq 2 ([M] + [N])
\]
(14.6.18)
Also, letting $\nu_{M+N}$ denote the measure obtained from the increasing function $[M+N]$ and $\nu_N, \nu_M$ defined similarly,
\[
\nu_{M+N} \leq 2 (\nu_M + \nu_N)
\]
(14.6.19)
on all Borel sets.
\textbf{Proof:} Since \((M,N) \to [M,N]\) is bilinear and satisfies
\[
[M,N] = [N,M] \\
[aM + bM_1, N] = a[M,N] + b[M_1, N] \\
[M,M] \geq 0
\]
which follows from Corollary 14.3.3, the usual Cauchy Schwartz inequality holds and so
\[
||[M,N]|| \leq [M]^{1/2} [N]^{1/2}
\]
Thus
\[
[M + N] = [M + N, M + N] = [M,M] + [N,N] + 2[M,N] \\
\leq [M] + [N] + 2[M]^{1/2} [N]^{1/2} = \left([M]^{1/2} + [N]^{1/2}\right)^2
\]
This proves 14.6.18. Now square both sides. Then the right side is no larger than
\[
2 ([M] + [N])
\]
and this shows 14.6.17.

Now consider the claim about the measures. It was just shown that
\[
[(M + N) - (M + N)^*] \leq 2 ([M - M^*] + [N - N^*])
\]
and from Corollary 14.3.3 this implies that for \(t > s\)
\[
[M + N] (t) - [M + N] (s \wedge t) \\
= [M + N] (t) - [M + N]^* (t) \\
= [M + N - (M^* + N^*)] (t) \\
= [M - M^* + (N - N^*)] (t) \\
\leq 2 [M - M^*] (t) + 2 [N - N^*] (t) \\
\leq 2 ([M] (t) - [M] (s)) + 2 ([N] (t) - [N] (s))
\]
Thus
\[
\nu_{M+N} ([s,t]) \leq 2 (\nu_M ([s,t]) + \nu_N ([s,t]))
\]
By regularity of the measures, this continues to hold with any Borel set \(F\) in place of \([s,t]\). \(\blacksquare\)

\textbf{Theorem 14.6.5} The integral is well defined and has a continuous version which is a local martingale. Furthermore it satisfies the Ito isometry,
\[
E \left( \int_0^t f dM \right)^2 = \int_0^T \int_0^t f_s^2 (s) d[M](s) dP
\]
Let the norm on \(\mathcal{G}_N \cap \mathcal{G}_M\) be the maximum of the norms on \(\mathcal{G}_N\) and \(\mathcal{G}_M\) and denote by \(\mathcal{E}_N\) and \(\mathcal{E}_M\) the elementary functions corresponding to the martingales \(N\) and \(M\) respectively. Define \(\mathcal{G}_{NM}\) as the closure in \(\mathcal{G}_N \cap \mathcal{G}_M\) of \(\mathcal{E}_N \cap \mathcal{E}_M\). Then for \(f,g \in \mathcal{G}_{NM}\),
\[
E \left( \left( \int_0^t f dM, \int_0^t g dN \right) \right) = \int_0^T \int_0^t f(t) g(t) d[M,N]. \tag{14.6.20}
\]

\textbf{Proof:} It is clear the definition is well defined because if \(\{f_n\}\) and \(\{g_n\}\) are two sequences of elementary functions converging to \(f\) in \(L^2 (\Omega; L^2 ([0,T], \nu (\cdot)))\) and if \(\int_0^T f dM\) is the integral which comes from \(\{g_n\}\),
\[
\int_\Omega \left\| \int_0^T f dM - \int_0^T f_n dM \right\|^2 dP \leq \lim_{n \to \infty} \int_\Omega \left\| \int_0^T g_n dM - \int_0^T f_n dM \right\|^2 dP = 0
\]
Consider the claim the integral has a continuous version. Recall Theorem 14.6.18, part of which is listed here for convenience.
Theorem 14.6.6 Let \( \{X(t)\} \) be a right continuous nonnegative submartingale adapted to the normal filtration \( \mathcal{F}_t \) for \( t \in [0,T] \). Let \( p \geq 1 \). Define

\[
X^* (t) \equiv \sup \{ X(s) : 0 < s < t \}, \quad X^* (0) \equiv 0.
\]

Then for \( \lambda > 0 \)

\[
P (\{X^*(T) > \lambda\}) \leq \frac{1}{\lambda^p} \int_\Omega X(T)^p \, dP \tag{14.6.21}
\]

Let \( \{f_n\} \) be a sequence of elementary functions converging to \( f \) in

\[
L^2 (\Omega; L^2 ([0,T], \nu (\cdot))).
\]

Then letting

\[
X^\tau_{l,n,m} (t) = \left\| \int_0^t (f_n - f_m) \, dM^\tau_l \right\|_U,
\]

\[
X_{n,m} (t) = \left\| \int_0^t (f_n - f_m) \, dM \right\|_U = \left\| \int_0^t f_n \, dM - \int_0^t f_m \, dM \right\|_U.
\]

It follows \( X^\tau_{l,n,m} \) is a continuous nonnegative submartingale and from Theorem 14.6.4 just listed,

\[
P (\{X^\tau_{l,n,m} (T) > \lambda\}) \leq \frac{1}{\lambda^2} \int_\Omega X^\tau_{l,n,m} (T)^2 \, dP
\]

\[
\leq \frac{1}{\lambda^2} \int_\Omega \int_0^T |f_n - f_m|^2 \, d[M^\tau_l] \, dP
\]

\[
\leq \frac{1}{\lambda^2} \int_\Omega \int_0^T |f_n - f_m|^2 \, d[M] \, dP
\]

Letting \( l \rightarrow \infty \),

\[
P (\{X^*_{n,m} (T) > \lambda\}) \leq \frac{1}{\lambda^2} \int_\Omega \int_0^T |f_n - f_m|^2 \, d[M] \, dP
\]

Therefore, there exists a subsequence, still denoted by \( \{f_n\} \) such that

\[
P (\{X^*_{n,n+1} (T) > 2^{-n}\}) < 2^{-n}
\]

Then by the Borel Cantelli lemma, the \( \omega \) in infinitely many of the sets

\[
[X^*_{n,n+1} (T) > 2^{-n}]
\]

has measure 0. Denoting this exceptional set as \( N \), it follows that for \( \omega \notin N \), there exists \( n (\omega) \) such that for \( n > n (\omega) \),

\[
\sup_{t \in [0,T]} \left\| \int_0^t f_n \, dM - \int_0^t f_{n+1} \, dM \right\| \leq 2^{-n}
\]

and this implies uniform convergence of \( \{\int_0^t f_n \, dM\} \). Letting

\[
G(t) = \lim_{n \rightarrow \infty} \int_0^t f_n \, dM,
\]

for \( \omega \notin N \) and \( G(t) = 0 \) for \( \omega \in N \), it follows that for each \( t \), the continuous adapted process \( G(t) \) equals \( \int_0^t f \, dM \) a.e. Thus \( \{\int_0^t f \, dM\} \) has a continuous version.
14.7. ANOTHER LIMIT FOR QUADRATIC VARIATION

It suffices to verify \[14.6.20\]. Let \( \{ f_n \} \) and \( \{ g_n \} \) be sequences of elementary functions converging to \( f \) and \( g \) in \( G_M \cap G_N \). By Lemma \[14.6.2\],

\[
E \left( \left( \int_0^t f_n dM, \int_0^t g_n dN \right)_U \right) = \int_\Omega \int_0^t f_n g_n d[M,N]
\]

Then by the Holder inequality and the above definition,

\[
\lim_{n \to \infty} E \left( \left( \int_0^t f_n dM, \int_0^t g_n dN \right)_U \right) = E \left( \left( \int_0^t f dM, \int_0^t g dN \right)_U \right)
\]

Consider the right side which equals

\[
\frac{1}{4} \int_\Omega \int_0^t f_n g_n d [M + N] dP - \frac{1}{4} \int_\Omega \int_0^t f_n g_n d [M - N] dP
\]

Now from Lemma \[14.6.4\],

\[
\left| \int_\Omega \int_0^t f_n g_n d [M + N] dP - \int_\Omega \int_0^t f g d [M + N] dP \right| = \left| \int_\Omega \int_0^t f_n g_n d\nu_{M+N} dP - \int_\Omega \int_0^t f g d\nu_{M+N} dP \right|
\]

\[
\leq 2 \left( \int_\Omega \int_0^t |f_n g_n - f g| d\nu_M dP + \int_\Omega \int_0^t |f_n g_n - f g| d\nu_N dP \right)
\]

and by the choice of the \( f_n \) and \( g_n \), these both converge to 0. Similar considerations apply to

\[
\left| \int_\Omega \int_0^t f_n g_n d [M - N] dP - \int_\Omega \int_0^t f g d [M - N] dP \right|
\]

and show

\[
\lim_{n \to \infty} \int_\Omega \int_0^t f_n g_n d [M,N] = \int_\Omega \int_0^t f g d [M,N] \tag*{\blacksquare}
\]

14.7 Another Limit For Quadratic Variation

The problem to consider first is to define an integral

\[
\int_0^t f dM
\]

where \( f \) has values in \( H' \) and \( M \) is a continuous martingale having values in \( H \). For the sake of simplicity assume \( M(0) = 0 \). The process of definition is the same as before. First consider an elementary function

\[
f(t) = \sum_{k=0}^{m-1} f_k \mathcal{X}_{[t_k, t_{k+1})}(t) \tag{14.7.22}
\]

where \( f_k \) is measurable into \( H' \) with respect to \( \mathcal{F}_{t_k} \). Then define

\[
\int_0^t f dM = \sum_{k=0}^{m-1} f_k (M(t \wedge t_{k+1}) - M(t \wedge t_k)) \in \mathbb{R} \tag{14.7.23}
\]

Lemma 14.7.1 The \( k^{th} \) term in the above sum is a martingale and the integral is also a martingale.

Proof: Let \( \sigma \) be a stopping time with two values. Then

\[
E (f_k (M(\sigma \wedge t_{k+1}) - M(\sigma \wedge t_k))) = E(E(f_k (M(\sigma \wedge t_{k+1}) - M(\sigma \wedge t_k))|\mathcal{F}_{t_k}))
\]

\[
= E(f_k E((M(\sigma \wedge t_{k+1}) - M(\sigma \wedge t_k))|\mathcal{F}_{t_k})) = 0
\]
Now what would it take for a mixed term. For \( j < k \), it follows from measurability considerations that

\[
E ((f_k (M (t \wedge t_{k+1}) - M (t \wedge t_k)) (f_j (M (t \wedge t_{j+1}) - M (t \wedge t_j))))
\]

\[
= E (E [(f_k (M (t \wedge t_{k+1}) - M (t \wedge t_k)) (f_j (M (t \wedge t_{j+1}) - M (t \wedge t_j))) | \mathcal{F}_{t_k}])
\]

\[
= E ((f_j (M (t \wedge t_{j+1}) - M (t \wedge t_j)) f_k E [(M (t \wedge t_{k+1}) - M (t \wedge t_k)) | \mathcal{F}_{t_k}]) = 0
\]

Therefore,

\[
E \left( \left| \int_0^t f dM \right|^2 \right) = E \left( \sum_{k=0}^{m-1} |f_k (M (t \wedge t_{k+1}) - M (t \wedge t_k))^2 \right)
\]

\[
\leq E \left( \sum_{k=0}^{m-1} ||f_k||^2 |M (t \wedge t_{k+1}) - M (t \wedge t_k)|^2 \right)
\]

\[
= E \left( \sum_{k=0}^{m-1} ||f_k||^2 (|M^{t_{k+1}} - M^{t_k}| (t) + N_k (t)) \right)
\]

\[
= E \left( \sum_{k=0}^{m-1} ||f_k||^2 (|M^{t_{k+1}}| (t) - |M^{t_k}| (t) + N_k (t)) \right)
\]

\[
= E \left( \sum_{k=0}^{m-1} ||f_k||^2 (|M| (t \wedge t_{k+1}) - |M| (t \wedge t_k) + N_k (t)) \right)
\]

where \( N_k \) is a martingale which equals 0 for \( t \leq t_k \). The above equals

\[
E \left( \int_0^t ||f||^2 d[M] \right) = E \left( \int_0^t ||f||^2 d\nu \right)
\]

the integral inside being the ordinary Lebesgue Stieltjes integral for the step function where \( \nu \) is the measure determined by the positive linear functional

\[
\Lambda g = \int_0^T g d [M]
\]

where the integral on the right is the ordinary Stieltjes integral. Thus, the following inequality is obtained.

\[
E \left( \left| \int_0^t f dM \right|^2 \right) \leq E \left( \int_0^t ||f||^2 d |M|, \right) \tag{14.7.24}
\]

Now what would it take for

\[
E \left( \left| \int_0^t f dM \right|^2 \right) \tag{14.7.25}
\]

to be well defined? A convenient condition would be to insist that each \( ||f_k|| M^f \) is in \( L^2 (\Omega) \) where

\[
M^f (\omega) = \sup_{t \in [0,T]} |M (t) (\omega)|_H
\]
Is this condition also sufficient for the above integral to be finite? From the above, that integral equals

\[ E \left( \sum_{k=0}^{m-1} \|f_k\|^{2} \left| M(t \wedge t_{k+1}) - M(t \wedge t_k) \right|^2 \right) \]

\[ \leq E \left( 4 \sum_{k=0}^{m-1} \|f_k\|^{2} (M_k^*)^2 \right) \]

Thus the condition that for each \( k, \|f_k\| M^* \in L^2(\Omega) \) is sufficient for all of the above to consist of real numbers and be well defined.

**Definition 14.7.2** A function \( f \) is called an elementary function if it is a step function of the form given in (14.7.22) where each \( f_k \) is \( F_{t_k} \) measurable and for each \( k, \|f_k\| M^* \in L^2(\Omega) \). Define \( G_M \) to be the collection of functions \( f \) having values in \( H' \) which have the property that there exists a sequence of elementary functions \( \{f_n\} \) with \( f_n \to f \) in the space

\[ L^2(\Omega; L^2([0,T],\nu)) \]

Then picking such an approximating sequence,

\[ \int_0^t f dM \equiv \lim_{n \to \infty} \int_0^t f_n dM \]

the convergence happening in \( L^2(\Omega) \).

The inequality (14.7.22) shows that this definition is well defined. So what are the properties of the integral just defined? Each \( \int_0^t f_n dM \) is a continuous martingale because it is the sum of continuous martingales. Since convergence happens in \( L^2(\Omega) \), it follows that \( \int_0^t f dM \) is also a martingale. Is it continuous? By the maximal inequality Theorem (14.3,13), it follows that

\[ P \left( \sup_{t \in [0,T]} \left| \int_0^t f_m dM - \int_0^t f_n dM \right| > \lambda \right) \leq \frac{1}{\lambda^2} E \left( \int_0^T (f_m - f_n) dM \right)^2 \]

\[ \leq \frac{1}{\lambda^2} E \left( \int_0^T \|f_m - f_n\|^2 d[M] \right) \]

and it follows that there exists a subsequence, still called \( n \) such that for all \( p \) positive,

\[ P \left( \sup_{t \in [0,T]} \left| \int_0^t f_{n+p} dM - \int_0^t f_n dM \right| > \frac{1}{n} \right) < 2^{-n} \]

By the Borel Cantelli lemma, there exists a set of measure zero \( N \) such that for \( \omega \notin N \), \( \left\{ \int_0^t f_n dM \right\} \) is a Cauchy sequence. Thus, what it converges to is continuous in \( t \) for each \( \omega \notin N \) and for each \( t \), it equals \( \int_0^t f dM \) a.e. Hence we can regard \( \int_0^t f dM \) as this continuous version.

What is an example of such a function in \( G_M \)?

**Lemma 14.7.3** Let \( R : H \to H' \) be the Riesz map.

\[ \langle Rf, g \rangle \equiv (f, g)_H \cdot \]

Also suppose \( M \) is a uniformly bounded continuous martingale with values in \( H \). Then \( RM \in G_M \).

**Proof:** I need to exhibit an approximating sequence of elementary functions as described above. Consider

\[ M_n(t) \equiv \sum_{i=0}^{m_n-1} M(t_i) \chi_{(t_i, t_{i+1}]}(t) \]

Then clearly \( RM_n(t) \ M^* \in L^\infty(\Omega) \) and so in particular it is in \( L^2(\Omega) \). Here

\[ \lim_{n \to \infty} \max \left\{ |t^n_i - t^n_{i+1}|, i = 0, \cdots, m_n \right\} = 0. \]
Say $M^*(\omega) \leq C$. Furthermore, I claim that
\[
\lim_{n \to \infty} E \left( \int_0^T \|RM_n - RM\|^2 \, d[M] \right) = 0. \tag{14.7.26}
\]
This requires a little proof. Recall the description of $[M](t)$. It was as follows. You considered
\[
P_n(t) \equiv 2 \sum_{k \geq 0} \left( (M(t \wedge \tau^0_{k+1}) - M(t \wedge \tau^0_k), M(t \wedge \tau^0_k) \right)
\]
where the stopping times were defined such that $\tau^0_{k+1}$ is the first time $t > \tau^0_k$ such that $|M(t) - M(\tau^0_k)|^2 = 2^{-n}$ and $\tau^0_0 = 0$. Recall that $\lim_{k \to \infty} \tau^0_k = \infty$ or $T$ in the way it was formulated earlier. Then it was shown that $P_n(t)$ converged to a martingale $P(t)$ in $L^1(\Omega)$. Then by the usual procedure using the Borel Cantelli lemma, a subsequence converges to $P(t)$ uniformly off a set of measure zero. It is easy to estimate $P_n(t)$.
\[
|P_n(t)| \leq \sum_{k \geq 0} |M(t \wedge \tau^0_{k+1})| - |M(t \wedge \tau^0_k)|^2 = |M(t)|^2 \leq M^*
\]
This follows from the observation that
\[
(M(t \wedge \tau^0_{k+1}), M(t \wedge \tau^0_k)) \leq \frac{1}{2} \left( |M(t \wedge \tau^0_{k+1})|^2 + |M(t \wedge \tau^0_k)|^2 \right)
\]
Then it follows that $\sup_{t \in [0,T]} |P(t)(\omega)| \leq M^*(\omega) \leq C$ for a.e. $\omega$. The quadratic variation $[M]$ was defined as
\[
|M(t)|^2 = P(t) + [M](t)
\]
Thus $[M](t) \leq 2(M^*)^2$. Now consider the above limit in $(14.7.26)$. From the assumption that $M$ is uniformly bounded,
\[
\int_0^T \|RM_n - RM\|^2 \, d[M] \leq \int_0^T 4C^2 \, d[M] = 4C^2 \, [M](T) \leq 4C^2 \, (2C^2) < \infty
\]
Also, by the continuity of the martingale, for each $\omega$,
\[
\lim_{n \to \infty} \|RM_n - RM\|^2 = 0
\]
By the dominated convergence theorem, and the fact that the integrand is bounded,
\[
\lim_{n \to \infty} \int_0^T \|RM_n - RM\|^2 \, d[M] = 0.
\]
Then from the above estimate and the dominated convergence theorem again, $(14.7.26)$ follows. Thus $RM \in \mathcal{G}_M$. ■

From the above lemma, it makes sense to speak of
\[
\int_0^t (RM) \, dM
\]
and this is a continuous martingale having values in $\mathbb{R}$. Also from the above argument, if $\{t^n_i\}_{k=0}^{m_n}$ is a sequence of partitions such that
\[
\lim_{n \to \infty} \max \{|t^n_i - t^n_{i+1}|, i = 0, \cdots, m_n\} = 0,
\]
then it follows that
\[
\sum_{i=0}^{m_n-1} RM(t_i) (M(t \wedge t_{i+1}) - M(t \wedge t_i)) \to \int_0^t (RM) \, dM
\]
in $L^2(\Omega)$, this for each $t \in [0, T]$.

Now here is the main result.
14.7. ANOTHER LIMIT FOR QUADRATIC VARIATION

**Theorem 14.7.4** Let $H$ be a Hilbert space and suppose $(M, \mathcal{F}_t), t \in [0, T]$ is a uniformly bounded continuous martingale with values in $H$. Also let $\{t^n_k\}_{k=1}^{m_n}$ be a sequence of partitions satisfying

$$
\lim_{n \to \infty} \max_i \{|t^n_i - t^n_{i+1}|, i = 0, \ldots, m_n\} = 0, \quad \{t^n_k\}_{k=1}^{m_n} \subseteq \{t^m_{k+1}\}_{k=1}^{m_{n+1}}.
$$

Then

$$
[M](t) = \lim_{n \to \infty} \sum_{k=0}^{m_{n-1}} \left|M(t \wedge t^n_{k+1}) - M(t \wedge t^n_k)\right|^2_H
$$

the limit taking place in $L^2(\Omega)$. In case $M$ is just a continuous local martingale, the above limit happens in probability.

**Proof:** First suppose $M$ is uniformly bounded.

$$
= \sum_{k=0}^{m_{n-1}} \left|M(t \wedge t^n_{k+1})\right|^2 - \left|M(t \wedge t^n_k)\right|^2 - 2 \sum_{k=0}^{m_{n-1}} (M(t \wedge t^n_k), M(t \wedge t^n_{k+1}) - M(t \wedge t^n_k))
$$

Then by Lemma [lem:lemma], the right side converges to

$$
|M(t)|_H^2 - 2 \int_0^t (RM) dM
$$

Therefore, in $L^2(\Omega)$,

$$
\lim_{n \to \infty} \sum_{k=0}^{m_{n-1}} \left|M(t \wedge t^n_{k+1}) - M(t \wedge t^n_k)\right|^2_H + 2 \int_0^t (RM) dM = |M(t)|_H^2
$$

That term on the left involving the limit is increasing and equal to 0 when $t = 0$. Therefore, it must equal $[M](t)$.

Next suppose $M$ is only a continuous local martingale. By Proposition [prop:prop], there exists an increasing localizing sequence $\{\tau_k\}$ such that $M^{\tau_k}$ is a uniformly bounded martingale. Then

$$
P(\bigcup_{k=1}^{\infty} [\tau_k = \infty]) = 1
$$

To save notation, let

$$
Q_n(t) \equiv \sum_{k=0}^{m_{n-1}} \left|M(t \wedge t^n_{k+1}) - M(t \wedge t^n_k)\right|^2_H
$$

Let $\eta, \varepsilon > 0$ be given. Then there exists $k$ large enough that $P([\tau_k = \infty]) > 1 - \eta/2$. This is because the sets $[\tau_k = \infty]$ increase to $\Omega$ other than a set of measure zero. Then,

$$
||Q_n^{\tau_k} - [M]^{\tau_k}(t)| > \varepsilon| \cap [\tau_k = \infty] = ||Q_n - [M](t)| > \varepsilon| \cap [\tau_k = \infty]
$$

Thus

$$
P(\{|Q_n - [M](t)| > \varepsilon\}) \leq P(\{|Q_n - [M](t)| > \varepsilon| \cap [\tau_k = \infty]\}) + P([\tau_k < \infty])
$$

$$
\leq P(\{|Q_n^{\tau_k} - [M]^{\tau_k}(t)| > \varepsilon\}) + \eta/2
$$

From the first part, the convergence in probability of $Q_n^{\tau_k}(t)$ to $[M]^{\tau_k}(t)$ follows from the convergence in $L^2(\Omega)$ and so if $n$ is large enough, the right side of the above inequality is less than $\eta/2 + \eta/2 = \eta$. Since $\eta$ was arbitrary, this proves convergence in probability.
Chapter 15

Gaussian Measures

15.1 Definitions And Basic Properties

First suppose $X$ is a random vector having values in $\mathbb{R}^n$ and its distribution function is $N(m, \Sigma)$ where $m$ is the mean and $\Sigma$ is the covariance. Then the characteristic function of $X$ or equivalently, the characteristic function of its distribution is

$$e^{it \cdot m - \frac{1}{2} t^* \Sigma t}$$

What is the distribution of $a \cdot X$ where $a \in \mathbb{R}^n$? In other words, if you take a linear functional and do it to $X$ to get a scalar valued random variable, what is the distribution of this scalar valued random variable? Let $Y = a \cdot X$. Then

$$E(e^{itY}) = E(e^{i t a \cdot X})$$

which from the above formula is

$$e^{i a \cdot m t - \frac{1}{2} a^* \Sigma a t^2}$$

which is the characteristic function of a random variable whose distribution is $N(a \cdot m, a^* \Sigma a)$. In other words, it is normally distributed having mean equal to $a \cdot m$ and variance equal to $a^* \Sigma a$. Obviously such a concept generalizes to a Banach space in place of $\mathbb{R}^n$ and this motivates the following definition.

**Definition 15.1.1** Let $E$ be a real separable Banach space. A probability measure, $\mu$ defined on $B(E)$ is called a Gaussian measure if for every $h \in E'$, the law of $h$ considered as a random variable defined on the probability space, $(E, B(E), \mu)$ is normal. That is, for $A \subseteq \mathbb{R}$ a Borel set,

$$\lambda_h (A) \equiv \mu (h^{-1} (A))$$

is given by

$$\int_A \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2\sigma^2} (x-m)^2} dx$$

for some $\sigma$ and $m$. A Gaussian measure is called symmetric if $m$ is always equal to 0.

There is another definition of symmetric. First here are a few simple conventions. For $f \in E'$, $x \rightarrow f(x)$ is normally distributed. In particular,

$$\int_E |f(x)| \, d\mu < \infty$$

and so it makes sense to define

$$m_\mu (f) \equiv \int_E f(x) \, d\mu.$$ 

Thus $m_\mu (f)$ is the mean of the random variable $x \rightarrow f(x)$. It is obvious that $f \rightarrow m_\mu (f)$ is linear. Also define the variance $\sigma^2 (f)$ by

$$\sigma^2 (f) \equiv \int_E (f(x) - m_\mu (f))^2 \, d\mu$$

This is finite because $x \rightarrow f(x)$ is normally distributed. The following lemma gives such an equivalent condition for $\mu$ to be symmetric.
Lemma 15.1.2 Let \( \mu \) be a Gaussian measure defined on \( \mathcal{B}(E) \). Then \( \mu(F) = \mu(-F) \) for all \( F \) Borel if and only if \( m_\mu(f) = 0 \) for all \( f \in E' \). Such a Gaussian measure is called symmetric.

Proof: Suppose first \( m_\mu(f) = 0 \) for all \( f \in E' \). Let
\[
G \equiv f_1^{-1}(F_1) \cap f_2^{-1}(F_2) \cap \cdots \cap f_m^{-1}(F_m)
\]
where \( F_i \) is a Borel set of \( \mathbb{R} \) and each \( f_i \in E' \). Since every linear combination of the \( f_i \) is in \( E' \), every such linear combination is normally distributed and so \( f \equiv (f_1, \ldots, f_m) \) is multivariate normal. That is, \( \lambda_f \) the distribution measure, is multivariate normal. Since each \( m_\mu(f) = 0 \), it follows
\[
\mu(G) = \lambda_f \left( \prod_{i=1}^m F_i \right) = \lambda_f \left( \prod_{i=1}^m -F_i \right) = \mu(-G)
\]
(15.1.1)

By Lemma 9.1.4 there exists a countable subset, \( D \equiv \{f_k\}_{k=1}^\infty \) of the closed unit ball such that for every \( x \in E \),
\[
||x|| = \sup_{f \in D} |f(x)|.
\]
Therefore, letting \( D(a, r) \) denote the closed ball centered at \( a \) having radius \( r \), it follows
\[
D(a, r) = \cap_{k=1}^\infty f_k^{-1}(D(f_k(a), r))
\]
Let
\[
D_n(a, r) = \cap_{k=1}^n f_k^{-1}(D(f_k(a), r))
\]
Then by Lemma 15.1.2
\[
\mu(D_n(a, r)) = \mu(-D_n(a, r))
\]
and letting \( n \to \infty \), it follows
\[
\mu(D(a, r)) = \mu(-D(a, r))
\]
Therefore the same is true with \( D(a, r) \) replaced with an open ball. Now consider
\[
D(a, r_1) \cap D(b, r_2) = \cap_{k=1}^\infty f_k^{-1}(D(f_k(a), r_1)) \cap \cap_{k=1}^\infty f_k^{-1}(D(f_k(b), r_2))
\]
The intersection of these two closed balls is the intersection of sets of the form
\[
\cap_{k=1}^n f_k^{-1}(D(f_k(a), r_1)) \cap \cap_{k=1}^n f_k^{-1}(D(f_k(b), r_2))
\]
to which Lemma 15.1.2 applies. Therefore, by continuing this way it follows that if \( G \) is any finite intersection of closed balls,
\[
\mu(G) = \mu(-G).
\]
Let \( \mathcal{K} \) denote the set of finite intersections of closed balls, a \( \pi \) system. Thus for \( G \in \mathcal{K} \) the above holds. Now let
\[
\mathcal{G} \equiv \{ F \in \sigma(\mathcal{K}) : \mu(F) = \mu(-F) \}
\]
Thus \( \mathcal{G} \) contains \( \mathcal{K} \) and it is clearly closed with respect to complements and countable disjoint unions. By the \( \pi \) system lemma, \( \mathcal{G} \supseteq \sigma(\mathcal{K}) \) but \( \sigma(\mathcal{K}) \) clearly contains the open sets since every open ball is the countable union of closed disks and every open set is the countable union of open balls. Therefore, \( \mu(G) = \mu(-G) \) for all Borel \( G \).

Conversely suppose \( \mu(G) = \mu(-G) \) for all \( G \) Borel. If for some \( f \in E' \), \( m_\mu(f) \neq 0 \), then
\[
\mu(f^{-1}(0, \infty)) = \lambda_f(0, \infty) \neq \lambda_f(-\infty, 0) = \mu(f^{-1}(-\infty, 0)) = \mu(-f^{-1}(0, \infty))
\]
a contradiction. This proves the lemma.

Lemma 15.1.3 Let \( \mu = \mathcal{L}(X) \) where \( X \) is a random variable defined on a probability space, \( (\Omega, \mathcal{F}, P) \) which has values in \( E \), a Banach space. Suppose also that for all \( \phi \in E' \), \( \phi \circ X \) is normally distributed. Then \( \mu \) is a Gaussian measure. Conversely, suppose \( \mu \) is a Gaussian measure on \( \mathcal{B}(E) \) and \( X \) is a random variable having values in \( E \) such that \( \mathcal{L}(X) = \mu \). Then for every \( h \in E' \), \( h \circ X \) is normally distributed.
15.2 Fernique’s Theorem

The following is an interesting lemma.

**Proof:** First suppose \( \mu \) is a Gaussian measure and \( X \) is a random variable such that \( \mathcal{L}(X) = \mu \). Then if \( F \) is a Borel set in \( \mathbb{R} \), and \( h \in E' \)

\[
P \left( (h \circ X)^{-1}(F) \right) = P \left( X^{-1}(h^{-1}(F)) \right) = \mu \left( h^{-1}(F) \right) = \frac{1}{\sqrt{2\pi\sigma}} \int_F e^{-\frac{|x-m|^2}{2\sigma^2}} \, dx
\]

for some \( m \) and \( \sigma^2 \) showing that \( h \circ X \) is normally distributed.

Next suppose \( h \circ X \) is normally distributed whenever \( h \in E' \) and \( \mathcal{L}(X) = \mu \). Then letting \( F \) be a Borel set in \( \mathbb{R} \), I need to verify

\[
\mu \left( h^{-1}(F) \right) = \frac{1}{\sqrt{2\pi\sigma}} \int_F e^{-\frac{|x-m|^2}{2\sigma^2}} \, dx.
\]

However, this is easy because

\[
\mu \left( h^{-1}(F) \right) = P \left( X^{-1}(h^{-1}(F)) \right) = P \left( (h \circ X)^{-1}(F) \right)
\]

which is given to equal

\[
\frac{1}{\sqrt{2\pi\sigma}} \int_F e^{-\frac{|x-m|^2}{2\sigma^2}} \, dx
\]

for some \( m \) and \( \sigma^2 \). This proves the lemma.

Here is another important observation. Suppose \( X \) is as just described, a random variable having values in \( E \) such that \( \mathcal{L}(X) = \mu \) and suppose \( h_1, \ldots, h_n \) are each in \( E' \). Then for scalars, \( t_1, \ldots, t_n \),

\[
t_1 h_1 \circ X + \cdots + t_n h_n \circ X = (t_1 h_1 + \cdots + t_n h_n) \circ X
\]

and this last is assumed to be normally distributed because \( (t_1 h_1 + \cdots + t_n h_n) \in E' \). Therefore, by Theorem 11.8.3

\[(h_1 \circ X, \ldots, h_n \circ X)\]

is distributed as a multivariate normal.

Obviously there exist examples of Gaussian measures defined on \( E \), a Banach space. Here is why. Let \( \xi \) be a random variable defined on a probability space, \((\Omega, \mathcal{F}, P)\) which is normally distributed with mean 0 and variance \( \sigma^2 \). Then let \( X(\omega) \equiv \xi(\omega) e \) where \( e \in E \). Then let \( \mu \equiv \mathcal{L}(X) \). For \( A \) a Borel set of \( \mathbb{R} \) and \( h \in E' \),

\[
\mu \left( [h(x) \in A] \right) = P \left( [X(\omega) \in [x : h(x) \in A]] \right) = P \left( [h \circ X \in A] \right) = P \left( [\xi(\omega) h(e) \in A] \right) = \frac{1}{|h(e)| \sigma \sqrt{2\pi}} \int_A e^{-\frac{|x|^2}{2|h(e)|^2 \sigma^2}} \, dx
\]

because \( h(e) \xi \) is a random variable which has variance \( |h(e)|^2 \sigma^2 \) and mean 0. Thus \( \mu \) is indeed a Gaussian measure. Similarly, one can consider finite sums of the form

\[
\sum_{i=1}^n \xi_i(\omega) e_i
\]

where the \( \xi_i \) are independent normal random variables having mean 0 for convenience. However, this is a rather trivial case.

15.2 Fernique’s Theorem

The following is an interesting lemma.
Lemma 15.2.1 Suppose \( \mu \) is a symmetric Gaussian measure on the real separable Banach space, \( E \). Then there exists a probability space, \((\Omega, \mathcal{F}, P)\) and independent random variables, \( X \) and \( Y \) mapping \( \Omega \) to \( E \) such that \( \mathcal{L}(X) = \mathcal{L}(Y) = \mu \). Also, the two random variables, 

\[
\frac{1}{\sqrt{2}}(X - Y), \frac{1}{\sqrt{2}}(X + Y)
\]

are independent and 

\[
\mathcal{L}\left( \frac{1}{\sqrt{2}}(X - Y) \right) = \mathcal{L}\left( \frac{1}{\sqrt{2}}(X + Y) \right) = \mu.
\]

Proof: Letting \( X' = \frac{1}{\sqrt{2}}(X + Y) \) and \( Y' = \frac{1}{\sqrt{2}}(X - Y) \), it follows from Theorem [11.5.2] on Page [189] that \( X' \) and \( Y' \) are independent if whenever \( h_1, \ldots, h_m \in E' \) and \( g_1, \ldots, g_k \in E' \), the two random vectors, 

\[
(h_1 \circ X', \ldots, h_m \circ X') \quad \text{and} \quad (g_1 \circ Y', \ldots, g_k \circ Y')
\]

are independent. Now consider linear combinations

\[
\sum_{j=1}^{m} t_j h_j \circ X' + \sum_{i=1}^{k} s_i g_i \circ Y'.
\]

This equals

\[
\frac{1}{\sqrt{2}} \sum_{j=1}^{m} t_j h_j (X) + \frac{1}{\sqrt{2}} \sum_{j=1}^{m} t_j h_j (Y)
\]

\[+ \frac{1}{\sqrt{2}} \sum_{i=1}^{k} s_i g_i (X) - \frac{1}{\sqrt{2}} \sum_{i=1}^{k} s_i g_i (Y)
\]

\[
= \frac{1}{\sqrt{2}} \left( \sum_{j=1}^{m} t_j h_j + \sum_{i=1}^{k} s_i g_i \right) (X)
\]

\[+ \frac{1}{\sqrt{2}} \left( \sum_{j=1}^{m} t_j h_j - \sum_{i=1}^{k} s_i g_i \right) (Y)
\]

and this is the sum of two independent normally distributed random variables so it is also normally distributed. Therefore, by Theorem [11.8.3]

\[
(h_1 \circ X', \ldots, h_m \circ X', g_1 \circ Y', \ldots, g_k \circ Y')
\]

is a random variable with multivariate normal distribution and by Theorem [11.8.7] the two random vectors

\[
(h_1 \circ X', \ldots, h_m \circ X') \quad \text{and} \quad (g_1 \circ Y', \ldots, g_k \circ Y')
\]

are independent if

\[
E ((h_i \circ X') (g_j \circ Y')) = 0
\]

for all \( i, j \). This is what I will show next.

\[
E ((h_i \circ X') (g_j \circ Y'))
\]

\[= \frac{1}{4} E ((h_i (X) + h_i (Y)) (g_j (X) - g_j (Y)))
\]

\[= \frac{1}{4} E (h_i (X) g_j (X)) - \frac{1}{4} E (h_i (X) g_j (Y))
\]

\[+ \frac{1}{4} E (h_i (Y) g_j (X)) - \frac{1}{4} E (h_i (Y) g_j (Y))
\]

(15.2.2)
Now from the above observation after the definition of Gaussian measure $h_i(X)g_j(X)$ and $h_i(Y)g_j(Y)$ are both in $L^1$ because each term in each product is normally distributed. Therefore, by Lemma 15.2.2, $\lambda$

$$E(h_i(X)g_j(X)) = \int_{\Omega} h_i(Y)g_j(Y)\,dP$$

$$= \int_E h_i(y)g_j(y)\,d\mu$$

$$= \int_{\Omega} h_i(X)g_j(X)\,dP$$

$$= E(h_i(Y)g_j(Y))$$

and so 15.2.2 reduces to $\frac{1}{4}(E(h_i(Y)g_j(X) - h_i(X)g_j(Y))) = 0$

because $h_i(X)$ and $g_j(Y)$ are independent due to the assumption that $X$ and $Y$ are independent. Thus

$$E(h_i(X)g_j(Y)) = E(h_i(X))E(g_j(Y)) = 0$$

due to the assumption that $\mu$ is symmetric which implies the mean of these random variables equals 0. The other term works out similarly. This has proved the independence of the random variables, $X'$ and $Y'$.

Next consider the claim they have the same law and it equals $\mu$. To do this, I will use Theorem 15.2.4 on Page 328. Thus I need to show

$$E\left(e^{ih(X')}\right) = E\left(e^{ih(Y')}\right) = E\left(e^{ih(X)}\right)$$

(15.2.3)

for all $h \in E'$. Pick such an $h$. Then $h \circ X$ is normally distributed and has mean 0. Therefore, for some $\sigma$,

$$E\left(e^{ithoX}\right) = e^{-\frac{1}{2}t^2\sigma^2}.$$

Now since $X$ and $Y$ are independent,

$$E\left(e^{ithoX'}\right) = E\left(e^{ith\left(\frac{1}{\sqrt{2}}\right)(X+Y)}\right)$$

$$= E\left(e^{ith\left(\frac{1}{\sqrt{2}}\right)}X\right)E\left(e^{ith\left(\frac{1}{\sqrt{2}}\right)}Y\right)$$

the product of two characteristic functions of two random variables, $\frac{1}{\sqrt{2}}X$ and $\frac{1}{\sqrt{2}}Y$. The variance of these two random variables which are normally distributed with zero mean is $\frac{1}{2}\sigma^2$ and so

$$E\left(e^{ithoX}\right) = e^{-\frac{1}{2}(\frac{t\sigma^2}{2}+\frac{1}{2}(\frac{t\sigma^2}{2}))} = e^{-\frac{1}{2}\sigma^2} = E\left(e^{ithoX}\right).$$

Similar reasoning shows $E\left(e^{ithoY'}\right) = E\left(e^{ithoY}\right) = E\left(e^{ithoX}\right)$. Letting $t = 1$, this yields 15.2.4. This proves the lemma.

With this preparation, here is an incredible theorem due to Fernique.

**Theorem 15.2.2** Let $\mu$ be a symmetric Gaussian measure on $B(E)$ where $E$ is a real separable Banach space. Then for $\lambda$ sufficiently small and positive,

$$\int_E e^{\lambda||x||^2}\,d\mu < \infty.$$

More specifically, if $\lambda$ and $r$ are chosen such that

$$\ln\left(\frac{\mu\left([x : ||x|| > r]\right)}{\mu\left(B(0,r)\right)}\right) + 25\lambda r^2 < -1,$$

then

$$\int_E e^{\lambda||x||^2}\,d\mu \leq \exp(\lambda r^2) + \frac{e^2}{e^2 - 1}.$$
Proof: Let $X, Y$ be independent random variables having values in $E$ such that $\mathcal{L}(X) = \mathcal{L}(Y) = \mu$. Then by Lemma 15.2.1

$$\frac{1}{\sqrt{2}} (X - Y), \frac{1}{\sqrt{2}} (X + Y)$$

are also independent and have the same law. Now let $0 \leq s \leq t$ and use independence of the above random variables along with the fact they have the same law as $X$ and $Y$ to obtain

$$P(||X|| \leq s, ||Y|| > t) = P(||X|| \leq s) P(||Y|| > t)$$

$$= P\left(\left\| \frac{1}{\sqrt{2}} (X - Y) \right\| \leq s \right) P\left(\left\| \frac{1}{\sqrt{2}} (X + Y) \right\| > t \right)$$

$$= P\left(\left\| \frac{1}{\sqrt{2}} (X - Y) \right\| \leq s, \left\| \frac{1}{\sqrt{2}} (X + Y) \right\| > t \right)$$

$$\leq P\left(\frac{1}{\sqrt{2}} ||X|| - ||Y|| \leq s, \frac{1}{\sqrt{2}} (||X|| + ||Y||) > t \right).$$

Now consider the following picture in which the region, $R$ represents the points, $||X||, ||Y||$ such that

$$\frac{1}{\sqrt{2}} ||X|| - ||Y|| \leq s \text{ and } \frac{1}{\sqrt{2}} (||X|| + ||Y||) > t.$$

Therefore, continuing with the chain of inequalities above,

$$P(||X|| \leq s) P(||Y|| > t)$$

$$\leq P\left(||X|| > \frac{t - s}{\sqrt{2}}, ||Y|| > \frac{t - s}{\sqrt{2}}\right)$$

$$= P\left(||X|| > \frac{t - s}{\sqrt{2}}\right)^2.$$

Since $X, Y$ have the same law, this can be written as

$$P(||X|| > t) \leq \frac{P\left(||X|| > \frac{t - s}{\sqrt{2}}\right)^2}{P(||X|| \leq s)}.$$

Now define a sequence as follows. $t_0 \equiv r > 0$ and $t_{n+1} \equiv r + \sqrt{2}t_n$. Also, in the above inequality, let $s \equiv r$ and then it follows

$$P(||X|| > t_{n+1}) \leq \frac{P\left(||X|| > \frac{t_{n+1} - r}{\sqrt{2}}\right)^2}{P(||X|| \leq r)}$$

$$= \frac{P(||X|| > t_n)^2}{P(||X|| \leq r)}.$$

Let

$$\alpha_n (r) \equiv \frac{P(||X|| > t_n)}{P(||X|| \leq r)}.$$
Then it follows
\[ \alpha_{n+1}(r) \leq \alpha_n(r)^2, \quad \alpha_0(r) = \frac{P(||X|| > r)}{P(||X|| \leq r)}. \]

Consequently, \( \alpha_n(r) \leq \alpha_0(r)^{2^n} \) and also
\[ P(||X|| > t_n) = \alpha_n(r) P(||X|| \leq r) \leq P(||X|| \leq r) \alpha_0(r)^{2^n} = P(||X|| \leq r) e^{ln(\alpha_0(r))2^n}. \]

Now using the distribution function technique and letting \( \lambda > 0 \),
\[ \int_E e^{\lambda||x||^2} d\mu = \int_0^\infty \mu \left( \left[ e^{\lambda||x||^2} > t \right] \right) dt = 1 + \int_1^\infty \mu \left( \left[ e^{\lambda||x||^2} > t \right] \right) dt = 1 + \int_1^\infty P \left( \left[ e^{\lambda||X||^2} > t \right] \right) dt. \]

From (15.2.4),
\[ P \left( \left[ \exp \left( \lambda ||X||^2 \right) > \exp \left( \lambda t_n^2 \right) \right] \right) \leq P \left( \left[ ||X|| \leq r \right] \right) e^{ln(\alpha_0(r))2^n}. \]

Now split the above improper integral into intervals, \( \exp(\lambda t_n^2) \), \( \exp(\lambda t_{n+1}^2) \) for \( n = 0, 1, \ldots \) and note that \( P \left( \left[ e^{\lambda||X||^2} > t \right] \right) \) is decreasing in \( t \). Then from (15.2.4),
\[ \int_E e^{\lambda||x||^2} d\mu \leq \exp(\lambda r^2) + \sum_{n=0}^{\infty} \int_{\exp(\lambda t_n^2)}^{\exp(\lambda t_{n+1}^2)} P \left( \left[ e^{\lambda||X||^2} > t \right] \right) dt \leq \exp(\lambda r^2) + \sum_{n=0}^{\infty} P \left( \left[ ||X|| \leq r \right] \right) e^{ln(\alpha_0(r))2^n} \exp(\lambda t_{n+1}^2) \leq \exp(\lambda r^2) + \sum_{n=0}^{\infty} e^{ln(\alpha_0(r))2^n} \exp(\lambda t_{n+1}^2). \]

It remains to estimate \( t_{n+1} \). From the description of the \( t_n \),
\[ t_n = \left( \sum_{k=0}^{n} (\sqrt{2})^k \right) r = r \left( \frac{\sqrt{2}}{\sqrt{2} - 1} \right)^{n+1} \leq \frac{\sqrt{2}}{\sqrt{2} - 1} r \left( \sqrt{2} \right)^n \]
and so
\[ t_{n+1} \leq 5r \left( \sqrt{2} \right)^n. \]

Therefore,
\[ \int_E e^{\lambda||x||^2} d\mu \leq \exp(\lambda r^2) + \sum_{n=0}^{\infty} e^{ln(\alpha_0(r))2^n + \lambda 25r^22^n}. \]

Now first pick \( r \) large enough that \( ln(\alpha_0(r)) < -2 \) and then let \( \lambda \) be small enough that \( 25\lambda r^2 < 1 \) or some such scheme and you obtain \( ln(\alpha_0(r)) + \lambda 25r^2 < -1 \). Then for this choice of \( r \) and \( \lambda \), or for any other choice which makes \( ln(\alpha_0(r)) + \lambda 25r^2 < -1 \),
\[ \int_E e^{\lambda||x||^2} d\mu \leq \exp(\lambda r^2) + \sum_{n=0}^{\infty} e^{-2n} \leq \exp(\lambda r^2) + \sum_{n=0}^{\infty} e^{-2n} = \exp(\lambda r^2) + \frac{e^2}{e^2 - 1}. \]
This proves the theorem.

15.3 Gaussian Measures For A Separable Hilbert Space

First recall the Kolmogorov extension theorem, Theorem 15.3.1 on Page 330 which is stated here for convenience. In this theorem, I is an ordered index set, possibly infinite, even uncountable.

**Theorem 15.3.1 (Kolmogorov extension theorem)** For each finite set

\[ J = (t_1, \cdots, t_n) \subseteq I, \]

suppose there exists a Borel probability measure, \( \nu_J = \nu_{t_1, \cdots, t_n} \) defined on the Borel sets of \( \prod_{t \in J} M_t \) where \( M_t = \mathbb{R}^n \) such that if

\[ (t_1, \cdots, t_n) \subseteq (s_1, \cdots, s_p), \]

then

\[ \nu_{t_1, \cdots, t_n} (F_{t_1} \times \cdots \times F_{t_n}) = \nu_{s_1, \cdots, s_p} (G_{s_1} \times \cdots \times G_{s_p}) \quad (15.3.6) \]

where if \( s_i = t_j \), then \( G_{s_i} = F_{t_j} \) and if \( s_i \) is not equal to any of the indices, \( t_k \), then \( G_{s_i} = M_{s_i} \). Then there exists a probability space, \( (\Omega, \mathcal{F}, \mathbb{P}) \) and measurable functions, \( \xi_k : \Omega \rightarrow M_t \) for each \( t \in I \) such that for each \( (t_1 \cdots t_n) \subseteq I, \)

\[ \nu_{t_1, \cdots, t_n} (F_{t_1} \times \cdots \times F_{t_n}) = \mathbb{P} (\{X_{t_1} \in F_{t_1} \cap \cdots \cap X_{t_n} \in F_{t_n}\}). \quad (15.3.7) \]

**Lemma 15.3.2** There exists a sequence, \( \{\xi_k\}_{k=1}^\infty \) of random variables such that

\[ \mathcal{L} (\xi_k) = N (0, 1) \]

and \( \{\xi_k\}_{k=1}^\infty \) is independent.

**Proof:** Let \( i_1 < i_2 \cdots < i_n \) be positive integers and define

\[ \mu_{i_1 \cdots i_n} (F_1 \times \cdots \times F_n) \equiv \frac{1}{(\sqrt{2\pi})^n} \int_{F_1 \times \cdots \times F_n} e^{-|x|^2/2} dx. \]

Then for the index set equal to \( \mathbb{N} \) the measures satisfy the necessary consistency condition for the Kolmogorov theorem above. Therefore, there exists a probability space, \( (\Omega, \mathcal{F}, \mathbb{P}) \) and measurable functions, \( \xi_k : \Omega \rightarrow \mathbb{R} \) such that

\[ P (\{\xi_{i_1} \in F_{i_1}\} \cap \{\xi_{i_2} \in F_{i_2}\} \cdots \cap \{\xi_{i_n} \in F_{i_n}\}) = \mu_{i_1 \cdots i_n} (F_1 \times \cdots \times F_n) = P (\{\xi_{i_1} \in F_{i_1}\}) \cdots P (\{\xi_{i_n} \in F_{i_n}\}) \]

which shows the random variables are independent as well as normal with mean 0 and variance 1. This proves the Lemma.

A random variable \( X \) defined on a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) is called Gaussian if

\[ P (\{X \in A\}) = \frac{1}{\sqrt{2\pi\sigma (v)^2}} \int_A e^{-\frac{1}{2\sigma (v)^2} (x-m(v))^2} dx \]

for all \( A \) a Borel set in \( \mathbb{R} \). Therefore, for the probability space \( (X, \mathcal{B} (X), \mu) \) it is natural to say \( \mu \) is a Gaussian measure if every \( x^* \) in the dual space \( X' \) is a Gaussian random variable. That is, normally distributed.

**Definition 15.3.3** Let \( \mu \) be a measure defined on \( \mathcal{B} (X) \), the Borel sets of \( X \), a separable Banach space. It is called a Gaussian measure if each of the functions in the dual space \( X' \) is normally distributed. As a special case, when \( X = U \) a separable real Hilberts space, \( \mu \) is called a Gaussian measure if for each \( v \in U \), the function \( u \rightarrow (u,v)_U \) is normally distributed. That is, denoting this random variable as \( v' \), it follows for \( A \) a Borel set in \( \mathbb{R} \)

\[ \lambda_{v'} (A) \equiv \mu (\{u : v' (u) \in A\}) = \frac{1}{\sqrt{2\pi\sigma (v)^2}} \int_A e^{-\frac{1}{2\sigma (v)^2} (x-m(v))^2} dx \]

in case \( \sigma (v) > 0 \). In case \( \sigma (v) = 0 \)

\[ \lambda_{v'} \equiv \delta_m(v) \]

In other words, the random variables \( v' \) for \( v \in U \) are all normally distributed on the probability space \( (U, \mathcal{B} (U), \mu) \).
15.3. GAUSSIAN MEASURES FOR A SEPARABLE HILBERT SPACE

Also recall the definition of the characteristic function of a measure.

**Definition 15.3.4** The Borel sets in a topological space $X$ will be denoted by $\mathcal{B}(X)$. For a Borel probability measure $\mu$ defined on $\mathcal{B}(U)$ for $U$ a real separable Hilbert space, define its characteristic function as follows.

$$\phi_\mu(u) \equiv \hat{\mu}(u) = \int_U e^{i(u,v)} d\mu(v)$$  \hspace{1cm} (15.3.8)

More generally, if $\mu$ is a probability measure defined on $\mathcal{B}(X)$ where $X$ is a separable Banach space, then the characteristic function is defined as

$$\phi_\mu(x^*) = \hat{\mu}(x^*) = \int_U e^{ix^*(x)} d\mu(x)$$

One can tell whether $\mu$ is a Gaussian measure by looking at its characteristic function. In fact you can show the following theorem. One part of this theorem is that if $\mu$ is Gaussian, then $m$ and $\sigma^2$ have a certain form.

**Theorem 15.3.5** A measure $\mu$ on $\mathcal{B}(U)$ is Gaussian if and only if there exists $m \in U$ and $Q \in \mathcal{L}(U)$ such that $Q$ is nonnegative symmetric with finite trace,

$$\sum_k (Qe_k, e_k) < \infty$$

for a complete orthonormal basis for $U$, and

$$\phi_\mu(u) = \hat{\mu}(u) = e^{i(m,u) - \frac{1}{2}(Qu,u)}$$  \hspace{1cm} (15.3.9)

In this case $\mu$ is called $N(m,Q)$ where $m$ is the mean and $Q$ is called the covariance. The measure $\mu$ is uniquely determined by $m$ and $Q$. Also for all $h,g \in U$

$$\int (x,h) d\mu(x) = (m,h)_U$$  \hspace{1cm} (15.3.10)

$$\int ((x,h) - (m,h))((x,g) - (m,g)) d\mu(x) = (Qh,g)$$  \hspace{1cm} (15.3.11)

$$\int ||x - m||^2 d\mu(x) = \text{trace}(Q).$$  \hspace{1cm} (15.3.12)

**Proof:** First of all suppose (15.3.9) holds. Why is $\mu$ Gaussian? Consider the random variable $u'$ defined by $u' (v) \equiv (v,u)$. Why is $\lambda_{u'}$ a Gaussian measure on $\mathbb{R}$? By the definition in (15.3.9),

$$\int_U e^{itu'(v)} d\mu(v) = \int_U e^{it(v,u)} d\mu(v) = \int_{\mathbb{R}} e^{ix} d\lambda_{u'}(x)$$

$$= \int_U e^{i(v,tu)} d\mu(v) = e^{it(m,u) - \frac{1}{2}t^2(Qu,u)}$$

and this is the characteristic equation for a random variable having mean $(m,u)$ and variance $(Qu,u)$. In case $(Qu,u) = 0$, you get $e^{it(m,u)}$ which is the characteristic function for a random variable having distribution $\delta_{(m,u)}$. Thus if (15.3.9) holds, then $u'$ is normally distributed as desired. Thus $\mu$ is Gaussian by definition.

The next task is to suppose $\mu$ is Gaussian and show the existence of $m,Q$ which have the desired properties. This involves the following lemma.

**Lemma 15.3.6** Let $U$ be a real separable Hilbert space and let $\mu$ be a probability measure defined on $\mathcal{B}(U)$. Suppose for some positive integer, $k$

$$\int_U |(x,z)|^k d\mu(x) < \infty$$

for all $z \in U$. Then the transformation,

$$(h_1, \cdots, h_k) \rightarrow \int_U (h_1,x) \cdots (h_k,x) d\mu(x)$$  \hspace{1cm} (15.3.13)

is a continuous $k$–linear form.
Proof: I need to show that for each $h \in U^k$, the integral in $(15.3.14)$ exists. From this it is obvious it is $k-$ linear, meaning linear in each argument. Then it is shown it is continuous.

First note

$$|(h_1, x) \cdots (h_k, x)| \leq |(h_1, x)|^k + \cdots + |(h_k, x)|^k$$

This follows from observing that one of $|(h_j, x)|$ is largest. Then the left side is smaller than $|(h_j, x)|^k$. Therefore, the above inequality is valid. This inequality shows the integral in $(15.3.14)$ makes sense.

I need to establish an estimate of the form

$$\int_U |(x, h)|^k \, d\mu(x) < C < \infty$$

for every $h \in U$ such that $||h||$ is small enough.

Let

$$U_n \equiv \left\{ z \in U : \int_U |(x, z)|^k \, d\mu(x) \leq n \right\}$$

Then by assumption $U = \cup_{n=1}^{\infty} U_n$ and it is also clear from Fatou’s lemma that each $U_n$ is closed. Therefore, by the Bair category theorem, at least one of these $U_{n_0}$ contains an open ball, $B(z_0, r)$. Then letting $|y| < r,$

$$\int_U |(x, z_0 + y)|^k \, d\mu(x) \leq n_0,$$

and so for such $y,$

$$\int_U |(x, y)|^k \, d\mu = \int_U |(x, z_0 + y) - (x, z_0)|^k \, d\mu$$

$$\leq \int_U 2^k |(x, z_0 + y)|^k + 2^k |(x, z_0)|^k \, d\mu(x)$$

$$\leq 2^k (n_0 + n_0) = 2^{k+1} n_0.$$

It follows that for arbitrary nonzero $y \in U$

$$\int_U \left( \frac{(r/2) y}{||y||} \right)^k \, d\mu \leq 2^{k+1} n_0$$

and so

$$\int_U |(x, y)|^k \, d\mu \leq (2^{k+2}/r) n_0 ||y||^k \equiv C ||y||^k.$$

Thus by Holder’s inequality,

$$\int_U |(h_1, x) \cdots (h_k, x)| \, d\mu(x) \leq \prod_{j=1}^k \left( \int_U |(h_j, x)|^k \, d\mu(x) \right)^{1/k}$$

$$\leq C \prod_{j=1}^k ||h_j||$$

This proves the lemma.

Now continue with the proof of the theorem. I need to identify $m$ and $Q$. It is assumed $\mu$ is Gaussian. Recall this means $h'$ is normally distributed for each $h \in U$. Then using

$$|x| \leq |x - m(h)| + |m(h)|$$

$$\int_U |(x, h)_t| \, d\mu(x) = \int_{\mathbb{R}} |x| \, d\lambda_{h'}(x)$$

$$= \frac{1}{\sqrt{2\pi}\sigma(h)} \int_{\mathbb{R}} |x - m(h)|^2 \, dx$$

$$\leq \frac{1}{\sqrt{2\pi}\sigma(h)} \int_{\mathbb{R}} |x - m(h)| \, e^{-\frac{(x-m(h))^2}{2\sigma^2(h)}} \, dx + |m(h)|$$
Then using the Cauchy Schwarz inequality, with respect to the probability measure
\[
\frac{1}{\sqrt{2\pi\sigma^2(h)}} e^{-\frac{1}{2\sigma^2}(x-m(h))^2} dx,
\]
\[
\leq \frac{1}{\sqrt{2\pi\sigma^2(h)}} \left( \int_{\mathbb{R}} |x - m(h)|^2 e^{-\frac{1}{2\sigma^2}(x-m(h))^2} dx \right)^{1/2} + |m(h)| < \infty
\]
Thus by Lemma 15.3.6
\[
h \to \int_{U} (x, h) d\mu(x)
\]
is a continuous linear transformation and so by the Riesz representation theorem, there exists a unique \( m \in U \) such that
\[
(h, m)_{U} = \int_{U} (h, x) d\mu(x)
\]
Also the above says \((h, m)\) is the mean of the random variable \( x \to (x, h) \) so in the above,
\[
m(h) = (h, m)_{U}.
\]
Next it is necessary to find \( Q \). To do this let \( Q \) be given by 15.3.11. Thus
\[
(Qh, g) = \int_{U} ((x, h) - (m, h)) ((x, g) - (m, g)) d\mu(x)
\]
\[
= \int_{U} (x - m, h) (x - m, g) d\mu(x)
\]
It is clear \( Q \) is linear and the above is a bilinear form (The integral makes sense because of the assumption that \( h', g' \) are normally distributed,) but is it continuous? Does \( (Qh, h) = \sigma^2(h) ? \)

First, the above equals
\[
\int_{U} (x, h) (x - m, g) d\mu(x) - \int_{U} (m, h) (x - m, g) d\mu(x)
\]
\[
= \int_{U} (x, h) (x - m, g) d\mu(x)
\]
(15.3.14)
because from the first part,
\[
\int_{U} (x - m, g) d\mu(x) = \int_{U} (x, g) d\mu(x) - (m, g)_{U} = 0.
\]
Now by the first part, the term in 15.3.11 is
\[
\int_{U} (x, h) (x, g) d\mu(x) - (m, g) \int_{U} (x, h) d\mu(x)
\]
\[
= \int_{U} (x, h) (x, g) d\mu(x) - (m, g) (m, h).
\]
Thus
\[
|(Qh, g)| \leq \int_{U} |(x, h)(x, g)| d\mu(x) + ||m||^2 ||h|| ||g||
\]
and since the random variables \( h' \) and \( g' \) given by \( x \to (x, h) \) and \( x \to (x, g) \) respectively are given to be normally distributed with variance \( \sigma^2(h) \) and \( \sigma^2(g) \) respectively, the above integral is finite. Also for all \( h \),
\[
\int_{U} |(x, h)|^2 d\mu(x) < \infty
\]
because the random variable \( h' \) is given to be normally distributed. Therefore from Lemma 15.3.11, there exists some constant \( C \) such that
\[
|(Qh, g)| \leq C \ ||h|| \ ||g||
\]
which shows \( Q \) is continuous as desired.
Why is $\sigma^2(h) = (Qh,h)$? This follows because from the above

$$
(Qh,h) \equiv \int_U (h, x - m)^2 d\mu(x) \\
= \int_U ((x, h) - (h, m))^2 d\mu(x) = \int_R (t - (h, m))^2 d\lambda_{h'}(t) \\
= \frac{1}{\sqrt{2\pi\sigma^2(h)}} \int_R (t - (h, m))^2 e^{-\frac{1}{2\sigma^2(h)}(t-(h, m))^2} dt = \sigma^2(h)
$$

from a standard result for the normal distribution function which follows from an easy change of variables argument.

Why must $Q$ have finite trace? For $h \in U$, it follows from the above that $h'$ is normally distributed with mean $(h, m)$ and variance $(Qh,h)$. Therefore, the characteristic function of $h'$ is known. In fact

$$
\int_U e^{it(h,x)} d\mu(x) = e^{it(h,m)} e^{-\frac{1}{2}(Qh,h)}
$$

Thus also

$$
\int_U e^{it(x-m,h)} d\mu(x) = e^{-\frac{1}{2}t^2(Qh,h)}
$$

and letting $t = 1$ this yields

$$
\int_U e^{ix-(m,h)} d\mu(x) = e^{-\frac{1}{2}(Qh,h)}
$$

From this it follows

$$
\int_U (1 - e^{ix-(m,h)}) d\mu(x) = 1 - e^{-\frac{1}{2}(Qh,h)}
$$

and since the right side is real, this implies

$$
\int_U (1 - \cos (x - (m, h))) d\mu(x) = 1 - e^{-\frac{1}{2}(Qh,h)}
$$

Thus

$$
1 - e^{-\frac{1}{2}(Qh,h)} \leq \int_{||x-m|| \leq c} (1 - \cos (x - m, h)) d\mu(x) \\
+ 2 \int_{||x-m|| > c} d\mu(x)
$$

Now it is routine to show

$$
1 - \cos t \leq \frac{1}{2} t^2
$$

and so

$$
1 - e^{-\frac{1}{2}(Qh,h)} \leq \frac{1}{2} \int_{||x-m|| \leq c} |(x - m, h)|^2 d\mu(x) \\
+ 2 \mu(||x-m|| > c)
$$

Pick $c$ large enough that the last term is smaller than $1/8$. This can be done because the sets decrease to $\emptyset$ as $c \to \infty$ and $\mu$ is given to be a finite measure. Then with this choice of $c$,

$$
\frac{7}{8} - \frac{1}{2} \int_{||x-m|| \leq c} |(x - m, h)|^2 d\mu(x) \leq e^{-\frac{1}{2}(Qh,h)} \quad (15.3.15)
$$

For each $h$ the integral in the above is finite. In fact

$$
\int_{||x-m|| \leq c} |(x - m, h)|^2 d\mu(x) \leq c^2 ||h||^2
$$

Let

$$
(Qc h, h_1) \equiv \int_{||x-m|| \leq c} (x - m, h) (x - m, h_1) d\mu(x)
$$
and let $A$ denote those $h \in U$ such that
\[(Q_c h, h) < 1.\]
Then from 15.3.15 it follows that for $h \in A$,
\[
\frac{3}{8} = \frac{7}{8} - \frac{1}{2} \leq \frac{7}{8} - \frac{1}{2} (Q_c h, h) \leq e^{-\frac{1}{2}(Q_c h, h)}
\]
Therefore, for such $h$,
\[
\frac{8}{3} \geq e^{\frac{1}{2}(Q_c h, h)} \geq 1 + \frac{1}{2} (Q h, h)
\]
and so for $h \in A$,
\[
(Q h, h) \leq \left(\frac{8}{3} - 1\right) \frac{2}{2} = \frac{10}{3}
\]
Now let $h$ be arbitrary. Then for each $\varepsilon > 0$
\[
\frac{h}{\varepsilon + \sqrt{(Q h, h)}} \in A
\]
and so
\[
\left(Q \left(\frac{h}{\varepsilon + \sqrt{(Q h, h)}}\right), \frac{h}{\varepsilon + \sqrt{(Q h, h)}}\right) \leq \frac{10}{3}
\]
which implies
\[
(Q h, h) \leq \frac{10}{3} \left(\varepsilon + \sqrt{(Q h, h)}\right)^2
\]
Since $\varepsilon$ is arbitrary,
\[
(Q h, h) \leq \frac{10}{3} (Q_c h, h). \tag{15.3.16}
\]
However, $Q_c$ has finite trace. To see this, let $\{e_k\}$ be an orthonormal basis in $U$. Then
\[
\sum_k (Q_c e_k, e_k) = \sum_k \int_{||x - m|| \leq c} |(x - m, e_k)|^2 d\mu (x)
\]
\[
= \int_{||x - m|| \leq c} \sum_k |(x - m, e_k)|^2 d\mu (x) = \int_{||x - m|| \leq c} ||x - m||^2 d\mu (x) \leq c^2
\]
It follows from 15.3.14 that $Q$ must also have finite trace.

That $\mu$ is uniquely determined by $m$ and $Q$ follows from Theorem 11.4.9. This proves the theorem.

Suppose you have a given $Q$ having finite trace and $m \in U$. Does there exist a Gaussian measure on $B(U)$ having these as the covariance and mean respectively?

**Proposition 15.3.7** Let $U$ be a real separable Hilbert space and let $m \in U$ and $Q$ be a positive, symmetric operator defined on $U$ which has finite trace. Then there exists a Gaussian measure with mean $m$ and covariance $Q$.

**Proof:** By Lemma 15.3.2, which comes from Kolmogorov’s extension theorem, there exists a probability space $(\Omega, \mathcal{F}, P)$ and a sequence $\{\xi_i\}$ of independent random variables which are normally distributed with mean 0 and variance 1. Then let
\[
X (\omega) = m + \sum_{j=1}^{\infty} \sqrt{\lambda_j} \xi_j (\omega) e_j
\]
where the $\{e_j\}$ are the eigenvectors of $Q$ such that $Q e_j = \lambda_j e_j$. The series in the above converges in $L^2(\Omega; U)$ because
\[
\left\| \sum_{j=m}^{n} \sqrt{\lambda_j} \xi_j e_j \right\|_{L^2(\Omega; U)}^2 = \int \left| \sum_{j=m}^{n} \lambda_j \xi_j (\omega) \right| dP = \sum_{j=m}^{n} \lambda_j
\]
and so the partial sums form a Cauchy sequence in $L^2(\Omega; U)$. 
Now if \( h \in U \), I need to show that \( \omega \rightarrow (X(\omega), h) \) is normally distributed. From this it will follow that \( \mathcal{L}(X) \) is Gaussian. A subsequence
\[
\left\{ m + \sum_{j=1}^{n_k} \sqrt{\lambda_j} \xi_j(\omega) e_j \right\} = \{S_{n_k}(\omega)\}
\]
of the above sequence converges pointwise a.e. to \( X \).

\[
E(\exp(it(X,h))) = \lim_{k \to \infty} E(\exp(it(S_{n_k},h)))
\]
\[
= \exp(it(m,h)) \lim_{k \to \infty} E\left( \exp\left( it \sum_{j=1}^{n_k} \sqrt{\lambda_j} \xi_j(\omega)(e_j, h) \right) \right)
\]
\[
= \exp(it(m,h)) \lim_{k \to \infty} \prod_{j=1}^{n_k} e^{-\frac{1}{2}t^2 \lambda_j(e_j, h)^2}
\]
\[
= \exp(it(m,h)) \lim_{k \to \infty} \exp\left( -\frac{1}{2} \sum_{j=1}^{n_k} \lambda_j(e_j, h)^2 \right).
\]
(15.3.17)

Now
\[
(Qh, h) = \left( \sum_{k=1}^{\infty} (e_k, h) e_k, \sum_{j=1}^{\infty} (e_j, h) e_j \right)
\]
\[
= \sum_{k=1}^{\infty} \lambda_k e_k, \sum_{j=1}^{\infty} (e_j, h) e_j
\]
\[
= \sum_{j=1}^{\infty} \lambda_j(e_j, h)^2
\]
and so, passing to the limit in (15.3.18) yields
\[
\exp(it(m,h)) \exp\left( -\frac{1}{2} t^2 (Qh, h) \right)
\]
(15.3.18)
which implies that \( \omega \rightarrow (X(\omega), h) \) is normally distributed with mean \( (m, h) \) and variance \( (Qh, h) \).

Now let \( \mu = \mathcal{L}(X) \). That is, for all \( B \in \mathcal{B}(U) \),
\[
\mu(B) \equiv P \left( \{ X \in B \} \right)
\]
In particular, \( B \) could be the cylindrical set
\[
B \equiv \{ x : (x, h) \in A \}
\]
for \( A \) a Borel set in \( \mathbb{R} \). Then by definition, if \( h \in U \), and \( A \) is a Borel set in \( \mathbb{R} \),
\[
\mu(B) = \mu(\{ x : (x, h) \in A \}) \equiv P(\{ \omega : (X(\omega), h) \in A \})
\]
\[
= \int_A \frac{1}{\sqrt{2\pi (Qh, h)}} e^{-\frac{(x-(m,h))^2}{2(Qh, h)}} dt
\]
and so \( x \rightarrow (x,h) \) is normally distributed. Therefore by definition, \( \mu \) is a Gaussian measure.

Letting \( t = 1 \) in (15.3.18) it follows
\[
\int_U e^{i(x,h)} d\mu(x) = \int_\Omega e^{i(X(\omega),h)} dP = \exp(i(m,h)) \exp\left( -\frac{1}{2} (Qh, h) \right)
\]
which is the characteristic function of a Gaussian measure on \( U \) having covariance \( Q \) and mean \( m \). This proves the proposition.
A real valued random variable $X$ is normally distributed with mean 0 and variance $\sigma^2$ if

$$P (X \in A) = \frac{1}{\sqrt{2\pi\sigma}} \int_A e^{-\frac{1}{2} \frac{x^2}{\sigma^2}} dx$$

Consider the characteristic function. By definition it is

$$\phi_X (\lambda) \equiv \int_{\mathbb{R}} e^{i\lambda x} d\lambda_X (x)$$

where $\lambda_X$ is the distribution measure for this random variable. Thus the characteristic function of this random variable is

$$\frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{i\lambda x} e^{-\frac{1}{2\sigma^2} \frac{x^2}{2}} dx$$

One can then show through routine arguments that this equals

$$\exp \left(-\frac{1}{2} \sigma \lambda^2 \right)$$

### 16.1 Real Wiener Processes

Here is the definition of a Wiener process.

**Definition 16.1.1** Let $W(t)$ be a stochastic process which has the properties that whenever $t_1 < t_2 < \cdots < t_m$, the increments $\{W(t_i) - W(t_{i-1})\}$ are independent and whenever $s < t$, it follows $W(t) - W(s)$ is normally distributed with variance $t - s$ and mean 0. Also $t \to W(t)$ is Holder continuous with every exponent $\gamma < 1/2$ and $W(0) = 0$. This is called a Wiener process.

Do Wiener processes exist? Yes, they do. First here is a simple lemma which has really been done before. It depends on the Kolmogorov extension theorem, Theorem 2.5.3 on Page 36.

**Lemma 16.1.2** There exists a sequence, $\{\xi_k\}_{k=1}^{\infty}$ of random variables such that

$$\mathcal{L} (\xi_k) = N (0, 1)$$

and $\{\xi_k\}_{k=1}^{\infty}$ is independent.

**Proof:** Let $i_1 < i_2 \cdots < i_n$ be positive integers and define

$$\mu_{i_1 \cdots i_n} (F_1 \times \cdots \times F_n) \equiv \frac{1}{(\sqrt{2\pi})^n} \int_{F_1 \times \cdots \times F_n} e^{-|x|^2/2} dx.$$

Then for the index set equal to $\mathbb{N}$ the measures satisfy the necessary consistency condition for the Kolmogorov theorem. Therefore, there exists a probability space, $(\Omega, P, \mathcal{F})$ and measurable functions, $\xi_k : \Omega \to \mathbb{R}$ such that

$$P (\bigcap \mathbb{N} (F_{i_1}) \cap [i_2 \in F_{i_2}] \cdots \cap [i_n \in F_{i_n}])$$

$$= \mu_{i_1 \cdots i_n} (F_1 \times \cdots \times F_n)$$

$$= P (\xi_{i_1} \in F_{i_1}) \cdots P (\xi_{i_n} \in F_{i_n})$$
which shows the random variables are independent as well as normal with mean 0 and variance 1.

Recall that the sum of independent normal random variables is normal. The Wiener process is just an infinite weighted sum of the above independent normal random variables, the weights depending on \( t \). Therefore, if the sum converges, it is not too surprising that the result will be normally distributed and the variance will depend on \( t \). This is the idea behind the following theorem.

**Theorem 16.1.3** There exists a real Wiener process as defined in Definition 16.1.1. Furthermore, the distribution of \( W(t) - W(s) \) is the same as the distribution of \( W(t-s) \) and \( W \) is Holder continuous with exponent \( \gamma \) for any \( \gamma < 1/2 \). Also for each \( \alpha > 1 \),

\[
E(|W(t) - W(s)|^\alpha) \leq C_\alpha |t-s|^{\alpha/2} E(|W(1)|^\alpha)
\]

**Proof:** Let \( \{g_m\}_{m=1}^\infty \) be a complete orthonormal set in \( L^2(0, \infty) \). Thus, if \( f \in L^2(0, \infty) \),

\[
f = \sum_{i=1}^\infty (f, g_i)_{L^2} g_i.
\]

The Wiener process is defined as

\[
W(t, \omega) = \sum_{i=1}^\infty (\mathcal{X}(0,t), g_i)_{L^2} \xi_i(\omega)
\]

where the random variables, \( \{\xi_i\} \) are as described in Lemma 16.1.2. The series converges in \( L^2(\Omega) \) where \( (\Omega, \mathcal{F}, P) \) is the probability space on which the random variables, \( \xi_i \), are defined. This will first be shown. Note first that from the independence of the \( \xi_i \),

\[
\int_{\Omega} \xi_i \xi_j dP = 0
\]

Therefore,

\[
\int_{\Omega} \left| \sum_{i=m}^n (\mathcal{X}(0,t), g_i)_{L^2} \xi_i(\omega) \right|^2 dP = \sum_{i=m}^n (\mathcal{X}(0,t), g_i)_{L^2}^2 \int_{\Omega} |\xi_i|^2 dP = \sum_{i=m}^n (\mathcal{X}(0,t), g_i)_{L^2}^2
\]

which converges to 0 as \( m, n \to \infty \). Thus the partial sums are a Cauchy sequence in \( L^2(\Omega, P) \).

It just remains to verify this definition satisfies the desired conditions. First I will show that \( \omega \to W(t, \omega) \) is normally distributed with mean 0 and variance \( t \). That it should be normally distributed is not surprising since it is just a sum of independent random variables which are this way. Selecting a suitable subsequence, \( \{n_k\} \) it can be assumed

\[
W(t, \omega) = \lim_{k \to \infty} \sum_{i=1}^{n_k} (\mathcal{X}(0,t), g_i)_{L^2} \xi_i(\omega) \text{ a.e.}
\]

and so from the dominated convergence theorem and the independence of the \( \xi_i \),

\[
E(\exp(i\lambda W(t))) = \lim_{k \to \infty} E\left(\exp\left(i\lambda \sum_{j=1}^{n_k} (\mathcal{X}(0,t), g_j)_{L^2} \xi_j(\omega)\right)\right)
\]

\[
= \lim_{k \to \infty} E\left(\prod_{j=1}^{n_k} \exp\left(i\lambda (\mathcal{X}(0,t), g_j)_{L^2} \xi_j(\omega)\right)\right)
\]

\[
= \lim_{k \to \infty} \prod_{j=1}^{n_k} E\left(\exp\left(i\lambda (\mathcal{X}(0,t), g_j)_{L^2} \xi_j(\omega)\right)\right)
\]

\[
= \lim_{k \to \infty} \prod_{j=1}^{n_k} e^{-\frac{1}{2} \lambda^2 (\mathcal{X}(0,t), g_j)_{L^2}^2}
\]

\[
= \lim_{k \to \infty} \exp\left(\sum_{j=1}^{n_k} -\frac{1}{2} \lambda^2 (\mathcal{X}(0,t), g_j)_{L^2}^2\right)
\]

\[
= \exp\left(-\frac{1}{2} \lambda^2 \|\mathcal{X}(0,t)\|_{L^2}^2\right) = \exp\left(-\frac{1}{2} \lambda^2 t\right),
\]
the characteristic function of a normally distributed random variable having variance \( t \) and mean 0.

It is clear \( W(0) = 0 \). It remains to verify the increments are independent. To do this, consider

\[
E \left( \exp \left( i \left[ \lambda (W(t) - W(s)) + \mu (W(s) - W(r)) \right] \right) \right) = \exp \left( -\frac{1}{2} \left[ \lambda^2 (t - s) + \mu^2 (s - r) \right] \right)
\]

which equals (16.1.1) and this shows the increments are independent. Obviously, this same argument shows this holds for any finite set of disjoint increments.

From the definition, if \( t > s \)

\[
W(t - s) = \sum_{k=1}^{\infty} (X_{(0,t-s)}, g_k)_{L^2} \xi_k
\]

while

\[
W(t) - W(s) = \sum_{k=1}^{\infty} (X_{(s,t)}, g_k)_{L^2} \xi_k.
\]

Then the same argument given above involving the characteristic function to show \( W(t) \) is normally distributed shows both of these random variables are normally distributed with mean 0 and variance \( t - s \) because they have the same characteristic function.
For example, ignoring the limit questions and proceeding formally,

\[ E(\exp(i\lambda (W(t) - W(s)))) = E\left(\exp\left(i\lambda \left(\sum_{k=1}^{\infty} (X(s,t), g_k)_{L^2} \xi_k\right)\right)\right) \]

\[ = E\left(\prod_{k=1}^{\infty} \exp\left(i\lambda (X(s,t), g_k)_{L^2} \xi_k\right)\right) \]

\[ = \prod_{k=1}^{\infty} E(\exp(i\lambda (X(s,t), g_k)_{L^2} \xi_k)) \]

\[ = \prod_{k=1}^{\infty} e^{-\frac{1}{2} \lambda^2 (X(s,t), g_k)^2_{L^2}} \]

\[ = \exp\left(-\frac{1}{2} \lambda^2 \sum_{k=1}^{\infty} (X(s,t), g_k)^2_{L^2}\right) \]

\[ = \exp\left(-\frac{1}{2} \lambda^2 (t-s)\right) \]

which is the characteristic function of a random variable having mean 0 and variance \(t-s\).

Finally note the distribution of \(W(t-s)\) is the same as the distribution of

\[ W(1)(t-s)^{1/2} = \sum_{k=1}^{\infty} (X(0,1), g_k)_{L^2} \xi_k (t-s)^{1/2} \]

because the characteristic function of this last random variable is the same as the characteristic function of \(W(t-s)\) which is \(e^{-\frac{1}{2} \lambda^2 (t-s)}\) which follows from a simple computation. Since \(W(1)\) is a normally distributed random variable with mean 0 and variance 1,

\[ E\left(\exp\left(i\lambda W(1)(t-s)^{1/2}\right)\right) = e^{-\frac{1}{2} \lambda^2 (t-s)} \]

which is the same as the characteristic function of \(W(t-s)\).

Hence for any positive \(\alpha\),

\[ E(\|W(t) - W(s)\|^\alpha) = E(\|W(t-s)\|^\alpha) \]

\[ = E\left(\left(\|t-s\|^{1/2} W(1)\right)^\alpha\right) \]

\[ = \|t-s\|^{\alpha/2} E(\|W(1)\|^\alpha) \] (16.1.3)

It follows from Theorem 16.3.2 that \(W(t)\) is H"older continuous with exponent \(\gamma\) where \(\gamma\) is any positive number less than \(\beta/\alpha\) where \(\alpha/2 = 1 + \beta\). Thus \(\gamma\) is any constant less than

\[ \frac{\frac{\alpha}{2} - 1}{\frac{\alpha}{2}} = \frac{\alpha - 2}{\alpha} \]

Thus \(\gamma\) is any constant less than \(\frac{1}{2}\). 

The proof of the theorem, which only depended on \(\{\xi_i\}_{i=1}^{\infty}\) being independent random variables each normal with mean 0 and variance 1, implies the following corollary.

**Corollary 16.1.4** Let \(\{\xi_i\}_{i=1}^{\infty}\) be independent random variables each normal with mean 0 and variance 1. Then

\[ W(t,\omega) \equiv \sum_{i=1}^{\infty} (X_{[0,t]}, g_i)_{L^2} \xi_i(\omega) \]

is a real Wiener process. Furthermore, the distribution of \(W(t) - W(s)\) is the same as the distribution of \(W(t-s)\) and \(W\) is H"older continuous with exponent \(\gamma\) for any \(\gamma < 1/2\). Also for each \(\alpha > 1\),

\[ E(\|W(t) - W(s)\|^\alpha) \leq C_\alpha |t-s|^{\alpha/2} E(\|W(1)\|^\alpha) \]
16.2 Nowhere Differentiability Of Wiener Processes

If $W(t)$ is a Wiener process, it turns out that $t \to W(t, \omega)$ is nowhere differentiable for a.e. $\omega$. This fact is based on the independence of the increments and the fact that these increments are normally distributed.

First note that $W(t) - W(s)$ has the same distribution as $(t - s)^{1/2} W(1)$. This is because they have the same characteristic function. Next it follows that because of the independence of the increments and what was just noted that,$$
P\left( \bigcap_{r=1}^{5} \left| W(t + r\delta) - W(t + (r - 1)\delta) \right| \leq K\delta \right)$$

$$= \prod_{r=1}^{5} \P\left( \left| W(t + r\delta) - W(t + (r - 1)\delta) \right| \leq K\delta \right)$$

$$= \prod_{r=1}^{5} P\left( \left| \delta^{1/2} W(1) \right| \leq K\delta \right) = \left( \frac{1}{\sqrt{2\pi}} \int_{-K\sqrt{\delta}}^{K\sqrt{\delta}} e^{-\frac{1}{2}t^{2}} \, dt \right)^{5} \leq C\delta^{5/2}. \tag{16.2.4}$$

With this observation, here is the proof which follows [32] and according to this reference is due to Payley, Wiener and Zygmund and the proof is like one given by Dvoretsky, Erdős and Kakutani.

**Theorem 16.2.1** Let $W(t)$ be a Wiener process. Then there exists a set of measure 0, $N$ such that for all $\omega \notin N$,

$$t \to W(t, \omega)$$

is nowhere differentiable.

**Proof:** Let $[0, a]$ be an interval. If for some $\omega, t \to W(t, \omega)$ is differentiable at some $s$, then for some $n, p > 0$,

$$\left| \frac{W(t, \omega) - W(s, \omega)}{t - s} \right| \leq p$$

whenever $|t - s| < 5a2^{-n} \equiv 5\delta_{n}$. Define $C_{np}$ by

$$\left\{ \omega : \text{for some } s \in [0, a], \left| \frac{W(t, \omega) - W(s, \omega)}{t - s} \right| \leq p \text{ if } |t - s| \leq 5\delta_{n} \right\}. \tag{16.2.5}$$

Thus $\cup_{n,p \in \mathbb{N}} C_{np}$ contains the set of $\omega$ such that $t \to W(t, \omega)$ is differentiable for some $s \in [0, a]$.

Now define uniform partitions of $[0, a), \{t_{k}^{n}\}_{k=0}^{2^n}$ such that

$$|t_{k}^{n} - t_{k-1}^{n}| = a2^{-n} \equiv \delta_{n}$$

Let

$$D_{np} \equiv \cup_{i=0}^{2^n-1} \left( \cap_{r=1}^{5} \left| W(t_{i}^{n} + r\delta_{n}) - W(t_{i}^{n} + (r - 1)\delta_{n}) \right| \leq 10p\delta_{n} \right)$$

If $\omega \in C_{np}$, then for some $s \in [0, a)$, the condition of [16.2.3] holds. Suppose $k$ is the number such that $s \in [t_{k-1}^{n}, t_{k}^{n})$. Then for $r \in \{1, 2, 3, 4, 5\}$,

$$\left| W(t_{k}^{n} + r\delta_{n}, \omega) - W(t_{k}^{n} + (r - 1)\delta_{n}, \omega) \right|$$

$$\leq \left| W(t_{k}^{n} + r\delta_{n}, \omega) - W(s, \omega) \right| + \left| W(s, \omega) - W(t_{k}^{n} + (r - 1)\delta_{n}, \omega) \right|$$

$$\leq 5p\delta_{n} + 5p\delta_{n} = 10p\delta_{n}$$

Thus $C_{np} \subseteq D_{np}$. Now from [16.2.4]

$$P(D_{np}) \leq 2^{n}C\delta^{5/2} = Ca^{5/2}2^{n} \left(2^{-n}\right)^{5/2} = C \left(\sqrt{a}\right)^{5}2^{-\frac{5}{2}n}. \tag{16.2.6}$$

Let

$$C_{p} = \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} C_{kp} \subseteq \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} D_{kp}.$$ 

It was just shown in [16.2.5] that $P(\cap_{k=n} D_{kp}) = 0$ and so $C_{p}$ has measure 0. Thus $\cup_{n=1}^{\infty} C_{np}$, the set of points, $\omega$ where $t \to W(t, \omega)$ could have a derivative has measure 0. Taking the union of the exceptional sets corresponding to intervals $[0, n)$ for $n \in \mathbb{N}$, this proves the theorem.
This theorem on nowhere differentiability is very important because it shows that one can define an integral \( \int f(s) dW(s) \) by simply fixing \( \omega \) and then doing some sort of Stieltjes integral in time. The reason for this is that the nowhere differentiability of \( W \) implies it is also not of bounded variation on any interval since if it were, it would equal the difference of two increasing functions and would therefore have a derivative at a.e. point.

I have presented the theorem on nowhere differentiability for one dimensional Wiener processes but the same proof holds with minor modifications if you have defined the Wiener process in \( \mathbb{R}^n \) or you could simply consider the components and apply the above result.

### 16.3 An Example Of Martingales, Independent Increments

Here is an interesting lemma.

**Lemma 16.3.1** Let \( (W(t), \mathcal{F}_t) \) be a stochastic process which has independent increments having values in \( E \) a real separable Banach space. Let

\[
A \in \mathcal{F}_s \equiv \sigma (W(u) - W(r): 0 \leq r < u \leq s)
\]

Suppose \( g(W(t) - W(s)) \in L^1(\Omega; E) \). Then the following formula holds.

\[
\int_\Omega \chi_{Ag}(W(t) - W(s)) \, dP = P(A) \int_\Omega g(W(t) - W(s)) \, dP \tag{16.3.7}
\]

**Proof:** Let \( \mathcal{G} \) denote the set of all \( A \in \mathcal{F}_s \) such that \( 16.3.7 \) holds. Then it is obvious \( \mathcal{G} \) is closed with respect to complements and countable disjoint unions. Let \( \mathcal{K} \) denote those sets which are finite intersections of the form

\[
A = \bigcap_{i=1}^m A_i
\]

where each \( A_i \) is in a set of \( \sigma (W(u_i) - W(r_i)) \) for some \( 0 \leq r_i < u_i \leq s \). For such \( A \), it follows

\[
A \in \sigma (W(u_i) - W(r_i), i = 1, \ldots, m).
\]

Now consider the random vector having values in \( E^{m+1} \),

\[
(W(u_1) - W(r_1), \ldots, W(u_m) - W(r_m), g(W(t) - W(s)))
\]

Let \( t^* \in (E^* )^m \) and \( s^* \in E^* \).

\[
t^* \cdot (W(u_1) - W(r_1), \ldots, W(u_m) - W(r_m))
\]

can be written in the form \( g^* \cdot (W(\tau_1) - W(\eta_1), \ldots, W(\tau_i) - W(\eta_i)) \) where the intervals, \((\eta_j, \tau_j)\) are disjoint and each \( \tau_j \leq s \). For example, suppose you have

\[
a(W(2) - W(1)) + b(W(2) - W(0)) + c(W(3) - W(1)),
\]

where obviously the increments are not disjoint. Then you would write the above expression as

\[
a(W(2) - W(1)) + b(W(2) - W(1)) + b(W(1) - W(0)) + c(W(3) - W(2)) + c(W(2) - W(1))
\]

and then you would collect the terms to obtain

\[
b(W(1) - W(0)) + (a + b + c)(W(2) - W(1)) + c(W(3) - W(2))
\]

and now these increments are disjoint.

Therefore, by independence of the disjoint increments,

\[
E(\exp i(t^* \cdot (W(u_1) - W(r_1), \ldots, W(u_m) - W(r_m)) + s^* (g(W(t) - W(s))))
\]

\[
= E(\exp i(g^* \cdot (W(\tau_1) - W(\eta_1), \ldots, W(\tau_i) - W(\eta_i)) + s^* (g(W(t) - W(s))))
\]

\[
= \prod_{j=1}^l E(\exp (i g_j (W(\tau_j) - W(\eta_j))) E(\exp (i s^* (g(W(t) - W(s))))
\]

\[
= E(\exp (i(t^* \cdot (W(u_1) - W(r_1), \ldots, W(u_m) - W(r_m)))) \cdot
E(\exp (i s^* (g(W(t) - W(s)))).
\]
By Theorem 16.3.2 it follows the vector \((W(u_1) - W(r_1), \ldots, W(u_m) - W(r_m))\) is independent of the random variable \(g(W(t) - W(s))\) which shows that for \(A \in K\), \(X_A\), measurable in \(\sigma(W(u_1) - W(r_1), \ldots, W(u_m) - W(r_m))\) is independent of \(g(W(t) - W(s))\). Therefore,

\[
\int_K X_A g(W(t) - W(s)) dP = \int_K X_A dP \int_K g(W(t) - W(s)) dP = P(A) \int g(W(t) - W(s)) dP
\]

Thus \(K \subseteq G\) and so by the lemma on \(\pi\) systems, Lemma 16.3.3 on Page 387, it follows \(G \supseteq \sigma(K) \supseteq F_s \supseteq G\). □

**Lemma 16.3.2** Let \(\{W(t)\}\) be a stochastic process having values in a separable Banach space which has the property that if \(t_1 < t_2 < \cdots < t_n\), then the increments,

\[
\{W(t_k) - W(t_{k-1})\}
\]

are independent and integrable and \(E(W(t) - W(s)) = 0\). Suppose also that \(W(t)\) is right continuous, meaning that for \(\omega\) off a set of measure zero, \(t \to W(t)(\omega)\) is right continuous. Also suppose that for some \(q > 1\)

\[
||W(t) - W(s)||_{L^q(\Omega)}
\]

is bounded independent of \(s \leq t\). Then \(\{W(t)\}\) is also a martingale with respect to the normal filtration defined by

\[
F_s = \cap_{1 > u} \sigma(W(u) - W(r) : 0 \leq r < u \leq t)
\]

where this denotes the intersection of the completions of the \(\sigma\) algebras

\[
\sigma(W(u) - W(r) : 0 \leq r < u \leq t)
\]

Also, in the same situation but without the assumption that \(E(W(t) - W(s)) = 0\), if \(t > s\) and \(A \in F_s\) it follows that if \(g\) is a continuous function such that

\[
||g(W(t) - W(s))||_{L^q(\Omega)}
\]

is bounded independent of \(s \leq t\) for some \(q > 1\) then for \(t > s\),

\[
\int_K X_A g(W(t) - W(s)) dP = P(A) \int g(W(t) - W(s)) dP.
\]

**Proof:** Consider first the claim, \([16.3.3]\). To begin with I show that if \(A \in F_s\) then for all \(\varepsilon\) small enough that \(t > s + \varepsilon,\)

\[
\int_K X_A g(W(t) - W(s + \varepsilon)) dP = P(A) \int g(W(t) - W(s + \varepsilon)) dP
\]

(16.3.10)

This will happen if \(X_A\) and \(g(W(t) - W(s + \varepsilon))\) are independent. First note that from the definition

\[
A \in \sigma(W(u) - W(r) : 0 \leq r < u \leq s + \varepsilon)
\]

and so from the process of completion of a measure space, there exists

\[
B \in \sigma(W(u) - W(r) : 0 \leq r < u \leq s + \varepsilon)
\]

such that \(B \supseteq A\) and \(P(B \setminus A) = 0\). Therefore, letting \(\phi \in E'\),

\[
E(\exp(i t X_A + i \phi (g(W(t) - W(s + \varepsilon)))) = E(\exp(i t X_B + i \phi (g(W(t) - W(s + \varepsilon)))) = E(\exp(i t X_B)) E(\exp(i \phi (g(W(t) - W(s + \varepsilon))))
\]

because \(X_B\) is independent of \(g(W(t) - W(s + \varepsilon))\) by Lemma 16.3.1 above. Then the above equals

\[
E(\exp(i t X_A)) E(\exp(i \phi (g(W(t) - W(s + \varepsilon))))
\]

\footnote{Note how the \(\sigma\) algebra \(F_s\) is defined, as the intersection of completions of \(\sigma\) algebras corresponding to \(t\) strictly larger than \(s\).}
Now by Theorem 16.1.3, 16.3.8 follows. Next pass to the limit in both sides of 16.3.8 as ε → 0. One can do this because of 16.3.5 which implies the functions in the integrands are uniformly integrable and Vitali’s convergence theorem, Theorem 11.3.8. This yields 16.4.2.

Now consider the part about the stochastic process being a martingale. Let g be the identity map. If A ∈ Fs, the above implies

\[ \int_A E(W(t) | F_s) dP = \int_A W(t) dP = \int_A (W(t) - W(s)) dP + \int_A W(s) dP \]

\[ = P(A) \int_\Omega (W(t) - W(s)) dP + \int_A W(s) dP = \int_A W(s) dP \]

and so since A is arbitrary, E(W(t) | Fs) = W(s). □

16.4 Hilbert Space Valued Wiener Processes

Next I will consider the case of Hilbert space valued Wiener processes. This will include the case of \( \mathbb{R}^n \) valued Wiener processes. Recall the definition of a real valued Wiener process.

**Definition 16.4.1** Let W(t) be a stochastic process which has the properties that whenever \( t_1 < t_2 < \cdots < t_m \), the increments \( \{W(t_i) - W(t_{i-1})\} \) are independent and whenever \( s < t \), it follows W(t) - W(s) is normally distributed with variance \( t - s \) and mean 0. Also W(0) = 0. This is called a Wiener process.

What follows depends on the following fundamental lemma.

**Lemma 16.4.2** There exists a sequence of real Wiener processes, \( \{\psi_k(t)\}_{k=1}^\infty \) which have the following properties. Let \( t_0 < t_1 < \cdots < t_n \) be an arbitrary sequence. Then the random variables

\[ \{\psi_k(t_q) - \psi_k(t_{q-1}) : (q, k) \in \{(1, 2, \cdots, n) \times (k_1, \cdots, k_m)\}\} \]

are independent. Also each \( \psi_k \) is Hölder continuous with exponent \( \gamma \) for any \( \gamma > 1/2 \) and for each \( m \in \mathbb{N} \) there exists a constant \( C_m \) independent of \( k \) such that

\[ \int_\Omega |\psi_k(t) - \psi_k(s)|^{2m} dP \leq C_m |t - s|^m \]

*Proof:* First, there exists a sequence \( \{\xi_{ij}\}_{(i, j) \in \mathbb{N} \times \mathbb{N}} \) such that the \( \{\xi_{ij}\} \) are independent and each normally distributed with mean 0 and variance 1. This follows from Lemma 11.3.3. Let \( \{\xi_{ij}\}_{i=1}^\infty \) be independent and normally distributed with mean 0 and variance 1. (Let \( \theta \) be a one to one and onto map from \( \mathbb{N} \) to \( \mathbb{N} \times \mathbb{N} \). Then define \( \xi_{ij} \equiv \xi_{\theta^{-1}(i,j)} \).

Let

\[ \psi_k(t) = \sum_{j=1}^\infty \langle \chi_{[0,t]}, g_j \rangle_{L^2} \xi_{kj} \]

(16.4.13)

where \( \{g_j\} \) is an orthonormal basis for \( L^2(0, \infty) \). By Corollary 11.3.12 this defines a real Wiener process satisfying 16.3.12. It remains to show that the random variables

\[ \psi_{kr}(t_q) - \psi_{kr}(t_{q-1}) \]

(16.4.14)

are independent.

Let

\[ P = \sum_{q=1}^n \sum_{r=1}^m s_{qr} (\psi_{kr}(t_q) - \psi_{kr}(t_{q-1})) \]

and consider \( E(e^{itP}) \). I want to use Proposition 16.5.6 on Page 163. To do this I need to show \( E(e^{itP}) \) equals

\[ \prod_{q=1}^n \prod_{r=1}^m E\left(\exp\left(is_{qr} (\psi_{kr}(t_q) - \psi_{kr}(t_{q-1}))\right)\right) \].
Using \( E(e^{itP}) \) equals
\[
E\left(\exp\left(i\sum_{q=1}^{n} \sum_{r=1}^{m} s_{qr} \sum_{j=1}^{\infty} (\mathcal{X}_{[t_{q-1}, t_q]}, g_j)_{L^2} \xi_{k_r,j}\right)\right)
\]

\[
= \lim_{N \to \infty} E\left(\exp\left(i\sum_{q=1}^{n} \sum_{r=1}^{m} s_{qr} \sum_{j=1}^{N} (\mathcal{X}_{[t_{q-1}, t_q]}, g_j)_{L^2} \xi_{k_r,j}\right)\right)
\]

Now the \( \xi_{k_r,j} \) are independent by construction. Therefore, the above equals
\[
= \lim_{N \to \infty} \prod_{q=1}^{n} \prod_{r=1}^{m} \prod_{j=1}^{N} E\left(\exp\left(i s_{qr} (\mathcal{X}_{[t_{q-1}, t_q]}, g_j)_{L^2} \xi_{k_r,j}\right)\right)
\]
\[
= \lim_{N \to \infty} \prod_{q=1}^{n} \prod_{r=1}^{m} \prod_{j=1}^{N} \exp\left(-\frac{1}{2} s_{qr}^2 (\mathcal{X}_{[t_{q-1}, t_q]}, g_j)_{L^2}^2\right)
\]
\[
= \prod_{q=1}^{n} \prod_{r=1}^{m} \exp\left(-\frac{1}{2} s_{qr}^2 (t_q - t_{q-1})\right)
\]
\[
= \prod_{q=1}^{n} \prod_{r=1}^{m} E\left(\exp\left(is_{qr} (\psi_{k_r}(t_q) - \psi_{k_r}(t_{q-1}))\right)\right)
\]

because \( \psi_{k_r}(t_q) - \psi_{k_r}(t_{q-1}) \) is normally distributed with variance \( t_q - t_{q-1} \) and mean 0. By Proposition 16.3.4 on Page 185, it follows the random variables of \( \{\xi_{k_r,j}\}_{j=1}^{\infty} \) are independent. Note that as a special case, this also shows the random variables, \( \{\psi_{k_r}(t)\}_{r=1}^{\infty} \) are independent due to the fact \( \psi_{k_r}(0) = 0 \). □

**Definition 16.4.3** Let \( W(t) \) be a stochastic process with values in \( H \), a real separable Hilbert space which has the properties that \( t \to W(t, \omega) \) is continuous, whenever \( t_1 < t_2 < \cdots < t_m \), the increments \( \{W(t_i) - W(t_{i-1})\} \) are independent, \( W(0) = 0 \), and whenever \( s < t \),
\[
\mathcal{L}(W(t) - W(s)) = N(0, (t - s)Q)
\]

which means that whenever \( h \in H \),
\[
\mathcal{L}(h, W(t) - W(s)) = N(0, (t - s)(Qh, h))
\]

Also
\[
E((h_1, W(t) - W(s))(h_2, W(t) - W(s))) = (Qh_1, h_2)(t - s).
\]

Here \( Q \) is a nonnegative trace class operator. Recall this means
\[
Q = \sum_{i=1}^{\infty} \lambda_i e_i \otimes e_i
\]
where \( \{e_i\} \) is a complete orthonormal basis, \( \lambda_i \geq 0 \), and
\[
\sum_{i=1}^{\infty} \lambda_i < \infty
\]

Such a stochastic process is called a \( Q \) Wiener process. In the case where these have values in \( \mathbb{R}^n \), \( tQ \) ends up being the covariance matrix of \( W(t) \).

With this and the definition, one can describe Hilbert space valued Wiener processes in a fairly general setting.
**Definition 16.4.4** Let $H$ and $G$ be two separable Hilbert spaces and let $T$ map $H$ to $G$ be linear. Then $T$ is called a Hilbert Schmidt operator if there exists some orthonormal basis for $H$, $\{e_j\}$ such that

$$
\sum_j ||Te_j||^2 < \infty.
$$

The collection of all such linear maps will be denoted by $L_2(H,G)$.

**Theorem 16.4.5** Let $U$ be a real separable Hilbert space and let $J : U_0 \to U$ be a Hilbert Schmidt operator where $U_0$ is a real separable Hilbert space. Then let $\{g_k\}$ be a complete orthonormal basis for $U_0$ and define for $t \in [0,T]$

$$
W(t) \equiv \sum_{k=1}^{\infty} \psi_k(t) Jg_k.
$$

Then $W(t)$ is a $Q$ Wiener process for $Q = JJ^*$ as in Definition 16.4.3. Furthermore, the distribution of $W(t) - W(s)$ is the same as the distribution of $W(t-s)$, and $W$ is Holder continuous with exponent $\gamma$ for any $\gamma < 1/2$. There also is a subsequence denoted by $N$ such that the convergence of the series

$$
\sum_{k=1}^{N} \psi_k(t) Jg_k
$$

is uniform for all $\omega$ not in some set of measure zero.

**Proof:** First it is necessary to show the series converges in $L^2(\Omega;U)$ for each $t$. For convenience I will consider the series for $W(t) - W(s)$. (Always, it is assumed $t > s$.) Then since $\psi_k(t) - \psi_k(s)$ is normal with mean 0 and variance $(t-s)$ and $\psi_k(t) - \psi_k(s)$ and $\psi_l(t) - \psi_l(s)$ are independent,

$$
\int_{\Omega} \left| \sum_{k=m}^{n} (\psi_k(t) - \psi_k(s)) Jg_k \right|^2 dP_U = \int_{\Omega} \sum_{k,l=m}^{n} ((\psi_k(t) - \psi_k(s)) Jg_k, (\psi_l(t) - \psi_l(s)) Jg_l)
$$

$$
= (t-s) \sum_{k=m}^{n} (Jg_k, Jg_k) = (t-s) \sum_{k=m}^{n} ||Jg_k||^2_U
$$

which converges to 0 as $m,n \to \infty$ thanks to the assumption that $J$ is Hilbert Schmidt. It follows the above sum converges in $L^2(\Omega;U)$. Now letting $m < n$, it follows by the maximal estimate, Theorem 16.4.2, and the above

$$
P \left( \sup_{t \in [0,T]} \left| \sum_{k=1}^{m} \psi_k(t) Jg_k - \sum_{k=1}^{n} \psi_k(t) Jg_k \right|_U \geq \lambda \right) \leq \frac{1}{\lambda^2} E \left( \sum_{k=m+1}^{n} \psi_k(T) Jg_k \right)^2_U \leq \frac{1}{\lambda^2} T \sum_{k=m}^{n} ||Jg_k||^2_U
$$

and so there exists a subsequence $n_l$ such that for all $p \geq 0$,

$$
P \left( \sup_{t \in [0,T]} \left| \sum_{k=1}^{n_l} \psi_k(t) Jg_k - \sum_{k=1}^{n_l+p} \psi_k(t) Jg_k \right|_U \geq 2^{-l} \right) < 2^{-l}
$$

Therefore, by Borel Cantelli lemma, there is a set of measure zero such that for $\omega$ not in this set,

$$
\lim_{l \to \infty} \sum_{k=1}^{n_l} \psi_k(t) Jg_k = \sum_{k=1}^{\infty} \psi_k(t) Jg_k
$$

is uniform on $[0,T]$. From now on denote this subsequence by $N$ to save on notation.
I need to consider the characteristic function of \((h, W(t) - W(s))_U\) for \(h \in U\). Then

\[
E \left( \exp \left( i r (h, (W(t) - W(s))_U) \right) \right)
= \lim_{N \to \infty} E \left( \exp \left( i r \left( \sum_{j=1}^{N} (\psi_j(t) - \psi_j(s)) (h, Jg_j) \right) \right) \right)
= \lim_{N \to \infty} E \left( \prod_{j=1}^{N} e^{ir(\psi_j(t) - \psi_j(s))(h,Jg_j)} \right)
\]

Since the random variables \(\psi_j(t) - \psi_j(s)\) are independent,

\[
= \lim_{N \to \infty} \prod_{j=1}^{N} E \left( e^{ir(h,Jg_j)(\psi_j(t) - \psi_j(s))} \right)
\]

Since \(\psi_j(t) - \psi_j(s)\) is a Gaussian random variable having mean 0 and variance \((t-s)\), the above equals

\[
= \lim_{N \to \infty} \prod_{j=1}^{N} e^{-\frac{1}{2} r^2 (h,Jg_j)^2 (t-s)}
\]

\[
= \lim_{N \to \infty} \exp \left( \sum_{j=1}^{N} -\frac{1}{2} r^2 (h,Jg_j)^2 (t-s) \right)
= \exp \left( -\frac{1}{2} r^2 (t-s) \sum_{j=1}^{\infty} (h,Jg_j)^2_U \right)
= \exp \left( -\frac{1}{2} r^2 (t-s) \sum_{j=1}^{\infty} (J^*h,Jg_j)^2_U \right)
= \exp \left( -\frac{1}{2} r^2 (t-s) \|J^*h\|^2_U \right) = \exp \left( -\frac{1}{2} r^2 (t-s) \|J^*h,h\|^2_U \right)
= \exp \left( -\frac{1}{2} r^2 (t-s) (Qh,h)_U \right)
\]

(16.4.15)

which shows \((h, W(t) - W(s))_U\) is normally distributed with mean 0 and variance \((t-s)(Qh,h)_U\) where \(Q \equiv J^*\). It is obvious from the definition that \(W(0) = 0\). Note that \(Q\) is of trace class because if \(\{e_k\}\) is an orthonormal basis for \(U\),

\[
\sum_k (Qe_k,e_k)_U = \sum_k \|J^*e_k\|^2_U = \sum_k \sum_l (J^*e_k,g_l)^2_U = \sum_k \sum_l (e_k,Jg_l)^2_U = \sum_l \|Jg_l\|^2_U < \infty
\]

To find the covariance, consider

\[
E \left( (h_1, W(t) - W(s)) (h_2, W(t) - W(s)) \right),
\]

This equals

\[
E \left( \sum_{k=1}^{\infty} (\psi_k(t) - \psi_k(s)) (h_1, Jg_k) \sum_{j=1}^{\infty} (\psi_j(t) - \psi_j(s)) (h_2, Jg_j) \right).
\]

Since the series converge in \(L^2(\Omega; U)\), the independence of the \(\psi_k(t) - \psi_k(s)\) implies the above equals

\[
= \lim_{n \to \infty} E \left( \sum_{k=1}^{n} (\psi_k(t) - \psi_k(s)) (h_1, Jg_k) \sum_{j=1}^{n} (\psi_j(t) - \psi_j(s)) (h_2, Jg_j) \right)
\]
Thus the above equals \( (t-s) \sum_{k=1}^{n} (h_1, Jg_k) (h_2, Jg_k) \)

\[
= \lim_{n \to \infty} (t-s) \sum_{k=1}^{n} (J^*h_1, g_k)_{U_0} (J^*h_2, g_k)_{U_0}
\]

\[
= (t-s) \sum_{k=1}^{\infty} (J^*h_1, g_k)_{U_0} (J^*h_2, g_k)_{U_0}
\]

\[
= (t-s) (J^*h_1, J^*h_2) = (t-s) (Qh_1, h_2).
\]

Next consider the claim that the increments are independent. Let \( W^N(t) \) be given by the appropriate partial sum and let \( \{h_j\}_{j=1}^{m} \) be a finite list of vectors of \( U \). Then from the independence properties of \( \psi_j \) explained above,

\[
E \left( \exp \sum_{j=1}^{m} i (h_j, W^N(t_j) - W^N(t_{j-1}))_{U} \right)
\]

\[
E \left( \exp \sum_{j=1}^{m} i (h_j, \sum_{k=1}^{N} Jg_k (\psi_k(t_j) - \psi_k(t_{j-1}))_{U} \right)
\]

\[
= E \left( \sum_{j=1}^{m} \sum_{k=1}^{N} i (h_j, Jg_k)_{U} (\psi_k(t_j) - \psi_k(t_{j-1})) \right)
\]

\[
= E \left( \prod_{j,k} \exp \left( i (h_j, Jg_k)_{U} (\psi_k(t_j) - \psi_k(t_{j-1})) \right) \right)
\]

This can be done because of the independence of the random variables \( \{\psi_k(t_j) - \psi_k(t_{j-1})\}_{j,k} \).

Thus the above equals

\[
\prod_{j,k} \exp \left( -\frac{1}{2} (h_j, Jg_k)_{U}^2 (t_j - t_{j-1}) \right)
\]

\[
= \prod_{j=1}^{m} \exp \left( -\frac{1}{2} \sum_{k=1}^{N} (h_j, Jg_k)_{U}^2 (t_j - t_{j-1}) \right)
\]

because \( \psi_k(t_j) - \psi_k(t_{j-1}) \) is normally distributed having variance \( t_j - t_{j-1} \). Now letting \( N \to \infty \), this implies

\[
E \left( \exp \sum_{j=1}^{m} i (h_j, W(t_j) - W(t_{j-1}))_{U} \right)
\]

\[
= \prod_{j=1}^{m} \exp \left( -\frac{1}{2} \sum_{k=1}^{\infty} (h_j, Jg_k)_{U}^2 (t_j - t_{j-1}) \right)
\]

\[
= \prod_{j=1}^{m} \exp \left( -\frac{1}{2} (t_j - t_{j-1}) \sum_{k=1}^{\infty} (J^*h_j, g_k)_{U_0}^2 \right)
\]

\[
= \prod_{j=1}^{m} \exp \left( -\frac{1}{2} (t_j - t_{j-1}) \|J^*h_j\|_{U_0}^2 \right)
\]

\[
= \prod_{j=1}^{m} \exp \left( -\frac{1}{2} (t_j - t_{j-1}) (Qh_j, h_j)_{U} \right)
\]

\[
= \prod_{j=1}^{m} \exp \left( i (h_j, W(t_j) - W(t_{j-1}))_{U} \right) \quad (16.4.16)
\]
from 16.4.15, letting \( r = 1 \), By Theorem 11.5.3 on Page 189, this shows the increments are independent.

It remains to verify the Hölder continuity. Recall

\[
W(t) = \sum_{k=1}^{\infty} Jg_k \psi_k(t)
\]

where \( \psi_k \) is a real Wiener process.

Next consider the claim about Hölder continuity. It was shown above that

\[
E(\exp(i r (h, (W(t) - W(s)))) U) = \exp\left(-\frac{1}{2} r^2 (t - s) (Qh, h)_U \right)
\]

Therefore, taking a derivative with respect to \( r \) two times yields

\[
E \left( \left( - (h, (W(t) - W(s)))_U \right)^2 \exp(i r (h, (W(t) - W(s)))) U \right) = -(t - s) (Qh, h) \exp\left(-\frac{1}{2} r^2 (t - s) (Qh, h)_U \right) + r^2 (t - s)^2 (Qh, h)_U^2 \exp\left(-\frac{1}{2} r^2 (t - s) (Qh, h)_U \right)
\]

Now plug in \( r = 0 \) to obtain

\[
E \left( (h, (W(t) - W(s)))^2 U \right) = (t - s) (Qh, h).
\]

Similarly, taking 4 derivatives, it follows that an expression of the following form holds.

\[
E \left( (h, (W(t) - W(s)))^4 U \right) = C_2 (Qh, h)^2 (t - s)^2,
\]

and in general,

\[
E \left( (h, (W(t) - W(s)))^{2m} U \right) = C_m (Qh, h)^m (t - s)^m.
\]

Now it follows from Minkowsky’s inequality applied to the two integrals \( \sum_{i=1}^{\infty} \) and \( \int_{\Omega} \) that

\[
\left[ E \left( |W(t) - W(s)|^{2m} \right) \right]^{1/m} = \left[ E \left( \left( \sum_{k=1}^{\infty} (e_k, W(t) - W(s))^2 \right)^m \right) \right]^{1/m} \leq \sum_{k=1}^{\infty} \left[ E \left( (e_k, W(t) - W(s))^{2m} \right) \right]^{1/m} \leq \sum_{k=1}^{\infty} [C_m (Qe_k, e_k)^m (t - s)^m]^{1/m} = C_m^{1/m} |t - s| \left( \sum_{k=1}^{\infty} (Qe_k, e_k) \right) \equiv C_m |t - s|.
\]

Hence there exists a constant \( C_m \) such that

\[
E \left( |W(t) - W(s)|^{2m} \right) \leq C_m |t - s|^m
\]

By the Kolmogorov Čentsov Theorem, Theorem 13.3.2, it follows that off a set of measure 0, \( t \to W(t, \omega) \) is Hölder continuous with exponent \( \gamma \) such that

\[
\gamma < \frac{m - 1}{2m}, \ m > 2.
\]

Finally, from 16.4.15 with \( r = 1 \),

\[
E (\exp i (h, W(t) - W(s))_U) = \exp\left(-\frac{1}{2} (t - s) (Qh, h) \right)
\]
which is the same as \( E (\exp i(h,W(t−s)))_U \) due to the fact \( W(0) = 0 \).

The above has shown that \( W(t) \) satisfies the conditions of Lemma 16.3.2 and so it is a martingale with respect to the filtration given there. What is its quadratic variation?

\[
E \left( ||W(t)||^2 \right) = \sum_{k=1}^{\infty} E ( (W(t),e_k)(W(t),e_k) ) = \sum_{k=1}^{\infty} (Qe_k,e_k) t = \text{trace}(Q)t
\]

Is it the case that \( ||W(t)|| = \text{trace}(Q)t \)? Let the filtration be as in Lemma 16.3.2 and let \( A \in \mathcal{F}_s \). Then using the result of that lemma,

\[
\int_A \left( ||W(t)||^2 − t \text{trace}(Q) |\mathcal{F}_s \right) dP
\]

\[
= \int_A \left( ||W(t)−W(s)||^2 + 2(W(t),W(s)) − ||W(s)||^2 \right) \left( −(t−s) \text{trace}Q − \text{trace}Qs|\mathcal{F}_s \right) dP
\]

\[
= P(A) \int_Q ||W(t)−W(s)||^2 − (t−s) \text{trace}QdP
\]

\[
+ \int_A \left( 2(W(t),W(s)) − ||W(s)||^2 − \text{trace}(Q)s|\mathcal{F}_s \right) dP
\]

Similar to the above, \( E \left( ||W(t)−W(s)||^2 − (t−s) \text{trace}Q \right) = 0 \) and so

\[
= \int_A \left( 2(W(s),E(W(t)|\mathcal{F}_s)) dP − \int_A ||W(s)||^2 dP − \int_A s \text{trace}QdP \right)
\]

\[
= \int_A \left( ||W(s)||^2 − s \text{trace}Q \right) dP
\]

Thus \( ||W(t)||^2 − t \text{trace}(Q) \) is a martingale so \( t \text{trace}(Q) \) must be the quadratic variation by uniqueness of the quadratic variation.

Note the characteristic function of a \( Q \) Wiener process is

\[
E \left( e^{ih,W(t)} \right) = e^{-\frac{1}{2}t^2(Qh,h)}
\]

By Theorem 16.3.2, if you simply say that the distribution measure of \( W(t) \) is Gaussian, then it follows there exists a trace class operator \( Q \) and \( m_t \in H \) such that this measure is \( N(m_t,Q_t) \). Thus for \( W(t) \) a Wiener process, \( Q_t = tQ \) and \( m_t = 0 \). In addition, the increments are independent so this is much more specific than the earlier definition of a Gaussian measure.

What is a \( Q \) Wiener process if the Hilbert space is \( \mathbb{R}^n \)? In particular, what is \( Q \)? It is given that

\[
\mathcal{L}((h,W(t) − W(s))) = N(0,(t−s)(Qh,h))
\]

In this case everything is a vector in \( \mathbb{R}^n \) and so for \( h \in \mathbb{R}^n \),

\[
E \left( e^{i\lambda(h,W(t)−W(s))} \right) = e^{-\frac{1}{2}\lambda^2(t−s)(Qh,h)}
\]

In particular, letting \( \lambda = 1 \) this shows \( W(t) − W(s) \) is normally distributed with covariance \( (t−s)Q \) because its characteristic function is \( e^{-\frac{1}{2}h^*(t−s)Qh} \).

Given a Hilbert Schmidt operator \( J : U_0 \rightarrow U \) one can define a \( Q \) Wiener process. The next theorem deals with the situation where you are given a trace class operator \( Q \) defined on \( U \). It shows that in this case you can obtain a \( Q \) Wiener process.

**Theorem 16.4.6** Let \( U \) be a real separable Hilbert space and let \( Q \) be a nonnegative trace class operator defined on \( U \). Then there exists a \( Q \) Wiener process as defined in Definition 16.4.2. Furthermore, the distribution of \( W(t) − W(s) \) is the same as the distribution of \( W(t−s) \) and \( W \) is Holder continuous with exponent \( \gamma \) for any \( \gamma < 1/2 \).
Proof: One can obtain this theorem as a corollary of Theorem 16.4.5 but this will not be done here. Let

$$Q = \sum_{i=1}^{\infty} \lambda_i e_i \otimes e_i$$

where \(\{e_i\}\) is a complete orthonormal set and \(\lambda_i \geq 0\) and \(\sum \lambda_i < \infty\). Now the definition of the \(Q\) Wiener process is

$$W(t) \equiv \sum_{k=1}^{\infty} \sqrt{\lambda_k} e_k \psi_k(t)$$  \hspace{1cm} (16.4.18)

where \(\{\psi_k(t)\}\) are the real Wiener processes defined in Lemma 16.4.2.

Now consider (16.4.18). From this formula, if \(s < t\)

$$W(t) - W(s) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} e_k (\psi_k(t) - \psi_k(s))$$  \hspace{1cm} (16.4.19)

First it is necessary to show this sum converges. Since \(\psi_j(t)\) is a Wiener process,

$$\int_{\Omega} \left| \sum_{j=m}^{n} \sqrt{\lambda_j} (\psi_j(t) - \psi_j(s)) e_j \right|^2 dP_U$$

$$= \int_{\Omega} \sum_{j=m}^{n} \lambda_j (\psi_j(t) - \psi_j(s))^2 dP$$

$$= (t-s) \sum_{j=m}^{n} \lambda_j$$

and this converges to 0 as \(m, n \to \infty\) because it was given that

$$\sum_{j=1}^{\infty} \lambda_j < \infty$$

so the series in (16.4.19) converges in \(L^2(\Omega; U)\).

Therefore, there exists a subsequence

$$\left\{ \sum_{k=1}^{N} \sqrt{\lambda_k} e_k (\psi_k(t) - \psi_k(s)) \right\}$$

which converges pointwise a.e. to \(W(t) - W(s)\) as well as in \(L^2(\Omega; U)\) as \(N \to \infty\). Then letting \(h \in U\),

$$(h, W(t) - W(s))_U = \sum_{k=1}^{\infty} \sqrt{\lambda_k} (\psi_k(t) - \psi_k(s)) (h, e_k)$$  \hspace{1cm} (16.4.20)

Then by the dominated convergence theorem,

$$E(\exp(\imath r (h, (W(t) - W(s))_U)))$$

$$= \lim_{N \to \infty} E \left( \exp \left( \imath r \left( \sum_{j=1}^{N} \sqrt{\lambda_j} (\psi_j(t) - \psi_j(s)) (h, e_j) \right) \right) \right)$$

$$= \lim_{N \to \infty} E \left( \prod_{j=1}^{N} e^{\imath r \sqrt{\lambda_j} (\psi_j(t) - \psi_j(s))(h, e_j)} \right)$$

Since the random variables \(\psi_j(t) - \psi_j(s)\) are independent,

$$= \lim_{N \to \infty} \prod_{j=1}^{N} E \left( e^{\imath r \sqrt{\lambda_j} (\psi_j(t) - \psi_j(s))(h, e_j)} \right)$$
Since \( \psi_j(t) \) is a real Wiener process,

\[
\lim_{N \to \infty} \prod_{j=1}^{N} e^{-\frac{1}{2}r^2\lambda_j(t-s)(h,e_j)^2}
\]

\[
= \lim_{N \to \infty} \exp \left( \sum_{j=1}^{N} -\frac{1}{2}r^2\lambda_j (t-s)(h,e_j)^2 \right)
\]

\[
= \exp \left( -\frac{1}{2}r^2 (t-s) \sum_{j=1}^{\infty} \lambda_j (h,e_j)^2 \right)
\]

\[
= \exp \left( -\frac{1}{2}r^2 (t-s) (Qh,h) \right)
\]

(16.4.21)

Thus \((h,W(t) - W(s))\) is normally distributed with mean 0 and variance \((t-s)(Qh,h)\). It is obvious from the definition that \(W(0) = 0\). Also to find the covariance, consider

\[
E ((h_1,W(t) - W(s)) (h_2,W(t) - W(s)))
\]

and use \(\text{Law of Large Numbers}\) to obtain this is equal to

\[
E \left( \sum_{k=1}^{\infty} \sqrt{\lambda_k} (\psi_k(t) - \psi_k(s)) (h_1,e_k) \sum_{j=1}^{\infty} \sqrt{\lambda_j} (\psi_j(t) - \psi_j(s)) (h_2,e_j) \right)
\]

\[
= \lim_{n \to \infty} E \left( \sum_{k=1}^{n} \sqrt{\lambda_k} (\psi_k(t) - \psi_k(s)) (h_1,e_k) \sum_{j=1}^{n} \sqrt{\lambda_j} (\psi_j(t) - \psi_j(s)) (h_2,e_j) \right)
\]

\[
= \lim_{n \to \infty} (t-s) \sum_{k=1}^{n} \lambda_k (h_1,e_k) (h_2,e_j) = (t-s) (Qh_1,h_2)
\]

(Recall \(Q \equiv \sum_k \lambda_k e_k \otimes e_k\).)

Next I show the increments are independent. Let \(N\) be the subsequence defined above and let \(W^N(t)\) be given by the appropriate partial sum and let \(\{h_j\}_{j=1}^{m}\) be a finite list of vectors of \(U\). Then from the independence properties of \(\psi_j\) explained above,

\[
E \left( \exp \sum_{j=1}^{m} i (h_j, W^N(t_j) - W^N(t_{j-1})) \right)
\]

\[
E \left( \exp \sum_{j=1}^{m} i (h_j, \sum_{k=1}^{N} \sqrt{\lambda_k} e_k (\psi_k(t_j) - \psi_k(t_{j-1})) ) \right)
\]

\[
= \prod_{j,k} \exp \left( i \sqrt{\lambda_k} (h_j,e_k) U (\psi_k(t_j) - \psi_k(t_{j-1})) \right)
\]

This can be done because of the independence of the random variables

\[
\{\psi_k(t_j) - \psi_k(t_{j-1})\}_{j,k}
\].
Thus the above equals

\[
= \prod_{j=1}^{m} \exp \left( -\frac{1}{2} \sum_{k=1}^{N} \lambda_k (h_j, e_k)^2 (t_j - t_{j-1}) \right)
\]

because \( \psi_k(t_j) - \psi_k(t_{j-1}) \) is normally distributed having variance \( t_j - t_{j-1} \) and mean 0. Now letting \( N \to \infty \), this implies

\[
E \left( \exp \sum_{j=1}^{m} i (h_j, W(t_j) - W(t_{j-1})) \right) = \prod_{j=1}^{m} \exp \left( -\frac{1}{2} (t_j - t_{j-1}) \sum_{k=1}^{\infty} \lambda_k (h_j, e_k)^2 U(t_j - t_{j-1}) \right)
\]

because of the fact shown above that \( (h, W(t) - W(s)) \) is normally distributed with mean 0 and variance \( t - s \) \( (Qh, h) \).

By Theorem 11.5.3 on Page 355, this shows the increments are independent.

Next consider the continuity assertion. Recall

\[
W(t) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} e_k \psi_k(t)
\]

where \( \psi_k \) is a real Wiener process. Therefore, letting \( 2m > 2, m \in \mathbb{N} \) and using Lemma 16.4.2 for \( \psi_k \) and Jensen’s inequality along with Lemma 16.4.3,

\[
E \left( \left| W(t) - W(s) \right|^{2m} \right) = E \left( \left| \sum_{k=1}^{\infty} \sqrt{\lambda_k} e_k (\psi_k(t) - \psi_k(s)) \right|^{2m} \right)
\]

\[
= E \left( \left( \sum_{k=1}^{\infty} \lambda_k \left| \psi_k(t) - \psi_k(s) \right|^{2m} \right)^{m} \right)
\]

\[
\leq E \left( \left( \sum_{k=1}^{\infty} \lambda_k \right)^{m-1} \sum_{k=1}^{\infty} \lambda_k \left| \psi_k(t) - \psi_k(s) \right|^{2m} \right)
\]

\[
\leq C_m \sum_{k=1}^{\infty} \lambda_k E \left( \left| \psi_k(t) - \psi_k(s) \right|^{2m} \right)
\]

\[
\leq C_m |t - s|^m
\]

By the Kolmogorov Čentsov Theorem, Theorem 13.3.2, it follows that off a set of measure 0, \( t \to W(t, \omega) \) is Hölder continuous with exponent \( \gamma \) such that

\[
\gamma < \frac{m - 1}{2m}.
\]

Finally, from 16.4.21 taking \( r = 1 \),

\[
E \left( \exp i (h, W(t) - W(s)) U \right) = \exp \left( -\frac{1}{2} (t - s) (Qh, h) \right)
\]

which is the same as \( E \left( \exp i (h, W(t - s)) U \right) \) due to the fact \( W(0) = 0 \). This proves the theorem.

The above shows there exists \( Q \) Wiener processes in any separable Hilbert space provided \( Q \) is a nonnegative trace class operator. Next I will show the way described above is the only way it can happen.
Theorem 16.4.7 Suppose \{W(t)\} is a Q Wiener process in U, a real separable Hilbert space. Then letting

$$Q = \sum_{k=1}^{\infty} \lambda_k e_k \otimes e_k$$

where the \{e_k\} are orthonormal, \(\lambda_k \geq 0\), and \(\sum_{k=1}^{\infty} \lambda_k < \infty\), it follows

$$W(t) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \psi_k(t) e_k$$

(16.4.25)

where

$$\psi_k(t) = \begin{cases} \frac{1}{\sqrt{\lambda_k}} (W(t), e_k)_U & \text{if } \lambda_k \neq 0 \\ 0 & \text{if } \lambda_k = 0 \end{cases}$$

then \{\psi_k(t)\} is a Wiener process and for \(t_0 < t_1 < \cdots < t_n\) the random variables

$$\{\psi_k(t_q) - \psi_k(t_{q-1}) : (q,k) \in (1,2,\ldots,n) \times (k_1,\ldots,k_m)\}$$

are independent. Furthermore, the sum in (16.4.25) converges uniformly for a.e. \(\omega\) on any closed interval, \([0,T]\).

Proof: First of all, the fact that \(W(t)\) has values in \(U\) and that \{\(e_k\)\} is an orthonormal basis implies the sum in (16.4.25) converges for each \(\omega\). Consider

$$E(\exp(it\psi_k(t))) = E\left(\exp\left(\frac{i}{\sqrt{\lambda_k}} (W(t), e_k)_U \right)\right)$$

Since \(W(t)\) is given to be a Q Wiener process, \((W(t), e_k)_U\) is normally distributed with mean 0 and variance \(t(Q\eta, \eta)\). Therefore, the above equals

$$e^{-\frac{1}{2}t^2 \frac{1}{\lambda_k} t(Q\eta, \eta)} = e^{-\frac{1}{2}t^2 \frac{1}{\lambda_k} \lambda_k} = e^{-\frac{1}{2}t^2},$$

the characteristic function for a random variable which is \(N(0,t)\). The independence of the increments for a given \(\psi_k(t)\) follows right away from the independence of the increments of \(W(t)\) and the distribution of the increments being \(N(0,(t-s))\) follows similarly to the above.

For \(t_1 < t_2 < \cdots < t_n\), why are the random variables,

$$\{(W(t_q), e_k)_U - (W(t_{q-1}), e_k)_U : (q,k) \in (1,2,\ldots,n) \times (k_1,\ldots,k_m)\}$$

(16.4.26)

independent? Let

$$P = \sum_{q=1}^{n} \sum_{j=1}^{m} s_{kj} \left((W(t_q), e_k)_U - (W(t_{q-1}), e_k)_U\right)$$

and consider \(E(e^{iP})\). This equals

$$e^{iP} = E\left(\exp\left(i \sum_{q=1}^{n} \sum_{j=1}^{m} s_{kj} \left((W(t_q), e_k)_U - (W(t_{q-1}), e_k)_U\right)\right)\right)$$

(16.4.27)

$$= E\left(\exp\left(i \sum_{q=1}^{n} \left((W(t_q), \sum_{j=1}^{m} s_{jq} e_{kj}_U - (W(t_{q-1}), \sum_{j=1}^{m} s_{jq} e_{kj})_U\right)\right)\right)$$

$$= E\left(\prod_{q=1}^{n} \exp\left(i \left(W(t_q) - W(t_{q-1}), \sum_{j=1}^{m} s_{jq} e_{kj}\right)_U\right)\right)$$

Now recall that by assumption the increments \(W(t) - W(s)\) are independent. Therefore, the above equals

$$\prod_{q=1}^{n} E\left(\exp\left(i \left(W(t_q) - W(t_{q-1}), \sum_{j=1}^{m} s_{jq} e_{kj}\right)_U\right)\right)$$
Recall that by assumption \((W(t) - W(s), h)_U\) is normally distributed with variance \((t - s)(Qh, h)\) and mean 0. Therefore, the above equals
\[
\prod_{q=1}^{n} \exp \left( -\frac{1}{2} (t_q - t_{q-1}) \left( \sum_{j=1}^{m} s_{qj} e_{k_j} \sum_{j=1}^{m} s_{qj} e_{k_j} \right) \right)
\]
\[
= \prod_{q=1}^{n} \exp \left( -\frac{1}{2} (t_q - t_{q-1}) \sum_{j=1}^{m} s_{qj}^2 \lambda_k \right)
\]
\[
\exp \left( -\frac{1}{2} \sum_{q=1}^{n} \sum_{j=1}^{m} (t_q - t_{q-1}) s_{qj}^2 \lambda_k \right)
\]
(16.4.28)

Also
\[
\prod_{q=1}^{n} \prod_{j=1}^{m} E \left( \exp \left( \text{i}s_{qj} \left( (W(t_q), e_{k_j})_U - (W(t_{q-1}), e_{k_j})_U \right) \right) \right)
\]
\[
= \prod_{q=1}^{n} \prod_{j=1}^{m} \exp \left( -\frac{1}{2} (t_q - t_{q-1}) s_{qj}^2 (Qe_{k_j}, e_{k_j}) \right)
\]
\[
= \prod_{q=1}^{n} \prod_{j=1}^{m} \exp \left( -\frac{1}{2} (t_q - t_{q-1}) s_{qj}^2 \lambda_k \right)
\]
\[
= \exp \left( -\frac{1}{2} \sum_{q=1}^{n} \sum_{j=1}^{m} (t_q - t_{q-1}) s_{qj}^2 \lambda_k \right)
\]
(16.4.29)

Therefore, \(e^{IP}\) equals the expression in (16.4.26) because both equal the expression in (16.4.28) and it follows from Proposition (16.4.26) on Page 357 that the random variables of (16.4.26) are independent.

What about the claim of uniform convergence? By the independence of the increments, it follows from Lemma (16.4.29) that \(\{W(t)\}\) is a martingale and each real valued function, \((W(t), e_k)_U\) is also a martingale. Therefore, Theorem (16.4.28) can be applied to conclude
\[
P \left( \left[ \sup_{t \in [0,T]} \left| \sum_{k=m}^{n} (W(t), e_k)_U e_k \right| \geq \alpha \right] \right) \leq \frac{1}{\alpha} \int_{\Omega} \left| \sum_{k=m}^{n} (W(T), e_k)_U e_k \right| dP
\]
\[
\leq \frac{1}{\alpha} \int_{\Omega} \left| \sum_{k=m}^{n} (W(T), e_k)_U e_k \right|^2 dP = \frac{1}{\alpha} \sum_{k=m}^{n} (W(T), e_k)_U^2 dP
\]
\[
= \frac{1}{\alpha} \sum_{k=m}^{n} (Qe_k, e_k) T = \frac{T}{\alpha} \sum_{k=m}^{n} \lambda_k \leq \frac{T}{\alpha} \sum_{k=m}^{\infty} \lambda_k
\]

Since \(\sum_{k=1}^{\infty} \lambda_k < \infty\), there exists a sequence, \(\{m_l\}\) such that if \(n > m_l\)
\[
P \left( \left[ \sup_{t \in [0,T]} \left| \sum_{k=m_l}^{n} (W(t), e_k)_U e_k \right| > 2^{-k} \right] \right) < 2^{-k}
\]
and so by the Borel Cantelli lemma, off a set of measure 0 the partial sums
\[
\left\{ \sum_{k=1}^{m} (W(t), e_k)_U e_k \right\}
\]
converge uniformly on \([0, T]\). This is very interesting but more can be said. In fact the original partial sums converge. Recall Lemma (16.4.28) stated below for convenience.
Lemma 16.4.8 Let \( \{ \zeta_k \} \) be a sequence of random variables having values in a separable real Banach space, \( E \) whose distributions are symmetric. Letting \( S_k \equiv \sum_{i=1}^{k} \zeta_i \), suppose \( \{ S_n \} \) converges a.e. Also suppose that for every \( m > n_k \),

\[
P \left( \| S_m - S_n \|_E > 2^{-k} \right) < 2^{-k}.
\]

(16.4.30)

Then in fact,

\[
S_k (\omega) \to S (\omega) \text{ a.e.} \omega
\]

(16.4.31)

Apply this lemma to the situation in which the Banach space, \( E \) is \( C ([0,T] ; U) \). Then you can conclude uniform convergence of the partial sums,

\[
\sum_{k=1}^{m} (W(t), e_k) U e_k.
\]

This proves the theorem.

Why is \( C ([0,T] ; E) \) separable? You can assume without loss of generality that the interval is \( [0,1] \) and consider the Bernstein polynomials

\[
p_n (t) \equiv \sum_{k=0}^{n} \left( \begin{array}{c} n \\ k \end{array} \right) f \left( \frac{k}{n} \right) t^k (1-t)^{n-k}
\]

These converge uniformly to \( f \). Now look at all polynomials of the form \( \sum_{k=0}^{n} a_k t^k (1-t)^k \) where the \( a_k \) is one of the countable dense set and \( n \in \mathbb{N} \). Each Bernstein polynomial uniformly close to one of these and also uniformly close to \( f \). Hence polynomials of this sort are countable and dense in \( C ([0,T] ; E) \).

16.5 Wiener Processes, Another Approach

16.5.1 Lots Of Independent Normally Distributed Random Variables

You can use the Kolmogorov extension theorem to prove the following corollary. It is Corollary \[\text{on Page} \ 216 \]\[\text{Page} \ 358 \]\[\text{CHAPTER 16. WIENER PROCESSES} \]

Corollary 16.5.1 Let \( H \) be a real Hilbert space. Then there exist real valued random variables \( W(h) \) for \( h \in H \) such that each is normally distributed with mean 0 and for every \( h, g \), \( (W(h), W(g)) \) is normally distributed and

\[
E (W(h) W(g)) = (h, g)_H
\]

Furthermore, if \( \{ e_i \} \) is an orthogonal set of vectors of \( H \), then \( \{ W(e_i) \} \) are independent random variables. Also for any finite set \( \{ f_1, f_2, \cdots, f_n \} \), \( (W(f_1), W(f_2), \cdots, W(f_n)) \) is normally distributed.

Corollary 16.5.2 The map \( h \mapsto W(h) \) is linear. Also, \( \{ W(h) : h \in H \} \) is a closed subspace of \( L^2 (\Omega, \mathcal{F}, P) \) where \( \mathcal{F} = \sigma (W(h) : h \in H) \).

Proof: This follows from the above description.

\[
E \left( (W(g+h) - (W(g) + W(h)))^2 \right) = E (W(g+h)^2)
\]

\[
+ E \left( (W(g) + W(h))^2 \right) - 2 E (W(g) + W(h)) (W(g) + W(h))
\]

\[
= |g+h|^2 + |g|^2 + |h|^2 + 2 (g, h) - 2 (g + h, g) - 2 (g + h, h)
\]

\[
= |g|^2 + |h|^2 + 2 (g, h) + +2 (g, h) + |g|^2
\]

\[
+ |h|^2 - 2 |g|^2 - 2 (g, h) - 2 (g, h) - 2 |h|^2 = 0
\]

Hence \( W(h + g) = W(g) + W(h) \).

\[
E \left( (W(\alpha f) - \alpha W(f))^2 \right) = E (W(\alpha f)^2) + E (\alpha^2 W(f)^2) - 2 E (W(\alpha f) \alpha W(f))
\]

\[
= \alpha^2 |f|^2 + \alpha^2 |f|^2 - 2 \alpha (\alpha f, f) = 0.
\]
Why is \( \{W(h) : h \in H\} \) a subspace? This is obvious because \( W \) is linear. Why is it closed? Say \( W(h_n) \to f \in L^2(\Omega) \). This requires that \( \{h_n\} \) is a Cauchy sequence. Thus \( h_n \to h \) and so

\[
E \left( |f - W(h)|^2 \right) \leq 2 \lim_{n \to \infty} E \left( |f - W(h_n)|^2 \right) + E \left( |W(h_n) - W(h)|^2 \right)
\]

\[
= 2 \lim_{n \to \infty} E \left( |W(h_n) - W(h)|^2 \right) = 2 \lim_{n \to \infty} |h_n - h|^2_H = 0
\]

and so \( f = W(h) \) showing that this is indeed a closed subspace. \( \blacksquare \)

Next is a technical lemma which will be of considerable use.

**Lemma 16.5.3** Let \( X \geq 0 \) and measurable. Also define a finite measure on \( B(\mathbb{R}^p) \)

\[ \nu(B) = \int_{\Omega} X_B(Y) \, dP \]

Then

\[ \int_{\Omega} X e^{\lambda Y} \, dP = \int_{\mathbb{R}^p} e^{\lambda Y} \nu(y) \]

where here \( Y \) is a measurable function with values in \( \mathbb{R}^p \). Formally, \( X \, dP = \nu \).

**Proof:** First say \( X = X_D \) and replace \( e^{\lambda Y} \) with \( X_{Y^{-1}(B)} \). Then

\[ \int_{\Omega} X_D X_{Y^{-1}(B)} \, dP = P(D \cap Y^{-1}(B)) \]

\[ \int_{\mathbb{R}^p} X_B(y) \, d\nu(y) = \nu(B) = \int_{\Omega} X_D X_B(Y) \, dP \]

\[ = \int_{\Omega} X_D X_{Y^{-1}(B)} \, dP = P(D \cap Y^{-1}(B)) \]

Thus

\[ \int_{\Omega} X_D X_{Y^{-1}(B)} \, dP = \int_{\Omega} X_D X_B(Y) \, dP = \int_{\mathbb{R}^p} X_B(y) \, d\nu(y) \]

Now let \( s_n(y) \uparrow e^{\lambda Y} \), and let \( s_n(y) = \sum_{k=1}^m c_k X_{B_k}(y) \) where \( B_k \) is a Borel set. Then

\[ \int_{\mathbb{R}^p} s_n(y) \, d\nu(y) = \int_{\mathbb{R}^p} \sum_{k=1}^m c_k X_{B_k}(y) \, d\nu(y) = \sum_{k=1}^m c_k \int_{\mathbb{R}^p} X_{B_k}(y) \, d\nu(y) \]

\[ = \sum_{k=1}^m c_k P(D \cap Y^{-1}(B_k)) \]

\[ \int_{\mathbb{R}^p} s_n(Y) X_D \, dP = \sum_{k=1}^m c_k \int_{\Omega} X_D X_{B_k}(Y) \, dP = \sum_{k=1}^m c_k P(D \cap Y^{-1}(B_k)) \]

which is the same thing. Therefore,

\[ \int_{\Omega} s_n(Y) X_D \, dP = \int_{\mathbb{R}^p} s_n(y) \, d\nu(y) \]

Now pass to a limit using the monotone convergence theorem to obtain

\[ \int_{\Omega} e^{\lambda Y} X_D \, dP = \int_{\mathbb{R}^p} e^{\lambda Y} \nu(y) \]

Next replace \( X_D \) with \( \sum_{k=1}^m d_k X_{D_k} \equiv s_n(\omega) \), a simple function.

\[ \int_{\Omega} e^{\lambda Y} \sum_{k=1}^m d_k X_{D_k} \, dP = \sum_{k=1}^m d_k \int_{\Omega} e^{\lambda Y} X_{D_k} \, dP \]
where \( \nu_k(B) = \int_\Omega \mathcal{X}_{D_k} \mathcal{X}_B(Y) \, dP \). Now let
\[
\nu_n(B) = \int_\Omega \sum_{k=1}^m d_k \mathcal{X}_{D_k} \mathcal{X}_B(Y) = \int_\Omega s_n \mathcal{X}_B(Y) \, dP
\]
Then
\[
\nu_n(B) = \sum_{k=1}^m d_k \int_\Omega \mathcal{X}_{D_k} \mathcal{X}_B(Y) \, dP = \sum_{k=1}^m d_k \nu_k(B)
\]
Hence
\[
\int_\Omega e^{\lambda Y} s_n dP = \int_\Omega e^{\lambda Y} \sum_{k=1}^m d_k \mathcal{X}_{D_k} \, dP = \sum_{k=1}^m d_k \int_\Omega e^{\lambda Y} \, d\nu_k
\]
Then let \( s_n(\omega) \uparrow X(\omega) \). Clearly \( \nu_n \ll \nu \) and so by the Radon Nikodym theorem \( d\nu_n = h_n \, d\nu \) where \( h_n \uparrow 1 \). It follows from the monotone convergence theorem that one can pass to a limit in the above and obtain
\[
\int_\Omega e^{\lambda Y} X dP = \int_\mathbb{R} e^{\lambda Y} d\nu \quad \blacksquare
\]
Note that \( e^{\lambda Y} \) could have been replaced in all of the above with \( f(Y) \) where \( f \) is Borel measurable and nonnegative.

**Lemma 16.5.4** Each \( e^{W(h)} \) is in \( L^p(\Omega) \) for every \( h \in H \) and for every \( p \geq 1 \). In fact,
\[
\int_\Omega \left( e^{W(h)} \right)^p \, dP = \int_\Omega e^{W(p|h|_H^2)} \, dP = e^{\frac{1}{2} |ph|_H^2}.
\]
In addition to this,
\[
\sum_{k=0}^n \frac{W(h)^k}{k!} \rightarrow e^{W(h)} \quad \text{in} \quad L^p(\Omega, \mathcal{F}, P), \quad p > 1
\]

**Proof:** It suffices to verify this for all positive integers \( p \). Let \( p \) be such an integer. Note that from the linearity of \( W \), \( (e^{W(h)})^p = e^{pW(h)} = e^{W(ph)} \) and so it suffices to verify that for each \( h \in H \), \( e^{W(h)} \) is in \( L^1(\Omega) \). From Lemma 16.5.3,
\[
\int_\Omega e^{W(h)} \, dP = \int_\mathbb{R} e^{y} d\nu(y)
\]
where \( \nu(B) = \int_\Omega \mathcal{X}_B(W(h)) \, dP = \int_\mathbb{R} \mathcal{X}_B(y) \, d\nu(y) \). Thus
\[
\int_\Omega e^{W(h)} \, dP = \int_0^{\infty} \nu(e^y > \lambda) \, d\lambda = \int_0^{\infty} \frac{1}{\sqrt{2\pi |h|}} \int_{|y| > \ln(\lambda)} e^{-\frac{1}{2} \frac{y^2}{|h|^2}} \, dy \, d\lambda
\]
\[
= \frac{1}{\sqrt{2\pi |h|}} \int_{-\infty}^{\infty} e^u \int_{-\infty}^{\infty} e^{-\frac{1}{2} \frac{u^2}{|h|^2}} \, dy \, du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^u e^{-\frac{1}{2} u^2} \, du
\]
\[
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{1}{2} v^2} e^{-\frac{1}{2} v^2} \, dv = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} v^2} e^{\frac{1}{2} |v|^2} \, dv
\]
\[
= \frac{1}{\sqrt{2\pi}} \sqrt{2\pi} e^{\frac{1}{2} |v|^2} = e^{\frac{1}{2} |h|^2} < \infty
\]
If \( h = 0 \), \( W(h) \) would be 0 because by the construction, \( E(W(0)^2) = (0,0) = 0 \). Then
\[
\int_\Omega e^{W(h)} \, dP = \int_\Omega e^0 \, dP = 1
\]
Consider the last claim. It is enough to assume \( p \) is an integer.

\[
\left| \sum_{k=0}^{n} \frac{W(h)^k}{k!} - e^{W(h)} \right| = \left| \sum_{k=n+1}^{\infty} \frac{W(h)^k}{k!} \right| = |W(h)^{n+1}| \left| \sum_{k=0}^{\infty} \frac{W(h)^k}{(n+1+k)!} \right|
\]
\[
= |W(h)^{n+1}| \left| \sum_{k=0}^{\infty} \frac{W(h)^k}{k!} \right| \left| \frac{k}{(n+1+k)!} \right|
\]
\[
\leq |W(h)^{n+1}| \left( \frac{1}{(n+1)!} \sum_{k=0}^{\infty} \frac{W(h)^k}{k!} \right) = |W(h)^{n+1}| e^{W(h)}
\]

This converges to 0 for each \( \omega \) because it says nothing more than that the \( n^{th} \) term of a convergent sequence converges to 0.

\[
\int_{\Omega} \left( \frac{W(h)^{n+1}}{(n+1)!} e^{W(h)} \right)^{2p} dP = \int_{\Omega} \left( \frac{W(h)^{n+1}}{(n+1)!} \right)^{2p} (e^{W(h)})^{2p} dP
\]
\[
= \left( \frac{1}{(n+1)!} \right)^{2p} \frac{1}{\sqrt{2\pi|h|}} \int_{\mathbb{R}} e^{-\frac{1}{2} \frac{x^2}{2|h|}} e^{2px x^{2p(n+1)}} dx
\]
\[
= \left( \frac{1}{(n+1)!} \right)^{2p} \frac{1}{\sqrt{2\pi|h|}} e^{2p|h|^2} \int_{\mathbb{R}} e^{-\frac{1}{2\pi|h|^2} (x-2p|h|^2)^2} dx
\]
\[
\leq \left( \frac{1}{(n+1)!} \right)^{2p} \frac{2^{2p(n+1)} |h|^{2p(n+1)} |h|^{2p|h|^2}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2\pi|h|^2} (x-2p|h|^2)^2} e^{(2p|h|^2)^{2p(n+1)}} dx
\]

The second term clearly converges to 0 as \( n \to \infty \). Consider the first term. To simplify, let \( t = \frac{x-2p|h|^2}{|h|^2} \). Then this term reduces to

\[
= 2 \left( \frac{1}{(n+1)!} \right)^{2p} \frac{2^{2p(n+1)} |h|^{2p(n+1)} |h|^{2p|h|^2}}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-u^{p(n+1)}} u^{-(1/2)} du
\]
\[
= 2 \left( \frac{1}{(n+1)!} \right)^{2p} \frac{2^{2p(n+1)} |h|^{2p(n+1)} |h|^{2p|h|^2}}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-u^{p(n+1)}} u^{-(1/2)} du
\]

Now let \( t^2 = u \). Then this becomes

\[
2 \left( \frac{1}{(n+1)!} \right)^{2p} \frac{2^{2p(n+1)} |h|^{2p(n+1)} |h|^{2p|h|^2}}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-u^{p(n+1)}} u^{-(1/2)} \frac{1}{2} du
\]
\[
= \left( \frac{1}{(n+1)!} \right)^{2p} \frac{2^{2p(n+1)} |h|^{2p(n+1)} |h|^{2p|h|^2}}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-u^{p(n+1)}} u^{-(1/2)} du
\]
\[
\leq C(h) \left( |h|^{2p(n+1)} \frac{1}{(n+1)!} \frac{1}{((n+1)!)^{2p-1}} \right) \Gamma \left( p(n+1) - \frac{1}{2} \right)
\]
\[
= C(h) \left( |h|^{2p(n+1)} \frac{1}{(n+1)!} \frac{1}{((n+1)!)^{2p-1}} \right) \Gamma \left( p(n+1) - \frac{1}{2} \right)
\]
\[
\leq C(h) \left( |h|^{2p(n+1)} \frac{1}{(n+1)!} \frac{1}{((n+1)!)^{2p-1}} \right) \Gamma \left( p(n+1) - \frac{1}{2} \right)
\]
\[
= C(h) \left( |h|^{2p(n+1)} \frac{1}{(n+1)!} \frac{1}{((n+1)!)^{2p-1}} \right) \Gamma \left( p(n+1) - \frac{1}{2} \right)
\]
this converges to 0 as \( n \to \infty \). This is obvious for \( \frac{2^p h^n}{(n+1)!} \). Consider \( \frac{(p(n+1))!}{(n+1)!} \). By the ratio test, \( \sum_n \frac{(p(n+1))!}{(n+1)!} < \infty \) so this also converges to 0. The details of this ratio test argument are as follows. The ratio, after simplifying is

\[
\frac{(pm + 2p)(pm + 2p - 1) \cdots (pm + 1)}{(n + 2)^{2p - 1}} \leq \frac{p^p (n + p)^p}{(n + 2)^{2p - 1}}
\]

which clearly converges to 0 since \( 2p - 1 > p \) since \( p \) is an integer larger than 1.

Therefore, \( \left\{ \frac{W(h)^{n+1}}{(n+1)!} e^{W(h)} \right\}_{n=1}^{\infty} \) is bounded in \( L^2(\Omega) \). Then

\[
\int_{\Omega} \left| \sum_{k=0}^{n} \frac{W(h)^k}{k!} - e^{W(h)} \right|^p dP \to 0
\]

because the integrand is bounded by \( \left( \left\{ \frac{W(h)^{n+1}}{(n+1)!} e^{W(h)} \right\}_{n=1}^{\infty} \right)^p \) and it was just shown that these functions are bounded in \( L^2(\Omega) \). Therefore, the claimed convergence follows from the Vitali convergence theorem.

The following lemma shows that the functions \( e^{W(h)} \) are dense in \( L^p(\Omega) \) for every \( p > 1 \).

**Lemma 16.5.5** Let \( \mathcal{F} \) be the \( \sigma \)-algebra determined by the random variables \( W(h) \). If \( X \in L^p(\Omega, \mathcal{F}, P) \), \( p > 1 \) and \( \int_{\Omega} X e^{W(h)} dP = 0 \) for every \( h \in H \), then \( X = 0 \).

**Proof:** Let \( h_1, \ldots, h_p \) be given. Then for \( t_i \in \mathbb{R} \),

\[
\sum_i t_i h_i \in H
\]

and so since \( W \) is linear,

\[
\int_{\Omega} X e^{t \cdot W(h)} dP = 0, \quad W(h) = (W(h_1), \ldots, W(h_p))
\]

Now by Lemma 16.5.3,

\[
\int_{\Omega} X^+ e^{(W(h_1), \ldots, W(h_p))} dP = \int_{\mathbb{R}^p} e^{t \cdot y} \nu_+(y)
\]

where \( \nu_+(B) = E(X^+ \mathcal{X}_B(W(h))) \). From Lemma 16.5.3, this function of \( t \) is finite for all \( t \in \mathbb{R}^p \). Similarly,

\[
\int_{\Omega} X^- e^{(W(h_1), \ldots, W(h_p))} dP = \int_{\mathbb{R}^p} e^{t \cdot y} \nu_-(y)
\]

where \( \nu_-(B) = E(X^- \mathcal{X}_B(W(h))) \). Thus for \( \nu \) equal to the signed measure \( \nu = \nu_+ - \nu_- \),

\[
f(t) \equiv \int_{\mathbb{R}^p} e^{t \cdot y} d\nu(y) = 0
\]

for \( t \in \mathbb{R}^p \). Also

\[
\int_{\Omega} X^+ e^{i t \cdot (W(h))} dP = \int_{\mathbb{R}^p} e^{i t \cdot y} \nu_+(y)
\]

with a similar formula holding for \( X^- \). Thus

\[
f(t) \equiv \int_{\mathbb{R}^p} e^{i t \cdot y} d\nu(y) \in \mathbb{C}
\]

is well defined for all \( t \in \mathbb{C}^p \). Consider

\[
\int_{\mathbb{R}^p} e^{i t \cdot y} d\nu_+(y)
\]

Is this function analytic in each \( t_k \)? Take a difference quotient. It equals for \( h \in \mathbb{C} \),

\[
\int_{\Omega} X^+ \frac{e^{i(t+h \cdot \mathcal{X}_B(W(h))} - e^{i(t \cdot W(h))}}{h} dP = \int_{\Omega} X^+ e^{i t \cdot W(h)} \frac{e^{i h \mathcal{X}_B(W(h))} - 1}{h} dP
\]
In case \( e_k \cdot W(h) = 0 \) there is nothing to show. Assume then that this is not 0. Then this equals

\[
\int_{\Omega} X^+ e_k \cdot (W(h)) e^{tW(h)} \left( \frac{e^{he_k \cdot (W(h))} - 1}{h (e_k \cdot (W(h)))} \right) dP
\]

Now

\[
\left| \frac{e^z - 1}{z} \right| = \left| \frac{1}{z} \sum_{k=1}^{\infty} \frac{z^k}{k!} \right| \leq \sum_{k=0}^{\infty} \leq |e^z|
\]

and so the integrand is dominated by

\[
\left| X^+ e_k \cdot (W(h)) e^{tW(h)} \left( \frac{e^{he_k \cdot (W(h))} - 1}{h (e_k \cdot (W(h)))} \right) \right| \leq X^+ \left| e_k \cdot (W(h)) e^{tW(h)} \left( \frac{e^{he_k \cdot (W(h))} - 1}{h (e_k \cdot (W(h)))} \right) \right|
\]

From Lemma 10.5.2 which says that \( e^{W(h)} \) is in \( L^q(\Omega) \) for each \( q > 1 \), this is in particular true for \( q = mp \) where \( m \) is an arbitrary positive integer satisfying

\[
p > \frac{m + 1}{m}
\]

Then the integrand is of the form \( fg_h \) where \( f \in L^p \) and \( g_h \) is bounded in \( L^{mp} \). Therefore,

\[
\alpha \equiv (pm) / (m + 1) > 1
\]

and

\[
\int_{\Omega} |fg_h|^\alpha dP = \int_{\Omega} |f|^\alpha |g_h|^\alpha dP \leq \left( \int_{\Omega} |f|^p dP \right)^{m/(m+1)} \left( \int_{\Omega} |g_h|^{pm} dP \right)^{1/(m+1)}
\]

which is bounded. By the Vitali convergence theorem,

\[
\lim_{h \to 0} \int_{\Omega} X^+ \left( \frac{e^{t+he_k \cdot W(h)} - e^{tW(h)}}{h} \right) dP = \int_{\Omega} X^+ e_k \cdot (W(h)) e^{tW(h)} dP
\]

and so this function of \( t_k \) is analytic. Similarly one can do the same thing for the integral involving \( X^- \). Thus

\[
0 = \int_{\mathbb{R}^p} e^{t \cdot \psi} d\nu(y)
\]

whenever \( t_j \in \mathbb{R} \) for all \( j \) and \( t_1 \to \int_{\mathbb{R}^p} e^{t \cdot \psi} d\nu(y) \) is analytic on \( \mathbb{C} \). Thus this analytic function of \( t_1 \) is zero for all \( t_1 \in \mathbb{C} \) since it is zero on a set which has a limit point, and in particular

\[
\int_{\mathbb{R}^p} e^{it_1 y_1 + it_2 y_2 + \cdots + it_p y_p} d\nu(y) = 0
\]

where each \( t_j \) is real. Now repeat the argument with respect to \( t_2 \) and conclude that

\[
\int_{\mathbb{R}^p} e^{it_1 y_1 + it_2 y_2 + \cdots + it_p y_p} d\nu(y) = 0,
\]

and continue this way to conclude that

\[
0 = \int_{\mathbb{R}^p} e^{it \cdot \psi} d\nu(y)
\]

which shows that the inverse Fourier transform of \( \nu \) is 0. Thus \( \nu = 0 \). To see this, let \( \psi \in \mathcal{S} \), the Schwartz class. Then neglecting troublesome constants in the Fourier transform,

\[
0 = \int_{\mathbb{R}^p} \psi(t) \int_{\mathbb{R}^p} e^{it \cdot \psi} d\nu(y) dt = \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \psi(t) e^{it \cdot \psi} dtd\nu(y) = \nu (F^{-1} \psi)
\]

Now \( F^{-1} \) maps \( \mathcal{S} \) onto \( \mathcal{S} \) and so this reduces to

\[
\int_{\mathbb{R}^p} \psi d\nu = 0
\]
for all $\psi \in \mathcal{S}$. By density of $\mathcal{S}$ in $C_0(\mathbb{R}^p)$, it follows that the above holds for all $\psi \in C_0(\mathbb{R}^p)$ and so $\nu = 0$.

It follows that for every $B$ Borel and for every such description of $W(h)$,

$$0 = \int_{\Omega} X\mathcal{X}_B(W(h))\,dP = \int_{\Omega} X\mathcal{X}_{W(h)^{-1}(B)}\,dP$$

Let $\mathcal{K}$ be sets of the form $W(h)^{-1}(B)$ where $B$ is of the form $B_1 \times \cdots \times B_p$, $B_i$ open, this for some $p$. Then this is clearly a $\pi$ system because the intersection of any two of them is another one and

$$\emptyset, \Omega = W(h)^{-1}(\mathbb{R}^p)$$

are both in $\mathcal{K}$. Also $\sigma(\mathcal{K}) = \mathcal{F}$. Let $\mathcal{G}$ be those sets $F$ of $\mathcal{F}$ such that

$$0 = \int_{\Omega} X\mathcal{X}_F\,dP \quad (16.5.32)$$

This is true for $F \in \mathcal{K}$. Now it is clear that $\mathcal{G}$ is closed with respect to complements and countable disjoint unions. It is closed with respect to complements because

$$\int_{\Omega} X\mathcal{X}_{F \cap} dP = \int_{\Omega} X (1 - \mathcal{X}_F) dP = \int_{\Omega} X dP - \int_{\Omega} X\mathcal{X}_F dP = 0$$

By Dynkin’s lemma, $\mathcal{G} = \mathcal{F}$ and so $(16.5.32)$ holds for all $F \in \mathcal{F}$ which requires $X = 0$. $\blacksquare$

### 16.5.2 The Wiener Processes

Recall the definition of the Wiener process.

**Definition 16.5.6** Let $W(t)$ be a stochastic process which has the properties that whenever $t_1 < t_2 < \cdots < t_m$, the increments $\{W(t_i) - W(t_{i-1})\}$ are independent and whenever $s < t$, it follows $W(t) - W(s)$ is normally distributed with variance $t - s$ and mean 0. Also $t \rightarrow W(t)$ is Holder continuous with every exponent $\gamma < 1/2$, $W(0) = 0$. This is called a Wiener process.

Now in the definition of $W$ above, you begin with a Hilbert space $H$. There exists a probability space $\left(\Omega, \mathcal{F}, P\right)$ and a linear mapping $W$ such that

$$E(W(f)W(g)) = \langle f, g \rangle$$

and $(W(f_1), W(f_2), \cdots, W(f_n))$ is normally distributed with mean 0. Next define $\mathcal{F} = \sigma(W(h) : h \in H)$.

Consider the special example where $H = L^2(0, \infty; \mathbb{R})$, real valued functions which are square integrable with respect to Lebesgue measure. Note that for each $t \in [0, \infty)$, $W(0, t) \in H$. Let

$$W(t) \equiv W(X_{0, t})$$

Then from definition, if $t_1 < t_2 < \cdots < t_m$, the increments $\{W(t_i) - W(t_{i-1})\}$ are independent. This is because, due to the linearity of $W$, each of these equals $W(X_{0, t_i} - X_{0, t_{i-1}}) = W(X_{t_{i-1}, t_i})$ and from Corollary 16.5.4, the random vector $(W(X_{t_1, t_2}), \cdots, W(X_{t_{m-1}, t_m}))$ is normally distributed with covariance equal to a diagonal matrix. Also $E\left(W(t)^2\right) = E\left(W(X_{0, t})^2\right) = \int_0^\infty X_{0, t}^2 ds = t$. More generally,

$$W(t) - W(s) = W(X_{0, t}) - W(X_{0, s}) = W(X_{s, t})$$

so both $W(t) - W(s)$ and $W(t - s)$ are normally distributed with mean 0 and variance $t - s$. What about the Holder continuity? The characteristic function of $W(t) - W(s)$ is

$$E\left(e^{i\lambda(W(t-s))}\right) = e^{i\lambda^2(t-s)}$$

Consider a few derivatives of the right side with respect to $\lambda$ and then let $\lambda = 0$. This will yield $E((W(t) - W(s))^n)$ for $n = 1, 2, 3, 4$.

$$0, |s - t|, 0, 3|s - t|^2$$
You see the pattern. By induction, you can show that $E \left( (W(t) - W(s))^{2m} \right) = C_m |t-s|^m$. By the Kolmogorov-Centsov theorem, Theorem 16.5.3,

$$E \left( \sup_{0 \leq s < t \leq T} \frac{||W(t) - W(s)||}{(t-s)^{1/2}} \right) \leq C_m$$

whenever $\gamma < \beta/\alpha = \frac{m-1}{2m}$. Thus the above is true whenever $\gamma < 1/2$. It follows that there exists a set of measure zero off which $t \rightarrow W(t)$ is Hölder continuous with exponent $\gamma < 1/2$.

Thus this gives a construction of the real Wiener process. Now consider the normal filtration

$$\mathcal{F}_s = \{ W(u) - W(r) : 0 \leq r < u \leq t \}$$

By Lemma 16.5.3, $\{W(t)\}$ is a martingale with respect to this filtration, because of the independence of the increments.

Of course you could also take an arbitrary $f \in L^2(0, \infty)$ and consider $W(t) = W(\mathcal{X}(0,t) f)$. You could consider this as an integral and write it in the notation

$$W(t) = \int_0^t f dW(t) = W(f \mathcal{X}(0,t))$$

Then from the construction,

$$E \left( \left( \int_0^t f dW(t) \right)^2 \right) = E \left( W(f \mathcal{X}(0,t))^2 \right) = \int_0^T f^2 \mathcal{X}(0,t) ds = \int_0^t |f|^2 ds = E \left( \int_0^t |f|^2 ds \right)$$

because $f$ does not depend on $\omega$. This of course is formally the Ito isometry.

### 16.5.3 $Q$ Wiener Processes In Hilbert Space

Now let $U$ be a real separable Hilbert space. Let an orthonormal basis for $U$ be $\{g_k\}$. Now let $L^2(0, \infty, U)$ be $H$ in the above construction. For $h, g \in L^2(0, \infty, U)$,

$$E(W(h) W(g)) = (h, g)_H \equiv (h, g)_U$$

Here each $W(g)$ will be a real valued normal random variable, the variance of $W(g)$ is $|g|^2_{L^2(0, \infty, U)}$ and its mean is $0$, every vector $(W(h_1), \ldots, W(h_n))$ being generalized multivariate normal. Let

$$\psi_k(t) = W(\mathcal{X}(0,t) g_k)$$

Then this is a real valued random variable. Disjoint increments are obviously independent in the same way as before. Also

$$E(\psi_k(t) \psi_j(s)) = E(W(\mathcal{X}(0,t) g_k) W(\mathcal{X}(0,s) g_j)) \equiv \int_0^\infty \mathcal{X}(0,t \wedge s) (g_k, g_j)_U dt = 0 \quad (16.5.33)$$

if $j \neq k$. Thus the random variables $\psi_k(t)$ and $\psi_j(s)$ are independent. This is because, from the construction, $(\psi_k(t), \psi_j(s))$ is normally distributed and the covariance is a diagonal matrix. Also

$$\psi_k(t) - \psi_k(s) = W(\mathcal{X}(0,t) J g_k) - W(\mathcal{X}(0,s) J g_k) = W(\mathcal{X}(s,t) J g_k)$$

$$\psi_k(t-s) \equiv W(\mathcal{X}(0,t-s) J g_k)$$

so $\psi_k(t-s)$ has the same mean, $0$ and variance, $|t-s|$, as $\psi_k(t) - \psi_k(s)$. Thus these have the same distribution because both are normally distributed.

Now let $J$ be a Hilbert Schmidt map from $U$ to $H$. Then consider

$$W(t) = \sum_k \psi_k(t) J g_k \quad (16.5.34)$$

This has values in $H$. It is shown below that the series converges in $L^2(\Omega; H)$. Recall the definition of a $Q$ Wiener process.
Definition 16.5.7 Let $W(t)$ be a stochastic process with values in $H$, a real separable Hilbert space which has the properties that $t \to W(t, \omega)$ is continuous, whenever $t_1 < t_2 < \cdots < t_m$, the increments $\{W(t_i) - W(t_{i-1})\}$ are independent, $W(0) = 0$, and whenever $s < t$,

\[ \mathcal{L}(W(t) - W(s)) = N(0, (t-s)Q) \]

which means that whenever $h \in H$,

\[ \mathcal{L}((h, W(t) - W(s))) = N(0, (t-s)(Qh, h)) \]

Also

\[ E((h_1, W(t) - W(s))(h_2, W(t) - W(s))) = (Qh_1, h_2)(t-s). \]

Here $Q$ is a nonnegative trace class operator. Recall this means

\[ Q = \sum_{i=1}^{\infty} \lambda_i e_i \otimes e_i \]

where \( \{e_i\} \) is a complete orthonormal basis, $\lambda_i \geq 0$, and

\[ \sum_{i=1}^{\infty} \lambda_i < \infty \]

Such a stochastic process is called a $Q$ Wiener process. In the case where these have values in $\mathbb{R}^n$, $tQ$ ends up being the covariance matrix of $W(t)$.

Proposition 16.5.8 The process defined in (16.5.34) is a $Q$ Wiener process in $H$ where $Q = JJ^*$.

Proof: First, why does the sum converge? Consider the sum for an increment in time. Let $t_{i-1} = 0$ to obtain the convergence of the sum for a given $t$. Consider the difference of two partial sums.

\[
E \left( \sum_{k,l=m}^{n} (\psi_k(t_i) - \psi_k(t_{i-1})) Jg_k, (\psi_l(t_i) - \psi_l(t_{i-1})) Jg_l \right)
\]

\[
= E \left( \sum_{k,l=m}^{n} (J^* Jg_k, g_l) (\psi_k(t_i) - \psi_k(t_{i-1})) (\psi_l(t_i) - \psi_l(t_{i-1})) \right)
\]

\[
= \sum_{k,l=m}^{n} (J^* Jg_k, g_l) E \left( (\psi_k(t_i) - \psi_k(t_{i-1})) (\psi_l(t_i) - \psi_l(t_{i-1})) \right)
\]

\[
= \sum_{k=m}^{n} (J^* Jg_k, g_k) E \left( \psi_k(t_i) - \psi_k(t_{i-1}) \right)^2 = \sum_{k=m}^{n} (J^* Jg_k, g_k)(t_i - t_{i-1})
\]

\[
= \sum_{k=m}^{n} |Jg_k|^2_H (t_i - t_{i-1})
\]

and this converges to 0 as $m, n \to \infty$ since $J$ is Hilbert Schmidt. Thus the sum converges in $L^2(\Omega, H)$. Why are the disjoint increments independent?

Let $\lambda_k \in H$. Consider $t_0 < t_1 < \cdots < t_n$.

\[
E \left( \exp i \sum_{k=1}^{n} (\lambda_k, W(t_k) - W(t_{k-1})) \right) = \prod_{k=1}^{n} E \left( \exp (i (\lambda_k, W(t_k) - W(t_{k-1}))) \right)? \quad (16.5.35)
\]

Start with the left. There are finitely many increments concerned and so it can be assumed that for each $k$ one can have $m \to \infty$ such that the partial sums up to $m$ in the definition of $W(t_k) - W(t_{k-1})$ converge pointwise a.e. Thus

\[
E \left( \exp i \sum_{k=1}^{n} (\lambda_k, W(t_k) - W(t_{k-1})) \right)
\]

\[
= \lim_{m \to \infty} E \left( \exp \sum_{k=1}^{n} \left( \lambda_k, \sum_{j=1}^{m} (\psi_j(t_k) - \psi_j(t_{k-1})) Jg_j \right) \right)
\]
16.5. WIENER PROCESSES, ANOTHER APPROACH

\[ \lim_{m \to \infty} E \left( \exp \sum_{k=1}^{n} \sum_{j=1}^{m} i \left( \lambda_k, (\psi_j(t_k) - \psi_j(t_{k-1})) Jg_j \right) \right) \]

\[ = \lim_{m \to \infty} E \left( \prod_{j=1}^{m} \exp \left( \sum_{k=1}^{n} i \left( \lambda_k, (\psi_j(t_k) - \psi_j(t_{k-1})) Jg_j \right) \right) \right) \]

Now from (16.5.35), \( \{ \sum_{k=1}^{m} i \left( \lambda_k, (\psi_j(t_k) - \psi_j(t_{k-1})) Jg_j \right) \} \) are independent. Hence the above equals

\[ = \lim_{m \to \infty} \prod_{j=1}^{m} E \left( \prod_{k=1}^{n} \exp \left( i \left( \lambda_k, (\psi_j(t_k) - \psi_j(t_{k-1})) Jg_j \right) \right) \right) \]

Now from independence of the increments for the \( \psi_j \), this equals

\[ = \lim_{m \to \infty} \prod_{j=1}^{m} \prod_{k=1}^{n} E \left( \exp \left( i \left( \lambda_k, (\psi_j(t_k) - \psi_j(t_{k-1})) Jg_j \right) \right) \right) \]

\[ = \lim_{m \to \infty} \prod_{j=1}^{m} \prod_{k=1}^{n} E \left( \exp \left( i \left( \lambda_k, Jg_j \right) (\psi_j(t_k) - \psi_j(t_{k-1})) \right) \right) \]

\[ = \lim_{m \to \infty} \prod_{j=1}^{m} \prod_{k=1}^{n} e^{-\frac{i}{2} (\lambda_k, Jg_j)^2 (t_k - t_{k-1})} = \lim_{m \to \infty} \prod_{j=1}^{m} e^{-\frac{1}{2} \sum_{k=1}^{n} (\lambda_k, Jg_j)^2 (t_k - t_{k-1})} \]

\[ = \lim_{m \to \infty} \exp \left( -\frac{1}{2} \sum_{j=1}^{m} \sum_{k=1}^{n} (\lambda_k, Jg_j)^2 (t_k - t_{k-1}) \right) \]

\[ = \exp \left( \frac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{\infty} (J^* \lambda_k, g_j)^2 (t_k - t_{k-1}) \right) \]  \hspace{1cm} (16.5.36)

What is the right side of (16.5.35)?

\[ \prod_{k=1}^{n} E \left( \exp(i \left( \lambda_k, W(t_k) - W(t_{k-1}) \right)) \right) = \prod_{k=1}^{n} E \left[ \exp \left( i \left( \lambda_k, \sum_{j=1}^{\infty} (\psi_j(t_k) - \psi_j(t_{k-1})) Jg_j \right) \right) \right] \]

\[ = \lim_{m \to \infty} \prod_{k=1}^{n} E \left[ \exp \left( i \left( \lambda_k, \sum_{j=1}^{m} (\psi_j(t_k) - \psi_j(t_{k-1})) Jg_j \right) \right) \right] \]

\[ = \lim_{m \to \infty} \prod_{k=1}^{n} E \left[ \exp \left( i \sum_{j=1}^{m} (Jg_j, \lambda_k) (\psi_j(t_k) - \psi_j(t_{k-1})) \right) \right] \]

\[ = \lim_{m \to \infty} \prod_{k=1}^{n} E \left( \prod_{j=1}^{m} i (J^* \lambda_k, g_j) (\psi_j(t_k) - \psi_j(t_{k-1})) \right) \]
and by independence,

\[
\lim_{m \to \infty} \prod_{k=1}^{n} \prod_{j=1}^{m} E \left[ i (J^* \lambda_k, g_j) (\psi_j (t_k) - \psi_j (t_{k-1})) \right] = \exp \left( -\frac{1}{2} \sum_{j=1}^{m} (J^* \lambda_k, g_j)^2 (t_k - t_{k-1}) \right)
\]

which is exactly the same thing as \[16.5.37]\) Thus the disjoint increments are independent.

You could also do something like the following. Let \(W_m (t)\) denote the partial sum for \(W (t)\) and since there are only finitely many increments, we can assume the partial sums converge a.e. Then we need to consider the random variables

\[
\{ (W_m (t_k) - W_m (t_{k-1})) \}_{k=1}^{m} = \left\{ \sum_{i=1}^{m} (\psi_i (t_k) - \psi_i (t_{k-1})) J g_i \right\}_{k=1}^{m}
\]

Then for any \(h \in H\), you could consider

\[
\left\{ \left( \sum_{i=1}^{m} (\psi_i (t_k) - \psi_i (t_{k-1})) (J g_i, h)_H \right) \right\}_{k=1}^{m}
\]

and the vector whose \(k^{th}\) component is \(\sum_{i=1}^{m} (\psi_i (t_k) - \psi_i (t_{k-1})) (J g_i, h)_H\) for \(k = 1, 2, \ldots, n\) is normally distributed and the covariance is a diagonal matrix. Hence these are independent random variables as hoped. Now you can pass to a limit as \(m \to \infty\). Since this is true for any \(h \in H\) that the random variables \((W (t_k) - W (t_{k-1}), h)_H\) are independent, it follows that the random variables \(W (t_k) - W (t_{k-1})\) are also.

What of the Holder continuity? In the above computation for independence, as a special case, for \(\lambda \in H\),

\[
E (\exp i (\lambda, W (t) - W (s))) = \exp \left( -\frac{1}{2} |J^* \lambda|_{U}^2 (t - s) \right)
\]

In particular, replacing \(\lambda\) with \(r \lambda\) for \(r\) real,

\[
E (\exp i r (\lambda, W (t) - W (s))) = \exp \left( -\frac{1}{2} r^2 |J^* \lambda|_{U}^2 (t - s) \right)
\]

Now we differentiate with respect to \(r\) and then take \(r = 0\) as before to obtain finally that

\[
E \left( (\lambda, W (t) - W (s))^{2m} \right) \leq C_m |J^* \lambda|_{U}^{2m} |t - s|^m = C_m (Q \lambda, \lambda)^m |t - s|^m
\]

Then letting \(\{h_k\}\) be an orthonormal basis for \(H\), and using the above inequality with Minkowski’s inequality,

\[
\left( E \left( |W (t) - W (s)|^{2m} \right) \right)^{1/m} = \left( E \left( \sum_{k=1}^{\infty} (W (t) - W (s), h_k)^2 \right)^{m} \right)^{1/m}
\]

\[
\leq \sum_{k=1}^{\infty} \left( E \left( (W (t) - W (s), h_k)^2 \right) \right)^{1/m} \leq \sum_{k=1}^{\infty} \left( C_m (t - s)^m |J^* h_k|_{U}^{2m} \right)^{1/m}
\]

\[
= C_m^{1/m} |t - s| \sum_{k=1}^{\infty} |J^* h_k|_{U}^{2} = C_m^{1/m} |t - s| \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} (J^* h_k, g_j)^2
\]

\[
= C_m^{1/m} |t - s| \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} (h_k, J g_j)^2 = |t - s| C_m^{1/m} \sum_{j=1}^{\infty} |J g_j|^2_{H}
\]
and since $J$ is Hilbert Schmidt, modifying the constant yields
\[ E \left( |W(t) - W(s)|^{2m} \right) \leq C_m |t-s|^m \]

By the Kolmogorov Centsov theorem, Theorem 16.5.3.
\[ E \left( \sup_{0 \leq s < t \leq T} \frac{\|W(t) - W(s)\|}{(t-s)^{1/2}} \right) \leq C_m \]
whenever $\gamma < \beta/\alpha = \frac{m-1}{2m}$. Thus the above is true whenever $\gamma < 1/2$. Hence off a set of measure zero, $t \to W(t)$ is Holder continuous.

What of the covariance condition? From 16.5.3a, letting $f, g$ be two elements of $H$,
\[ E(\exp i(\alpha f + \beta g, W(t) - W(s))) = \exp \left( -\frac{1}{2} (Q(\alpha f + \beta g), \alpha f + \beta g)(t-s) \right) \]

Differentiate with respect to $\alpha$
\[ E \left( i(f, W(t) - W(s)) \exp i(\alpha f + \beta g, W(t) - W(s)) \right) = -[Qf, g](t-s) \exp \left( -\frac{1}{2} (Q(\alpha f + \beta g), \alpha f + \beta g)(t-s) \right) \]

Let $\alpha = 0$.
\[ E \left( i(f, W(t) - W(s)) \exp i(\beta g, W(t) - W(s)) \right) = -[Qf, \beta g](t-s) \exp \left( -\frac{1}{2} (Q(\beta g), \beta g)(t-s) \right) \]

Now differentiate with respect to $\beta$
\[ E \left( - (f, W(t) - W(s))(g, W(t) - W(s)) \exp i(\beta g, W(t) - W(s)) \right) = -[(Qf, g)](t-s) \exp \left( -\frac{1}{2} (Q(\beta g), \beta g)(t-s) \right) + -[(Qf, \beta g)](t-s)(\text{something}) \]

Now let $\beta = 0$.
\[ E \left( (f, W(t) - W(s))(g, W(t) - W(s)) \right) = (Qf, g)(t-s) \]

Finally, $Q = JJ^*$. It is self adjoint and nonnegative and so there is a complete orthonormal basis $\{e_i\}$ such that $Qe_i = \lambda_i e_i$. Then $\lambda_i = (Qe_i, e_i)_H$ and so
\[ \sum_i \lambda_i = \sum_i (Qe_i, e_i) = \sum_i |J^* e_i|_U^2 < \infty \]
because $J$ and hence $J^*$ are both Hilbert Schmidt operators. 

Recall the notion of the Hilbert space $LU$ in Definition 16.5.1.

What if you have a given $Q \in \mathcal{L}(H, H)$ which is trace class, $Q = Q^*$, and nonnegative. Does there exist a $Q$ Wiener process of the sort just described? It appears this amounts to obtaining a Hilbert Schmidt map $J$ from some Hilbert space $U$ to $H$ such that $Q = JJ^*$.

Since $Q$ is trace class and is self adjoint, it follows that there is an orthonormal basis $\{e_i\}$, $Qe_i = \lambda_i e_i$, where $\lambda_i$ is positive for $i \leq L$ or positive for all $i$. Then
\[ Q^{1/2} = \sum_{i=1}^L \sqrt{\lambda_i} e_i \otimes e_i \]
and
\[ Q^{1/2} e_i = \sqrt{\lambda_i} e_i. \]

Then also on $Q^{1/2} H$,
\[ (Q^{1/2} e_i, Q^{1/2} e_j)_{Q^{1/2} H} \equiv (e_i, e_j)_H \]
and so an orthonormal basis in \( Q^{1/2}H \) is \( \{ \sqrt{\lambda_i} e_i \}_{i=1}^L \). Then define \( J : Q^{1/2}H \to H \)

\[
Jx = \sum_{k=1}^L \left( x, \sqrt{\lambda_k} e_k \right)_{Q^{1/2}H} \sqrt{\lambda_k} e_k
\]

It follows from the above that

\[
J e_j = \sum_{k=1}^L \frac{1}{\sqrt{\lambda_j}} \left( \sqrt{\lambda_j} e_j, \sqrt{\lambda_k} e_k \right)_{Q^{1/2}H} \sqrt{\lambda_k} e_k = e_j
\]

Then

\[
\sum_{i=1}^L |J \sqrt{\lambda_i} e_i|_H^2 = \sum_{i=1}^L \left| \sqrt{\lambda_i} e_i, \sqrt{\lambda_k} e_k \right|_{Q^{1/2}H} \sqrt{\lambda_k} e_k \right|_H^2 = \sum_{i=1}^L \sqrt{\lambda_i} < \infty
\]

Thus it is clear that \( J \) is Hilbert Schmidt. Is \( JJ^* = Q ? \) For \( y \in Q^{1/2}H, x \in H, \)

\[
(J^* x, y)_{Q^{1/2}H} = (x, J(y))_H = \left( x, \sum_{k=1}^L \left( y, \sqrt{\lambda_k} e_k \right)_{Q^{1/2}H} \sqrt{\lambda_k} e_k \right)_H = \sum_{k=1}^L \left( x, \sqrt{\lambda_k} e_k \right)_H \left( y, \sqrt{\lambda_k} e_k \right)_{Q^{1/2}H}
\]

Thus for \( y \in H, x \in H, \)

\[
(J^* x, J^* y)_{Q^{1/2}H} = \sum_{k=1}^L \left( x, \sqrt{\lambda_k} e_k \right)_H \left( J^* y, \sqrt{\lambda_k} e_k \right)_{Q^{1/2}H} = \sum_{k=1}^L \left( x, \sqrt{\lambda_k} e_k \right)_H \left( y, \sqrt{\lambda_k} J e_k \right)_H = \sum_{k=1}^L \lambda_k (x, e_k)_H (y, e_k)_H = (Q x, y)
\]

and so \( (J^* x, y) = (Q x, y) \) showing that \( J J^* = Q \). This shows the following.

**Proposition 16.5.9** Let \( Q \in \mathcal{L}(H, H) \) where \( H \) is a real separable Hilbert space and \( (Q x, x) \geq 0 \) and is trace class. Then there exists a one to one Hilbert Schmidt map \( J : Q^{1/2}H \to H \) such that \( J J^* = Q \). Then the Q Wiener process is \( W(t) = \sum_{k=1}^\infty \psi_k(t) J g_k \) where \( \{g_k\} \) is a complete orthonormal basis for the Hilbert space \( Q^{1/2}H \).

Note that in case \( H \) is \( \mathbb{R}^p \) and \( Q \) is any symmetric \( p \times p \) matrix, having nonnegative eigenvalues, this is automatically trace class and so the above conclusion holds. In particular, the covariance condition says in this case that

\[
E ((e_i, W(t) - W(s))(e_j, W(t) - W(s))) = E ((W_i(t) - W_i(s))(W_j(t) - W_j(s))) = (Q e_i, e_j) = Q_{ij}
\]

This is a \( p \) dimensional Wiener process.
Chapter 17

Stochastic Integration

17.1 Integrals Of Elementary Processes

Stochastic integration starts with a $Q$ Wiener process having values in a separable Hilbert space $U$. Thus it satisfies the following definition.

**Definition 17.1.1** Let $W(t)$ be a stochastic process with values in $U$, a real separable Hilbert space which has the properties that $t \rightarrow W(t, \omega)$ is continuous. Whenever $t_1 < t_2 < \cdots < t_m$, the increments $\{W(t_i) - W(t_{i-1})\}$ are independent, $W(0) = 0$, and whenever $s < t$,

$$\mathcal{L}(W(t) - W(s)) = N(0, (t - s)Q)$$

which means that whenever $h \in H$,

$$\mathcal{L}((h, W(t) - W(s))) = N(0, (t - s)(Qh, h))$$

Also

$$E((h_1, W(t) - W(s))(h_2, W(t) - W(s))) = (Qh_1, h_2)(t - s).$$

Here $Q$ is a nonnegative trace class operator. Recall this means

$$Q = \sum_{i=1}^{\infty} \lambda_i e_i \otimes e_i$$

where $\{e_i\}$ is a complete orthonormal basis, $\lambda_i \geq 0$, and

$$\sum_{i=1}^{\infty} \lambda_i < \infty$$

Such a stochastic process is called a $Q$ Wiener process.

Recall that such Wiener processes are always of the form

$$\sum_{k=1}^{\infty} \psi_k(t) Jg_k$$

where $J$ is a Hilbert Schmidt operator from a suitable space $U_0$ to $U$ and the $\psi_k$ are real independent Wiener processes described earlier. This follows from Theorem [16.4.7] where you let $U_0 \subseteq U$ be such that for $J$ the inclusion map, $Je_k = \sqrt{\lambda_k}e_k$ for $Q = \sum_k \lambda_k e_k \otimes e_k$, the $e_k$ an orthonormal set in $U$. Thus

$$(Qx, y) = \left(\sum_k \lambda_k e_k(x, e_k), y\right) = \sum_k \left(x, \sqrt{\lambda_k}e_k\right) \left(y, \sqrt{\lambda_k}e_k\right) = \sum_k (x, Je_k) (y, Je_k) = \sum_k (J^*x, e_k)(J^* y, e_k) = (J^*x, J^* y) = (JJ^*x, y)$$

371
so it follows that \( Q = J J^* \). Of course in finite dimensions, there is no issue because the identity map is Hilbert Schmidt.

Recall the definition of \( \mathcal{L}_2 (U, H) \equiv \mathcal{L}_2 \) the space of Hilbert Schmidt operators. \( \Psi \in \mathcal{L}_2 (U, H) \) means \( \Psi \) has the property that for some (equivalently all) orthonormal basis of \( U \) \( \{ e_k \} \), it follows

\[
\sum_{k=1}^{\infty} ||\Psi (e_k)||^2 < \infty
\]

and the inner product for two of these, \( \Psi, \Phi \) is given by

\[
(\Psi, \Phi)_{\mathcal{L}_2} \equiv \sum_k (\Psi (e_k), \Phi (e_k))
\]

Then for such a Hilbert Schmidt operator, the norm in \( \mathcal{L}_2 \) is given by

\[
\left( \sum_{k=1}^{\infty} ||\Psi (e_k)||^2 \right)^{1/2} \equiv ||\Psi||_{\mathcal{L}_2}.
\]

Note this is the same as

\[
\left( \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} (\Psi (e_k), f_j)^2 \right)^{1/2} \quad (17.1.1)
\]

where \( \{ f_j \} \) is an orthonormal basis for \( H \). This is the analog of the Frobenius norm for matrices obtained as

\[
\text{trace} (MM^*)^{1/2} = \left( \sum_i (MM^*)_{ii} \right)^{1/2} = \left( \sum_{i,j} M_{ij}^2 \right)^{1/2}
\]

Also (17.1.1) shows right away that if \( \Psi \in \mathcal{L}_2 (U, H) \), then

\[
||\Psi||_{\mathcal{L}_2(U,H)}^2 = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} (e_k, f_j)^2_H
\]

\[
= \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} (e_k, \Psi^* f_j)^2_U = ||\Psi^*||_{\mathcal{L}_2(H,U)}^2
\]

and that \( \Psi \) and \( \Psi^* \) are Hilbert Schmidt together.

The filtration will continue to be denoted by \( \mathcal{F}_t \). It will be defined as the following normal filtration in which

\[
\sigma (W(s) - W(r) : 0 \leq r < s \leq u)
\]

is the completion of \( \sigma (W(s) - W(r) : 0 \leq r < s \leq u) \).

\[
\mathcal{F}_t \equiv \cap_{u > t} \sigma (W(s) - W(r) : 0 \leq r < s \leq u)
\]

(17.1.2)

and \( \sigma (W(s) - W(r) : 0 \leq r < s \leq u) \) denotes the \( \sigma \) algebra of all sets of the form

\[
(W(s) - W(r))^{-1} \text{(Borel)}
\]

where \( 0 \leq r < s \leq u \).

**Definition 17.1.2** Let \( \Phi (t) \in \mathcal{L} (U, H) \) be constant on each interval, \( (t_m, t_{m+1}] \) determined by a partition of \([a, T] \), \( 0 \leq a = t_0 < t_1 \cdots < t_n = T \). Then \( \Phi (t) \) is said to be elementary if also \( \Phi (t_m) \) is \( \mathcal{F}_{t_m} \) measurable and \( \Phi (t_m) \) equals a sum of the form

\[
\Phi (t_m) (\omega) = \sum_{j=1}^{m} \Phi_j \chi_{A_j}\]
where \( \Phi_j \in \mathcal{L}(U, H) \), \( A_j \in \mathcal{F}_{t_m} \). What does the measurability assertion mean? It means that if \( O \) is an open (Borel) set in the topological space \( \mathcal{L}(U, H) \), \( \Phi(t_m)^{-1}(O) \in \mathcal{F}_{t_m} \). Thus an elementary function is of the form

\[
\Phi(t) = \sum_{k=0}^{n-1} \Phi(t_k) \chi_{(t_k, t_{k+1})}(t).
\]

Then for \( \Phi \) elementary, the stochastic integral is defined by

\[
\int_a^t \Phi(s) \, dW(s) \equiv \sum_{k=0}^{n-1} \Phi(t_k) (W(t \wedge t_{k+1}) - W(t \wedge t_k)).
\]

It is also sometimes denoted by \( \Phi \cdot W(t) \).

The following lemma will be useful.

**Lemma 17.1.3** Let \( f, g \in L^2(\Omega; H) \) and suppose \( g \) is \( \mathcal{G} \) measurable and \( f \) is \( \mathcal{F} \) measurable where \( \mathcal{F} \supseteq \mathcal{G} \). Then

\[
E((f,g)_H|\mathcal{G}) = \langle E(f|\mathcal{G}), g \rangle_H \text{ a.e.}
\]

Similarly if \( \Phi \) is \( \mathcal{G} \) measurable as a map into \( \mathcal{L}(U, H) \) with

\[
\int_\Omega ||\Phi||^2 \, dP < \infty
\]

and \( f \) is \( \mathcal{F} \) measurable as a map into \( U \) such that \( f \in L^2(\Omega; H) \), then

\[
E(\Phi f|\mathcal{G}) = \Phi E(f|\mathcal{G}).
\]

**Proof:** Let \( A \in \mathcal{G} \). Let \( \{g_n\} \) be a sequence of simple functions, measurable with respect to \( \mathcal{G} \),

\[
g_n(\omega) = \sum_{k=1}^{m_n} a_k^n \chi_{E_k^n}(\omega)
\]

which converges in \( L^2(\Omega; H) \) and pointwise to \( g \). Then

\[
\int_A (E(f|\mathcal{G}), g)_H \, dP = \lim_{n \to \infty} \int_A (E(f|\mathcal{G}), g_n)_H \, dP
\]

\[
= \lim_{n \to \infty} \int_A \sum_{k=1}^{m_n} (E(f|\mathcal{G}), a_k^n \chi_{E_k^n})_H \, dP = \lim_{n \to \infty} \int_A \sum_{k=1}^{m_n} E(f, a_k^n)_H \chi_{E_k^n} \, dP
\]

\[
= \lim_{n \to \infty} \int_A \sum_{k=1}^{m_n} E(f, a_k^n \chi_{E_k^n})_H \, dP = \lim_{n \to \infty} \int_A \left( f, \sum_{k=1}^{m_n} a_k^n \chi_{E_k^n} \right)_H \, dP
\]

\[
= \lim_{n \to \infty} \int_A E(f(\cdot, g_n)_H|\mathcal{G}) \, dP = \lim_{n \to \infty} \int_A (f, g_n)_H \, dP = \int_A (f, g)_H \, dP
\]

which shows

\[
(E(f|\mathcal{G}), g)_H = E((f,g)_H|\mathcal{G})
\]

as claimed.

Consider the other claim. Let

\[
\Phi_n(\omega) = \sum_{k=1}^{m_n} \Phi_k^n \chi_{E_k^n}(\omega), \ E_k^n \in \mathcal{G}
\]
where \( \Phi_n \in \mathcal{L}(U, H) \) be such that \( \Phi_n \) converges to \( \Phi \) pointwise in \( \mathcal{L}(U, H) \) and also

\[
\int_{\Omega} \|\Phi_n - \Phi\|^2 \, dP \to 0.
\]

Then letting \( A \in \mathcal{G} \) and using Corollary \( \text{[16.24]} \) as needed,

\[
\int_A \Phi E(f|\mathcal{G}) \, dP
= \lim_{n \to \infty} \int_A \Phi_n E(f|\mathcal{G}) \, dP
= \lim_{n \to \infty} \sum_{k=1}^m \Phi_n^k E(f|\mathcal{G}) X_{E_k} \, dP
= \lim_{n \to \infty} \sum_{k=1}^m \Phi_n^k \int_A E(f|\mathcal{G}) X_{E_k} \, dP
= \lim_{n \to \infty} \sum_{k=1}^m \Phi_n^k \int_A E(X_{E_k}^* f|\mathcal{G}) \, dP
= \lim_{n \to \infty} \int_A \Phi_n f \, dP = \lim_{n \to \infty} \int_A \Phi f \, dP = \int_A E(\Phi f|\mathcal{G}) \, dP
\]

Since \( A \in \mathcal{G} \) is arbitrary, this proves the lemma. \( \blacksquare \)

**Lemma 17.1.4** Let \( J : U_0 \to U \) be a Hilbert Schmidt operator and let \( W(t) \) be the resulting Wiener process

\[
W(t) = \sum_{k=1}^\infty \psi_k(t) Jg_k
\]

where \( \{g_k\} \) is an orthonormal basis for \( U_0 \). Let \( f \in H \). Then considering one of the terms of the integral defined above,

\[
E \left( \left( \Phi(t_k) (W(t \wedge t_{k+1}) - W(t \wedge t_k)) , f \right)^2 \right)
= E \left( \left( (W(t \wedge t_{k+1}) - W(t \wedge t_k)) , \Phi(t_k)^* f \right)^2 \right)
= (t \wedge t_{k+1} - t \wedge t_k) E \left( \|J^* \Phi(t_k)^* f\|_{U_0}^2 \right).
\]

**Proof:** For simplicity, write \( \Delta W_k(t) \) for \( W(t \wedge t_{k+1}) - W(t \wedge t_k) \) and \( \Delta_k(t) = (t \wedge t_{k+1}) - (t \wedge t_k) \). If \( \Phi(t_k) \) were a constant, then the result would follow right away from the fact that \( W(t) \) is a Wiener process. Therefore, suppose for disjoint \( E_i \),

\[
\Phi(t_k)(\omega) = \sum_{i=1}^m \Phi_i X_{E_i}(\omega)
\]

where \( \Phi_i \in \mathcal{L}(U, H) \) and \( E_i \in \mathcal{F}_{t_k} \). Then, since the \( E_i \) are disjoint,

\[
E \left( \left( \Phi(t_k) (W(t \wedge t_{k+1}) - W(t \wedge t_k)) , f \right)^2 \right)
= \sum_{i=1}^m E \left( \left( (\Delta_k W(t)) , \Phi_i^* f X_{E_i} \right)^2 \right)
= \sum_{i=1}^m \int_{\Omega} \left( X_{E_i}^* ((\Delta_k W(t)) , \Phi_i^* f \right)^2 \, dP
\]

Thus each \( E_i \) is \( \mathcal{F}_{t_k} \) measurable. By Lemma \( \text{[16.32]} \) and the properties of the Wiener process, this equals

\[
\sum_{i=1}^m P(E_i) \int_{\Omega} \left( ((\Delta_k W(t)) , \Phi_i^* f \right)^2 \, dP = \sum_{i=1}^m P(E_i) \Delta_k t \left( Q \Phi_i^* f , \Phi_i^* f \right)_U
\]

where \( Q = JJ^* \). Then the above reduces to

\[
(t \wedge t_{k+1} - t \wedge t_k) E \left( \|J^* \Phi(t_k)^* f\|_{U_0}^2 \right). \quad \blacksquare
\]

Now here is a major result on the integral of elementary functions. The last assertion in the following proposition is called the Ito isometry.
Proposition 17.1.5  Let \( \Phi (t) \) be an elementary process as defined in Definition 17.1.2 and let \( W(t) \) be a Wiener process.

\[
W(t) = \sum_{k=1}^{\infty} \psi_k(t) Jg_k
\]

where \( J : U_0 \to U \) is Hilbert Schmidt and the \( \psi_k \) are real independent Wiener processes as described above.

\[
U_0 \xrightarrow{g_k} U \xrightarrow{W(t)} H
\]

Then \( \int_a^t \Phi(s) \, dW \) is a continuous square integrable \( H \) valued martingale with respect to the \( \sigma \) algebras of 17.1.2 on \([0,T]\) and

\[
E \left( \left| \int_a^t \Phi(s) \, dW \right|_H^2 \right) = \int_a^t E \left( \left| \Phi \circ J \right|_{L^2(U_0,H)}^2 \right) \, ds
\]

Proof: Start with the left side. Denote by \( \Delta_k W(t) \equiv W(t \wedge t_{k+1}) - W(t \wedge t_k) \). Then

\[
E \left( \left| \int_a^t \Phi(s) \, dW \right|_H^2 \right) = E \left( \left| \sum_{k=0}^{n-1} \Phi(t_k) \Delta_k W(t) \right|_H^2 \right).
\]

Consider a mixed term for \( j < k \). Using Lemma 17.1.3 and the fact that \( W(t) \) is a martingale,

\[
E \left( \left( \Phi(t_k) \Delta_k W(t), \Phi(t_j) \Delta_j W(t) \right)_H \right)
\]

\[
= E \left( E \left( \left( \Phi(t_k) \Delta_k W(t), \Phi(t_j) \Delta_j W(t) \right)_H \right| \mathcal{F}_{t_k} \right)
\]

\[
= E \left( \left( \Phi(t_j) \Delta_j W(t), \Phi(t_k) \Delta_k W(t) \right| \mathcal{F}_{t_k} \right)
\]

\[
= E \left( \left( \Phi(t_j) \Delta_j W(t), \Phi(t_k) \Delta_k W(t) \right| \mathcal{F}_{t_k} \right)
\]

Then, from Lemma 17.1.3 and letting \( \{f_j\} \) be an orthonormal basis for \( H \),

\[
E \left( \left| \int_a^t \Phi(s) \, dW \right|_H^2 \right) = \sum_{k=0}^{n-1} E \left( \left( \Phi(t_k) \Delta_k W(t), \Phi(t_k) \Delta_k W(t) \right)_H \right)
\]

\[
= \sum_{k=0}^{n-1} \left( \sum_{j=1}^{\infty} \left( \Phi(t_k) \Delta_k W(t), f_j \right)_H^2 \right) = \sum_{k=0}^{n-1} \sum_{j=1}^{\infty} \left( \Phi(t_k) \Delta_k W(t), f_j \right)^2
\]

\[
= \sum_{k=0}^{n-1} \sum_{j=1}^{\infty} (t \wedge t_{k+1} - t \wedge t_k) E \left( \left| J^* \Phi(t_k)^* f_j \right|_{U_0}^2 \right)
\]

\[
= \sum_{k=0}^{n-1} (t \wedge t_{k+1} - t \wedge t_k) E \left( \left| J^* \Phi(t_k)^* \right|_{L^2(H,U_0)}^2 \right)
\]

\[
= \sum_{k=0}^{n-1} (t \wedge t_{k+1} - t \wedge t_k) E \left( \left| \Phi(t_k) J \right|_{L^2(U_0,H)}^2 \right)
\]

\[
= \int_a^t E \left( \left| \Phi \circ J \right|_{L^2(U_0,H)}^2 \right) \, ds
\]

It is obvious that \( \int_a^t \Phi(s) \, dW \) is a continuous square integrable martingale from the definition, because it is just a finite sum of such things.

Of course this is a version of the Ito isometry. The presence of the \( J \) is troublesome but it is hidden in the definition of \( W \) on the left side of the conclusion of the proposition. In finite dimensions one could just let \( J = I \) and this fussy detail would not be there to cause confusion. The next task is to generalize the above integral to a more general class of functions and obtain a process which is not explicitly dependent on \( J \).
17.2 Different Definition Of Elementary Functions

What if elementary functions had been defined in terms of \( \mathcal{X}_{[t_k,t_{k+1}]} \)? That is, what if the elementary functions had been of the form
\[
\Phi (t) = \sum_{k=0}^{n-1} \Phi (t_k) \mathcal{X}_{[t_k,t_{k+1}]} (t)
\]
Would anything change? If you go over the arguments given, it is clear that nothing would change at all. Furthermore, this elementary function equals the one described above off a finite set of mesh points so the convergence properties in \( L^2 ([0,T] \times \Omega, \mathcal{L}_2 (Q^{1/2} U, H) ) \), which will be important in what follows are exactly the same. Thus it does not matter whether we give elementary functions in this form or in the form described above. However, some arguments given later about localization depend on it being in the earlier form.

17.3 Approximating With Elementary Functions

Here is a really surprising result about approximating with step functions which is due to Doob. See \( \text{[17]} \) which is where I found this lemma.

**Lemma 17.3.1** Let \( \Phi : [0,T] \times \Omega \to E \), be \( B ([0,T]) \times \mathcal{F} \) measurable and suppose
\[
\Phi \in K \equiv L^p ([0,T] \times \Omega; E), \ p \geq 1
\]
Then there exists a sequence of nested partitions, \( \mathcal{P}_k \subseteq \mathcal{P}_{k+1} \),
\[
\mathcal{P}_k \equiv \{ t_0^k, \cdots , t_{m_k}^k \}
\]
such that the step functions given by
\[
\Phi^k \equiv \sum_{j=1}^{m_k} \Phi (t_j^k) \mathcal{X}_{[t_{j-1}^k,t_j^k]} (t)
\]
both converge to \( \Phi \) in \( K \) as \( k \to \infty \) and
\[
\lim_{k \to \infty} \max \{ |t_j^k - t_{j+1}^k| : j \in \{ 0, \cdots , m_k \} \} = 0.
\]

Also, each \( \Phi (t_j^k) \) and \( \Phi (t_{j-1}^k) \) is in \( L^p (\Omega; E) \). One can also assume that \( \Phi (0) = 0 \). The mesh points \( \{ t_j^k \}_{j=0}^{m_k} \) can be chosen to miss a given set of measure zero. In addition to this, we can assume that
\[
|t_j^k - t_{j-1}^k| = 2^{-nk}
\]
except for the case where \( j = 1 \) or \( j = m_{nk} \) when this is so, you could have \( |t_j^k - t_{j-1}^k| < 2^{-nk} \).

Note that it would make no difference in terms of the conclusion of this lemma if you defined
\[
\Phi^k \equiv \sum_{j=1}^{m_k} \Phi (t_{j-1}^k) \mathcal{X}_{[t_{j-1}^k,t_j^k]} (t)
\]
because the modified function equals the one given above off a countable subset of \( [0,T] \), the union of the mesh points. One could make a similar change to \( \Phi^k \) with no change in the conclusion.

**Proof:** For \( t \in \mathbb{R} \) let \( \gamma_n (t) \equiv k/2^n, \delta_n (t) \equiv (k+1)/2^n \), where \( t \in (k/2^n,(k+1)/2^n) \), and \( 2^{-n} < T/4 \). Also suppose \( \Phi \) is defined to equal 0 on \( [0,T] \times \Omega \). There exists a set of measure zero \( N \) such that for \( \omega \notin N, t \to \| \Phi (t, \omega) \| \) is in \( L^p (\mathbb{R}) \). Therefore by continuity of translation, as \( n \to \infty \) it follows that for \( \omega \notin N \), and \( t \in [0,T] \),
\[
\int_{\mathbb{R}} \| \Phi (\gamma_n (t) + s) - \Phi (t + s) \|_E^p ds \to 0
\]
The above is dominated by
\[
\int_{-2T}^{2T} 2^{p-1} (||\Phi(s)||^p + ||\Phi(s)||^p) \mathcal{X}_{[-2T,2T]}(s) \, ds
\]
\[
= \int_{-2T}^{2T} 2^{p-1} (||\Phi(s)||^p + ||\Phi(s)||^p) \, ds < \infty
\]

Consider
\[
\int_{\Omega} \int_{-2T}^{2T} \left( \int_{\mathbb{R}} ||\Phi(\gamma_n(t) + s) - \Phi(t + s)||_E^p \, ds \right) dt \, dP
\]

By the dominated convergence theorem, this converges to 0 as \( n \to \infty \). This is because the integrand with respect to \( \omega \) is dominated by
\[
\int_{-2T}^{2T} \left( \int_{\mathbb{R}} 2^{p-1} (||\Phi(s)||^p + ||\Phi(s)||^p) \mathcal{X}_{[-2T,2T]}(s) \, ds \right) dt
\]
and this is in \( L^1(\Omega) \) by assumption that \( \Phi \in K \). Now Fubini. This yields
\[
\int_{\Omega} \int_{\mathbb{R}} \int_{-2T}^{2T} ||\Phi(\gamma_n(t) + s) - \Phi(t + s)||_E^p \, dt \, ds \, dP
\]

Change the variables on the inside.
\[
\int_{\Omega} \int_{\mathbb{R}} \int_{-2T}^{2T} ||\Phi(\gamma_n(t - s) + s) - \Phi(t)||_E^p \, dt \, ds \, dP
\]

Now by definition, \( \Phi(t) \) vanishes if \( t \notin [0,T] \), thus the above reduces to
\[
\int_{\Omega} \int_{\mathbb{R}} \int_{0}^{T} ||\Phi(\gamma_n(t - s) + s) - \Phi(t)||_E^p \, dt \, ds \, dP + \int_{\Omega} \int_{\mathbb{R}} \int_{-2T+s}^{2T+s} \mathcal{X}_{[0,T]}(\gamma_n(t - s) + s) ||\Phi(t)||_E^p \, dtds \, dP
\]
\[
= \int_{\Omega} \int_{\mathbb{R}} \int_{0}^{T} ||\Phi(\gamma_n(t - s) + s) - \Phi(t)||_E^p \, dt \, ds \, dP + \int_{\Omega} \int_{\mathbb{R}} \int_{-2T+s}^{2T+s} \mathcal{X}_{[0,T]}(\gamma_n(t - s) + s) ||\Phi(t)||_E^p \, dtds \, dP
\]

Also by definition, \( \gamma_n(t - s) + s \) is within \( 2^{-n} \) of \( t \) and so the integrand in the integral on the right equals 0 unless \( t \in [-2^{-n} - T, T + 2^{-n}] \subseteq [-2T,2T] \). Thus the above reduces to
\[
\int_{\Omega} \int_{\mathbb{R}} \int_{-2T}^{2T} ||\Phi(\gamma_n(t - s) + s) - \Phi(t)||_E^p \, dt \, ds \, dP.
\]

Now Fubini again.
\[
\int_{\mathbb{R}} \int_{\Omega} \int_{-2T}^{2T} ||\Phi(\gamma_n(t - s) + s) - \Phi(t)||_E^p \, dt \, ds \, dP \, ds
\]

This converges to 0 as \( n \to \infty \) as was shown above. Therefore,
\[
\int_{0}^{T} \int_{\Omega} \int_{0}^{T} ||\Phi(\gamma_n(t - s) + s) - \Phi(t)||_E^p \, dt \, ds \, dP
\]
also converges to 0 as \( n \to \infty \). The only problem is that \( \gamma_n(t - s) + s \geq t - 2^{-n} \) and so \( \gamma_n(t - s) + s \) could be less than 0 for \( t \in [0, 2^{-n}] \). Since this is an interval whose measure converges to 0 it follows
\[
\int_{0}^{T} \int_{\Omega} \int_{0}^{T} \left||\Phi(\gamma_n(t - s) + s) + \Phi(t)\right|_E^p \, dt \, ds \, dP
\]
converges to 0 as \( n \to \infty \). Let
\[
m_n(s) = \int_{\Omega} \int_{0}^{T} \left||\Phi(\gamma_n(t - s) + s) + \Phi(t)\right|_E^p \, dt \, ds
\]
Then letting \( \mu \) denote Lebesgue measure,
\[
\mu\left(\{m_n(s) > \lambda\}\right) \leq \frac{1}{\lambda} \int_{0}^{T} m_n(s) \, ds.
\]
It follows there exists a subsequence $n_k$ such that

$$
\mu\left(\left\{ m_{n_k}(s) > \frac{1}{k} \right\}\right) < 2^{-k}
$$

Hence by the Borel Cantelli lemma, there exists a set of measure zero $N$ such that for $s \notin N$,

$$
m_{n_k}(s) \leq 1/k
$$

for all $k$ sufficiently large. Pick such an $s$. Then consider $t \to \Phi\left(\left(\gamma_{n_k}(t-s)+s\right)^+\right)$. For $n_k$, $t \to \left(\gamma_{n_k}(t-s)+s\right)^+$ has jumps at points of the form $0, s + l2^{-m}$ where $l$ is an integer. Thus $P_{n_k}$ consists of points of $[0, T]$ which are of this form and these partitions are nested. Define $\Phi_k(0) \equiv 0$, $\Phi_k(t) \equiv \Phi\left(\left(\gamma_{n_k}(t-s)+s\right)^+\right)$. Now suppose $N_1$ is a set of measure zero. Can $s$ be chosen such that all jumps for all partitions occur off $N_1$? Let $(a, b)$ be an interval contained in $[0, T]$. Let $S_j$ be the points of $(a, b)$ which are translations of the measure zero set $N_1$ by $t_j$ for some $j$. Thus $S_j$ has measure 0. Now pick $s \in (a, b) \setminus \cup_j S_j$.

It will be assumed that all these mesh points miss the set of all $t$ such that $\omega \to \Phi(t, \omega)$ is not in $L^p(\Omega; E)$. To get the other sequence of step functions, the right step functions, just use a similar argument with $\delta_{n_k}$ in place of $\gamma_n$. Just apply the argument to a subsequence of $n_k$ so that the same $s$ can hold for both.

The following proposition says that elementary functions can be used to approximate progressively measurable functions under certain conditions.

**Proposition 17.3.2** Let $\Phi \in L^p([0, T] \times \Omega, E)$, $p \geq 1$, be progressively measurable. Then there exists a sequence of elementary functions which converges to $\Phi$ in

$L^p([0, T] \times \Omega, E)$.

These elementary functions have values in $E_0$, a dense subset of $E$. If $\varepsilon_n \to 0$, and

$$
\Phi_n(t) = \sum_{k=1}^{m_n} \Psi^n_k \chi_{(t_{k-1}, t_k]}(t)
$$

$\Psi^n_k$ having values in $E_0$, it can be assumed that

$$
\sum_{k=1}^{m_n} \left\| \Psi^n_k - \Phi(t_k) \right\|_{L^p(\Omega; E)} < \varepsilon_n.
$$

**Proof:** By Lemma 17.3.1 there exists a sequence of step functions

$$
\Phi_k(t) = \sum_{j=1}^{m_k} \Phi(t_{j-1}^k) \chi_{(t_{j-1}^k, t_j^k]}(t)
$$

which converges to $\Phi$ in $L^p([0, T] \times \Omega, E)$ where at the left endpoint $\Phi(0)$ can be modified as described above. Now each $\Phi(t_{j-1}^k)$ is in $L^p(\Omega, E)$ and is $\mathcal{F}(t_{j-1}^k)$ measurable and so it can be approximated as closely as desired in $L^p(\Omega)$ with a simple function

$$
s(t_{j-1}^k) \equiv \sum_{i=1}^{m_k} c_i^k \chi_{F_i}(\omega), \quad F_i \in \mathcal{F}(t_{j-1}^k).
$$

Furthermore, by density of $E_0$ in $E$, it can be assumed each $c_i^k \in E_0$ and the condition 17.3.3 holds. Replacing each $\Phi(t_{j-1}^k)$ with $s(t_{j-1}^k)$, the result is an elementary function which approximates $\Phi_k$.

Of course everything in the above holds with obvious modifications replacing $[0, T]$ with $[a, T]$ where $a < T$.

Here is another interesting proposition about the time integral being adapted.

**Proposition 17.3.3** Suppose $f \geq 0$ is progressively measurable and $\mathcal{F}_t$ is a filtration. Then

$$
\omega \to \int_a^t f(s, \omega) \, ds
$$

is $\mathcal{F}_t$ adapted.

**Proof:** This follows right away from the fact $\chi_{[0,t]} f$ is $\mathcal{B}([a, t]) \times \mathcal{F}_t$ measurable. This is just product measure and so the integral from $a$ to $t$ is $\mathcal{F}_t$ measurable. See also Proposition 17.3.1.
17.4 Some Hilbert Space Theory

Recall the following definition which makes $LU$ into a Hilbert space where $L \in \mathcal{L}(U, H)$.

**Definition 17.4.1** Let $L \in \mathcal{L}(U, H)$, the bounded linear maps from $U$ to $H$ for $U, H$ Hilbert spaces. For $y \in L(U)$, let $L^{-1}y$ denote the unique vector in $\{ x : Lx = y \}$ which is closest in $U$ to 0.

Note this is a good definition because $\{ x : Lx = y \}$ is closed thanks to the continuity of $L$ and it is obviously convex. Thus Theorem 6.0.9 applies. With this definition define an inner product on $L(U)$ as follows. For $y, z \in L(U)$,

$$ (y, z)_{L(U)} \equiv (L^{-1}y, L^{-1}z)_U $$

Thus it is obvious that $L^{-1} : LU \to U$ is continuous. The notation is abominable because $L^{-1}(y)$ is the normal notation for $M_y$.

With this definition, here is one of the main results. It is Theorem 6.1.3 proved earlier.

**Theorem 17.4.2** Let $U, H$ be Hilbert spaces and let $L \in \mathcal{L}(U, H)$. Then Definition 17.4.1 makes $L(U)$ into a Hilbert space. Also $L : U \to L(U)$ is continuous and $L^{-1} : L(U) \to U$ is continuous. Furthermore there is a constant $C$ independent of $x \in U$ such that

$$ \|L\|_{\mathcal{L}(U, H)} \|Lx\|_{L(U)} \geq \|Lx\|_H \quad (17.4.5) $$

If $U$ is separable, so is $L(U)$. Also $(L^{-1}(y), x) = 0$ for all $x \in \ker(L)$, and $L^{-1} : L(U) \to U$ is linear. Also, in case that $L$ is one to one, both $L$ and $L^{-1}$ preserve norms.

Let $U$ be a separable Hilbert space and let $Q$ be a positive self adjoint operator. Then consider

$$ J : Q^{1/2}U \to U_1, $$

a one to one Hilbert Schmidt operator, where $U_1$ is a separable real Hilbert space. First of all, there is the obvious question whether there are any examples.

**Lemma 17.4.3** Let $A \in \mathcal{L}(U, U)$ be a bounded linear transformation defined on $U$ a separable real Hilbert space. There exists a one to one Hilbert Schmidt operator $J : AU \to U_1$ where $U_1$ is a separable real Hilbert space. In fact you can take $U_1 = U$.

**Proof:** Let $\alpha_k > 0$ and $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$. Then let $\{g_k\}_{k=1}^{L}$ be an orthonormal basis for $AU$, the inner product and norm given in Definition 17.4.1 above, and let

$$ Jx \equiv \sum_{k=1}^{L} (x, g_k)_{AU} \alpha_k g_k. $$
Then it is clear that $J \in \mathcal{L}(AU, U)$. This is because,
\[
\|Jx\|_U \leq \sum_{k=1}^L |(x, g_k)_{AU}| \alpha_k \|g_k\|_U
\]
\[
\leq C \sum_{k=1}^L |(x, g_k)_{AU}| \alpha_k \|g_k\|_{AU}^{1/2}
\]
\[
\leq C \left( \sum_{k=1}^L |(x, g_k)_{AU}|^2 \right)^{1/2} \left( \sum_{k=1}^L \alpha_k^2 \right)^{1/2}
\]
\[
= C \left( \sum_{k=1}^L \alpha_k^2 \right)^{1/2} \|x\|_{AU}
\]

Also, from the definition, $Jg_j = \alpha_j g_j$. Say $g_j = A f_j$ where $f_j \in U$ and $1 = \|g_j\|_{AU} = \|f_j\|_U$. Since $A$ is continuous,
\[
\|g_j\|_U = \|Af_j\|_U \leq \|A\| \|f_j\|_U = \|A\| \|g_j\|_{AU} = \|A\| \equiv C^{1/2}
\]
Thus
\[
\sum_{j=1}^L \|Jg_j\|_U^2 = \sum_{j=1}^L \alpha_j^2 \|g_j\|_{U}^2 \leq C \sum_{j=1}^L \alpha_j^2 < \infty
\]
and so $J$ is also a Hilbert Schmidt operator which maps $AU$ to $U$. It is clear that $J$ is one to one because each $\alpha_k > 0$. If $AU$ is finite dimensional, $L < \infty$ and so the above sum is finite. ■

**Definition 17.4.4** Let $U_1, U, H$ be real separable Hilbert spaces and let $Q$ be a nonnegative self adjoint operator, $Q \in \mathcal{L}(U, U)$. Let $Q^{1/2}U$ be the Hilbert space described in Definition [17.4.1]. Let $J$ be a one to one Hilbert Schmidt map from $Q^{1/2}U$ to $U_1$.
\[U_1 \xleftarrow{J} Q^{1/2}U \xrightarrow{\Phi} H\]
Then denote by $\mathcal{L}(U_1, H)_0$ the space of restrictions of elements of $\mathcal{L}(U_1, H)$ to the Hilbert space $JQ^{1/2}U \subseteq U_1$.

Here is a diagram to keep this straight.

\[\begin{array}{c}
U_1 \supseteq JQ^{1/2}U \xleftarrow{J} Q^{1/2}U \\
\Phi \downarrow \\
H
\end{array}\]

**Lemma 17.4.5** In the context of the above definition, $\mathcal{L}(U_1, H)_0$ is dense in
\[\mathcal{L}_2\left(JQ^{1/2}U, H\right),\]
the Hilbert Schmidt operators from $JQ^{1/2}U$ to $H$. That is, if $f \in \mathcal{L}_2\left(JQ^{1/2}U, H\right)$, there exists $g \in \mathcal{L}(U_1, H)_0$, $\|g - f\|_{\mathcal{L}_2(JQ^{1/2}U, H)} < \varepsilon$.

**Proof:** The operator $JJ^* \equiv Q_1 : U_1 \rightarrow U_1$ is self adjoint and nonnegative. It is also compact because $J$ is Hilbert Schmidt. Therefore, by Theorem 6.2.3 on Page 172,
\[Q_1 = \sum_{k=1}^L \lambda_k e_k \otimes e_k\]
where the $\lambda_k$ are decreasing and positive, the $\{e_k\}$ are an orthonormal basis for $U_1$, and $\lambda_L$ is the last positive $\lambda_j$. (This is a lot like the singular value matrix in linear algebra.) Thus also
\[Q_1 e_k = \lambda_k e_k\]
If the \( \lambda_k \) are all positive, then \( L \equiv \infty \). Then for \( k \leq L \) if \( L < \infty \), \( k < \infty \) otherwise,

\[
\left( \frac{J^*e_j}{\sqrt{\lambda_j}} \right)_U = \left( \frac{JJ^*e_j}{\sqrt{\lambda_j}} \right)_U = \left( \frac{\lambda_k e_k}{\sqrt{\lambda_k}} \right)_U = \frac{\sqrt{\lambda_k}}{\sqrt{\lambda_j}} \delta_{kj} = \delta_{jk}
\]

Now in case \( L < \infty \), \( J \left( Q^{1/2} (U) \right) \subseteq \text{span} (e_1, \ldots, e_L) \). Here is why. First note that \( Q_1 \) is one to one on \( \text{span} (e_1, \ldots, e_L) \) and maps this space onto itself because \( Q_1 \) maps \( e_k \) to a nonzero multiple of \( e_k \). Hence its restriction to this subspace has an inverse which does the same. It also maps all of \( U \) to \( \text{span} (e_1, \ldots, e_L) \). This follows from the definition of \( Q_1 \) given in the above sum. For \( x \in Q^{1/2} (U) \), \( Jx \in U \) and so

\[ JJ^*Jx = Q_1 (Jx) \in \text{span} (e_1, \ldots, e_L) \]

Hence \( Jx \in Q_1^{-1} \left( \text{span} (e_1, \ldots, e_L) \right) \in \text{span} (e_1, \ldots, e_L) \). Recall that \( J \) is one to one so there is only one element of \( J^{-1}x \).

Then for \( x \in Q^{1/2}U \),

\[
\sum_{j=1}^{L} \sqrt{\lambda_j} e_j \otimes Q^{1/2}U \frac{J^*e_j}{\sqrt{\lambda_j}} (x) = \sum_{j=1}^{L} \sqrt{\lambda_j} e_j \left( \frac{J^*e_j}{\sqrt{\lambda_j}} , x \right)_{Q^{1/2}U}
\]

\[
= \sum_{j=1}^{L} \sqrt{\lambda_j} e_j \left( \frac{e_j}{\sqrt{\lambda_j}} , Jx \right)_{U_1} = \sum_{j=1}^{L} e_j (e_j , Jx)_{U_1}
\]

\[
= \sum_{j=1}^{\infty} e_j (e_j , Jx)_{U_1} = Jx. \quad (J \left( Q^{1/2} (U) \right) \subseteq \text{span} (e_1, \ldots, e_L) \text{ if } L < \infty)
\]

Thus,

\[ J = \sum_{j=1}^{L} \sqrt{\lambda_j} e_j \otimes Q^{1/2}U \frac{J^*e_j}{\sqrt{\lambda_j}} \]

It follows that an orthonormal basis in \( JQ^{1/2}U \) is \( \left\{ \frac{J^*e_j}{\sqrt{\lambda_j}} \right\}_{j=1}^{L} \). This is because an orthonormal basis for \( Q^{1/2}U \)

\[ \left\{ \frac{J^*e_j}{\sqrt{\lambda_j}} \right\}_{j=1}^{L} \]

is \( \left\{ \frac{J^*e_j}{\sqrt{\lambda_j}} \right\}_{j=1}^{L} \). Since \( J \) is one to one, it preserves norms between \( Q^{1/2}U \) and \( JQ^{1/2}U \). Let \( \Phi \in L_2 \left( JQ^{1/2}U , H \right) \). Then by the discussion of Hilbert Schmidt operators given earlier, in particular the demonstration that these operators are compact,

\[ \Phi = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \phi_{ij} f_i \otimes JQ^{1/2}U \frac{J^*e_j}{\sqrt{\lambda_j}} \]

where \( \{ f_i \} \) is an orthonormal basis for \( H \). In fact, \( \left\{ f_i \otimes \frac{J^*e_j}{\sqrt{\lambda_j}} \right\}_{i,j} \) is an orthonormal basis for \( L_2 \left( JQ^{1/2}U , H \right) \) and

\[ \sum_{i} \sum_{j} \phi_{ij}^2 < \infty \], the \( \phi_{ij} \) being the Fourier coefficients of \( \Phi \). Then consider

\[ \Phi_n = \sum_{i=1}^{n} \sum_{j=1}^{n} \phi_{ij} f_i \otimes JQ^{1/2}U \frac{J^*e_j}{\sqrt{\lambda_j}} \tag{17.4.6} \]

Consider one of the finitely many operators in this sum. For \( x \in JQ^{1/2}U \), since \( J \) preserves norms,

\[
f_i \otimes JQ^{1/2}U \frac{J^*e_j}{\sqrt{\lambda_j}} (x) = f_i \left( \frac{J^*e_j}{\sqrt{\lambda_j}} , x \right)_{JQ^{1/2}U} = f_i \left( \frac{J^*e_j}{\sqrt{\lambda_j}} , J^{-1}x \right)_{Q^{1/2}U} = f_i \left( \frac{e_j}{\sqrt{\lambda_j}} , J^{-1}x \right)_{U_1} = \lambda_{ij} (x)
\]
Recall how, since \( J \) is one to one, it preserves norms and inner products. Now \( \Lambda_{ij} \) makes sense from the above formula for all \( x \in U_1 \) and is also a continuous linear map from \( U_1 \) to \( H \) because

\[
\left\| f_1 \left( \frac{e_j}{\sqrt{\lambda_j}}, x \right) \right\|_{U_1 H} \leq \|f_1\|_H \frac{1}{\sqrt{\lambda_j}} \|x\|_{U_1}
\]

Thus each term in the finite sum of \( \left\| f_1 \left( \frac{e_j}{\sqrt{\lambda_j}}, x \right) \right\|_{U_1 H} \) is in \( \mathcal{L}(U_1, H)_0 \) and this proves the lemma. ■

It is interesting to note that \( Q_{1/2} U_1 = J \left( Q_{1/2} (U) \right) \).

\[
\sum_{j=1}^{L} \sqrt{\lambda_j} e_j \left( \frac{J^* e_j}{\sqrt{\lambda_j}}, x \right)_{Q^{1/2}(U)} = J x
\]

and \( \left\{ \frac{J^* e_j}{\sqrt{\lambda_j}} \right\} \) are an orthonormal set in \( Q^{1/2} (U) \). Therefore, the sum of the squares of \( \left( \frac{J^* e_j}{\sqrt{\lambda_j}}, x \right)_{Q^{1/2}(U)} \) is finite. Hence you can define \( y \in U_1 \) by

\[
y \equiv \sum_{j=1}^{L} \left( \frac{J^* e_j}{\sqrt{\lambda_j}}, x \right)_{Q^{1/2}(U)} e_j
\]

Also

\[
\sum_{i=1}^{L} \sqrt{\lambda_i} e_i \otimes e_i (y) = \sum_{i=1}^{L} \sqrt{\lambda_i} e_i \left( \frac{J^* e_i}{\sqrt{\lambda_i}}, x \right)_{Q^{1/2}(U)}
\]

\[
= \sum_{i=1}^{L} e_i (e_i, Jx)_{U_1} = Jx
\]

Now you can show that \( Q_{1/2} = \sum_{i=1}^{L} \sqrt{\lambda_i} e_i \otimes e_i \). You do this by showing that it works and commutes with every operator which commutes with \( Q_1 \). Thus \( Jx = Q_{1/2} y \). This shows that \( J \left( Q_{1/2} (U) \right) \subseteq Q_{1/2} (U_1) \). However, you can also turn the inclusion around. Thus if you start with \( y \in U_1 \) and form

\[
Q_{1/2} y = \sum_{i=1}^{L} \sqrt{\lambda_i} e_i \otimes e_i (y) = \sum_{i=1}^{L} \sqrt{\lambda_i} e_i (y, e_i),
\]

then the \( (y, e_i)_U_1 \) has a finite sum because the \( \{e_i\} \) are orthonormal. Thus you can form

\[
x \equiv \sum_{i=1}^{L} (y, e_i)_U_1 \frac{J^* e_i}{\sqrt{\lambda_i}} \in Q^{1/2} (U)
\]

Then since the \( \left\{ \frac{J^* e_i}{\sqrt{\lambda_i}} \right\} \) are orthonormal,

\[
J (x) = \sum_{j=1}^{L} \sqrt{\lambda_j} e_j \left( \frac{J^* e_j}{\sqrt{\lambda_j}}, x \right)_{Q^{1/2}(U)} = \sum_{j=1}^{L} \sqrt{\lambda_j} e_j (y, e_j)_{U_1}
\]

\[
= \sum_{j=1}^{L} \sqrt{\lambda_j} e_j \otimes e_j (y) = Q_{1/2} (y)
\]

It follows that \( Q_{1/2} (U_1) \subseteq J \left( Q_{1/2} (U) \right) \).

One can also show that \( W(t) \equiv \sum_{k=1}^{K} \psi_k (t) Jg_k \) where the \( \psi_k (t) \) are the real Wiener processes described earlier and \( \{g_k\} \) is an orthonormal basis for \( Q_{1/2} (U) \), is a \( Q_1 \) Wiener process. To see this, recall the above definition of a Wiener process in terms of Hilbert Schmidt operators, the convergence happening in \( U_1 \) in this case. Then by independence of the \( \psi_j \),

\[
E \left( \left( h, \sum_{k=1}^{L} \psi_k (t-s) Jg_k \right) \left( l, \sum_{j=1}^{L} \psi_j (t-s) Jg_j \right) \right) = E \left( \sum_k (h, Jg_k) (l, Jg_j) \psi_k (t-s) \psi_j (t-s) \right)
\]
Now since \( \Phi \) is Hilbert Schmidt, it follows that \( \Phi^* \) is Hilbert Schmidt if and only if \( \Phi \) is progressively measurable with respect to the usual filtration associated with the Wiener process \( U \).

Since \( \Phi \) is given to be progressively measurable, so is \( \Phi^* \). Thus \( \Phi^* \) is Hilbert Schmidt.

There exists a sequence of elementary functions follows from Proposition 17.4.5.

It is time to generalize the integral. The following diagram illustrates the ingredients of the next lemma.

\[
\begin{array}{c}
W(t) \in U_1 \overset{J}{\to} Q^{1/2}U \overset{\Phi^*}{\to} H \\
U_1 \ni JQ^{1/2}U \overset{J_{i=1}}{\to} Q^{1/2}U \\
\Phi_n \ni \Phi \\
H
\end{array}
\]

**Lemma 17.5.1** Let \( \Phi \in L^2 ([a,T] \times \Omega; \mathcal{L}_2 (Q^{1/2}U, H)) \) and suppose also that \( \Phi \) is progressively measurable with respect to the usual filtration associated with the Wiener process

\[
W(t) = \sum_{k=1}^{L} \psi_k(t) Jg_k
\]

which has values in \( U_1 \) for \( U_1 \) a separable real Hilbert space such that \( J : Q^{1/2}U \to U_1 \) is Hilbert Schmidt and one to one, \( \{g_k\} \) an orthonormal basis in \( Q^{1/2}U \). Then letting \( J^{-1} : JQ^{1/2}U \to Q^{1/2}U \) be the map described in Definition 17.4.4, it follows that

\[
\Phi \circ J^{-1} \in L^2 ([a,T] \times \Omega; \mathcal{L}_2 (JQ^{1/2}U, H))
\]

Also there exists a sequence of elementary functions \( \{\Phi_n\} \) having values in \( L(U_1, H)_0 \) which converges to \( \Phi \circ J^{-1} \) in \( L^2 ([a,T] \times \Omega; \mathcal{L}_2 (JQ^{1/2}U, H)) \).

**Proof:** First, why is \( \Phi \circ J^{-1} \in L^2 ([a,T] \times \Omega; \mathcal{L}_2 (JQ^{1/2}U, H)) \)? This follows from the observation that \( A \) is Hilbert Schmidt if and only if \( A^* \) is Hilbert Schmidt. In fact, the Hilbert Schmidt norms of \( A \) and \( A^* \) are the same. Now since \( \Phi \) is Hilbert Schmidt, it follows that \( \Phi^* \) is also Hilbert Schmidt. Also letting \( \mathcal{L}_2 \) be the appropriate space of Hilbert Schmidt operators,

\[
|||(J^{-1})^*||| \leq |||\Phi^*|||_{\mathcal{L}_2} \leq |||\Phi \circ J^{-1}|||_{\mathcal{L}_2} = |||\Phi \circ J^{-1}|||_{\mathcal{L}_2}
\]

Thus \( \Phi \circ J^{-1} \) has values in \( \mathcal{L}_2 (JQ^{1/2}U, H) \). This also shows that

\[
\Phi \circ J^{-1} \in L^2 ([a,T] \times \Omega; \mathcal{L}_2 (JQ^{1/2}U, H))
\]

Since \( \Phi \) is given to be progressively measurable, so is \( \Phi \circ J^{-1} \). Therefore, the existence of the desired sequence of elementary functions follows from Proposition 17.4.5 and Lemma 17.5.1.

**Definition 17.5.2** Let \( \Phi \in L^2 ([a,T] \times \Omega; \mathcal{L}_2 (Q^{1/2}U, H)) \) and be progressively measurable where \( Q \) is a self-adjoint nonnegative operator defined on \( U \). Let \( J : Q^{1/2}U \to U_1 \) be Hilbert Schmidt. Then the stochastic integral

\[
\int_a^t \Phi dW
\]

is defined as

\[
\lim_{n \to \infty} \int_a^t \Phi_n dW \text{ in } L^2 (\Omega; H)
\]
where $W(t)$ is a Wiener process

$$
\sum_{k=1}^{\infty} \psi_k(t) Jg_k, \{g_k\} \text{ orthonormal basis in } Q^{1/2} U,
$$

and $\Phi_n$ is an elementary function which has values in $\mathcal{L}(U_1,H)$ and converges to $\Phi \circ J^{-1}$ in

$$
L^2 \left( \{a,T\} \times \Omega; \mathcal{L}_2 \left( JQ^{1/2} U, H \right) \right),
$$
such a sequence exists by Lemma 17.4.3 and Proposition 17.3.2.

It is necessary to show that this is well defined and does not depend on the choice of $U_1$ and $J$.

**Theorem 17.5.3** The stochastic integral $\int_{a}^{T} \Phi(s) dW$ is well defined. It also is a continuous martingale and does not depend on the choice of $J$ and $U_1$. Furthermore,

$$
E \left( \left| \int_{a}^{T} \Phi(s) dW \right|^2 \right) = \int_{a}^{T} E \left( \|\Phi\|^2_{\mathcal{L}_2(Q^{1/2} U, H)} \right) ds
$$

**Proof:** First of all, it is obvious that it is well defined in the sense that the same stochastic process is obtained from two different sequences of elementary functions. This follows from the isometry of Proposition 17.1.3 with $U_1$ in place of $U$ and $Q^{1/2} U$ in place of $U_0$. Thus if $\{\Psi_n\}$ and $\{\Phi_n\}$ are two sequences of elementary functions converging to $\Phi \circ J^{-1}$ in $L^2 \left( \{a,T\} \times \Omega; \mathcal{L}_2 \left( JQ^{1/2} U, H \right) \right)$,

$$
E \left( \left| \int_{a}^{T} (\Phi_n(s) - \Psi_n(s)) dW \right|^2 \right) = \int_{a}^{T} E \left( \|\Phi_n - \Psi_n\|_{\mathcal{L}_2(Q^{1/2} U, H)}^2 \right) ds
$$

(17.5.8)

Now for $\Phi \in \mathcal{L}_2(U_1,H)$ and $\{g_k\}$ an orthonormal basis for $Q^{1/2} U$,

$$
\|\Phi \circ J\|^2_{\mathcal{L}_2(Q^{1/2} U, H)} = \sum_{k=1}^{\infty} \|\Phi(J(g_k))\|^2_{H} = \|\Phi\|^2_{\mathcal{L}_2(JQ^{1/2} U, H)}
$$

because, by definition, $\{Jg_k\}$ is an orthonormal basis in $JQ^{1/2} U$. Hence (17.5.8) reduces to

$$
\int_{a}^{T} E \left( \|\Phi_n - \Psi_n\|_{\mathcal{L}_2(JQ^{1/2} U, H)}^2 \right) ds
$$

which is given to converge to 0. This reasoning also shows that the sequence $\left\{ \int_{a}^{T} \Phi_n dW \right\}$ is indeed a Cauchy sequence in $L^2(\Omega,H)$.

Why is $\int_{a}^{T} \Phi dW$ a continuous martingale? The integrals $\int_{a}^{T} \Phi_n dW$ are martingales and so, by the maximal estimate of Theorem 13.6.3,

$$
P \left( \sup_{t \in [a,T]} \left( \int_{a}^{t} \Phi_n dW - \int_{a}^{t} \Phi_m dW \right) \geq \lambda \right) \leq \frac{1}{\lambda^2} E \left( \left| \int_{a}^{T} (\Phi_n - \Phi_m) dW \right|^2 \right)
$$

$$
= \frac{1}{\lambda^2} \int_{a}^{T} E \left( \|\Phi_n - \Phi_m\|_{\mathcal{L}_2(Q^{1/2} U, H)}^2 \right) ds
$$
must converge in $L^2$. Therefore, there exists a subsequence $\{n_k\}$ such that
\[
P \left( \sup_{t \in [a,s]} \left| \int_a^t \Phi_{n_k} dW - \int_a^t \Phi_{n_{k+1}} dW \right|_H \geq 2^{-k} \right) \leq 2^{-k}.
\]
Consequently, by the Borel Cantelli lemma, there is a set of measure zero $N$ such that if $\omega \notin N$, then the convergence of $\int_a^t \Phi_{n_k} dW$ to $\int_a^t \Phi dW$ is uniform on $[a, T]$. Hence $t \to \int_a^t \Phi dW$ is continuous as claimed.

Why is it a martingale? Let $s < t$ and $A \in \mathcal{F}_s$. Then
\[
\int_A \left( \int_a^t \Phi dW \right) dP = \lim_{n \to \infty} \int_A \left( \int_a^t \Phi_n dW \right) dP = \lim_{n \to \infty} \int_A E \left( \left( \int_a^t \Phi_n dW \right) | \mathcal{F}_s \right) dP
\]
\[
= \lim_{n \to \infty} \int_A \left( \int_a^s \Phi_n dW \right) dP = \int_A \left( \int_a^s \Phi dW \right) dP
\]
Hence this is a martingale as claimed.

It remains to verify that the stochastic process does not depend on $J$ and $U_1$. Let the approximating sequence of elementary functions be
\[
\Phi_n(t) = \sum_{j=0}^{m_n} f^n_j \chi_{\{t^n_j, t^n_{j+1}\}}(t)
\]
where $f^n_j$ is $\mathcal{F}_{t^n_j}$ measurable and has finitely many values in $L^1(U_1, H)_0$, the restrictions of things in $L^1(U_1, H)$ to $JQ^{1/2}U$. These are the elementary functions which converge to $\Phi \circ J^{-1}$. Also let the partitions be such that
\[
\Phi^n \circ J^{-1} = \sum_{j=0}^{m_n} \Phi \left( t^n_j \right) \circ J^{-1} \chi_{\{t^n_j, t^n_{j+1}\}}
\]
converges to $\Phi \circ J^{-1}$ in $L^2([a, T] \times \Omega; L^2(JQ^{1/2}(U), H))$. Then by definition,
\[
\int_a^t \Phi_n dW = \sum_{j=0}^{m_n} f^n_j \left( W(t \wedge t^n_{j+1}) - W(t \wedge t^n_j) \right)
\]
\[
= \sum_{j=0}^{m_n} f^n_j \sum_{k=1}^{\infty} \left( \psi_k \left( t \wedge t^n_{j+1} \right) - \psi_k \left( t \wedge t^n_j \right) \right) g_k
\]
where $\{g_k\}$ is an orthonormal basis for $Q^{1/2}U$. The infinite sum converges in $L^2(\Omega; U_1)$ and $f^n_j$ is continuous on $U_1$. Therefore, $f^n_j$ can go inside the infinite sum, and this last expression equals
\[
= \sum_{j=0}^{m_n} \sum_{k=1}^{\infty} \left( \psi_k \left( t \wedge t_{k+1} \right) - \psi_k \left( t \wedge t_k \right) \right) f^n_j (t) g_k,
\]
the infinite sum converging in $L^2(\Omega, H)$.

Now consider the left sum. Since $\Phi \left( t^n_j \right) \in L^2(Q^{1/2}U, H)$, it follows that the sum
\[
\sum_{k=1}^{\infty} \left( \psi_k \left( t \wedge t^n_{j+1} \right) - \psi_k \left( t \wedge t^n_j \right) \right) \Phi \left( t^n_j \right) g_k
\]
\[
= \sum_{k=1}^{\infty} \left( \psi_k \left( t \wedge t^n_{j+1} \right) - \psi_k \left( t \wedge t^n_j \right) \right) \Phi \left( t^n_j \right) \circ J^{-1} \left( J g_k \right)
\]
must converge in $L^2(\Omega, H)$. Let us review why this is.

**Diversion** The reason the series converges goes as follows. Estimate
\[
\mathbb{E} \left( \sum_{k=p}^{q} \left( \psi_k \left( t \wedge t^n_{j+1} \right) - \psi_k \left( t \wedge t^n_j \right) \right) \Phi \left( t^n_j \right) g_k \right)^2
\]
First consider the mixed terms. Let \( \Delta \psi_k = \psi_k (t \wedge t_{j+1}^n) - \psi_k (t \wedge t_j^n) \). For \( l < k \),
\[
E \left( (\Delta \psi_k \Phi (t_j^n) g_k, \Delta \psi_l \Phi (t_j^n) g_l) \right)
\]
\[
= E \left( \Delta \psi_k \Delta \psi_l (\Phi (t_j^n) g_k, \Phi (t_j^n) g_l) \right)
\]
Now by independence, this equals
\[
E (\Delta \psi_k \Delta \psi_l) E \left( (\Phi (t_j^n) g_k, \Phi (t_j^n) g_l) \right)
\]
\[
= E (\Delta \psi_k) E (\Delta \psi_l) E \left( (\Phi (t_j^n) g_k, \Phi (t_j^n) g_l) \right) = 0
\]
Thus you only need to consider the non mixed terms, and the thing you want to estimate is of the form
\[
\sum_{k=p}^{q} E \left( (\psi_k (t \wedge t_{j+1}^n) - \psi_k (t \wedge t_j^n)) \Phi (t_j^n) g_k \right)^2
\]
Now by independence again, this equals
\[
\sum_{k=p}^{q} E \left( (\Delta \psi_k \Phi (t_j^n) g_k, \Delta \psi_k \Phi (t_j^n) g_k) \right)
\]
\[
= \sum_{k=p}^{q} E (\Delta \psi_k^2) E (\Phi (t_j^n) g_k) E (\Phi (t_j^n) g_k)
\]
\[
= \sum_{k=p}^{q} E (\Delta \psi_k^2) E (\Phi (t_j^n) g_k) E (\Phi (t_j^n) g_k)
\]
\[
= (t \wedge t_{j+1}^n) - (t \wedge t_j^n) \sum_{k=p}^{q} E \left( (\Phi (t_j^n) g_k)^2 \right)
\]
and this sum is just a part of the convergent infinite sum for
\[
\int_{\Omega} \| \Phi (t_j^n) \|_{L^2(Q^{1/2} U, H)}^2 dP < \infty
\]
Therefore, this converges to 0 as \( p, q \to \infty \) and so the sum converges in \( L^2 (\Omega, H) \) as claimed.

**End of diversion**

The \( J \) and the \( J^{-1} \) cancel in \( \{ U, \hat{U} \} \) because \( J \) is one to one. It follows that \( \{ U, \hat{U} \} \) equals
\[
\sum_{j=0}^{m_n} \sum_{k=1}^{\infty} (\psi_k (t \wedge t_{j+1}^n) - \psi_k (t \wedge t_j^n)) \Phi (t_j^n) g_k + \sum_{j=0}^{m_n} \sum_{k=1}^{\infty} (\psi_k (t \wedge t_{j+1}^n) - \psi_k (t \wedge t_j^n)) ((f_j^n - \Phi (t_j^n) \circ J^{-1}) (J g_k))
\]
The first expression does not depend on \( J \) or \( U_1 \). I need only argue that the second expression converges to 0 as \( n \to \infty \). The infinite sum converges in \( L^2 (\Omega; H) \) and also, as in the above diversion, the independence of the \( \psi_k \) implies that
\[
E \left( \sum_{j=0}^{m_n} \sum_{k=1}^{\infty} (\psi_k (t \wedge t_{j+1}^n) - \psi_k (t \wedge t_j^n)) ((f_j^n - \Phi (t_j^n) \circ J^{-1}) (J g_k))^2 \right)_{H}
\]
\[
= \sum_{j=0}^{m_n} (t \wedge t_{j+1}^n - t \wedge t_j^n) \sum_{k=1}^{\infty} E \left( (f_j^n - \Phi (t_j^n) \circ J^{-1}) (J g_k))^2 \right)_{H}
\]
\[
= \sum_{j=0}^{m_n} (t \wedge t_{j+1}^n - t \wedge t_j^n) E \left( (f_j^n - \Phi (t_j^n) \circ J^{-1}) \right)_{L^2(Q^{1/2} U, H)}^2
\]
\[
= \int_{0}^{t} E \left( (\Phi_n - \Phi^n \circ J^{-1} \right)_{L^2(Q^{1/2} U, H)}^2 ds
\]
which is given to converge to 0 since both converge to \( \Phi \circ J^{-1} \). Consequently, the stochastic integral defined above does not depend on \( J \) or \( U_1 \).

It is interesting to note that in the above definition, the approximate problems do appear to depend on \( J \) and \( U_1 \) but the limiting stochastic process does not. Since it is the case that the stochastic integral is independent of \( U_1 \) and \( J \), it can only be dependent on \( Q^{1/2}U \) and \( U \), and so we refer to \( W(t) \) as a cylindrical process on \( U \). By Lemma 17.5.3 you can take \( U_1 = U \) and so you can consider the finite sums defining the Wiener process to be in \( U \) itself. From the proof of this lemma, you can even have \( J \) being the identity on the span of the first \( n \) vectors in the orthonormal basis for \( Q^{1/2}U \). The case where \( Q \) is trace class follows in the next section. In this case, \( W \) is an actual \( Q \) Wiener process on \( U \).

The following corollary follows right away from the above theorem.

**Corollary 17.5.4** Let \( \Phi, \Psi \in L^2 ([a,T] \times \Omega; \mathcal{L}_2 (Q^{1/2}U, H)) \) and suppose they are both progressively measurable. Then

\[
E \left( \left( \int_a^t \Phi dW, \int_a^t \Psi dW \right)_H \right) = E \left( \int_a^t (\Phi, \Psi)_{\mathcal{L}_2 (Q^{1/2}U, H)} \, ds \right)
\]

Also if \( L \) is in \( L^\infty (\Omega, \mathcal{L} (H, H)) \) and is \( \mathcal{F}_a \) measurable, then

\[
L \int_a^t \Phi dW = \int_a^t L \Phi dW \tag{17.5.13}
\]

and

\[
E \left( \left( L \int_a^t \Phi dW, \int_a^t \Psi dW \right)_H \right) = E \left( \int_a^t (L \Phi, \Psi)_{\mathcal{L}_2 (Q^{1/2}U, H)} \, ds \right). \tag{17.5.14}
\]

**Proof:** First note that

\[
\left( \int_a^t \Phi dW, \int_a^t \Psi dW \right)_H = \frac{1}{4} \left[ \left| \int_a^t (\Phi + \Psi) dW \right|_H^2 - \left| \int_a^t (\Phi - \Psi) dW \right|_H^2 \right]
\]

and so from the above theorem,

\[
E \left( \left( \int_a^t \Phi dW, \int_a^t \Psi dW \right)_H \right) =
\]

\[
= E \left( \frac{1}{4} \left[ \left| \int_a^t (\Phi + \Psi) dW \right|_H^2 - \left| \int_a^t (\Phi - \Psi) dW \right|_H^2 \right] \right)
\]

\[
\frac{1}{4} E \left( \int_a^t \| \Phi + \Psi \|^2_{\mathcal{L}_2 (Q^{1/2}U, H)} \, ds \right) + \frac{1}{4} E \left( \int_a^t \| \Phi - \Psi \|^2_{\mathcal{L}_2 (Q^{1/2}U, H)} \, ds \right)
\]

\[
= E \left( \int_a^t \frac{1}{4} \left[ \| \Phi + \Psi \|^2_{\mathcal{L}_2 (Q^{1/2}U, H)} + \| \Phi - \Psi \|^2_{\mathcal{L}_2 (Q^{1/2}U, H)} \right] ds \right)
\]

\[
= E \left( \int_a^t (\Phi, \Psi)_{\mathcal{L}_2 (Q^{1/2}U, H)} \, ds \right).
\]

Now consider the last claim. First suppose \( L = lX_A \) where \( A \in \mathcal{F}_a \), and \( l \in \mathcal{L} (H, H) \). Also suppose \( \Phi \) is an elementary function

\[
\Phi = \sum_{i=0}^n \psi_i X_{(s_i, s_{i+1})}
\]

Then

\[
L \int_a^t \Phi dW = lX_A \sum_{i=0}^n \psi_i (W (t \wedge s_{i+1}) - W (t \wedge s_i))
\]

\[
= \sum_{i=0}^n lX_A \psi_i (W (t \wedge s_{i+1}) - W (t \wedge s_i))
\]
Thus (17.13) also holds for $L$ a simple function which is $\mathcal{F}_t$ measurable. For general $L \in L^\infty (\Omega, \mathcal{L}(H, H))$, approximating with a sequence of such simple functions $L_n$ yields

$$L \int_a^t \Phi dW = \lim_{n \to \infty} L_n \int_a^t \Phi dW = \lim_{n \to \infty} \int_a^t L_n \Phi dW = \int_a^t L \Phi dW$$

because $L_n \Phi \to L \Phi$ in $L^2 ([a, T] \times \Omega; \mathcal{L}_2 (Q^{1/2}U, H))$. Now what about general $\Phi$? Let $\{\Phi_n\}$ be elementary functions converging to $\Phi \circ J^{-1}$ in $L^2 ([a, T] \times \Omega; \mathcal{L}_2 (J Q^{1/2}U, H))$. Then by definition of the integral,

$$L \int_a^t \Phi dW = \lim_{n \to \infty} L \int_a^t \Phi_n dW = \lim_{n \to \infty} \int_a^t L \Phi_n dW = \int_a^t L \Phi dW$$

The remaining claim now follows from the first part of the proof. ■

The above has discussed the integral of $\Phi \in L^2 ([a, T] \times \Omega; \mathcal{L}_2 (Q^{1/2}U, H))$. An obvious case to consider is when

$$\Phi = \sum_{k=0}^{n-1} \Phi_k \lambda'_{\{t_k, t_{k+1}\}} (t)$$

and $\Phi_k \in L^2 (\Omega; \mathcal{L}_2 (Q^{1/2}U, H))$ with $\Phi_k$ measurable with respect to $\mathcal{F}_{t_k}$. What is $\int_a^t \Phi dW$? First note that $\Phi_k \circ J^{-1} \in L^2 (\Omega; \mathcal{L}_2 (J Q^{1/2}U, H))$. Let $\lim_{n \to \infty} \Phi_n \to \Phi_k \circ J^{-1}$ in $L^2 (\Omega; \mathcal{L}_2 (J Q^{1/2}U, H))$ where $\Phi_n$ is $\mathcal{F}_{t_k}$ measurable and is a simple function having values in $\mathcal{L}(U, H)$. Thus

$$\Phi_m \equiv \sum_{k=0}^{n-1} \Phi_k \lambda'_{\{t_k, t_{k+1}\}} (t)$$

is an elementary function and it converges to $\Phi \circ J^{-1}$ in $L^2 ([a, T] \times \Omega; \mathcal{L}_2 (J Q^{1/2}U, H))$. It follows that

$$\int_a^t \Phi dW = \lim_{m \to \infty} \int_a^t \Phi_m dW = \lim_{m \to \infty} \sum_{k=0}^{n-1} \Phi_k (W(t \wedge t_{k+1}) - W(t \wedge t_k))$$

$$= \sum_{k=0}^{n-1} \Phi_k \circ J^{-1} (W(t \wedge t_{k+1}) - W(t \wedge t_k)).$$

Note again how it appears to depend on $J$ but really doesn’t because there is a $J$ in the definition of $W$.

### 17.6 The Case That $Q$ Is Trace Class

In this special case, you have a $Q$ Wiener process with values in $U$ and still you have

$$\Phi \in L^2 ([a, T] \times \Omega; \mathcal{L}_2 (Q^{1/2}U, H))$$

with $\Phi$ progressively measurable. The difference here is that in fact, $Q$ is trace class.

$$Q = \sum_{i=1}^L \lambda_i e_i \otimes e_i$$

where $\lambda_i > 0$, $\sum_i \lambda_i < \infty$, and the $e_i$ form an orthonormal set of vectors. $L$ is either a positive integer or $\infty$. Then let $U_0 = Q^{1/2}U$. Then $Q^{1/2} = \sum_{i=1}^L \lambda_i e_i \otimes e_i$ because this works, and the square root is unique. Hence $Q^{1/2}e_i = \sqrt{\lambda_i}e_i$ and so an orthonormal basis for $U_0 = Q^{1/2}U$ is $\{\lambda_i e_i\}_{i=1}^L$. Now consider $J = \sum_{i=1}^L \lambda_i (e_i \otimes \sqrt{\lambda_i} e_i)$, $J : U_0 \to U$, where the tensor product is defined in the usual way;

$$u \otimes v (w) \equiv u (w, v)_{U_0}.$$

Then $J^* = \sum_{i=1}^L \sqrt{\lambda_i} (\sqrt{\lambda_i} e_i \otimes e_i)$ and $JJ^* = \sum_{i=1}^L \lambda_i e_i \otimes e_i = Q$. Also, $J$ is a Hilbert Schmidt map into $U$ from $U_0$.

$$\sum_{i=1}^L \|J (\sqrt{\lambda_i} e_i)\|^2_U = \sum_{i=1}^L \|\sqrt{\lambda_i} e_i\|^2_U = \sum_{i=1}^L \lambda_i < \infty$$
17.7 A Short Comment On Measurability

It will also be important to consider the composition of functions. The following is the main result. With the explanation of progressively measurable given, it says the composition of progressively measurable functions is progressively measurable.

**Proposition 17.7.1** Let $A : [a, T] \times V \times \Omega \to U$ where $V, U$ are topological spaces and suppose $A$ satisfies its restriction to $[a, t] \times V \times \Omega$ is $B \langle [a, t] \rangle \times B (V) \times \mathcal{F}_t$ measurable. This will be referred to as $A$ is progressively measurable. Then if $X : [a, T] \times \Omega \to V$ is progressively measurable, then so is the map

$$(t, \omega) \to A (t, X (t, \omega), \omega)$$

**Proof:** Consider the restriction of this map to $[a, t_0] \times \Omega$. For such $(t, \omega)$, to say

$$A (t, X (t, \omega), \omega) \in O$$

for $O$ a Borel set in $U$ is to say that

$$X (t, \omega) \in \{ v : (t, v, \omega) \in A^{-1} (O), t \leq t_0 \} \equiv A^{-1} (O)_{t_0}$$

Consider the set

$$\{(t, \omega) \in [a, t_0] \times \Omega : X (t, \omega) \in A^{-1} (O)_{t_0}\}$$

Is this in $B ([a, t_0]) \times \mathcal{F}_{t_0}$? This is what needs to be checked. Since $A$ is progressively measurable,

$$A^{-1} (O) \cap [a, t_0] \times V \times \Omega \in B ([a, t_0]) \times B (V) \times \mathcal{F}_{t_0} \equiv \mathcal{P}_{t_0}$$

because $A^{-1} (O)$ is a progressively measurable set. So let

$$\mathcal{G} \equiv \{ S \in \mathcal{P}_{t_0} : \{(t, \omega) \in [a, t_0] \times \Omega : X (t, \omega) \in S_{t_0}\} \in B ([a, t_0]) \times \mathcal{F}_{t_0}\}$$

It is clear that $\mathcal{G}$ contains the $\pi$ system composed of sets of the form $I \times B \times W$ where $I$ is an interval in $[a, t_0]$, $B$ is Borel, and $W \in \mathcal{F}_{t_0}$. This is because for $S$ of this form, $S_{t_0} = B$ or $\emptyset$. Thus if not empty,

$$\{ (t, \omega) \in [a, t_0] \times \Omega : X (t, \omega) \in S_{t_0}\} = X^{-1} (B) \cap [0, t_0] \times \Omega \in B ([a, t_0]) \times \mathcal{F}_{t_0}$$

because $X$ is given to be progressively measurable. Now if $S \in \mathcal{G}$, what about $S^C$? You have $(S^C)_{t_0} = (S_{t_0})^C$ thus

$$\{ (t, \omega) \in [a, t_0] \times \Omega : X (t, \omega) \in (S^C)_{t_0}\} = \{ (t, \omega) \in [a, t_0] \times \Omega : X (t, \omega) \in (S^C)_{t_0}\}$$

which is the complement with respect to $[a, t_0] \times \Omega$ of a set in $B ([a, t_0]) \times \mathcal{F}_{t_0}$. Therefore, $\mathcal{G}$ is closed with respect to complements. It is clearly closed with respect to countable disjoint unions. It follows, $\mathcal{G} = \mathcal{P}_{t_0}$. Thus

$$\{ (t, \omega) \in [a, t_0] \times \Omega : X (t, \omega) \in S_{t_0}\} \in B ([a, t_0]) \times \mathcal{F}_{t_0}$$

where $S = A^{-1} (O) \cap [a, t_0] \times V \times \Omega$. In other words,

$$\{ (t, \omega), t \leq t_0 : A (t, X (t, \omega), \omega) \in O\} \in B ([0, t_0]) \times \mathcal{F}_{t_0}$$

and so $(t, \omega) \to A (t, X (t, \omega), \omega)$ is progressively measurable. ■
17.8 Localization For Elementary Functions

It is desirable to extend everything to stochastically square integrable functions. This will involve localization using a suitable stopping time. First it is necessary to understand localization for elementary functions. As above, we are in the situation described by the following diagram.

\[
\begin{array}{c}
U_1 \supseteq JQ^{1/2}U \\
\Phi_n \downarrow \Phi \\
\end{array}
\]

The elementary functions \( \{\Phi_n\} \) have values in \( \mathcal{L}(U_1, H) \), meaning they are restrictions of functions in \( \mathcal{L}(U_1, H) \) to \( JQ^{1/2}U \) and converge to \( \Phi \circ J^{-1} \) in \( L^2 \left( \left[ a, T \right] \times \Omega; \mathcal{L}_2 \left( JQ^{1/2}U, H \right) \right) \) where \( \Phi \in L^2 \left( \left[ a, T \right] \times \Omega; \mathcal{L}_2 \left( Q^{1/2}U, H \right) \right) \) is given. Let \( \Phi \) be an elementary function. In particular, let \( \Phi(t_k) \) be \( \mathcal{F}_{t_k} \)-measurable as a map into \( \mathcal{L}(U_1, H) \), and has finitely many values. As just mentioned, the topic of interest is the elementary functions \( \Phi_n \) in the above diagram. Thus \( \Phi \) will be one of these elementary functions.

Let \( \tau \) be a stopping time having values from the set of mesh points \( \{t_k\} \) for the elementary function. Then from the definition of the integral for elementary functions,

\[
\int_a^{t \wedge \tau} \Phi dW = \sum_{k=0}^{n-1} \Phi(t_k) \left( W(t \wedge t_{k+1}) - W(t \wedge t_k) \right)
\]

If \( \omega \) is such that \( \tau(\omega) = t_j \), then to get something nonzero, you must have \( t_j > t_k \) so \( k \leq j - 1 \). Thus the above on the right reduces to

\[
\sum_{k=0}^{j-1} \Phi(t_k) \left( W(t \wedge t_{k+1}) - W(t \wedge t_k) \right)
\]

It clearly is 0 if \( j = 0 \). Define \( \sum_{k=0}^{-1} \equiv 0 \). Thus the integral equals

\[
\sum_{j=0}^{n} \chi_{[\tau = t_j]} \sum_{k=0}^{j-1} \Phi(t_k) \left( W(t \wedge t_{k+1}) - W(t \wedge t_k) \right)
\]

Interchanging the order of summation, \( k \leq j - 1 \) so \( j \geq k + 1 \) and this equals

\[
\sum_{k=0}^{n-1} \sum_{j=k+1}^{n} \chi_{[\tau = t_j]} \Phi(t_k) \left( W(t \wedge t_{k+1}) - W(t \wedge t_k) \right)
\]

\[
= \sum_{k=0}^{n-1} \chi_{[\tau > t_k]} \Phi(t_k) \left( W(t \wedge t_{k+1}) - W(t \wedge t_k) \right)
\]

Therefore

\[
\int_a^{t \wedge \tau} \Phi dW = \int_a^{a} \sum_{k=0}^{n-1} \chi_{[\tau > t_k]} \Phi(t_k) \chi(t_k, t_{k+1}) dW \tag{17.8.15}
\]
Now observe

\[
X_{[a,\tau]}(t) \Phi(t) = \sum_{k=0}^{n-1} X_{[a,\tau]}(t) \Phi(t_k) X_{(t_k,t_{k+1}]}(t)
\]

\[
= \sum_{k=0}^{n-1} \chi_{[\tau > t_k]} \Phi(t_k) X_{(t_k,t_{k+1}]}(t)
\]

\[
= \sum_{k=0}^{n-1} \chi_{[\tau > t_k]} \Phi(t_k) X_{(t_k,t_{k+1}]}(t)
\quad (17.8.16)
\]

The last step occurs because of the following reasoning. The \( k^{th} \) term of the sum in the middle expression above equals \( \Phi(t_k) \) if and only if \( t > t_k \) and \( \tau \geq t \). If the two conditions do not hold, then the \( k^{th} \) term equals 0. As to the third line, if \( \tau > t_k \) and \( t \in (t_k,t_{k+1}] \), then \( \tau \geq t_{k+1} \geq t \) which is the same as the situation in the second line. The term equals \( \Phi(t_k) \). Note that \( \chi_{[\tau > t_k]}(\omega) \) is \( F_{t_k} \) measurable, because \( [\tau > t_k] \) is the complement of \( [\tau \leq t_k] \). Therefore, this is an elementary function. Thus, from \( \{17.8.15, 17.8.16\} \), \( X_{[a,\tau]}(t) \Phi(t) \) is an elementary function and

\[
\int_a^{t \land \tau} \Phi dW = \int_a^t \sum_{k=0}^{n-1} \chi_{[\tau > t_k]} \Phi(t_k) X_{(t_k,t_{k+1}]}(t) \ dW = \int_a^t X_{[a,\tau]}(t) \Phi(t) \ dW
\]

From Proposition \( \{17.8.16, 17.8.17\} \), if you have \( \Phi, \Psi \) two of these elementary functions

\[
E \left( \left\| \int_a^t X_{[a,\tau]}(t) \Phi(t) \ dW - \int_a^t X_{[a,\tau]}(t) \Psi(t) \ dW \right\|_H^2 \right) = 0
\]

\[
\int_a^t \int_\Omega |X_{[a,\tau]}(t)||(|\Phi(s) - \Psi(s)) \circ J||^2_{L^2(Q^{1/2}U,H)} \ dPds 
\]

\[
\leq \int_a^t \int_\Omega |(|\Phi(s) - \Psi(s)) \circ J||^2_{L^2(Q^{1/2}U,H)} \ dPds
\quad (17.8.17)
\]

### 17.9 Localization In General

Next, what about the general case where \( \Phi \in L^2([a,T] \times \Omega; L_2(Q^{1/2}U,H)) \) and is progressively measurable? Is it the case that for an arbitrary stopping time \( \tau \),

\[
\int_a^{t \land \tau} \Phi dW = \int_a^\tau X_{[a,\tau]} \Phi dW?
\]

This is the sort of thing which would be expected for an ordinary Stieltjes integral which of course this isn’t. Let

\[
L^2([a,T] \times \Omega; L_2(JQ^{1/2}U,H)) = K
\]

From Doob’s result Proposition \( \{17.8.20\} \) and Lemma \( \{17.8.21\} \), there exists a sequence of elementary functions \( \{\Phi_k\} \)

\[
\Phi_k(t) = \sum_{j=0}^{m_k-1} \Phi(t_k^j) X_{(t_k^j,t_k^{j+1}]}(t)
\]

which converges to \( \Phi \circ J^{-1} \) in \( K \) where also the lengths of the sub intervals converge uniformly to 0 as \( k \to \infty \).

Now let \( \tau \) be an arbitrary stopping time. The partition points corresponding to \( \Phi_k \) are \( \{t_k^j\}_{j=0}^{m_k} \). Let \( \tau_k = t_{j+1}^k \) on \( \tau^{-1}(t_j^k,t_{j+1}^k] \). Then \( \tau_k \) is a stopping time because

\[
[\tau_k \leq t] \in F_t
\]

Here is why. If \( t \in (t_j^k,t_{j+1}^k] \), then if \( t = t_{j+1}^k \), it would follow that \( \tau_k(\omega) \leq t \) would be the same as saying \( \omega \in [\tau \leq t_{j+1}^k] = [\tau \leq t] \in F_t \). On the other hand, if \( t < t_{j+1}^k \), then \( [\tau_k \leq t] = [\tau \leq t_j^k] \in F_{t_j^k} \subseteq F_t \) because \( \tau_k \) can only take the values \( t_j^k \). 
Consider $\mathcal{X}_{[a,\tau_k]}\Phi_k$. It is given that $\Phi_k \to \Phi \circ J^{-1}$ in $K$. Does it follow that $\mathcal{X}_{[a,\tau_k]}\Phi_k \to \mathcal{X}_{[a,\tau]}\Phi \circ J^{-1}$ in $K$? Consider first the indicator function. Let $\tau(\omega) \in (t_j, t_{j+1}^k]$. Fixing $t$, if $\mathcal{X}_{[a,\tau]}(t) = 1$, then also $\mathcal{X}_{[a,\tau_k]}(t) = 1$ because $\tau_k \geq \tau$. Therefore, in this case $\lim_{k \to \infty} \mathcal{X}_{[a,\tau_k]}(t) = \mathcal{X}_{[a,\tau]}(t)$. Next suppose $\mathcal{X}_{[a,\tau]}(t) = 0$ so that $\tau(\omega) < t$. Since the intervals defined by the partition points have lengths which converge to zero, it follows that for all $k$ large enough, $\tau_k(\omega) < t$ also and so $\mathcal{X}_{[a,\tau_k]}(t) = 0$. Therefore,

$$
\lim_{k \to \infty} \mathcal{X}_{[a,\tau_k]}(\omega)(t) = \mathcal{X}_{[a,\tau]}(\omega)(t).
$$

It follows that $\mathcal{X}_{[a,\tau_k]}\Phi_k \to \mathcal{X}_{[a,\tau]}\Phi \circ J^{-1}$ in $K$. Now from 17.8.10, the function $\mathcal{X}_{[a,\tau]}\Phi_k$ is progressively measurable. Therefore, the same is true of $\mathcal{X}_{[a,\tau]}\Phi \circ J^{-1}$.

From the proof of Theorem 17.5.3, the part depending on maximal estimates and the fact that $\int_a^t \mathcal{X}_{[a,\tau]}\Phi_k \, dW$ is a continuous martingale, there is a set of measure zero $N$, such that off this set, a suitable subsequence satisfies

$$
\int_a^t \mathcal{X}_{[a,\tau]}\Phi_k \, dW \to \int_a^t \mathcal{X}_{[a,\tau]} \Phi \, dW
$$

uniformly on $[a,T]$. But also, since $\Phi_k \to \Phi \circ J^{-1}$ in $K$, a suitable subsequence satisfies,

$$
\int_a^t \Phi_k \, dW \to \int_a^t \Phi \, dW
$$

uniformly on $[a,T]$ a.e. $\omega$. In particular, $\int_a^{t \land \tau_k} \Phi_k \, dW \to \int_a^{t \land \tau} \Phi \, dW$. Therefore,

$$
\int_a^t \mathcal{X}_{[a,\tau]} \Phi \, dW = \lim_{k \to \infty} \int_a^t \mathcal{X}_{[a,\tau_k]} \Phi_k \, dW

= \lim_{k \to \infty} \int_a^{t \land \tau_k} \Phi_k \, dW

= \int_a^{t \land \tau} \Phi \, dW
$$

This has proved the following major localization lemma. This is a marvelous result. It says that the stochastic integral acts algebraically like an ordinary Stieltjes integral, one for each $\omega$ off a set of measure zero.

**Lemma 17.9.1** Let $\Phi$ be progressively measurable and in

$$
L^2 \left( [a,T] \times \Omega; \mathcal{L}_2 \left( Q^{1/2}U, H \right) \right)
$$

Let $W(t)$ be a cylindrical Wiener process as described above. Then for $\tau$ a stopping time, $\mathcal{X}_{[a,\tau]} \Phi$ is progressively measurable, in $K$, and

$$
\int_a^{t \land \tau} \Phi \, dW = \int_a^t \mathcal{X}_{[a,\tau]} \Phi \, dW.
$$

### 17.10 The Stochastic Integral As A Local Martingale

With Lemma 17.9.1 it becomes possible to define the stochastic integral on functions which are only stochastically square integrable.

**Definition 17.10.1** $\Phi$ is stochastically square integrable in $\mathcal{L}_2 \left( Q^{1/2}U, H \right)$ if $\Phi$ is progressively measurable and

$$
P \left( \left[ \int_a^T \| \Phi(s) \|^2_{\mathcal{L}_2 \left( Q^{1/2}U, H \right)} \, ds < \infty \right] \right) = 1
$$

Thus equivalently, there exists $N$ such that $P(N) = 0$ and for $\omega \notin N$,

$$
\int_a^T \| \Phi(s, \omega) \|^2_{\mathcal{L}_2 \left( Q^{1/2}U, H \right)} \, ds < \infty.
$$
Lemma 17.10.2 Suppose $\Phi$ is $\mathcal{L}_2\left(Q^{1/2}U,H\right)$ progressively measurable and
\[
P\left(\left[\int_a^T \|\Phi\|^2_{\mathcal{L}_2(Q^{1/2}U,H)} \, ds < \infty\right]\right) = 1.
\]
Define
\[
\tau_n(\omega) \equiv \inf\left\{ t \in [a,T] : \int_a^t \|\Phi\|^2_{\mathcal{L}_2(Q^{1/2}U,H)} \, ds \geq n \right\}.
\]
By convention, let $\inf\emptyset = \infty$. Then $\tau_n$ is a stopping time. Furthermore, $\tau_n$ has the following properties.

1. $\{\tau_n\}$ is an increasing sequence and for $\omega$ outside a set of measure zero $N$, for every $t \in [a,T]$ there exists $n$ such that $\tau_n(\omega) > t$. (It is a localizing sequence of stopping times.)

2. For each $n$, $X_{[a,\tau_n]} \Phi$ is progressively measurable and
\[
E\left(\int_a^T \|X_{[a,\tau_n]} \Phi\|^2_{\mathcal{L}_2(Q^{1/2}U,H)} \, dt\right) < \infty
\]

Proof: It follows from Proposition 13.8.2 that $\tau_n$ is a stopping time because it is the first hitting time of a closed set by an adapted continuous process.

It remains to verify the two claims. There exists a set of measure 0, $N$ such that for $\omega \notin N$
\[
\int_a^T \|\Phi\|^2_{\mathcal{L}_2(Q^{1/2}U,H)} \, dt < \infty
\]
Therefore, for such $\omega$, there exists $n$ large enough that
\[
\int_a^t \|\Phi\|^2_{\mathcal{L}_2(Q^{1/2}U,H)} \, ds < n
\]
and so $\tau_n(\omega) \geq t$. Now consider the second claim.
\[
E\left(\int_a^T \|X_{[a,\tau_n]} \Phi\|^2_{\mathcal{L}_2(Q^{1/2}U,H)} \, dt\right) = E\left(\int_{\tau_n(\omega)}^{\tau_n(\omega)\wedge T} \|\Phi\|^2_{\mathcal{L}_2(Q^{1/2}U,H)} \, dt\right) \leq E(n) = n. \blacksquare
\]

With this lemma, it is possible to give the following definition.

Definition 17.10.3 Suppose $\Phi$ is $\mathcal{L}_2\left(Q^{1/2}U,H\right)$ progressively measurable and
\[
P\left(\left[\int_a^T \|\Phi\|^2_{\mathcal{L}_2(Q^{1/2}U,H)} \, ds < \infty\right]\right) = 1. \tag{17.10.18}
\]
More generally, suppose there exists a localizing sequence of stopping times $\tau_n$ having the two properties of Lemma 17.10.2. Then for all $\omega$ not in the exceptional set $N$.
\[
\int_a^t \Phi dW \equiv \lim_{n \to \infty} \int_a^t X_{[a,\tau_n]} \Phi dW
\]

Lemma 17.10.4 The above definition is well defined. For all $\omega$ not in a set of measure zero,
\[
\int_a^t \Phi dW(\omega) \equiv \lim_{n \to \infty} \int_a^t X_{[a,\tau_n]} \Phi dW(\omega)
\]
the function on the right being constant for all $n$ large enough for a given $\omega$. The random variable $\int_a^t \Phi dW$ is also $\mathcal{F}_t$ adapted.
The next lemma says that even when \( \int_a^t \Phi (s) \, dW (s) \) is only a local martingale relative to a suitable localizing sequence, it is still the case that
\[
\int_a^t \Phi (s) \, dW (s) = \int_a^t \chi_{[a, \tau_n]} \Phi (s) \, dW (s)
\]
for all \( \omega \), in particular for the given \( \omega \). Therefore, for the particular \( \omega \) of interest,
\[
\int_a^t \chi_{[a, \tau_n]} \Phi (s) \, dW (s) = \int_a^t \chi_{[a, \tau_m]} \Phi (s) \, dW (s)
\]
Thus the limit exists because for all \( n \) large enough, the integral is eventually constant. Then \( \int_a^t \Phi (s) \, dW (s) \) is \( \mathcal{F}_t \) adapted because for \( U \) an open set in \( H \),
\[
\left( \int_a^t \Phi (s) \, dW (s) \right)^{-1} (U) = \bigcup_{n=1}^{\infty} \left( \left( \int_a^t \chi_{[a, \tau_n]} \Phi (s) \, dW (s) \right)^{-1} (U) \cap [\tau_n > t] \right) \in \mathcal{F}_t.
\]

**Lemma 17.10.5** Let \( \Phi \) be progressively measurable and suppose there exists the localizing sequence described above. Then if \( \sigma \) is a stopping time,
\[
\int_a^t \chi_{[a, \sigma]} \Phi (s) \, dW (s) = \int_a^t \chi_{[a, \sigma]} \Phi (s) \, dW (s)
\]

**Proof:** Let \( \{ \tau_n \} \) be the localizing sequence described above for which, when the local martingale is stopped, it results in a martingale, (satisfying \( \mathbf{1} \) and \( \mathbf{4} \) on Page 393). Then by definition,
\[
\int_a^{t \land \sigma} \Phi (s) \, dW (s) = \lim_{n \to \infty} \int_a^{t \land \tau_n \land \sigma} \Phi (s) \, dW (s)
\]
for all \( \omega \), in particular for the given \( \omega \). Therefore, for the particular \( \omega \) of interest,
\[
\int_a^{t \land \sigma} \chi_{[a, \sigma]} \Phi (s) \, dW (s) = \int_a^t \chi_{[a, \sigma]} \Phi (s) \, dW (s)
\]
Since \( t \land \tau_n = t \) for all \( n \) large enough.

### 17.11 The Quadratic Variation Of The Stochastic Integral

An important corollary of Lemma 17.10.4 concerns the quadratic variation of \( \int_a^t \Phi (s) \, dW (s) \). It is convenient here to use the notation \( \int_a^t \Phi (s) \, dW (s) = \Phi \cdot W (t) \). Recall this is a local submartingale \( [\Phi \cdot W] \) such that
\[
\| \Phi \cdot W (t) \|_H^2 = [\Phi \cdot W] (t) + N (t)
\]
where \( N \) is a local martingale. Recall the quadratic variation is unique so that if it acts like the quadratic variation, then it is the quadratic variation. Recall also why this was so. If you have a local martingale equal to the difference of increasing adapted processes which equals 0 when \( t = 0 \), then the local martingale was equal to 0. Of course you can substitute \( a \) for 0.
Corollary 17.11.1 Suppose $\Phi$ is $\mathcal{L}_2 \left(Q^{1/2}U,H\right)$ progressively measurable and has the localizing sequence with the two properties in Lemma 17.10.3. Then the quadratic variation, $[\Phi \cdot W]$ is given by the formula

$$[\Phi \cdot W] (t) = \int_a^t \|\Phi (s)\|^2_{\mathcal{L}_2 (Q^{1/2}U,H)} \, ds$$

**Proof:** By the above discussion, $\int_a^t \Phi dW$ is a local martingale. Let $\{\tau_n\}$ be a localizing sequence for which the stopped local martingale is a martingale and $\Phi \mathcal{X}_{[a,\tau_n]}$ is in $L^2 \left([0,T] \times \Omega, \mathcal{L}_2 \left(Q^{1/2}U,H\right)\right)$. Also let $\sigma$ be a stopping time with two values no larger than $T$. Then from Lemma 17.10.3,

$$E \left( \left| \int_a^{T \wedge \tau_n \wedge \sigma} \Phi dW \right|^2 - \left| \int_a^{\tau_n \wedge \sigma} \Phi (s)\right|^2_{\mathcal{L}_2 (Q^{1/2}U,H)} \right) = 0$$

thanks to the Ito isometry. There is also no change in letting $\sigma = t$. You still get 0. It follows from Lemma 17.10.3, the lemma about recognizing a martingale when you see one, that

$$t \rightarrow \left| \int_a^{t \wedge \tau_n} \Phi dW \right|^2 - \left| \int_a^{t \wedge \tau_n} \Phi (s)\right|^2_{\mathcal{L}_2 (Q^{1/2}U,H)} \, ds$$

is a martingale. Therefore,

$$\left| \int_a^t \Phi dW \right|^2 - \int_a^t \|\Phi (s)\|^2_{\mathcal{L}_2 (Q^{1/2}U,H)} \, ds$$

is a local martingale and so, by uniqueness of the quadratic variation,

$$[\Phi \cdot W] (t) = \int_a^t \|\Phi (s)\|^2_{\mathcal{L}_2 (Q^{1/2}U,H)} \, ds$$

Here is an interesting little lemma which seems to be true.

**Lemma 17.11.2** Let $\Phi, \Phi_n$ all be in $L^2 \left([0,T] \times \Omega, \mathcal{L}_2 (Q^{1/2}U,H)\right)$ off some set of measure zero. These are all progressively measurable. Thus there are all stochastically square integrable.

$$P \left( \int_0^T \|\Phi\|^2 \, ds \right) = 1$$

Suppose also that for each $\omega \notin N$, the exceptional set,

$$\int_0^T \|\Phi_n - \Phi\|^2_{\mathcal{L}_2} \, dt \to 0$$

Then there exists a set of measure zero, still denoted as $N$ and a subsequence, still denoted as $n$ such that for each $\omega \notin N$,

$$\lim_{n \to \infty} \int_0^T \Phi_n dW = \int_0^T \Phi dW$$
**Proof:** Define stopping times

\[ \tau_{np} \equiv \inf \left\{ t \in [0, T] : \int_0^t \| \Phi_n \|^2 ds > p \right\} \]

Let \( \tau_p \) be similar but defined with reference to \( \Phi \). Then by Itô isometry,

\[
\begin{align*}
E \left( \left| \int_0^T X_{[0, \tau_{np}]} \Phi_n dW - \int_0^T X_{[0, \tau_p]} \Phi dW \right|^2 \right) \\
= E \left( \int_0^T \| X_{[0, \tau_{np}]} \Phi_n - X_{[0, \tau_p]} \Phi \|_{L^2}^2 \right) \\
\end{align*}
\]

The integrand in the right side is bounded by \( 2p^2 \). Also this integrand converges to 0 for each \( \omega \) as \( n \to \infty \). This is shown next.

\[
\begin{align*}
\int_0^T \| X_{[0, \tau_{np}]} \Phi_n - X_{[0, \tau_p]} \Phi \|_{L^2}^2 \ dt \\
\leq 2 \int_0^T \left( \| X_{[0, \tau_{np}]} \Phi_n - X_{[0, \tau_{np}]} \Phi \|_{L^2}^2 + \| X_{[0, \tau_{np}]} \Phi - X_{[0, \tau_p]} \Phi \|_{L^2}^2 \right) \ dt \\
\leq 2 \int_0^T \| \Phi_n - \Phi \|_{L^2}^2 \ dt + 2 \int_0^T \| X_{[0, \tau_{np}]} (t) - X_{[0, \tau_{np}]} (t) \|_{L^2}^2 \ dt \\
\end{align*}
\]

The first converges to 0 by assumption. Problem is, it does not look like this second integral converges to 0. We do know that \( \int_0^t \| \Phi_n \|^2 ds \to \int_0^t \| \Phi \|^2 ds \) uniformly so \( \tau_{np} \to \tau_p \) is likely. However, this does not imply \( X_{[0, \tau_{np}]} \to X_{[0, \tau_p]} \). However, it would converge in \( L^2 (0, T) \) and so there is a subsequence such that convergence takes place a.e. \( t \). Then restricting to this subsequence, the second integral converges to 0. Actually, it may be easier than this. \( X_{[0, \tau_p]} \) has a single point of discontinuity and convergence takes place at every other point. Thus it appears that the integrand in (17.11.19) converges to 0 for each \( \omega \). Thus, by dominated convergence theorem the whole expectation converges to 0.

Now consider

\[
P \left( \int_0^T X_{[0, \tau_{np}]} \Phi_n dW - \int_0^T X_{[0, \tau_p]} \Phi dW \|^2 > \lambda \right) \\
\leq E \left( \int_0^T X_{[0, \tau_{np}]} \Phi_n dW - \int_0^T X_{[0, \tau_p]} \Phi dW \|^2 \right) \frac{1}{\lambda}
\]

and so, there exists a subsequence, still denoted as \( n \) such that

\[
P \left( \int_0^T X_{[0, \tau_{np}]} \Phi_n dW - \int_0^T X_{[0, \tau_p]} \Phi dW \|^2 > \frac{1}{n} \right) < 2^{-k}
\]

It follows that \( N \) can be enlarged so that for \( \omega \notin N_p \)

\[
\int_0^T X_{[0, \tau_{np}]} \Phi_n dW - \int_0^T X_{[0, \tau_p]} \Phi dW \|^2 \leq \frac{1}{n}
\]

for all \( n \) large enough. Now obtain a succession of subsequences for \( p = 1, 2, \cdots \), each a subsequence of the preceding one such that the above convergence takes place and let \( N \) include \( \cup_p N_p \). Then for \( \omega \notin N \), and letting \( n \) denote the diagonal sequence, it follows that for all \( p \),

\[
\lim_{n \to \infty} \int_0^T X_{[0, \tau_{np}]} \Phi_n dW - \int_0^T X_{[0, \tau_p]} \Phi dW = 0
\]

For \( \omega \notin N \), there is a \( p \) such that \( \tau_p = \infty \). Then this means \( \int_0^T \| \Phi \|^2 ds < p \). It follows that the same is true for \( \Phi_n \) for all \( n \) large enough. Hence \( \tau_{np} = \infty \) also. Thus, for large enough \( n \),

\[
\int_0^T \Phi_n dW - \int_0^T \Phi dW = \left| \int_0^T X_{[0, \tau_{np}]} \Phi_n dW - \int_0^T X_{[0, \tau_p]} \Phi dW \right|
\]

and the latter was just shown to converge to 0. \( \blacksquare \)
17.12 The Holder Continuity Of The Integral

Let $\Phi \in L^2 ([0, T] \times \Omega, \mathcal{L}_2 (Q^{1/2}U, H))$. Then you can consider the stochastic integral as described above and it yields a continuous function off a set of measure zero. What if $\Phi \in L^\infty ([0, T] \times \Omega, \mathcal{L}_2 (Q^{1/2}U, H))$? Can you say more? The short answer is yes. You obtain a Holder condition in addition to continuity. This is a consequence of the Burkholder Davis Gundy inequality and Corollary 17.11.1 above. Let $\alpha > 2$. Let $\|\Phi\|_\infty$ denote the norm in $L^\infty ([0, T] \times \Omega, \mathcal{L}_2 (Q^{1/2}U, H))$. By the Burkholder Davis Gundy inequality,

$$\int_\Omega \left( \int_s^t \Phi dW \right)^\alpha dP \leq \int_\Omega \left( \sup_{r \in [s,t]} \int_s^r \Phi dW \right)^\alpha dP \leq C \|\Phi\|_\infty^\alpha \int_\Omega \left( \int_s^t \|\Phi\|^2 d\tau \right)^{\alpha/2} dP \leq C \|\Phi\|_\infty^\alpha \int_\Omega \left( \int_s^t \|\Phi\|^2 d\tau \right)^{\alpha/2} dP \leq C \|\Phi\|_\infty^\alpha |t-s|^{\alpha/2}$$

By the Kolmogorov Čentsov theorem, Theorem 13.3.2, this shows that $t \rightarrow \int_0^t \Phi dW$ is Holder continuous with exponent

$$\gamma < \frac{(\alpha/2) - 1}{\alpha} = \frac{1}{2} - \frac{1}{\alpha}$$

Since $\alpha > 2$ is arbitrary, this shows that for any $\gamma < 1/2$, the stochastic integral is Holder continuous with exponent $\gamma$. This is exactly the same kind of continuity possessed by the Wiener process.

**Theorem 17.12.1** Suppose $\Phi \in L^\infty ([0, T] \times \Omega, \mathcal{L}_2 (Q^{1/2}U, H))$ and is progressively measurable. Then if $\gamma < 1/2$, there exists a set of measure zero such that off this set,

$$t \rightarrow \int_0^t \Phi dW$$

is Holder continuous with exponent $\gamma$.

17.13 Taking Out A Linear Transformation

When is $L \int_a^T \Phi dW = \int_a^T L\Phi dW$?

It is assumed $L \in \mathcal{L}(H, H_1)$ where $H_1$ is another separable real Hilbert space. First of all, here is a lemma which shows $\int_a^t L\Phi dW$ at least makes sense.

**Proposition 17.13.1** Suppose $\Phi$ is $\mathcal{L}_2 (Q^{1/2}U, H)$ progressively measurable and

$$P \left( \int_a^T \|\Phi\|^2_{\mathcal{L}_2(Q^{1/2}U,H)} ds < \infty \right) = 1.$$  

Then the same is true of $L\Phi$. Furthermore, for each $t \in [a, T]$

$$\int_a^t L\Phi dW = L \int_a^t \Phi dW$$

**Proof:** First note that if $\Phi \in \mathcal{L}_2 (Q^{1/2}U, H)$, then $L\Phi \in \mathcal{L}_2 (Q^{1/2}U, H_1)$ and that the map $\Phi \rightarrow L\Phi$ is continuous. It follows $L\Phi$ is $\mathcal{L}_2 (Q^{1/2}U, H_1)$ progressively measurable. All that remains is to check the appropriate integral.

$$\int_a^T \|L\Phi\|^2_{\mathcal{L}_2(Q^{1/2}U,H_1)} dt \leq \int_a^T \|L\|^2 \|\Phi\|^2_{\mathcal{L}_2(Q^{1/2}U,H)} dt$$

and so this proves $L\Phi$ satisfies the same conditions as $\Phi$, being stochastically square integrable.
It follows one can consider
\[ \int_{a}^{T} L \Phi dW. \]

Assume to begin with that \( \Phi \in L^2 \left( [a, T] \times \Omega; \mathcal{L}_2 \left( Q^{1/2} U, H \right) \right) \). Next recall the situation in which the definition of the integral is considered.

\[ U \uparrow Q^{1/2} \]
\[ U_1 \supseteq JQ^{1/2} U \]
\[ \Phi_n \searrow \Phi \]
\[ H \]

Letting \( \{ \Phi_n \} \) be an approximating sequence of elementary functions satisfying
\[ E \left( \int_{a}^{T} \left| \Phi_n - \Phi \circ J^{-1} \right|^2_{L^2(Q^{1/2} U, H)} dt \right) \to 0, \]
it is also the case that
\[ E \left( \int_{a}^{T} \left| L\Phi_n - L\Phi \circ J^{-1} \right|^2_{L^2(Q^{1/2} U, H_1)} dt \right) \to 0. \]

By the definition of the integral, for each \( t \)
\[ \int_{a}^{t} L \Phi dW = \lim_{n \to \infty} \int_{a}^{t} L\Phi_n dW = \lim_{n \to \infty} L \int_{a}^{t} \Phi_n dW \]
\[ = L \lim_{n \to \infty} \int_{a}^{t} \Phi_n dW = \int_{a}^{t} \Phi dW \]

The second equality is obvious for elementary functions.

Now consider the case where \( \Phi \) is only stochastically square integrable so that all is known is that
\[ P \left( \int_{a}^{T} \left| \Phi \right|^2_{L^2(Q^{1/2} U, H)} dt < \infty \right) = 1. \]

Then define \( \tau_n \) as above
\[ \tau_n = \inf \left\{ t : \int_{a}^{t} \left| \Phi \right|^2_{L^2(Q^{1/2} U, H)} dt \geq n \right\} \]

This sequence of stopping times works for \( L\Phi \) also. Recall there were two conditions the sequence of stopping times needed to satisfy. The first is obvious. Here is why the second holds.

\[ \int_{a}^{T} \left| X_{[a, \tau_n]} L\Phi \right|^2_{L^2(Q^{1/2} U, H_1)} dt \leq \left| L \right|^2 \int_{a}^{T} \left| X_{[a, \tau_n]} \Phi \right|^2_{L^2(Q^{1/2} U, H)} dt \]
\[ = \left| L \right|^2 \int_{a}^{\tau_n} \left| \Phi \right|^2_{L^2(Q^{1/2} U, H)} dt \leq \left| L \right|^2 n \]

Then let \( t \) be given and pick \( n \) such that \( \tau_n (\omega) \geq t \). Then from the first part, for that \( \omega \),
\[ L \int_{a}^{t} \Phi dW = L \int_{a}^{t} X_{[a, \tau_n]} \Phi dW \]
\[ = \int_{a}^{t} L X_{[a, \tau_n]} \Phi dW \]
\[ = \int_{a}^{t} X_{[a, \tau_n]} L \Phi dW \equiv \int_{a}^{t} L \Phi dW \]
17.14 A Technical Integration By Parts Result

Let $Z \in L^2 \left( [0, T] \times \Omega, \mathcal{L}_2 \left( Q^{1/2} U, H \right) \right)$ where this has reference to the usual diagram

$$
\begin{array}{c}
U_1 \supseteq JQ^{1/2} U \xrightarrow{j=1} Q^{1/2} U \\
\Phi_n \searrow \Phi
\end{array}
$$

Also suppose $X \in L^2 \left( [0, T] \times \Omega, H \right)$, both $X$ and $Z$ being progressively measurable. Let $\{ t^n_J \}_{j=1}^{m_n}$ denote a sequence of partitions of the sort discussed earlier where

$$
X_n (t) = \sum_{j=0}^{m_n-1} X (t^n_J) \mathcal{X} \left( t^n_J, t^n_{j+1} \right) (t)
$$

converges to $X$ in $L^2 \left( [0, T] \times \Omega, H \right)$. Thus $X_n (t)$ is right continuous. Let

$$
\tau^n_p = \inf \{ t : |X_n (t)|_H > p \}.
$$

This is the first hitting time of a right continuous adapted process so it is a stopping time. Also there exists a set of measure zero $N$ such that for $\omega \notin N$, then given $t$,

$$
\tau^n_p \geq t
$$

if $p$ is large enough because of the assumption on $X$. Here is why. There exists a set of measure 0 $N$ such that if $\omega \notin N$, then

$$
\int_0^T |X_n (t)|^2_H dt = \sum_{j=0}^{m_n-1} \left| X (t^n_J) \right|^2_H \left( t^n_{j+1} - t^n_J \right) < \infty.
$$

It follows that there exists an upper bound, depending on $\omega$ which dominates each of the values $|X (t^n_J)|^2_H$. Then if $p$ is larger than this upper bound, $\tau^n_p = \infty > t$.

Next consider the expression

$$
\sum_{j=0}^{m-1} \left( \int_{t^n_J \wedge t}^{t^n_{j+1} \wedge t} Z (u) dW, X (t^n_J) \right)_H.
$$

This expression is a function of $\omega$.

I want to write this in the form of a stochastic integral. To begin with, consider one of the terms. For simplicity of notation, consider

$$
\left( \int_a^b Z (u) dW, X (a) \right)_H
$$

where $Z \in L^2 \left( [a, b] \times \Omega, \mathcal{L}_2 \left( Q^{1/2} U, H \right) \right)$ and $X (a) \in L^2 (\Omega, H)$. Also assume the function of $\omega$, $|X (a)|_H$, is bounded. There is an Ito integral involved in the above. Let $\{ Z_n \}_{n=1}^{\infty}$ be a sequence of elementary functions defined on $[a, b]$ which converges to $Z \circ J^{-1}$ in $L^2 \left( [a, b] \times \Omega, \mathcal{L}_2 \left( JQ^{1/2} U, H \right) \right)$. Then by the definition of the integral,

$$
\left\| \int_a^t Z (u) dW - \int_a^t Z_n (u) dW \right\|_{L^2 (\Omega, H)} \to 0
$$

Also, by the use of a maximal inequality and the fact that the two integrals above are martingales, there is a subsequence, still called $n$ and a set of measure zero $N$ such that for $\omega \notin N$, the convergence

$$
\int_a^t Z_n (u) dW (\omega) \to \int_a^t Z (u) dW (\omega)
$$

is uniform for $t \in [a, b]$. Therefore, for such $\omega$,

$$
\left( \int_a^t Z (u) dW, X (a) \right)_H = \lim_{n \to \infty} \left( \int_a^t Z_n (u) dW, X (a) \right)_H
$$
Say $Z_n(u) = \sum_{k=0}^{m_n-1} Z^u_k X_{t^u_k,t^u_{k+1}}(u)$ where $Z^u_k$ has finitely many values in $L(U_1, H)$, the restrictions of $L(U_1, H)$ to $JQ^{1/2}U$. Then the inner product in the above formula on the right is of the form

$$\sum_{k=0}^{m_n-1} (Z^u_k (W(t \wedge t^u_{k+1}) - W(t \wedge t^u_k)), X((a)))_H$$

$$= \sum_{k=0}^{m_n-1} ((W(t \wedge t^u_{k+1}) - W(t \wedge t^u_k)), (Z^u_k)^* X((a)))_{U_1}$$

$$= \sum_{k=0}^{m_n-1} \mathcal{R}((Z^u_k)^* X((a))) (W(t \wedge t^u_{k+1}) - W(t \wedge t^u_k))$$

$$= \int_a^t \mathcal{R}(Z^u_n X((a))) dW$$

where $\mathcal{R}$ is the Riesz map from $U_1$ to $U'$. Note that $\mathcal{R}(Z^u_n X((a)))$ has values in $L(U_1, \mathbb{R}) \subseteq L_2(JQ^{1/2}U, \mathbb{R})$.

Now let $\{g_i\}$ be an orthonormal basis for $Q^{1/2}U$, so it follows that $\{Jg_i\}$ is an orthonormal basis for $JQ^{1/2}U$. Then

$$\sum_i |\mathcal{R}(Z^u_n X((a)) - (Z \circ J^{-1})^* X((a)))(Jg_i)|^2$$

$$\equiv \sum_i \left| \left( Z^u_n X((a)) - (Z \circ J^{-1})^* X((a)), Jg_i \right)_{U_1} \right|^2 = \sum_i |(X((a)), (Z_n - Z \circ J^{-1}) Jg_i)|_H^2$$

$$\leq \sum_i |X((a))|^2_H \| (Z_n - Z \circ J^{-1}) Jg_i \|^2_H = |X((a))|^2_H \| Z_n - Z \circ J^{-1} \|^2_{L_2(JQ^{1/2}U, H)}$$

When integrated over $[a, b] \times \Omega$, it is given that this converges to 0. This has shown that

$$\mathcal{R}(Z^u_n X((a))) \rightarrow \mathcal{R}((Z \circ J^{-1})^* X((a)))$$

in $L_2(JQ^{1/2}U, \mathbb{R})$. In other words

$$\mathcal{R}(Z^u_n X((a))) \rightarrow \mathcal{R}((Z \circ J^{-1})^* X((a)) \circ J^{-1}) \circ J^{-1}$$

It follows that

$$\left( \int_a^t Z((u)) dW, X((a)) \right)_H = \int_a^t \mathcal{R}((Z \circ J^{-1})^* X((a)) \circ J) dW$$

From localization,

$$\left( \int_{a \wedge \tau_p}^{b \wedge \tau_p} Z((u)) dW, X((a)) \right)_H = \left( \int_{a \wedge \tau_p}^{b} X_{[0, \tau_p]} Z((u)) dW, X((a)) \right)_H$$

$$= \int_{a \wedge \tau_p}^{b} X_{[0, \tau_p]} \mathcal{R}((Z \circ J^{-1})^* X((a)) \circ J) dW$$

Then it follows that, using the stopping time,

$$\sum_{j=0}^{m-1} \left( \int_{t^u_{j+1} \wedge \tau_p \wedge t^u_j}^{t^u_j \wedge \tau_p \wedge t^u_j} Z((u)) dW, X((u)) \right)_H = \sum_{j=0}^{m-1} \int_{t^u_{j+1} \wedge \tau_p \wedge t^u_j}^{t^u_j \wedge \tau_p \wedge t^u_j} \mathcal{R}((Z \circ J^{-1})^* X_n(t^u_n)) \circ J dW$$

$$= \int_0^{t^u_m \wedge \tau_p \wedge t} \mathcal{R}((Z \circ J^{-1})^* (X^t_n)) \circ J dW$$
where $X^l_n$ is the step function

$$X^l_n(t) \equiv \sum_{k=0}^{m_n-1} X(t^u_k) \chi_{[t^u_k,t^l_{k+1})}(t)$$

By localization, this is

$$\int_0^t X_{[0,\tau^T_p]}^n \mathcal{R} \left( \left( Z \circ J^{-1} \right)^* (X^l_n) \right) \circ J \, dW$$

If $\omega$ is not in a suitable set of measure zero, then $\tau^T_p(\omega) \geq t$ provided $p$ is large enough. Thus, for such $\omega$, if $p$ is large enough,

$$\sum_{j=0}^{m_n-1} \left( \int_{t^l_{j+1} \wedge \tau^T_p \wedge t}^{t^l_j \wedge \tau^T_p \wedge t} Z(u) \, dW, X(t^l_j) \right)_H = \int_0^t X_{[0,\tau^T_p]}^n \mathcal{R} \left( \left( Z \circ J^{-1} \right)^* (X^l_n) \right) \circ J \, dW$$

This shows that the expression is a local martingale. Also note that the expression on the left does not depend on $J$ or $U_1$ so the same must be true of the expression on the right although it does not look that way. This has proved the following important theorem.

**Theorem 17.14.1** Let $Z \in L^2([0,T] \times \Omega, \mathcal{L}_2(Q^{1/2} U, H))$ and let $X \in L^2([0,T] \times \Omega, H)$, both $X, Z$ progressively measurable. Also let $\{t^l_j\}_{j=1}^{m_n}$ be a sequence of partitions of $[0,T]$ such that each $X(t^l_j)$ is in $L^2(\Omega, H)$. Then

$$\sum_{j=0}^{m_n-1} \left( \int_{t^l_j \wedge \tau^T_p \wedge t}^{t^l_{j+1} \wedge \tau^T_p \wedge t} Z(u) \, dW, X(t^l_j) \right)_H$$

is a stochastic integral of the form

$$\int_0^t \mathcal{R} \left( \left( Z \circ J^{-1} \right)^* (X^l_n) \right) \circ J \, dW$$

where $\{\tau^T_p\}_{p=1}^{\infty}$ is a localizing sequence used to define the above integral whose integrand is only stochastically square integrable. Here $X^l_n$ is the step function defined by

$$X^l_n(t) \equiv \sum_{k=0}^{m_n-1} X(t^u_k) \chi_{[t^u_k,t^l_{k+1})}(t)$$

In particular, [17.14.2] is a local martingale.

Of course it would be very interesting to see what happens in the case where $X^l_n \to X$ in $L^2([0,T] \times \Omega, H)$. Is it the case that convergence to

$$\int_0^t \mathcal{R} \left( \left( Z \circ J^{-1} \right)^* (X) \right) \circ J \, dW$$

(17.14.22)

happens in some sense? Also, does the above stochastic integral even make sense? First of all, consider the question whether it makes sense. It would be nice to define a stopping time

$$\tau_n \equiv \inf \{ t : |X(t)|_H > n \}$$

because then $X_{[0,\tau_n]} \mathcal{R} \left( \left( Z \circ J^{-1} \right)^* (X) \right) \circ J$ would end up being integrable in the right way and you could define the stochastic integral provided $\tau_n > t$ whenever $n$ is large enough. However, this is problematic because $t \to X(t)$ is not known to be continuous. Therefore, some other condition must be assumed.

**Lemma 17.14.2** Suppose $t \to X(t)$ is weakly continuous into $H$ for a.e. $\omega$, and that $X$ is adapted. Then the $\tau_n$ described above is a stopping time.
Proof: Let \( B \equiv \{ x \in H : |x| > n \} \). Then the complement of \( B \) is a closed convex set. It follows that \( B^C \) is also weakly closed. Hence \( B \) must be weakly open. Now \( t \to X(t) \) is adapted as a function mapping into the topological space consisting of \( H \) with the weak topology because it is in fact adapted into the strong topology. Therefore, the above \( \tau_n \) is just the first hitting time of an open set by a continuous process so \( \tau_n \) is a stopping time by Proposition 17.3.2. Also, by the assumption that \( t \to X(t) \) is weakly continuous, it follows that \( X(t) \) for \( t \in [0,T] \) is weakly bounded. Hence, for each \( \omega \) off a set of measure zero, \( |X(t)| \) is bounded for \( t \in [0,T] \). This follows from the uniform boundedness theorem. It follows that \( \tau_n = \infty \) for \( n \) large enough.

Hence the weak continuity of \( t \to X(t) \) suffices to define the stochastic integral in \( \mathbb{R} \). It remains to verify some sort of convergence in the case that

\[
\lim_{n \to \infty} \left[ \max_{j \leq m_n - 1} (t^n_{j+1} - t^n_j) \right] = 0
\]

Lemma 17.14.3 Let \( X(s) - X^*_k(s) \equiv \Delta_k(s) \). Here \( Z \in L^2 ([0,T] \times \Omega, \mathcal{L}_2 \left( Q^{1/2}U, H \right) ) \) and let \( X \in L^2 ([0,T] \times \Omega, H) \) with both \( X \) and \( Z \) progressively measurable, \( t \to X(t) \) being weakly continuous into \( H \),

\[
\lim_{k \to \infty} \| X - X^*_k \|_{L^2([0,T] \times \Omega, H)} = 0
\]

Then the integral

\[
\int_0^t \mathcal{R} \left( (Z \circ J^{-1})^* (X) \right) \circ JdW
\]

exists as a local martingale and the following limit occurs for a suitable subsequence, still called \( k \).

\[
\lim_{k \to \infty} P \left( \left[ \sup_{t \in [0,T]} \left| \int_0^t \mathcal{R} \left( (Z \circ J^{-1})^* (X(s) - X^*_k(s)) \right) \circ JdW(s) \right| \geq \varepsilon \right] \right) = 0. \tag{17.14.23}
\]

That is,

\[
\sup_{t \in [0,T]} \left| \int_0^t \mathcal{R} \left( (Z \circ J^{-1})^* (X(s) - X^*_k(s)) \right) \circ JdW(s) \right|
\]

converges to 0 in probability.

Proof: Let \( k \) denote a subsequence for which \( X^*_k \) also converges pointwise to \( X \).

The existence of the integral follows from Lemma 17.14.2. From the assumption of weak continuity, \( \sup_{t \in [0,T]} |X(t)| \leq C(\omega) \) for \( a.e. \omega \). For the first part of the argument, assume \( C \) does not depend on \( \omega \) off a set of measure zero. Let

\[
M(t) \equiv \int_0^t ZdW
\]

Let \( \{ e_i \} \) be an orthonormal basis for \( H \) and let \( P_n \) be the orthogonal projection onto span \( \langle e_1, \cdots, e_n \rangle \). For each \( e_i \)

\[
\lim_{k \to \infty} \left| \langle X(s) - X^*_k(s), e_i \rangle \right| = 0
\]

and so, by weak continuity,

\[
\lim_{k \to \infty} P_n (X(s) - X^*_k(s)) = 0 \text{ for } a.e. \omega
\]

Then

\[
\lim_{k \to \infty} \int_0^T \int_\Omega |P_n (X(s) - X^*_k(s))|^2 \| Z(s) \|_{L_2}^2 d\sigma dP = 0
\]

because you can apply the dominated convergence theorem with respect to the measure \( \| Z(s) \|_{L_2}^2 d\sigma dP \).

Therefore,

\[
\lim_{k \to \infty} P \left( \left[ \sup_{t \in [0,T]} \left| \int_0^t \mathcal{R} \left( (Z \circ J^{-1})^* P_n \Delta_k(s) \right) \circ JdW(s) \right| \geq \varepsilon/2 \right] \right) = 0 \tag{17.14.24}
\]

Here is why. By the Burkholder Davis Gundy theorem, Theorem 17.3.2 and Corollary 17.4.2 which describes the quadratic variation of the stochastic integral,

\[
\int_\Omega \left( \sup_{t \in [0,T]} \left| \int_0^t \mathcal{R} \left( (Z \circ J^{-1})^* P_n \Delta_k(s) \right) dW(s) \right| \right) dP
\]
It follows that for $k > K$, the first term is no larger than

$$C \int_{\Omega} \left( \int_0^T \left| P_n (X(s) - X_k^I(s)) \right|^2 \left\| Z(s) \right\|^2_{L^2} ds \right)^{1/2} dP$$

Consider the following two probabilities.

$$P \left( \left[ \sup_{t \in [0,T]} \left| \int_0^t \mathcal{R} \left( (Z(s) \circ J^{-1})^* (I - P_n) X(s) \right) \circ JdW(s) \right| \geq \varepsilon / 2 \right] \right) \quad (17.14.25)$$

$$P \left( \left[ \sup_{t \in [0,T]} \left| \int_0^t \mathcal{R} \left( (Z(s) \circ J^{-1})^* (I - P_n) X_k^I(s) \right) \circ JdW(s) \right| \geq \varepsilon / 2 \right] \right) \quad (17.14.26)$$

By Corollary 17.14.25 which depends on the Burkholder Davis Gundy inequality and Corollary 17.14.26 which describes the quadratic variation of the stochastic integral, 17.14.27 is dominated by

$$\frac{C}{\varepsilon} E \left( \left( \int_0^T \left| Z(s) \right|^2 \left| (I - P_n) X(s) \right|^2 ds \right)^{1/2} \wedge \delta \right) + P \left( \left( \int_0^T \left| Z(s) \right|^2 \left| (I - P_n) X(s) \right|^2 ds \right)^{1/2} > \delta \right)$$

$$\leq \frac{C \delta}{\varepsilon} + P \left( \left( \int_0^T \left| Z(s) \right|^2 \left| (I - P_n) X(s) \right|^2 ds \right)^{1/2} > \delta \right) \quad (17.14.27)$$

Let $\eta > 0$ be given. Then let $\delta$ be small enough that the first term is less than $\eta$. Fix such a $\delta$. Consider the second of the above terms.

$$P \left( \left( \int_0^T \left| Z(s) \right|^2 \left| (I - P_n) X(s) \right|^2 ds \right)^{1/2} > \delta \right)$$

$$\leq \frac{1}{\delta} \left( E \left( \int_0^T \left| Z(s) \right|^2 \left| (I - P_n) X(s) \right|^2 ds \right) \right)^{1/2}$$

and this converges to 0 because $(I - P_n) X(s)$ is assumed to be bounded and converges to 0. Next consider 17.14.26. By similar reasoning, we end up with having to estimate

$$\frac{1}{\delta} \left( E \left( \int_0^T \left| Z(s) \right|^2 \left| (I - P_n) X_k^I(s) \right|^2 ds \right) \right)^{1/2} \cdot$$

But this is dominated by

$$\frac{2}{\delta} \left( E \left( \int_0^T \left| Z(s) \right|^2 \left| X_k^I(s) - X(s) \right|^2 ds \right) \right)^{1/2} + \frac{2}{\delta} \left( E \left( \int_0^T \left| Z(s) \right|^2 \left| (I - P_n) X(s) \right|^2 ds \right) \right)^{1/2}$$

The first term is no larger than $\eta$ provided $k$ is large enough, independent of $n$ thanks to the pointwise convergence and the assumption that $X$ is bounded. Thus, there exists $K$ such that if $k > K$, then the term in 17.14.27 is dominated by

$$2\eta + \frac{2}{\delta} \left( E \left( \int_0^T \left| Z(s) \right|^2 \left| (I - P_n) X(s) \right|^2 ds \right) \right)^{1/2}$$

It follows that for $k > K$, the sum of 17.14.26 and 17.14.27 is dominated by

$$3\eta + \frac{3}{\delta} \left( E \left( \int_0^T \left| Z(s) \right|^2 \left| (I - P_n) X(s) \right|^2 ds \right) \right)^{1/2}$$
This is then no larger than $4\eta$ provided $n$ is large enough. Pick such an $n$. Then for all $k > K$, this has shown that
\[
\mathbb{P}\left(\left[\sup_{t \in [0,T]} \left| \int_0^t \mathcal{R}\left(\left( Z(s) \circ J^{-1}\right)^* \Delta_k(s) \right) \circ JdW(s) \right| \geq \varepsilon \right] \right)
\]
\[
\leq \mathbb{P}\left(\left[\sup_{t \in [0,T]} \left| \int_0^t \mathcal{R}\left(\left( Z(s) \circ J^{-1}\right)^* P_n \Delta_k(s) \right) \circ JdW(s) \right| \geq \varepsilon/2 \right] \right) +
\]
\[
\mathbb{P}\left(\left[\sup_{t \in [0,T]} \left| \int_0^t \mathcal{R}\left(\left( Z(s) \circ J^{-1}\right)^* \left( I - P_n \right) \Delta_k(s) \right) \circ JdW(s) \right| \geq \varepsilon/2 \right] \right)
\]
\[
\leq \mathbb{P}\left(\left[\sup_{t \in [0,T]} \left| \int_0^t \mathcal{R}\left(\left( Z(s) \circ J^{-1}\right)^* P_n \Delta_k(s) \right) \circ JdW(s) \right| \geq \varepsilon/2 \right] \right) + 4\eta
\]

By (17.14.28) this whole thing is less than $5\eta$ provided $k$ is large enough. This has proved that under the assumption that $X$ is bounded uniformly off a set of measure zero,
\[
\lim_{k \to \infty} \mathbb{P}\left(\left[\sup_{t \in [0,T]} \left| \int_0^t \mathcal{R}\left(\left( Z(s) \circ J^{-1}\right)^* \Delta_k(s) \right) \circ JdW(s) \right| \geq \varepsilon \right] \right) = 0
\]

This is what was desired to show. It remains to remove the extra assumption that $X$ is bounded.

Now to finish the argument, define the stopping time
\[
\tau_m \equiv \inf \{ t > 0 : |X(t)|_H > m \}.
\]

As observed in Lemma (17.14.24) this is a valid stopping time. Also define $\Delta_k^m \equiv X^{\tau_m} - \langle X_k \rangle^{\tau_m}$. Using this stopping time on $X$ and $X_k$ does not affect the pointwise convergence to 0 as $k \to \infty$ of $\Delta_k^m$ on which the above argument depends.

Consider
\[
A_{k\varepsilon} \equiv \left[\sup_{t \in [0,T]} \left| \int_0^t \mathcal{R}\left(\left( Z(s) \circ J^{-1}\right)^* \Delta_k(s) \right) \circ JdW(s) \right| \geq \varepsilon \right]
\]

Then
\[
\mathbb{P}(A_{k\varepsilon} \cap [\tau_m = \infty]) \leq \mathbb{P}\left(\left[\sup_{t \in [0,T]} \left| \int_0^t \mathcal{R}\left(\left( Z(s) \circ J^{-1}\right)^* \Delta_k^m(s) \right) \circ JdW(s) \right| \geq \varepsilon \right] \right)
\]

which converges to 0 as $k \to \infty$ by the first part of the argument. This is because $|X^{\tau_m}|$ and $\langle X_k^{\tau_m} \rangle$ are both bounded by $m$ and the same pointwise convergence condition still holds. Now
\[
A_{k\varepsilon} = \bigcup_{m=1}^{\infty} A_{k\varepsilon} \cap ([\tau_m = \infty] \setminus [\tau_{m-1} < \infty])
\]

Thus
\[
\mathbb{P}(A_{k\varepsilon}) = \sum_{m=1}^{\infty} \mathbb{P}(A_{k\varepsilon} \cap ([\tau_m = \infty] \setminus [\tau_{m-1} < \infty]))
\]

(17.14.28)

Also
\[
\mathbb{P}(A_{k\varepsilon} \cap ([\tau_m = \infty] \setminus [\tau_{m-1} < \infty])) \leq \mathbb{P}( [\tau_m = \infty] \setminus [\tau_{m-1} < \infty])
\]

which is summable because these are disjoint sets. Hence one can apply the dominated convergence theorem in (17.14.29) and conclude
\[
\lim_{k \to \infty} \mathbb{P}(A_{k\varepsilon}) = \sum_{m=1}^{\infty} \lim_{k \to \infty} \mathbb{P}(A_{k\varepsilon} \cap ([\tau_m = \infty] \setminus [\tau_{m-1} < \infty])) = 0 \quad \blacksquare
Chapter 18

The Integral $\int_0^t (Y, dM)_H$

First the integral is defined for elementary functions.

**Definition 18.0.4** Let an elementary function be one which is of the form

$$\sum_{i=0}^{m-1} Y_i \mathcal{X}_{[t_i, t_{i+1}]}(t)$$

where $Y_i$ is $\mathcal{F}_{t_i}$ measurable with values in $H$ a separable real Hilbert space for $0 = t_0 < t_1 < \cdots < t_m = T$.

**Definition 18.0.5** Now let $M$ be a $H$ valued continuous local martingale, $M(0) = 0$. Then for $Y$ a simple function as above,

$$\int_0^t (Y, dM)_H \equiv \sum_{i=0}^{m-1} (Y_i, M(t \wedge t_{i+1}) - M(t \wedge t_i))_H$$

**Assumption 18.0.6** We will always assume that $d[M]$ is absolutely continuous with respect to Lebesgue measure. Thus $d[M] = kdt$ where $k \geq 0$ and is in $L^1([0, T] \times \Omega)$. This is done to avoid technical questions related to whether $t \to \int_0^t d[M]$ is continuous and also to make it easier to get examples of a certain class of functions.

This includes the usual stochastic integral $M(t) = \int_0^t \Phi dW$ where $[M](t) = \int_0^t \|\Phi\|^2 ds$ so $d[M] = \|\Phi\|^2 dt$.

Next is to consider how this relates to stopping times which have values in the $\{t_i\}$. Let $\tau$ be a stopping time which takes the values $\{t_i\}_{i=0}^m$. Then

$$\int_0^{t \wedge \tau} (Y, dM)_H \equiv \sum_{i=0}^{m-1} (Y_i, M(t \wedge t_{i+1} \wedge \tau) - M(t \wedge t_i \wedge \tau))_H \quad \text{(18.0.1)}$$

Now consider $\mathcal{X}_{[0, \tau]} Y$. Is it also an elementary function?

$$\mathcal{X}_{[0, \tau]} Y = \sum_{i=0}^{m-1} \mathcal{X}_{[0, \tau]}(t) Y_i \mathcal{X}_{[t_i, t_{i+1}]}(t)$$

To get the $i^{th}$ term to be non zero, you must have $\tau \geq t$ and $t \in (t_i, t_{i+1}]$. Thus it must be the case that $\tau > t_i$. Also, if $\tau > t_i$ and $t \in (t_i, t_{i+1}]$, then $\tau \geq t_{i+1}$ because $\tau$ has only the values $t_i$. Hence also $\tau \geq t$. Thus the above sum reduces to

$$\sum_{i=0}^{m-1} \mathcal{X}_{[\tau > t_i]}(\omega) Y_i \mathcal{X}_{[t_i, t_{i+1}]}(t)$$

This shows that $\mathcal{X}_{[0, \tau]} Y$ is of the right sort, the sum of $\mathcal{F}_{t_i}$ measurable functions times $\mathcal{X}_{[t_i, t_{i+1}]}(t)$. Thus from the definition of this funny integral,

$$\int_0^t (\mathcal{X}_{[0, \tau]} Y, dM)_H \equiv \sum_{i=0}^{m-1} \left( \mathcal{X}_{[\tau > t_i]}(\omega) Y_i, M(t \wedge t_{i+1}) - M(t \wedge t_i) \right)_H \quad \text{(18.0.2)}$$

405
Are the right sides of \textbf{18.0.1} and \textbf{18.0.2} equal? 

Begin with the right side of \textbf{18.0.1} and consider \( \tau = t_j \). Then to get something nonzero in the terms of the sum in \textbf{18.0.1}, you would need to have \( t_j \geq t_{i+1} \). Otherwise, \( t_j \leq t_i \) and the difference involving \( M \) would go 0. Hence, for such \( \omega \) you would need to have the term in \textbf{18.0.1} equal to

\[
\sum_{i=0}^{j-1} (Y_i, M(t \land t_{i+1}) - M(t \land t_i))_H
\]

Thus this sum in \textbf{18.0.1} equals

\[
\sum_{j=0}^{m} \sum_{i=0}^{j-1} (Y_i, M(t \land t_{i+1}) - M(t \land t_i))_H
\]

Of course when \( j = 0 \) the term in the sum in \textbf{18.0.1} equals 0 so there is no harm in defining \( \sum_{i=0}^{m-1} = 0 \). Then from the sum, you have \( i \leq j - 1 \) and so when you interchange the order, you get that \( \int_0^{t \land \tau} (Y, dM) = \)

\[
\sum_{i=0}^{m-1} \sum_{j=i+1}^{m} (Y_i, M(t \land t_{i+1}) - M(t \land t_i))_H = \sum_{i=0}^{m-1} (Y_i, M(t \land t_{i+1}) - M(t \land t_i))_H
\]

Thus the right side of \textbf{18.0.1} equals the right side of \textbf{18.0.2}

\[
\int_0^t (X[0,r] Y, dM) = \sum_{i=0}^{m-1} (X[\tau > t_i], Y, M(t \land t_{i+1}) - M(t \land t_i)) = \int_0^{t \land \tau} (Y, dM)
\]

This has proved the first part of the following lemma.

**Lemma 18.0.7** For an elementary function \( Y \), and a stopping time \( \tau \) having values in the \( \{t_i\} \), the points of discontinuity of \( Y \), it follows that \( X[0,\tau] Y \) is also an elementary function and

\[
\int_0^{t \land \tau} (Y, dM) = \int_0^t (X[0,\tau] Y, dM) = \int_0^t (Y, dM^\tau)
\]

**Proof:** Consider the second equal sign. By definition,

\[
\int_0^{t \land \tau} (Y, dM) = \sum_{i=0}^{m-1} (Y_i, M(t \land t_{i+1} \land \tau) - M(t \land t_i \land \tau))_H
\]

\[
= \sum_{i=0}^{m-1} (Y_i, M^\tau(t \land t_{i+1}) - M^\tau(t \land t_i))_H = \int_0^t (Y, dM^\tau)
\]

Next is another lemma about these integrals of elementary functions. First recall the following definition

\[
M^* \equiv \sup \{|M(t)| : t \in [0,T]\}
\]

**Lemma 18.0.8** Let \( M \) be a local martingale on \([0,T]\) where \( M(0) = 0 \) and \( M \) is continuous. Let \( 0 < r < s < T \) and consider \( Y, (M^{r,s} - M^{r,r}) (t) \) where \( Y (M^{r,s})^* \in L^2(\Omega) \) and \( Y \) is \( F_r \) measurable and \( \tau_p \) is a localizing sequence of stopping times for which \( M^{r,s} \) is a \( L^2 \) martingale. Then this is a martingale on \([0,T]\) which equals \( 0 \) at \( t = 0 \) and

\[
[(Y, (M^{r,s} - M^{r,r})) (t)] \leq |Y|^2 (M^{r,s} - M^{r,r}) (t)
\]

\[
= |Y|^2 (M^{r,s})^* (t) - (M^{r,r})^* (t)
\]

\[
= |Y|^2 (M^{r,s} (t \land s) - M^{r,s} (t \land r))
\]

It follows that for \( Y \) an elementary function where each \( Y_i (M^{r,s})^* \) is in \( L^2(\Omega) \),

\[
\int_0^t (Y, dM)
\]

is a local martingale.
Proof: To save notation, $M$ is written in place of $M^\tau r$. It is clear that $(Y, (M^s - M^r) (t)) = 0$ if $t \leq r$. Is it a martingale?

\[
E ( (Y, (M^s - M^r) (t)) ) = E ( E ((Y, (M^s - M^r) (t)) | \mathcal{F}_r) ) = E ( (Y, E ((M (s \wedge t) - M (r \wedge t)) | \mathcal{F}_r) ) ) = 0
\]

because $M$ is a martingale. Now let $\sigma$ be a bounded stopping time with two values. Then using the optional sampling theorem where needed,\n
\[
E ( (Y, (M^s - M^r) (\sigma)) ) = E ( E ((Y, (M^s - M^r) (\sigma)) | \mathcal{F}_r) ) = E ( (Y, E ((M (s \wedge \sigma) - M (r \wedge \sigma)) | \mathcal{F}_r) ) ) = E ( (Y, M (\sigma \wedge r) - M (r \wedge \sigma)) ) = 0
\]

It follows that this is indeed a martingale as claimed.

By the definition of the quadratic variation,

\[
\| (Y, (M^s - M^r) (t)) \|^2 \leq \| Y \|^2 \| (M^s - M^r) (t)) \|^2
\]

\[
= \| Y \|^2 [(M^s - M^r)] (t) + \| Y \|^2 \hat{N}(t)
\]

where $\hat{N}(t)$ is a martingale. It equals 0 if $t \leq r$. By similar reasoning to the above, $\| Y \|^2 \hat{N}(t)$ is a martingale. To see this,

\[
E ( \| Y \|^2 \hat{N}(\sigma) ) = E ( E (\| Y \|^2 \hat{N}(\sigma) | \mathcal{F}_r ) ) = E ( \| Y \|^2 E ( \hat{N}(\sigma) | \mathcal{F}_r ) ) = E ( \| Y \|^2 N (\sigma \wedge r) ) = 0
\]

One also sees that $E ( \| Y \|^2 \hat{N}(t) ) = 0$.

Now it follows from Corollary [406] that

\[
[(M^s - M^r)] = [M^s] - [M^r] = [M]^s - [M]^r
\]

Hence

\[
[(Y, (M^s - M^r))] (t) \leq \| Y \|^2 [M^s - M^r] (t) = \| Y \|^2 ( [M]^s (t) - [M]^r (t) )
\]

as claimed.

The last claim is easy. Let $\tau_p$ be a localizing sequence for which $M^\tau r$ is a martingale. Then

\[
\int_0^{t \wedge \tau_p} (Y, dM) = \sum_{i=0}^{m-1} (Y_t, M(t \wedge t_{i+1} \wedge \tau_p) - M(t \wedge t_i \wedge \tau_p))_H
\]

\[
= \sum_{i=0}^{m-1} (Y_t, M^\tau r (t \wedge t_{i+1}) - M^\tau r (t \wedge t_i))_H
\]

a finite sum of martingales. \[ \square \]

Note that this is just a definition and did not use the above localization lemma. In particular, $\tau_p$ is not restricted to having only the partition points as values.

Next one needs to generalize past the elementary functions.

Continue writing $M$ in place of $M^\tau r$ in what follows. Consider an elementary function

\[
Y \equiv \sum_{k=0}^{m-1} Y_{k,t_k,t_{k+1}} (t)
\]

where $Y_k M^s \in L^2(\Omega)$. Consider

\[
\int_0^t (Y, dM) = \sum_{k=0}^{m-1} (Y_k, M(t \wedge t_{k+1}) - M(t \wedge t_k))
\]

(18.0.3)
Then it is routine to verify that
\[
E \left( \left( \sum_{k=0}^{m_n-1} (Y_k, M (t \wedge t_{k+1}) - M (t \wedge t_k))_H \right)^2 \right) = \sum_{k=0}^{m_n-1} E \left( (Y_k, M (t \wedge t_{k+1}) - M (t \wedge t_k))_H^2 \right)
\]

This is because the mixed terms all vanish. This follows from the following reasoning. Let \( t_j < t_k \)
\[
E \left( \left( \sum_{k=0}^{m_n-1} (Y_k, M (t \wedge t_{k+1}) - M (t \wedge t_k))_H \right)^2 \right) = \sum_{k=0}^{m_n-1} E \left( (Y_k, (M^{t_{k+1}} - M^t) (t))_H^2 \right)
\]

It follows from \[[18.0.3]\]
\[
E \left( \left( \sum_{k=0}^{m_n-1} (Y_k, M (t \wedge t_{k+1}) - M (t \wedge t_k))_H \right)^2 \right) = \sum_{k=0}^{m_n-1} E \left( (Y_k, (M^{t_{k+1}} - M^t) (t))_H^2 \right)
\]

where \( N_k \) is a martingale equal to 0 for \( t \leq t_k \). Then this equals
\[
\sum_{k=0}^{m_n-1} E \left( (Y_k, (M^{t_{k+1}} - M^t) (t)) \right)
\]

From Lemma \[[18.0.6]\]
\[
\leq E \left( \sum_{k=0}^{m_n-1} \|Y_k\|_H^2 \left( [M]^{t_{k+1}} (t) - [M]^t (t) \right) \right)
\]

\[
= E \left( \sum_{k=0}^{m_n-1} \|Y_k\|_H^2 \left( [M] (t_{k+1}^n \wedge t) - [M] (t_k^n \wedge t) \right) \right)
\]

Note that everything makes sense because it is assumed that \( \|Y_k\| M^* \in L^2 (\Omega) \). This proves the following lemma.

**Lemma 18.0.9** Let \( \|Y (t)\| (M^{\tau_p})^* \in L^2 (\Omega) \) for each \( t \), where \( Y \) is an elementary function and let \( \tau_p \) be a stopping time for which \( M^{\tau_p} \) is a \( L^2 \) martingale. Then

\[
E \left( \left| \int_0^t Y (dM^{\tau_p}) \right|^2 \right) \leq E \left( \int_0^t \|Y\|_H^2 d[M]^{\tau_p} \right)
\]

The condition that \( \|Y (t)\| (M^{\tau_p})^* \in L^2 (\Omega) \) ensures that
\[
E \left( (Y_k, M^{\tau_p} (t \wedge t_{k+1}) - M^{\tau_p} (t \wedge t_{k+1}))_H^2 \right)
\]

always is finite.
**Definition 18.0.10** Let $\mathcal{G}$ denote those functions $Y$ which are adapted and have the property that for each $p,$

$$\lim_{n \to \infty} E \left( \int_0^T \|Y - Y^n\|_H^2 d[M]^T_p \right) = 0$$

for some sequence $Y^n$ of elementary functions for which $\|Y^n(t)\|_H^2 \in L^2(\Omega)$ for each $t$. Here $d[M]^T_p$ signifies the Lebesgue-Stieltjes measure determined by the increasing function $t \to [M]^T(t)$. Let $M^T_p$ be an $L^2$ martingale. Recall that $\tau_p$ is just a localizing sequence for the local martingale $M$.

It is not known whether this increasing function is absolutely continuous.

**Definition 18.0.11** Let $Y \in \mathcal{G}$. Then

$$\int_0^t (Y, dM^T_p) \equiv \lim_{n \to \infty} \int_0^t (Y^n, dM^T_p) \text{ in } L^2(\Omega)$$

For example, suppose $Y$ is a bounded continuous process having values in $H$. Then you could look at the left step functions

$$Y^n(t) \equiv \sum_{i=0}^{m_n-1} Y(t_i) \mathcal{X}_{[t_i, t_{i+1})}(t)$$

The $Y^n$ would converge to $Y$ pointwise on $[0, T]$ for each $\omega$ and these $Y^n$ are bounded. In fact, in this case, these converge uniformly to $Y$ on $[0, T]$. Thus this is an example of the situation in the above definition. In this case, the integrand would be bounded by $C$ for some $C$ and

$$E \left( \int_0^T Cd[M]^T_p \right) = E \left( [M]^T_p (T) \right) = E \left( \|M^T_p (T)\|^2 \right) < \infty$$

by assumption. Hence, by the dominated convergence theorem,

$$\lim_{n \to \infty} E \left( \int_0^T \|Y - Y^n\|^2_H d[M]^T_p \right) = 0.$$

What if $[M]^T_p$ were bounded and absolutely continuous with respect to Lebesgue measure? This could be the case if you had $\tau_p$ a stopping time of the form

$$\tau_p = \inf \{ t : [M](t) > p \}$$

Then if $Y \in L^2([0,T] \times \Omega, H)$, and progressively measurable there are left step functions which converge to $Y$ in $L^2([0,T] \times \Omega, H).$ Say $d[M^T] = k(t, \omega) dm$ where $k$ is bounded. Then

$$E \left( \int_0^T \|Y - Y^n\|^2_H d[M]^T_p \right) = E \left( \int_0^T \|Y - Y^n\|^2_H k dt \right) \to 0$$

**Lemma 18.0.12** The above definition is well defined. Also, $\int_0^t (Y, dM^T_p)$ is a continuous martingale. The inequality

$$E \left( \left| \int_0^t (Y, dM^T_p) \right|^2 \right) \leq E \left( \int_0^t \|Y\|_H^2 d[M]^T_p \right)$$

is also valid. For any sequence of elementary functions $\{Y^n\}, \|Y^n(t)\|_H^2 \in L^2(\Omega),$ 

$$\|Y^n - Y\|_{L^2([0,T];L^2(H^2))} \to 0$$

there exists a subsequence, still denoted as $\{Y^n\}$ of elementary functions for which $\int_0^t (Y^n, dM^T_p)$ converges uniformly to $\int_0^t (Y, dM^T_p)$ on $[0, T]$ for $\omega$ off some set of measure zero.
Theorem 18.0.13 Let $H$ be a Hilbert space and suppose $(M, F_t), t \in [0, T]$ is a uniformly bounded continuous martingale with values in $H$. Also let $\{t^n_k\}_{k=1}^{m} \subset [0, T]$ be a sequence of partitions satisfying
\[
\lim_{n \to \infty} \max_i \left\{ |t^n_i - t^n_{i+1}|, i = 0, \ldots, m_n \right\} = 0, \quad \{t^n_k\}_{k=1}^{m} \leq \left\{t^{n+1}_k\right\}_{k=1}^{m+1}.
\]
Then
\[
[M](t) = \lim_{n \to \infty} \sum_{k=0}^{m_n-1} \|M(t \wedge t^n_{k+1}) - M(t \wedge t^n_k)\|^2_H
\]
the limit taking place in $L^2(\Omega)$. In case $M$ is just a continuous local martingale, the above limit happens in probability.
In the above Lemma, you would find the quadratic variation according to this theorem as follows.

\[
\left[ \int_0^{\cdot} (Y, dM^{\tau_p}) \right] (t) = \lim_{n \to \infty} \sum_{k=0}^{m_n-1} \left| \int_{t \wedge \tau_{n+1}^k} (Y, dM^{\tau_p}) \right|_H^2
\]

where the limit is in probability. Thus

\[
\lim_{n \to \infty} P \left( \left| \left[ \int_0^{\cdot} (Y, dM^{\tau_p}) \right] (t) - \sum_{k=0}^{m_n-1} \left| \int_{t \wedge \tau_{n+1}^k} (Y, dM^{\tau_p}) \right|_H^2 \right| \geq \varepsilon \right) = 0
\]

Then you can obtain from this and the usual appeal to the Borel Cantelli lemma a set of measure zero \( N_t \) and a subsequence still denoted with \( n \) satisfying that for all \( \omega \notin N_t \) and \( n \) large enough,

\[
\left| \left[ \int_0^{\cdot} (Y, dM^{\tau_p}) \right] (t) - \sum_{k=0}^{m_n-1} \left| \int_{t \wedge \tau_{n+1}^k} (Y, dM^{\tau_p}) \right|_H^2 \right| \leq \frac{1}{n}
\]

Hence

\[
\left[ \int_0^{\cdot} (Y, dM^{\tau_p}) \right] (t) \leq \frac{1}{n} + \sum_{k=0}^{m_n-1} \int_{t \wedge \tau_{n+1}^k} \| Y \|_H^2 d[M]^{\tau_p}
\]

\[
= \frac{1}{n} + \int_0^t \| Y \|_H^2 d[M]^{\tau_p}
\]

Then for that \( t \), you have on taking a limit as \( n \to \infty \),

\[
\left[ \int_0^{\cdot} (Y, dM^{\tau_p}) \right] (t) \leq \int_0^t \| Y \|_H^2 d[M]^{\tau_p}
\]

Now take the union of \( N_t \) for \( t \in \mathbb{Q} \cap [0, T] \). Denote this as \( N \). Then if \( \omega \notin N \), the above shows that for such \( t \),

\[
\left[ \int_0^{\cdot} (Y, dM^{\tau_p}) \right] (t) \leq \int_0^t \| Y \|_H^2 d[M]^{\tau_p}
\]

But both sides are continuous in \( t \) and so this inequality holds for all \( t \in [0, T] \). Thus the following corollary is obtained.

**Corollary 18.0.14** Let \( M \) be a continuous local martingale and \( \tau_p \) a localizing sequence which makes \( M^{\tau_p} \) an \( L^2 \) martingale and assume that \( Y \in \mathcal{G} \). Then the quadratic variation of this martingale satisfies

\[
\left[ \int_0^{\cdot} (Y, dM^{\tau_p}) \right] (t) \leq \int_0^t \| Y \|_H^2 d[M]^{\tau_p} \leq \int_0^t \| Y \|_H^2 d[M]
\]

for \( \omega \) off a set of measure zero.

Does the localization stuff hold for an arbitrary stopping time? Let \( \{t^k\} \) denote the \( k \)th partition of a sequence of nested partitions whose maximum length between successive points converges to 0. Let \( \tau \) be a stopping time and let \( \tau_k = t^k_{j+1} \) on \( \tau^{-1}(t^k_j, t^k_{j+1}) \). Then \( \tau_k \) is a stopping time because

\[
[\tau_k \leq t] \in \mathcal{F}_t
\]

Here is why. If \( t \in (t^k_j, t^k_{j+1}) \), then if \( t = t^k_{j+1} \), it would follow that \( \tau_k (\omega) \leq t \) would be the same as saying \( \omega \in [\tau \leq t^k_{j+1}] = [\tau \leq t] \in \mathcal{F}_t \). On the other hand, if \( t < t^k_{j+1} \), then \( [\tau_k \leq t] = [\tau \leq t^k_j] \in \mathcal{F}_{t^k_j} \subseteq \mathcal{F}_t \) because \( \tau_k \) can only take the values \( t^k_j \).

Let \( Y \) be one of those elementary functions which is in \( \mathcal{G} \), \( \| Y (t) \| M^* \in L^2 (\Omega) \).

\[
Y (t) = \sum_{i=0}^{m_k-1} Y_i \chi_{(t^k_i, t^k_{i+1})} (t)
\]
and consider $\chi_{[0,\tau_k]} Y$. Here $Y$ will be always the same for the different partitions. It is just that some of the $Y_i$ will be repeated on smaller and smaller intervals. Does it follow that $\chi_{[0,\tau_k]} Y \to \chi_{[0,\tau]} Y$ for each fixed $\omega$? This depends only on the indicator function. Let $\tau(\omega) \in (t_k, t_{k+1})$. Fixing $t$, if $\chi_{[0,\tau]} (t) = 1$, then also $\chi_{[0,\tau_k]} (t) = 1$ because $\tau_k \geq \tau$. Therefore, in this case $\lim_{k \to \infty} \chi_{[0,\tau_k]} (t) = \chi_{[0,\tau]} (t)$. Next suppose $\chi_{[0,\tau]} (t) = 0$ so that $\tau(\omega) < t$. Since the intervals defined by the partition points have lengths which converge to 0, it follows that for all $k$ large enough, $\tau_k(\omega) < t$ also and so $\chi_{[0,\tau_k]} (t) = 0$. Therefore, 

$$\lim_{k \to \infty} \chi_{[0,\tau_k]} (\omega) (t) = \chi_{[0,\tau]} (\omega) (t).$$

It follows that $\chi_{[0,\tau_k]} Y \to \chi_{[0,\tau]} Y$. Also it is clear from the dominated convergence theorem, 

$$\|\chi_{[0,\tau_k]} Y - \chi_{[0,\tau]} Y\|_H^2 \leq 4 \|Y\|_H^2,$$

that 

$$\lim_{k \to \infty} E \left( \int_0^T \|\chi_{[0,\tau_k]} Y - \chi_{[0,\tau]} Y\|_H^2 \ d [M^\tau] \right) = 0$$

Thus $\chi_{[0,\tau]} Y \in \mathcal{G}$. By Lemma 18.0.12, there is a subsequence, still denoted as $\chi_{[0,\tau_k]} Y$ such that off a set of measure zero, 

$$\int_0^t \chi_{[0,\tau_k]} Y, dM^\tau \to \int_0^t \chi_{[0,\tau]} Y, dM^\tau$$

uniformly on $[0, T]$. Therefore, from the localization for elementary functions and this uniform convergence, 

$$\int_0^t \chi_{[0,\tau]} Y, dM^\tau = \lim_{n \to \infty} \int_0^t \chi_{[0,\tau_n]} Y, dM^\tau = \lim_{n \to \infty} \int_0^{t \wedge \tau_n} (Y, dM^\tau) = \int_0^{t \wedge \tau} (Y, dM^\tau)$$

This proves most of the following lemma.

**Lemma 18.0.15** Let $Y$ be an elementary function. Then if $\tau$ is any stopping time, then off a set of measure zero, 

$$\int_0^{t \wedge \tau} (Y, dM^\tau) = \int_0^t \chi_{[0,\tau]} Y, dM^\tau = \int_0^t (Y, dM^{\tau \wedge \tau})$$

**Proof:** It remains to prove the second equation.

$$\int_0^{t \wedge \tau} (Y, dM^\tau) = \sum_{i=0}^{m-1} (Y_i, M^\tau (t \wedge t_{i+1} \wedge \tau) - M^\tau (t \wedge t_i \wedge \tau)) = \sum_{i=0}^{m-1} (Y_i, M^{\tau \wedge \tau} (t \wedge t_{i+1}) - M^\tau (t \wedge t_i))$$

$$= \int_0^t (Y, dM^{\tau \wedge \tau}) \quad \blacksquare$$

**Lemma 18.0.16** Let $Y \in \mathcal{G}$. Then for any stopping time $\tau$, 

$$\int_0^{t \wedge \tau} (Y, dM^\tau) = \int_0^t \chi_{[0,\tau]} Y, dM^\tau = \int_0^t (Y, dM^{\tau \wedge \tau})$$

for $\omega$ off some set of measure zero.

**Proof:** From Lemma 18.0.12, there exists a sequence of elementary functions $Y^n$ such that $t \to \int_0^t (Y^n, dM)$ converges uniformly to $t \to \int_0^t (Y, dM^\tau)$ on $[0, T]$ for each $\omega \notin N$, a set of measure zero. Then 

$$\int_0^{t \wedge \tau} (Y, dM^\tau) = \lim_{n \to \infty} \int_0^{t \wedge \tau} (Y^n, dM^\tau)$$

$$= \lim_{n \to \infty} \int_0^t \chi_{[0,\tau]} Y^n, dM^\tau = \int_0^t \chi_{[0,\tau]} Y, dM^\tau$$
The last claim needs a little clarification. As shown in the above discussion proving Lemma 18.0.12, while $X_{[0,\tau]} Y^n$ is no longer obviously an elementary function due to the fact that $\tau$ has values which are not partition points, it is still the limit of a sequence of elementary functions $X_{[0,\tau_n]} Y^n$ and so the integral makes sense. Then from the inequality of Lemma 18.0.12,

$$E \left( \left| \int_0^t (X_{[0,\tau]} Y^n, dM^\tau) - \int_0^t (X_{[0,\tau]} Y, dM^\tau) \right|^2 \right) \leq E \left( \int_0^T ||Y^n - Y||_H^2 d[M]^\tau \right)$$

and so by the same Borel Cantelli argument of that lemma, there is a further subsequence for which the convergence is uniform off a set of measure zero as $n \to \infty$. (Actually, the same subsequence as in the first part of the argument works.) Therefore, the conclusion follows.

What of the second equation? Let $\{Y^n\}$ be as above where uniform convergence takes place for the stochastic integrals. Then from Lemma 18.0.15

$$\int_0^t (X_{[0,\tau]} Y^n, dM^\tau) = \int_0^t (Y^n, dM^{\tau_\omega})$$

Hence

$$E \left( \left| \int_0^t (Y^n, dM^{\tau_\omega}) - \int_0^t (X_{[0,\tau]} Y, dM^\tau) \right|^2 \right) \leq E \left( \int_0^T ||Y^n - Y||_H^2 d[M]^\tau \right)$$

Now by the usual application of the Borel Canelli lemma, there is a subsequence and a set of measure zero off which $\int_0^t (Y^n, dM^{\tau_\omega})$ converges uniformly to $\int_0^t (Y, dM^{\tau_\omega})$ on $[0, T]$ and as $n \to \infty$, and also

$$\int_0^t (Y^n, dM^{\tau_\omega}) \to \int_0^t (X_{[0,\tau]} Y, dM^\tau)$$

uniformly on $t \in [0, T]$. Then from the above,

$$\int_0^t (Y^n, dM^{\tau_\omega}) \to \int_0^t (X_{[0,\tau]} Y, dM^\tau) = \int_0^{t \land \tau} (Y, dM^\tau)$$

uniformly. Thus $\int_0^t (Y, dM^{\tau_\omega}) = \int_0^{t \land \tau} (Y, dM^\tau)$. 

**Definition 18.0.17** Let $\tau_p$ be an increasing sequence of stopping times for which $\lim_{p \to \infty} \tau_p = \infty$ and such that $M^\tau$ is a $L^2$ martingale and $X_{[0,\tau_p]} Y \in G$. Then the definition of $\int_0^t (Y, dM)$ is as follows. For each $\omega$,

$$\int_0^t (Y, dM) \equiv \lim_{p \to \infty} \int_0^t (X_{[0,\tau_p]} Y, dM^\tau)$$

In fact, this is well defined.

**Theorem 18.0.18** The above definition is well defined. Also this makes $\int_0^t (Y, dM)$ a local martingale. In particular,

$$\int_0^{t \land \tau} (Y, dM) = \int_0^t (X_{[0,\tau]} Y, dM^\tau)$$

In addition to this, if $\sigma$ is any stopping time,

$$\int_0^{t \land \sigma} (Y, dM) = \int_0^t (X_{[0,\sigma]} Y, dM)$$

In this last formula, $X_{[0,\sigma]} X_{[0,\tau]} Y \in G$. In addition, the following estimate holds for the quadratic variation.

$$\left\{ \int_0^t (Y, dM) \right\} (t) \leq \int_0^t ||Y||_H^2 d[M]$$
**Proof:** Suppose for some \( \omega, t < \tau_p < \tau_q \). Let \( \omega \) be such that both \( \tau_p, \tau_q \) are larger than \( t \). Then for all \( \omega \), and \( \tau \) a stopping time,
\[
\int_0^{t \land \tau} (X_{[0, \tau_q]} Y, dM^{\tau_p}) = \int_0^t (X_{[0, \tau_q]} Y, d((M^{\tau_p})^{\tau}))
\]
In particular, for the given \( \omega \),
\[
\int_0^t (X_{[0, \tau_q]} Y, d(M^{\tau_p})^\tau) = \int_0^t (X_{[0, \tau_q]} Y, dM^{\tau_p})
\]
For the particular \( \omega \), this equals
\[
\int_0^{t \land \tau_p} (X_{[0, \tau_q]} Y, dM^{\tau_p})
\]
Now for all \( \omega \) including the particular one, this equals
\[
\int_0^t (X_{[0, \tau_q]} Y, d((M^{\tau_p})^{\tau})) = \int_0^t (X_{[0, \tau_q]} Y, dM^{\tau_p})
\]
For the \( \omega \) of interest, this is
\[
\int_0^{t \land \tau_p} (X_{[0, \tau_q]} Y, dM^{\tau_p})
\]
and for all \( \omega \), including the one of interest, the above equals
\[
\int_0^t (X_{[0, \tau_p]} X_{[0, \tau_q]} Y, dM^{\tau_p}) = \int_0^t (X_{[0, \tau_q]} Y, dM^{\tau_p})
\]
thus for this particular \( \omega \), you get the same for both \( p \) and \( q \). Thus the definition is well defined because for a given \( \omega \), \( \int_0^t (X_{[0, \tau_p]} Y, dM^{\tau_p}) \) is constant for all \( p \) large enough.

Next consider the claim about this process being a local martingale. Is
\[
\int_0^{t \land \tau_p} (Y, dM)
\]
is a martingale? From the definition,
\[
\int_0^{t \land \tau_p} (Y, dM) = \lim_{q \to \infty} \int_0^{t \land \tau_p} (X_{[0, \tau_q]} Y, dM^{\tau_p})
\]
\[
= \lim_{q \to \infty} \int_0^t (X_{[0, \tau_q]} Y, d(M^{\tau_p})^\tau) = \lim_{q \to \infty} \int_0^{t \land \tau_p} (X_{[0, \tau_q]} Y, dM^{\tau_p})
\]
\[
= \lim_{q \to \infty} \int_0^t (X_{[0, \tau_p]} X_{[0, \tau_q]} Y, dM^{\tau_p}) = \int_0^t (X_{[0, \tau_q]} Y, dM^{\tau_p})
\]
which is known to be a martingale since \( X_{[0, \tau_q]} Y \in \mathcal{G} \). This is what it means to be a local martingale. You localize and get a martingale.

Next consider the claim about an arbitrary stopping time. Why is \( X_{[0, \sigma]} Y_{[0, \tau_q]} Y \in \mathcal{G} \)? This is part of a more general question. Suppose \( \dot{Y} \in \mathcal{G} \). Then why is \( X_{[0, \sigma]} \dot{Y} \in \mathcal{G} \). It suffices to show this. Let \( \{Y^n\} \) be the sequence of elementary functions which converge to \( \dot{Y} \) as in the definition. Also let \( \sigma_n \) be the stopping time with discrete values which equals \( t_{k+1}^n \) when \( \sigma \in (t_k^n, t_{k+1}^n] \). \( \{t_k^n\}_{k=0}^{n-1} \) being the partition associated with \( Y^n \). Then, as explained earlier, \( X_{[0, \sigma_n]} Y^n \) is an acceptable elementary function and also
\[
\left\{ E \left( \int_0^T \|X_{[0, \sigma_n]} Y^n - X_{[0, \sigma_n]} \dot{Y}\|^2 d[M] \right) \right\}^{1/2} \leq \left\{ E \left( \int_0^T \|X_{[0, \sigma_n]} Y^n - X_{[0, \sigma_n]} \dot{Y}\|^2 d[M] \right) \right\}^{1/2}
\]
\[
+ \left\{ E \left( \int_0^T \|X_{[0, \sigma_n]} \dot{Y}\|^2 d[M] \right) \right\}^{1/2}
\]
which converges to 0 from the definition of $\hat{Y} \in \mathcal{G}$ and the dominated convergence theorem. Thus $X_{[0,\tau]}\hat{Y} \in \mathcal{G}$.

From the above definition, for each $\omega$ off a suitable set of measure zero, from Lemma 18.0.16,

$$\int_{0}^{t \wedge \sigma} (Y, dM) \equiv \lim_{p \to \infty} \int_{0}^{t \wedge \sigma} (X_{[0,\tau_{p}]}Y, dM^{\tau_{p}}) = \lim_{p \to \infty} \int_{0}^{t} (X_{[0,\tau_{p}]}Y, dM^{\tau_{p}}) \equiv \int_{0}^{t} (X_{[0,\sigma]}Y, dM)$$

Finally, consider the claim about the quadratic variation. Using Lemma 18.0.16,

$$\left[ \left( \int_{0}^{(c)} (Y, dM) \right)^{\tau_{p}} (t) \right] = \left[ \left( \int_{0}^{(c)} (Y, dM) \right)^{\tau_{p}} \right] = \left[ \int_{0}^{(c)} (X_{[0,\tau_{p}]}Y, dM^{\tau_{p}}) \right] (t)$$

$$\leq \int_{0}^{t} \|X_{[0,\tau_{p}]}Y\|^2 d[M]^{\tau_{p}} \leq \int_{0}^{t} \|Y\|^2 d[M]$$

Now letting $\tau_{p} \to \infty$,

$$\left[ \left( \int_{0}^{(c)} (Y, dM) \right)^{\tau_{p}} \right] (t) \leq \int_{0}^{t} \|Y\|^2 d[M] \blacksquare$$

Next is the case in which $Y$ is continuous in $t$ but not necessarily bounded nor assumed to be in any kind of $L^2$ space either.

**Definition 18.0.19** Let $Y$ be continuous in $t$ and adapted. Let $M$ be a continuous local martingale $M(0) = 0$. Then the definition of a local martingale $\int_{0}^{t} (Y, dM)$ is as follows. Let $\tau_{p}$ be an increasing sequence of stopping times for which $[M]^{\tau_{p}}, \|M^{\tau_{p}}\|, \|X_{[0,\tau_{p}]}Y\|$ are all bounded by $p$. Then

$$\int_{0}^{t} (Y, dM) \equiv \lim_{p \to \infty} \int_{0}^{t} (X_{[0,\tau_{p}]}Y, dM^{\tau_{p}})$$

Then it is clear that $X_{[0,\tau_{p}]}Y \in \mathcal{G}$. Therefore, the above Theorem yields the following corollary.

**Corollary 18.0.20** The above definition is well defined. Also this makes $\int_{0}^{t} (Y, dM)$ a local martingale. In particular,

$$\int_{0}^{t \wedge \sigma} (Y, dM) = \int_{0}^{t} (X_{[0,\tau_{p}]}Y, dM^{\tau_{p}})$$

In addition to this, if $\sigma$ is any stopping time,

$$\int_{0}^{t \wedge \sigma} (Y, dM) = \int_{0}^{t} (X_{[0,\sigma]}Y, dM)$$

In this last formula, $X_{[0,\sigma]}Y$ has the same properties as $Y$, being the pointwise limit on $[0,T]$ of a bounded sequence of elementary functions for each $\omega$. In addition to this, there is an estimate for the quadratic variation

$$\left[ \int_{0}^{(c)} (Y, dM) \right] (t) \leq \int_{0}^{t} \|Y\|^2 d[M]$$

Of course there is no change in anything if $M$ has its values in a Hilbert space $W$ while $Y$ has its values in its dual space. Then one defines $\int_{0}^{t} (Y, dM)_{W',W}$ by analogy to the above for $Y$ an elementary function, step function which is adapted.

We use the following definition.

**Definition 18.0.21** Let $\tau_{p}$ be an increasing sequence of stopping times for which $M^{\tau_{p}}$ is a $L^2$ martingale. If $M$ is already an $L^2$ martingale, simply let $\tau_{p} \equiv \infty$. Let $\mathcal{G}$ denote those functions $Y$ which are adapted and for which there is a sequence of elementary functions $\{Y^{n}\}$ satisfying $\|Y^{n}(t)\|_{W}, M^{\tau} \in L^2(\Omega)$ for each $t$ with

$$\lim_{n \to \infty} E \left( \int_{0}^{T} \|Y - Y^{n}\|^2_{W',W} d[M]^{\tau_{p}} \right) = 0$$

for each $\tau_{p}$.
Then exactly the same arguments given above yield the following simple generalizations.

**Definition 18.0.22** Let \( Y \in \mathcal{G} \). Then
\[
\int_0^t \langle Y, dM^{\tau_p} \rangle_{W', W} \equiv \lim_{n \to \infty} \int_0^t \langle Y^n, dM^{\tau_p} \rangle_{W', W} \text{ in } L^2(\Omega)
\]

**Lemma 18.0.23** The above definition is well defined. Also, \( \int_0^t \langle Y, dM^{\tau_p} \rangle_{W', W} \) is a continuous martingale. The inequality
\[
E \left( \left( \int_0^t \langle Y, dM^{\tau_p} \rangle_{W', W} \right)^2 \right) \leq E \left( \int_0^t \|Y\|^2_{W}, d[M^{\tau_p}] \right)
\]
is also valid. For any sequence of elementary functions \( \{Y^n\}, \|Y^n(t)\|_{W}, M^* \in L^2(\Omega), \)
\[
\|Y^n - Y\|_{L^2(\Omega; L^2([0,T]; W', [dM^{\tau_p}]))) \to 0
\]
there exists a subsequence, still denoted as \( \{Y^n\} \) of elementary functions for which \( \int_0^t \langle Y^n, dM^{\tau_p} \rangle_{W', W} \) converges uniformly to \( \int_0^t \langle Y, dM^{\tau_p} \rangle_{W', W} \) on \([0,T]\) for \( \omega \) off some set of measure zero. In addition, the quadratic variation satisfies the following inequality.
\[
\left[ \int_0^t \langle Y, dM^{\tau_p} \rangle_{W', W} \right](t) \leq \int_0^t \|Y\|^2_{W}, d[M^{\tau_p}] \leq \int_0^t \|Y\|^2_{W'}, d[M]
\]
As before, you can consider the case where you only know \( \mathcal{X}_{[0, \tau_p]}Y \in \mathcal{G} \). This yields a local martingale as before.

**Definition 18.0.24** Let \( \tau_p \) be an increasing sequence of stopping times for which \( \lim_{p \to \infty} \tau_p = \infty \) and such that \( M^{\tau_p} \) is a martingale and \( \mathcal{X}_{[0, \tau_p]}Y \in \mathcal{G} \). Then the definition of \( \int_0^t \langle Y, dM \rangle_{W', W} \) is as follows. For each \( \omega \) off a set of measure zero,
\[
\int_0^t \langle Y, dM \rangle_{W', W} \equiv \lim_{p \to \infty} \int_0^t \langle \mathcal{X}_{[0, \tau_p]}Y, dM^{\tau_p} \rangle_{W', W}
\]
where \( \int_0^t \langle \mathcal{X}_{[0, \tau_p]}Y, dM^{\tau_p} \rangle_{W', W} \) is a martingale.

In fact, this is well defined.

**Theorem 18.0.25** The above definition is well defined. Also this makes \( \int_0^t \langle Y, dM \rangle_{W', W} \) a local martingale. In particular,
\[
\int_0^{t \land \tau_p} \langle Y, dM \rangle_{W', W} = \int_0^t \langle \mathcal{X}_{[0, \tau_p]}Y, dM^{\tau_p} \rangle_{W', W}
\]
In addition to this, if \( \sigma \) is any stopping time,
\[
\int_0^{t \land \sigma} \langle Y, dM \rangle_{W', W} = \int_0^t \langle \mathcal{X}_{[0, \sigma]}Y, dM \rangle_{W', W}
\]
In this last formula, \( \mathcal{X}_{[0, \sigma]}\mathcal{X}_{[0, \tau_p]}Y \in \mathcal{G} \). In addition, the following estimate holds for the quadratic variation.
\[
\left[ \int_0^t \langle Y, dM \rangle_{W', W} \right](t) \leq \int_0^t \|Y\|^2_{W'}, d[M]
\]
Note that from Definition 18.0.24 it is also true that
\[
\int_0^t \langle Y, dM \rangle_{W', W} \equiv \lim_{p \to \infty} \int_0^t \langle \mathcal{X}_{[0, \tau_p]}Y, dM^{\tau_p} \rangle_{W', W}
\]
in probability. In addition, since \( \tau_p \to \infty \), it follows that for each \( \omega \), eventually \( \tau_p > T \). Therefore, \( t \to \int_0^t \langle Y, dM \rangle_{W', W} \) is continuous, being equal to \( \int_0^t \langle \mathcal{X}_{[0, \tau_p]}Y, dM^{\tau_p} \rangle_{W', W} \) for that \( \omega \).
Chapter 19

A Different Kind Of Stochastic Integration

For more on this material, see [63] which is what this is based on.

19.1 Hermite Polynomials

Consider

$$\exp \left( tx - \frac{t^2}{2} \right) = \exp \left( \frac{x^2}{2} - \frac{1}{2} (x - t)^2 \right)$$

Now the Hermite polynomials are the coefficients of the power series of this function expanded in powers of $t$. Thus the $n^{th}$ one of these is

$$H_n(x) = \exp \left( \frac{x^2}{2} \right) \frac{1}{n!} \frac{d^n}{dt^n} \left( \exp \left( -\frac{1}{2} (x - t)^2 \right) \right) \bigg|_{t=0} \quad (19.1.1)$$

and

$$\exp \left( tx - \frac{t^2}{2} \right) = \sum_{n=0}^{\infty} H_n(x) t^n \quad (19.1.2)$$

Note that $H_0(x) = 1$,

$$H_1(x) = \exp \left( \frac{x^2}{2} \right) \frac{d}{dt} \left( \exp \left( -\frac{1}{2} (x - t)^2 \right) \right) \bigg|_{t=0} = -e^{-\frac{1}{2}(t-x)^2} e^{\frac{1}{2}x^2} (t-x) \bigg|_{t=0} = x$$

From (19.1.2), differentiating both sides formally with respect to $x$,

$$t \exp \left( tx - \frac{t^2}{2} \right) = \sum_{n=1}^{\infty} H'_n(x) t^n$$

and so

$$\sum_{n=0}^{\infty} H_n(x) t^n = \exp \left( tx - \frac{t^2}{2} \right) = \sum_{n=1}^{\infty} H'_n(x) t^n - 1 = \sum_{n=0}^{\infty} H'_{n+1}(x) t^n$$

showing that

$$H'_n(x) = H_{n-1}(x), \quad n \geq 1, \quad H_0(x) = 0, \quad H_1(x) = x$$

which could have been obtained with more work from (19.1.1). Also, differentiating both sides of (19.1.2) with respect to $t$,

$$- \exp \left( tx - \frac{t^2}{2} \right) (t-x) = \sum_{n=0}^{\infty} nH_n(x) t^{n-1}$$

Thus

$$(x-t) \sum_{n=0}^{\infty} H_n(x) t^n = \sum_{n=0}^{\infty} nH_n(x) t^{n-1} = \sum_{n=0}^{\infty} (n+1) H_{n+1}(x) t^n$$
and so
$$\sum_{n=0}^{\infty} xH_n(x) t^n - \sum_{n=0}^{\infty} H_n(x) t^{n+1} = \sum_{n=0}^{\infty} (n+1) H_{n+1}(x) t^n$$
and so
$$\sum_{n=0}^{\infty} xH_n(x) t^n - \sum_{n=1}^{\infty} H_{n-1}(x) t^n = \sum_{n=0}^{\infty} (n+1) H_{n+1}(x) t^n$$
Thus for \( n \geq 1 \),
$$xH_n(x) - H_{n-1}(x) = (n+1) H_{n+1}(x)$$
Now also
$$\exp\left( t(-x) - \frac{t^2}{2}\right) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} t^n$$
and taking successive derivatives with respect to \( t \) of the left side and evaluating at \( t = 0 \) yields
$$H_n(-x) = (-1)^n H_n(x).$$
Summarizing these as in Lemma 19.1.1,
$$H'_n(x) = H_{n-1}(x), \quad n \geq 1, \quad H_0(x) = 0, H_1(x) = x$$
$$xH_n(x) - H_{n-1}(x) = (n+1) H_{n+1}(x), \quad n \geq 1$$
$$H_n(-x) = (-1)^n H_n(x) \quad (19.1.3)$$
Clearly, these relations show that all of these \( H_n \) are polynomials. Also the degree of \( H_n(x) \) is \( n \) and the coefficient of \( x^n \) is \( 1/n! \).

**Lemma 19.1.1** Say \((X,Y)\) is generalized normally distributed and \( E(X) = E(Y) = 0, \ E(X^2) = E(Y^2) = 1. \) Then for \( m, n \geq 0 \),
$$E\left( H_n(X) H_m(Y) \right) = \begin{cases} 0 & \text{if } n \neq m \\ \frac{1}{m} \left( E(XY) \right)^n & \text{if } n = m \end{cases}$$

**Proof:** By assumption, \( sX + tY \) is normal distributed with mean 0. This follows from Theorem 11.8.4. Also
$$\sigma^2 \equiv E \left( (sX + tY)^2 \right) = s^2 + t^2 + 2E(XY)st$$
and so its characteristic function is
$$E \left( \exp(i \lambda (sX + tY)) \right) = \phi_{sX+tY}(\lambda) = e^{-\frac{1}{2} \sigma^2 \lambda^2} = e^{-\frac{1}{2} (s^2 + t^2) \lambda^2} e^{-E(XY)st \lambda^2}$$
So let \( \lambda = -i \). You can do this because both sides are analytic in \( \lambda \in \mathbb{C} \) and they are equal for real \( \lambda \), a set with a limit point. This leads to
$$E \left( \exp(sX + tY) \right) = e^{\frac{1}{2} (s^2 + t^2)} e^{E(XY)st}$$
Hence, multiplying both sides by \( e^{-\frac{1}{2} (s^2 + t^2)} \),
$$e^{\frac{1}{2} (s^2 + t^2)} E \left( \exp(sX + tY) \right) = E \left( \exp \left( sX - \frac{s^2}{2} \right) \exp \left( tY - \frac{t^2}{2} \right) \right)$$
$$= \exp(stE(XY))$$
Now take \( \frac{\partial^{n+m}}{\partial^n \partial^m} \) of both sides. Recall the description of the Hermite polynomials given above
$$n! H_n(x) = \frac{d^n}{dt^n} \exp\left( tx - \frac{t^2}{2}\right) |_{t=0}$$
Thus
$$E \left( n! H_n(X) m! H_m(Y) \right) = \frac{\partial^{n+m}}{\partial^n \partial^m} \exp(stE(XY)) |_{s=t=0}$$
Consider \( m < n \)
$$\frac{\partial^{n+m}}{\partial^n \partial^m} \exp(stE(XY)) = \frac{\partial^m}{\partial t^m} ((tE(Y))^n \exp(stE(XY)))$$
You have something like
\[ \frac{\partial^n}{\partial t^n} \left[ t^n \left( (E (XY))^n \exp \left( s t E (XY) \right) \right) \right] \]
and \( m < n \) so when you take partial derivatives with respect to \( t \), \( m \) times and set \( s, t = 0 \), you must have 0. Hence, if \( n > m \),
\[ E (n!H_n (X) m!H_m (Y)) = 0 \]
Similarly this equals 0 if \( m > n \). So assume \( m = n \). Then you will go through the same process just described but this time at the end you will have something of the form
\[ n!E (XY)^n + \text{terms multiplied by } s \text{ or } t \]
Hence, in this case,
\[ E (n!H_n (X) n!H_n (Y)) = n!E (XY)^n \]
and so
\[ E (H_n (X) H_n (Y)) = \frac{1}{n!} E (XY)^n \quad \blacksquare \]
Let \( W \) be the function defined above, \( W (h) \) is normally distributed with mean 0 and variance \( |h|^2 \) and
\[ E (W (h) W (g)) = (h, g)_H \]
Then from Lemma [M3.14, M3.15],
\[ E (H_n (W (h)) H_m (W (g))) = \begin{cases} 0 & \text{if } n \neq m \\ \frac{1}{n!} (E (W (h) W (g)))^n & \text{if } n = m \end{cases} \]
\[ = \begin{cases} 0 & \text{if } n \neq m \\ \frac{1}{n!} (h, g)_H^n & \text{if } n = m \end{cases} \]
This is a really neat result. From definition of \( W \);
\[ E \left( (W (h) W (g))^1 \right) = (h, g)_H \]
Note this is a special case of the above result because \( H_1 (x) = x \). However, we don’t know that \( E ((W (h) W (g))^n) \) is equal to something times \( (h, g)_H^n \) but we know that this is true of some \( n^{th} \) degree polynomials in \( W (h) \) and \( W (g) \).

**Definition 19.1.2** Let \( \mathcal{H}_n \equiv \text{span} \{ H_n (W (h)) : h \in H, |h|_H = 1 \} \).

Thus \( \mathcal{H}_n \) is a closed subspace of \( L^2 (\Omega, \mathcal{F}) \). Recall \( \mathcal{F} \equiv \sigma \{ W (h) : h \in H \} \). This subspace \( \mathcal{H}_n \) is called the Wiener chaos of order \( n \).

**Theorem 19.1.3** \( L^2 (\Omega, \mathcal{F}, P) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n \). The symbol denotes the infinite orthogonal sum of the closed subspaces \( \mathcal{H}_n \). That is, if \( f \in L^2 (\Omega) \), there exists \( f_n \in \mathcal{H}_n \) and constants such that \( f = \sum_n c_n f_n \) and if \( f \in \mathcal{H}_n, g \in \mathcal{H}_m \), then \( (f, g)_{L^2 (\Omega)} = 0 \).

**Proof:** Clearly each \( \mathcal{H}_n \) is a closed subspace. Also, if \( f \in \mathcal{H}_n \) and \( g \in \mathcal{H}_m \) for \( n \neq m \), what about \( (f, g)_{L^2 (\Omega)} \)?
\[ (f, g)_{L^2 (\Omega)} = \lim_{l \to \infty} \frac{1}{M_l} \left( \sum_{k=1}^{M_l} a_k H_n (W (h_k)) \right) \left( \sum_{j=1}^{M_p} a_j H_m (W (h_j)) \right) \]
\[ = \lim_{l \to \infty} \sum_{k,j} a_k a_j E (H_n (W (h_k)) H_m (W (h_j))) = 0 \]
Thus these are orthogonal subspaces. Clearly \( L^2 (\Omega) \supseteq \bigoplus \mathcal{H}_n \). Suppose \( X \) is orthogonal to each \( \mathcal{H}_n \). Is \( X = 0 \)? Each \( x^n \) can be obtained as a linear combination of the \( H_k (x) \) for \( k \leq n \). This is clear because the space of polynomials of degree \( n \) is of dimension \( n + 1 \) and \( \{ H_0 (x), H_1 (x), \ldots, H_n (x) \} \) is independent on \( \mathbb{R} \).

This is easily seen as follows. Suppose
\[ \sum_{k=0}^{n} c_k H_k (x) = 0 \]
and that not all $c_k = 0$. Let $m$ be the smallest index such that
\[ \sum_{k=0}^{m} c_k H_k(x) = 0 \]
with $c_m \neq 0$. Then just differentiate both sides and obtain
\[ \sum_{k=1}^{m} c_k H_{k-1}(x) = 0 \]
contradicting the choice of $m$.

Therefore, each $x^n$ is really a linear combination of the $H_k$ as claimed. Say
\[ x^n = \sum_{k=0}^{n} c_k H_k(x) \]
Then for $|h| = 1$,\[ W(h)^n = \sum_{k=0}^{n} c_k H_k(W(h)) \in \mathcal{H}_n \]
Hence $(X, W(h)^n)_{L^2(\Omega)} = 0$ whenever $|h| = 1$. It follows that for $h \in H$ arbitrary, and the fact that $W$ is linear,
\[ (X, W(h)^n)_{L^2} = \left( X, \left( |h| W \left( \frac{h}{|h|} \right) \right)^n \right) = |h|^n \left( X, W \left( \frac{h}{|h|} \right)^n \right) = 0 \]
Therefore, $X$ is perpendicular to $e^{W(h)}$ for every $h \in H$ and so from Lemma 16.5.3, $X = 0$. Thus $\oplus \mathcal{H}_n$ is dense in $L^2(\Omega)$.

Note that from Lemma 16.5.3, every polynomial in $W(h)$ is in $L^p(\Omega)$ for all $p > 1$. Now what is next is really tricky.

**Corollary 19.1.4** Let $\mathcal{P}_n$ denote all polynomials of the form
\[ p(W(h_1), \ldots, W(h_k)), \text{ degree of } p \leq n, \text{ some } h_1, \ldots, h_k \]
Also let $\mathcal{P}_n$ denote the closure in $L^2(\Omega, \mathcal{F}, P)$ of $\mathcal{P}_n^0$. Then
\[ \mathcal{P}_n = \oplus_{i=0}^{n} \mathcal{H}_i \]

**Proof:** It is obvious that $\mathcal{P}_n \supseteq \oplus_{i=0}^{n} \mathcal{H}_i$ because the thing on the right is just the closure of a set of polynomials of degree no more than $n$, a possibly smaller set than the polynomials used to determine $\mathcal{P}_n^0$ and hence $\mathcal{P}_n$. If $\mathcal{P}_n$ is orthogonal to $\mathcal{H}_m$ for all $m \geq n$, then you must have $\mathcal{P}_n \subseteq \oplus_{i=0}^{n} \mathcal{H}_i$. So consider $H_m(W(h))$. Recall that $H_m$ is the closure of the span of things like this for $|h|_H = 1$. Thus we need to consider
\[ E(p(W(h_1), \ldots, W(h_k)) H_m(W(h))), \text{ } |h|_H = 1, \]
and show that this is 0. Now here is the tricky part. Let $\{e_1, \ldots, e_s, h\}$ be an orthonormal basis for
\[ \text{span}(h_1, \ldots, h_k, h) \]
Then since $W$ is linear, there is a polynomial $q$ of degree no more than $n$ such that
\[ p(W(h_1), \ldots, W(h_k)) = q(W(e_1), \ldots, W(e_s), W(h)) \]
Then consider a term of $q(W(e_1), \ldots, W(e_s), W(h)) H_m(W(h))$
\[ aW(e_1)^{r_1} \cdots W(e_s)^{r_s} W(h)^{r} H_m(W(h)) \]
Now from Corollary 16.5.3 these random variables $\{W(e_1), \ldots, W(e_s), W(h)\}$ are independent due to the fact that the vector $(W(e_1), \ldots, W(e_s), W(h))$ is multivariate normally distributed and the covariance is diagonal. Therefore,
\[ E(aW(e_1)^{r_1} \cdots W(e_s)^{r_s} W(h)^{r} H_m(W(h))) \]
Now since \( r \leq n \), \( W (h)^r = \sum_{k=1}^{r} c_k H_k (W (h)) \) for some choice of scalars \( c_k \). By Lemma 19.1.4, this last term,
\[
E (W (h)^r H_m (W (h))) = \sum_k c_k E (H_k (W (h)) H_m (W (h))) = 0
\]
since each \( k < m \).

Note how remarkable this is. \( P_n^d \) includes all polynomials in \( W (h_1), \ldots, W (h_k) \) some \( h_1, \ldots, h_k \), of degree no more than \( n \), including those which have mixed terms but a typical thing in \( \oplus_{i=1}^{n} H_i \) is a sum of Hermite polynomials in \( W (h_k) \). It is not the case that you would have terms like \( W (h_1) W (h_2) \) as could happen in the case of \( P_n \).

Obviously it would be a good idea to obtain an orthonormal basis for \( L^2 (\Omega, \mathcal{F}, P) \). This is done next. Let \( \Lambda \) be the multiindices, \( (a_1, a_2, \ldots) \) each \( a_k \) a nonnegative integer. Also in the description of \( \Lambda \) assume that \( a_k = 0 \) for all \( k \) large enough. For such a multiindex \( a \in \Lambda \),
\[
a! \equiv \prod_{i=1}^{\infty} a_i!, \quad |a| \equiv \sum_{i} a_i
\]
Also for \( a \in \Lambda \), define
\[
H_a (x) \equiv \prod_{i=1}^{\infty} H_{a_i} (W (e_i)) \in L^2 (\Omega)
\]
This is well defined because \( H_0 (x) = 1 \) and all but finitely many terms of this infinite product are therefore equal to 1. Now let \( \{ e_i \} \) be an orthonormal basis for \( H \). For \( a \in \Lambda \),
\[
\Phi_a \equiv \sqrt{a!} \prod_{i=1}^{\infty} H_{a_i} (W (e_i)) \in L^2 (\Omega)
\]
Suppose \( a, b \in \Lambda \).
\[
\int_{\Omega} \Phi_a \Phi_b dP = \sqrt{a!} \sqrt{b!} \int_{\Omega} \prod_{i=1}^{\infty} H_{a_i} (W (e_i)) H_{b_i} (W (e_i)) dP
\]
Now recall from Corollary 19.5.1 the random variables \( \{ W (e_i) \} \) are independent. Therefore, the above equals
\[
\sqrt{a!} \sqrt{b!} \prod_{i=1}^{\infty} \int_{\Omega} H_{a_i} (W (e_i)) H_{b_i} (W (e_i)) dP = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}
\]
Thus \( \{ \Phi_a : a \in \Lambda \} \) is an orthonormal set in \( L^2 (\Omega) \).

**Lemma 19.1.5** If \( s_k \to h \), then for \( n \in \mathbb{N} \), there is a subsequence, still called \( s_k \) for which \( W (s_k)^n \to W (h)^n \) in \( L^2 (\Omega) \).

**Proof:** If \( s_k \to h \), does \( W (s_k)^n \to W (h)^n \) in \( L^2 (\Omega) \) for some subsequence? First of all,
\[
||W (h) - W (s_k)||_{L^2(\Omega)} = |s_k - h|^2 \to 0
\]
and so there is a subsequence, still called \( k \) such that \( W (s_k) (\omega) \to W (h) (\omega) \) for a.e. \( \omega \). Consider
\[
\int_{\Omega} |W (h) - W (s_k)|^2 dP \tag{19.1.4}
\]
Does this converge to 0? The integrand is bounded by \( 2 \left( W (h)^{2n} + W (s_k)^{2n} \right) \). Since \( W (h), W (s_k) \) are symmetric,
\[
\int_{\Omega} \left( 2 \left( W (h)^{2n} + W (s_k)^{2n} \right) \right)^2 dP \leq 8 \int_{\Omega} \left( W (h)^{4n} + W (s_k)^{4n} \right) dP
\]
\[
= 16 \int_{\Omega \cap \{ W (h) \geq 0 \}} e^{4nW (h)} dP + 16 \int_{\Omega \cap \{ W (s_k) \geq 0 \}} e^{4nW (s_k)} dP
\]
\[
\leq 16 \int_{\Omega} e^{4nW (h)} dP + 16 \int_{\Omega} e^{4nW (s_k)} dP
\]
\[
\leq 16 e^{\frac{4n}{2} |h|} + 16 e^{\frac{4n}{2} |s_k|}
\]
which is bounded independent of \( k \), the last step following from Lemma \([16.5.3]\). Therefore, the Vitali convergence theorem applies in \([19.1.3]\). ■

Given an \( h \in H \), let \( s_k = \sum_{j=1}^k (h, e_j) e_j \), the \( k^{th} \) partial sum in the Fourier series for \( h \).

\[ W(s_k)^m = \left( \sum_{j=1}^k (h, e_j) W(e_j) \right)^m = p(W(e_1), \cdots, W(e_k)) \]

where \( p \) is a homogeneous polynomial of degree \( m \). Now this equals

\[ q(H_0(W(e_1)), \cdots, H_0(W(e_k)) \cdots H_m(W(e_1)), \cdots, H_m(W(e_k))) \]

where \( q \) is a polynomial. This is because each \( W(e_j)^r \) is a linear combination of \( H_s(W(e_j)) \) for \( s \leq r \). Now you look at terms of this polynomial. They are all of the form \( c \Phi_a \) for some constant \( c \) and \( a \in A \). Therefore, if \( X \in L^2(\Omega) \), there is a subsequence, still denoted as \( \{s_k\} \) such that

\[ E(W(h)^n X) = \lim_{k \to \infty} E(W(s_k)^n X) \]

Now if \( X \) is orthogonal to each \( \Phi_a \), then for any \( h \) and \( n \), there is a subsequence still denoted with \( k \) such that

\[ E(W(h)^n X) = \lim_{k \to \infty} E(W(s_k)^n X) = 0 \]

It follows from Lemma \([16.5.4]\), the part about the convergence of the partial sums to \( e^{W(h)} \) that \( X \) is orthogonal to \( e^{W(h)} \) for any \( h \). Here are the details. From the lemma, for large \( n \),

\[ \left| E(e^{W(h)} X) - E\left( \sum_{j=0}^n \frac{W(h)^j}{j!} X \right) \right| < \varepsilon, \]

Also for large \( k \),

\[ \left| E\left( \sum_{j=0}^n \frac{W(h)^j}{j!} X \right) - E\left( \sum_{j=0}^n \frac{W(s_k)^j}{j!} X \right) \right| = \left| E\left( \sum_{j=0}^n \frac{W(h)^j}{j!} X \right) \right| < \varepsilon \]

Therefore,

\[ \left| E(e^{W(h)} X) \right| < 2\varepsilon \]

Since \( \varepsilon \) is arbitrary, this proves the desired result. By Lemma \([16.5.2]\), \( X = 0 \) and this shows that \( \{\Phi_a, a \in A\} \) is complete.

**Proposition 19.1.6** \( \{\Phi_a : a \in A\} \) is a complete orthonormal set for \( L^2(\Omega, F, P) \).

### 19.2 A Multiple Integral

Now in the above construction, from now on, unless indicated otherwise, \( H = L^2(T) \) where the measure is ordinary Lebesgue measure on \( T = [0, T] \) or \([0, \infty)\) or some other interval of time. However, it could be more general, but for the sake of simplicity let it be Lebesgue measure. Generalities appear to be nothing but identifying that which works in the case of Lebesgue measure. If \( \mu \ll m \) everything would work also. A careful description of what kind of measures work is in \([103]\). Also, for \( A \) a Borel set having finite Lebesgue measure,

\[ W(A) \equiv W(\chi_A), \]

This is a random variable, and as explained earlier, since any finite set of these is normally distributed, if all the sets are pairwise disjoint, the random variables are independent because the covariance is a diagonal matrix.

**Lemma 19.2.1** Let \( \{A_1, \cdots, A_m\} \) be pairwise disjoint sets in \( \mathcal{B}(T) \) each having finite measure. Then the products \( A_{i_1} \times \cdots \times A_{i_n} \) are pairwise disjoint. Also to say that the function

\[ \sum_i c_i \chi_{A_{i_1} \times \cdots \times A_{i_n}} \]

equals 0 whenever some \( t_j = t_i, i \neq j \) is to say that \( c_i = 0 \) whenever there is a repeated index in \( i \).
Proof: Suppose the condition that the $A_k$ are pairwise disjoint holds and consider two of these products, $A_{i_1} \times \cdots \times A_{i_n}$ and $A_{j_1} \times \cdots \times A_{j_m}$. If the two ordered lists $(i_1, \cdots , i_n), (j_1, \cdots , j_m)$ are different, then since the $A_k$ are disjoint the two products have empty intersection because they differ in some position.

Now suppose that $c_i = 0$ whenever there is a repeated index. Then the sum is taken over all permutations of $n$ things taken from $\{1, \cdots , m\}$ and so if some $t_r = t_s$ for $r \neq s$, all terms of the sum equal zero because $X_{A_{i_1} \times \cdots \times A_{i_n}} \neq 0$ only if $t \in A_{i_1} \times \cdots \times A_{i_n}$ and since $t_r = t_s$ and the sets $\{A_k\}$ are disjoint, there must be the same set in positions $r$ and $s$ so $c_i = 0$. Hence the function equals 0.

Conversely, suppose the sum $\sum_i c_i X_{A_{i_1} \times \cdots \times A_{i_n}}$ equals zero whenever some $t_r = t_s$ for $s \neq r$. Does it follow that $c_i = 0$ whenever some $t_r = t_s$? The value of this function at $t \in A_{i_1} \times \cdots \times A_{i_n}$ is $c_i$ because for any other ordered list of indices, the resulting product has empty intersection with $A_{i_1} \times \cdots \times A_{i_n}$. Thus, $t_r = t_s$, it is given that this function equals 0 which equals $c_i$.

This says that when you consider such a function $\sum_i c_i X_{A_{i_1} \times \cdots \times A_{i_n}}$ with the $A_k$ pairwise disjoint, then to say that it equals 0 whenever some $t_i = t_j$ is to say that it is really a sum over all permutations of $n$ indices taken from $\{1, \cdots , m\}$. Thus there are $C(n, m) m! = P(n, m)$ possible non zero terms in this sum.

Lemma 19.2.2 Consider the set of all ordered lists of $n$ indices from $\{1, 2, \cdots , m\}$. Thus two lists are the same if they consist of the same numbers in the same positions. We denote by $i$ or $j$ such an index, $i$ from $\{1, \cdots , m\}$ and $j$ from $\{1, \cdots , q\}$. Also let $A_1, \cdots , A_m, B_1, \cdots , B_q$ are two lists of pairwise disjoint Borel sets from $T$ having finite Lebesgue measure. Also suppose

$$\sum_i c_i X_{A_{i_1} \times \cdots \times A_{i_n}} = \sum_j d_j X_{B_{j_1} \times \cdots \times B_{j_m}}$$

Then

$$\sum_i c_i \prod_{k=1}^n W(A_{i_k}) = \sum_j d_j \prod_{k=1}^n W(B_{j_k})$$

Proof: Suppose that $n = 1$ first. Then you have

$$\sum_i c_i X_{A_i} = \sum_j d_j X_{B_j} \tag{19.2.5}$$

where the sets $\{A_i\}$ and $\{B_j\}$ are disjoint. Clearly

$$A_i \supseteq \cup_j A_i \cap B_j \tag{19.2.6}$$

Consider

$$c_i X_{A_i} = \sum_j c_j X_{A_i \cap B_j} \tag{19.2.7}$$

If strict inequality holds in $(19.2.6)$, then you must have a point in $A_i \setminus \cup_j A_i \cap B_j$ where the left side of $\sum_i c_i X_{A_i}$ equals $c_i$ but the right side would equal 0. Hence $c_i = 0$ and so $\sum_i c_i X_{A_i \cap B_j} = 0$ which shows that the two expressions in $(19.2.7)$ are equal. If $A_i = \cup_j A_i \cap B_j$, it is also true that the two expressions in $(19.2.7)$ are equal. Thus

$$\sum_i c_i X_{A_i} = \sum_i \sum_j c_j X_{A_i \cap B_j}$$

Similar considerations apply to the right side. Thus

$$\sum_j \sum_i c_i X_{A_i \cap B_j} = \sum_j \sum_i d_j X_{A_i \cap B_j}$$

$$\sum_{i,j} (c_i - d_j) X_{A_i \cap B_j} = 0$$

hence if $W(A_i \cap B_j) \neq 0$, then, since these sets are disjoint, $c_i - d_j = 0$. It follows that

$$\sum_{i,j} (c_i - d_j) W(A_i \cap B_j) = 0$$
and so
\[ \sum_i c_i W(A_i) = \sum_i \sum_j c_i W(A_i \cap B_j) = \sum_j \sum_i d_j W(A_i \cap B_j) = \sum_j d_j W(B_j) \]

This proves the theorem if \( n = 1 \). Consider the general case. Let \( i' \) be
\[ (i_1, \ldots, i_{n-1}), i_k \leq m \]
\[ \sum_{i_n=1}^m \sum_{i'} c(i', i_n) \mathcal{X}_{A_{i_n}}(t_n) \mathcal{X}_{A_1 \times \cdots \times A_{i_n-1}} = \sum_i c_i \mathcal{X}_{A_1 \times \cdots \times A_n} \]
\[ = \sum_j d_j \mathcal{X}_{B_1 \times \cdots \times B_{j_n}} = \sum_{j_n=1}^m \sum_{j'} d(j', j_n) \mathcal{X}_{B_{j_n}}(t_n) \mathcal{X}_{B_1 \times \cdots \times B_{j_n-1}} \]

Now pick \((t_1, \ldots, t_{n-1})\). The above is then
\[ \sum_{i_n=1}^m \left( \sum_{i'} c(i', i_n) \mathcal{X}_{A_i} \times \cdots \times A_{i_n-1} (t_1, \ldots, t_{n-1}) \right) \mathcal{X}_{A_{i_n}}(t_n) \]
\[ = \sum_{j_n=1}^m \left( \sum_{j'} d(j', j_n) \mathcal{X}_{B_i} \times \cdots \times B_{j_n-1} (t_1, \ldots, t_{n-1}) \right) \mathcal{X}_{B_{j_n}}(t_n) \]

and by what was just shown for \( n = 1 \), for each such choice,
\[ \sum_{i_n} \left( \sum_{i'} c(i', i_n) \mathcal{X}_{A_i} \times \cdots \times A_{i_n-1} \right) W(A_{i_n}) \]
\[ = \sum_{j_n} \left( \sum_{j'} d(j', j_n) \mathcal{X}_{B_i} \times \cdots \times B_{j_n-1} \right) W(B_{j_n}) \]

Then
\[ \sum_{i'} \left( \sum_{i_n} W(A_{i_n}) c(i', i_n) \right) \mathcal{X}_{A_1 \times \cdots \times A_{i_n-1}} = \]
\[ \sum_{j'} \left( \sum_{j_n} W(B_{j_n}) d(j', j_n) \right) \mathcal{X}_{B_1 \times \cdots \times B_{j_n-1}} \]

Pick \( \omega = \omega_0 \). Then by induction,
\[ \sum_{i'} \left( \sum_{i_n} W(A_{i_n}) (\omega_0) c(i', i_n) \right) W(A_1) \cdots W(A_{i_n-1}) \]
\[ = \sum_{j'} \left( \sum_{j_n} W(B_{j_n}) (\omega_0) d(j', j_n) \right) W(B_1) \cdots W(B_{j_n-1}) \]

and this reduces to what was to be shown because \( \omega_0 \) was arbitrary.

In what follows it will be assumed \( c_i = 0 \) if any two of the \( i_k \) are equal. That is
\[ \sum_i c_i \mathcal{X}_{A_1 \times \cdots \times A_n} (t_1, \cdots, t_n) = 0 \]

if any \( t_i = t_j \).

**Definition 19.2.3** Let \( \mathcal{E}_n \) be functions of the form
\[ f(t_1, \cdots, t_n) \equiv \sum_i c_i \mathcal{X}_{A_i} \times \cdots \times A_n (t_1, \cdots, t_n) \]
where the \( A_k \) come from some list of the form \( \{A_1, A_2, \ldots, A_n\} \) where this list of sets is pairwise disjoint, each \( A_k \neq \emptyset \) and \( c_i = 0 \) whenever two indices are equal. By Lemma 19.2.4 this is the same as saying that \( f = 0 \) if \( t_i = t_j \) for some \( i \neq j \). A function of \( n \) variables \( f \) is symmetric means that for \( \sigma \) a permutation,

\[
f( t_1, \cdots, t_n ) = f( t_{\sigma(1)}, \cdots, t_{\sigma(n)} )
\]

**Lemma 19.2.4** Let \( f(t_1, \cdots, t_n) = \sum c_i X_{A_i} (t_1, \cdots, t_n) \). Then \( f \) is symmetric if and only if for all \( \{c_i, \ldots, c_n\} \)

\[
c_{i_1, \ldots, i_n} = c_{i_{\sigma(1)}, \ldots, i_{\sigma(n)}},
\]

**Proof:** First of all, every \( c_i = 0 \) if there are repeated indices so it suffices to consider only the case where all indices are distinct.

Consider all the terms associated with a particular set of indices \( \{i_1, \ldots, i_n\} \). Then, since these sets \( A_{i_k} \) are disjoint, the function \( f \) is symmetric if and only if the part of the sum in the definition of \( f \) associated with each such set of indices is symmetric. To save on notation, denote such a list by \( \{1, 2, \cdots, n\} \). It suffices then to show that

\[
f( t_1, \cdots, t_n ) = \sum_{\sigma \in S_n} c_{\sigma(1) \cdots \sigma(n)} X_{A_{\sigma(1)} \times \cdots \times A_{\sigma(n)}} (t_1, \cdots, t_n)
\]

is symmetric if and only if for all \( \sigma, c_{\sigma(1) \cdots \sigma(n)} = c_{1 \cdots n} \). Suppose then that \( f \) is symmetric. Then

\[
f( t_{\beta(1)}, \cdots, t_{\beta(n)} ) = \sum_{\sigma \in S_n} c_{\sigma(1) \cdots \sigma(n)} X_{A_{\sigma(1)} \times \cdots \times A_{\sigma(n)}} (t_{\beta(1)}, \cdots, t_{\beta(n)})
\]

\[
= \sum_{\sigma \in S_n} c_{\sigma(1) \cdots \sigma(n)} X_{A_{\beta^{-1} \sigma(1)} \times \cdots \times A_{\beta^{-1} \sigma(n)}} (t_1, \cdots, t_n)
\]

(19.2.8)

However,

\[
\sum_{\sigma \in S_n} c_{\sigma(1) \cdots \sigma(n)} X_{A_{\sigma(1)} \times \cdots \times A_{\sigma(n)}} (t_{\beta(1)}, \cdots, t_{\beta(n)})
\]

\[
= \sum_{\sigma \in S_n} c_{\sigma(1) \cdots \sigma(n)} X_{A_{\beta^{-1} \sigma(1)} \times \cdots \times A_{\beta^{-1} \sigma(n)}} (t_1, \cdots, t_n)
\]

(19.2.9)

It is supposed to equal

\[
\sum_{\sigma \in S_n} c_{\sigma(1) \cdots \sigma(n)} X_{A_{\sigma(1)} \times \cdots \times A_{\sigma(n)}} (t_1, \cdots, t_n)
\]

\[
= \sum_{\sigma \in S_n} c_{\beta^{-1} \sigma(1) \cdots \beta^{-1} \sigma(n)} X_{A_{\beta^{-1} \sigma(1)} \times \cdots \times A_{\beta^{-1} \sigma(n)}} (t_1, \cdots, t_n)
\]

(19.2.10)

Thus

\[
\sum_{\sigma \in S_n} c_{\sigma(1) \cdots \sigma(n)} X_{A_{\beta^{-1} \sigma(1)} \times \cdots \times A_{\beta^{-1} \sigma(n)}} (t_1, \cdots, t_n)
\]

\[
= \sum_{\sigma \in S_n} c_{\beta^{-1} \sigma(1) \cdots \beta^{-1} \sigma(n)} X_{A_{\beta^{-1} \sigma(1)} \times \cdots \times A_{\beta^{-1} \sigma(n)}} (t_1, \cdots, t_n)
\]

Since the sets \( A_k \) are distinct, as explained above, this requires that

\[
X_{A_{\sigma(1)} \times \cdots \times A_{\sigma(n)}} \neq X_{A_{\beta^{-1} \sigma(1)} \times \cdots \times A_{\beta^{-1} \sigma(n)}}
\]

if \( \alpha \neq \sigma \). Therefore, \( \text{Lemma 19.2.4} \) requires that for all \( \beta \) and each \( \sigma \),

\[
c_{\sigma(1) \cdots \sigma(n)} = c_{\beta^{-1} \sigma(1) \cdots \beta^{-1} \sigma(n)}
\]

In particular, this is true if \( \beta = \sigma \) and so \( c_{\sigma(1) \cdots \sigma(n)} = c_{1 \cdots n} \).
The converse of this is also clear. If \( c_{\sigma(1)\ldots\sigma(n)} = c_{1\ldots n} \) for each \( \sigma \), then
\[
\sum_{\sigma \in S_n} c_{\sigma(1)\ldots\sigma(n)} X_{A_{\sigma(1)}\ldots A_{\sigma(n)}} (t_{\beta(1)}, \ldots, t_{\beta(n)})
\]
\[
= \sum_{\sigma \in S_n} c_{\sigma(1)\ldots\sigma(n)} X_{A_{\beta(1)}\ldots A_{\beta(n)}} (t_{\beta(1)}, \ldots, t_{\beta(n)})
\]
\[
= \sum_{\sigma \in S_n} c_{\sigma(1)\ldots\sigma(n)} X_{A_{\beta(1)}\ldots A_{\beta(n)}} (t_{\beta(1)}, \ldots, t_{\beta(n)})
\]
\[
= \sum_{\sigma \in S_n} c_{\sigma(1)\ldots\sigma(n)} X_{A_{\alpha(1)}\ldots A_{\alpha(n)}} (t_{1}, \ldots, t_{n})
\]
\[
= \sum_{\sigma \in S_n} c_{\alpha(1)\ldots\alpha(n)} X_{A_{\alpha(1)}\ldots A_{\alpha(n)}} (t_{1}, \ldots, t_{n})
\]
\[
= f (t_{1}, \ldots, t_{n}) \quad \blacksquare
\]

Observe that \( \mathcal{E}_n \) is a vector space because if you have two such functions
\[
\sum_{i} c_i X_{A_{i_{1}}\ldots A_{i_{n}}} (t_{1}, \ldots, t_{n}), \sum_{i} d_i X_{B_{i_{1}}\ldots B_{i_{n}}} (t_{1}, \ldots, t_{n})
\]
where the \( A_{i_{k}} \) are from \( \{ A_1, A_2, \ldots, A_m \} \) and the \( B_{i_{k}} \) are from \( \{ B_1, B_2, \ldots, B_q \} \). Then consider the single list consisting of the sets of the form \( A_k \cap B_j \). You could write each of these functions in terms of indicator functions of products of these disjoint sets. Thus the sum of the two functions can be written in the desired form. Since each of the elementary functions, \( c_i = 0 \), and the \( A_k \) are all disjoint,
\[
I_n (f) = \sum_{i} c_i W (A_{i_{1}}) \ldots W (A_{i_{n}})
\]

**Lemma 19.2.5** \( I_n \) is linear on \( \mathcal{E}_n \). If \( f \in \mathcal{E}_n \) and \( \sigma \) is a permutation of \( (1, \ldots, n) \) and
\[
f_{\sigma} (t_{1}, \ldots, t_{n}) \equiv f (t_{\sigma(1)}, \ldots, t_{\sigma(n)})
\]
so \( f \) is symmetric, then
\[
I_n (f_{\sigma}) = I_n (f)
\]
For \( f = \sum_{i} c_i X_{A_{i_{1}}\ldots A_{i_{n}}} \), one can conclude that
\[
I_n (f) = n! \sum_{i_{1} < i_{2} < \ldots < i_{n}} c_{i_{1}, \ldots, i_{n}} \prod_{k} W (A_{i_{k}})
\]
(19.2.11)

Also, the following holds for the expectation. For \( f, g \in \mathcal{E}_n, \mathcal{E}_m \) respectively,
\[
E (I_n (f) I_m (g)) = \begin{cases} 
0 & \text{if } n \neq m \\
\frac{1}{n!} (\bar{f}, \bar{g})_{L^2 (\mathcal{F}^n)} & \text{if } n = m
\end{cases}
\]
where \( \bar{f} \) denotes the symmetrization of \( f \) given by
\[
\bar{f} (t_{1}, \ldots, t_{n}) = \frac{1}{n!} \sum_{\sigma \in S_n} f (t_{\sigma(1)}, \ldots, t_{\sigma(n)})
\]
Proof: It is clear from the definition being well defined that $I_n$ is linear. In particular, consider

$$I_n \left( a \sum_i c_i \chi_{A_{i_1} \times \cdots \times A_{i_n}} + b \sum_j d_j \chi_{B_{j_1} \times \cdots \times B_{j_n}} \right).$$

As explained above in the observation that $E_n$ is a vector space, it can be assumed that the $A_{i_k}$ and $B_{j_k}$ are all from a single set of disjoint Borel sets of $T$. Then the above is of the form

$$a \sum_i c_i \prod_k W( \chi_{A_{i_k}} ) + b \sum_j d_j \prod_k W( \chi_{B_{j_k}} )$$

$$= a I_n \left( \sum_i c_i \chi_{A_{i_1} \times \cdots \times A_{i_n}} \right) + b I_n \left( \sum_j d_j \chi_{B_{j_1} \times \cdots \times B_{j_n}} \right)$$

Next consider

$$f_\sigma (t_1, \cdots, t_n) = \sum_i c_i \chi_{A_{i_1} \times \cdots \times A_{i_n}} (t_{\sigma(1)}, \cdots, t_{\sigma(n)})$$

$$= \sum_i c_i \chi_{A_{i_1-1(1)} \times \cdots \times A_{i_n-1(n)}} (t_1, \cdots, t_n)$$

Thus, it appears that $f_\sigma \neq f$. However,

$$I_n (f_\sigma) = \sum_i c_i \prod_{k=1}^n W( \chi_{A_{i_{\sigma(k)}}} ) = I_n (f)$$

because one just considers the factors in a different order than the other. Let

$$\bar{f} (t_1, \cdots, t_n) = \frac{1}{n!} \sum_\sigma f_\sigma (t_1, \cdots, t_n)$$

Then $I_n (\bar{f}) = I_n (f)$ and $\bar{f} (t_{\sigma(1)}, \cdots, t_{\sigma(n)}) = \bar{f} (t_1, \cdots, t_n)$. This is called the symetrization of $f$. If

$$f (t_{\sigma(1)}, \cdots, t_{\sigma(n)}) = f (t_1, \cdots, t_n)$$

then $\bar{f} = f$. From the above, $\bar{f}$ equals

$$\frac{1}{n!} \sum_\sigma \sum_i c_i \chi_{A_{i_{\sigma(1)}} \times \cdots \times A_{i_{\sigma(n)}}} (t_1, \cdots, t_n)$$

Note that implies that

$$I_n (f) = I_n (\bar{f}) = n! \sum_{i_1 < i_2 < \cdots < i_n} c_{i_1, \cdots, i_n} \prod_k W( \chi_{A_{i_k}} )$$

Now consider

$$\bar{f} = \sum_i c_i \chi_{A_{i_1} \times \cdots \times A_{i_n}}$$

and $\bar{g} = \sum_i d_i \chi_{A_{i_1} \times \cdots \times A_{i_n}}$

where without loss of generality, these sets $A_k$ come from a single list of disjoint sets.

Consider

$$E (I_n (f) I_n (g)) = E (I_n (\bar{f}) I_n (\bar{g})).$$

From the above, $E (I_n (\bar{f}) I_n (\bar{g})) =$

$$E \left( (n!)^2 \sum_{i_1 < \cdots < i_n} c_{i_1, \cdots, i_n} \prod_k W( \chi_{A_{i_k}} ) \prod_l W( \chi_{A_{j_l}} ) \right)$$

$$= (n!)^2 \sum_{i_1 < \cdots < i_n} c_{i_1, \cdots, i_n} \prod_k W( \chi_{A_{i_k}} ) \prod_l W( \chi_{A_{j_l}} )$$

$$= (n!)^2 \sum_{i_1 < \cdots < i_n} c_{i_1, \cdots, i_n} \prod_k W( \chi_{A_{i_k}} ) W( \chi_{A_{j_k}} )$$

(19.2.14)
That product is of independent random variables. Recall any collection of the \( W(A_k) \) are normally distributed and also the covariance is diagonal and so these will all be independent random variables. If any one of them is not repeated, say \( W(A_{i_k}) \), then

\[
E \left( \prod_k W(A_{i_k}) W(A_{j_k}) \right) = E \left( W(A_{i_k}) \right) (\text{stuff}) = 0
\]

It follows that to get something nonzero out of this, all \( A_{i_k} \) are repeated. That is, you must have \( j = i \) and \( n \) reduces to \( E(I_n(f) I_n(g)) = \)

\[
(nt)^2 \sum_{i_1 < \cdots < i_n} c_{i_1, \ldots, i_n} d_{i_1, \ldots, i_n} E \left( \prod_k W(A_{i_k})^2 \right) = \sum_{i_1 < \cdots < i_n} c_{i_1, \ldots, i_n} d_{i_1, \ldots, i_n} \prod_k m(A_{i_k})
\]

By Lemma 19.2.14, used at the end of the following string of equalities, and the observation that

\[
X_{A_{i_1} \times \cdots \times A_{i_n}} X_{A_{j_1} \times \cdots \times A_{j_n}} = 0
\]
to eliminate mixed terms,

\[
(f, \bar{g})_{L^2(T^n)} = \int_0^\infty \cdots \int_0^\infty \left( \sum_i c_i X_{A_{i_1} \times \cdots \times A_{i_n}} \right) \left( \sum_i d_i X_{A_{i_1} \times \cdots \times A_{i_n}} \right) dt \cdots dt
\]

\[
= \int_0^\infty \cdots \int_0^\infty \left( \sum_i c_i d_i X_{A_{i_1} \times \cdots \times A_{i_n}} \right) dt \cdots dt
\]

\[
= \sum_i c_i d_i \prod_k m(A_{i_k}) = n! \sum_{i_1 < \cdots < i_n} c_{i_1, \ldots, i_n} d_{i_1, \ldots, i_n} \prod_k m(A_{i_k})
\]

Now it follows from this and \( L^2(T^n) \) that

\[
E(I_n(f) I_n(g)) = n! (f, \bar{g})_{L^2(T^n)}
\]

What happens if you consider \( E(I_n(f) I_m(g)) \) where \( m < n \)? You would still get

\[
E(I_n(f) I_m(g)) = E(I_n(f) I_m(g)) = E \left( \prod_k W(A_{i_k}) \right)
\]

Then at least one of the \( W(A_{i_k}) \) is not repeated. This is because \( n > m \). That product is a product of independent random variables at least one of which is of the form \( W(A_{i_k}) \). Hence when you take the expectation of the product it is of the form \( E(W(A_{i_k})) \) (Other terms) = 0. Thus if \( n \neq m \), the result is 0 as claimed.

An integral has now been defined on the functions of the form

\[
f(t_1, \cdots, t_n) \equiv \sum_i c_i X_{A_{i_1} \times \cdots \times A_{i_n}} (t_1, \cdots, t_n)
\]

which equals 0 if the \( A_k \) are disjoint and \( f = 0 \) if any \( t_i = t_j \) for \( i \neq j \). This integral defined on these elementary functions is interesting because for such functions \( f, g \)

\[
E(I_n(f) I_m(g)) = \begin{cases} 0 & \text{if } n \neq m \\ n! (f, \bar{g})_{L^2(T^n)} & \text{if } n = m \end{cases}
\]

where \( \bar{f} \) is the symmetrization of \( f \). It is desired to extend this integral to \( L^2(T^n) \). Simple functions are always dense in \( L^2(T^n) \). Also, there is an easy lemma which can be concluded for \( L^2(T^n) \).
Lemma 19.2.6 Let $\mathcal{B}_0(T)$ be the Borel sets having finite measure. Linear combinations of functions of the form

$$\mathcal{X}_{A_1 \times \cdots \times A_n}$$

where $A_i \in \mathcal{B}_0(T)$ are dense in $L^2(T, \mathcal{B}^n)$ where of course $\mathcal{B}^n$ refers to the product $\sigma$ algebra.

Proof: If you have $U = A_1 \times \cdots \times A_n$ in $T^n$ one can approximate $\mathcal{X}_{U \cap R_p}$ for $R_p \equiv (-p, p)^n$ in $L^2$ with linear combinations of sets of the desired form. In fact, you just consider $\mathcal{X}_{A_1 \cap (-p, p) \times \cdots \times A_n(-p, p)}$ and you get equality. Now let $\mathcal{K}$ denote the $\pi$ system of sets of this sort. Let $\mathcal{G}$ denote those Borel sets $G$ such that there exists a sequence of linear combinations of sets of the form $\mathcal{X}_A, A = A_1 \times \cdots \times A_n$ which converges to $\mathcal{X}_{U \cap R_p}$ in $L^2(T^n)$. Thus $\mathcal{G} \supseteq \mathcal{K}$.

Let $\{G_k\}$ be a disjoint sequence of sets of $\mathcal{G}$. Is $G \equiv \bigcup_k G_k \in \mathcal{G}$? By monotone convergence theorem,

$$\left\| \mathcal{X}_{G \cap R_p} - \sum_{k=1}^m \mathcal{X}_{G_k \cap R_p} \right\|_{L^2(T^n)} < \varepsilon$$

provided $m$ is large enough. Now by definition of $\mathcal{G}$ there exists $L_k$ a linear combination of these special sets such that

$$\left\| \mathcal{X}_{G_k \cap R_p} - L_k \right\|_{L^2(T^n)} < \frac{\varepsilon}{m}$$

It follows that

$$\left\| \mathcal{X}_{G \cap R_p} - \sum_{k=1}^m L_k \right\|_{L^2} \leq \left\| \mathcal{X}_{G \cap R_p} - \sum_{k=1}^m \mathcal{X}_{G_k \cap R_p} \right\|_{L^2}$$

$$+ \left\| \sum_{k=1}^m \mathcal{X}_{G_k \cap R_p} - \sum_{k=1}^m L_k \right\| < \varepsilon + \varepsilon$$

and so it follows that $G \in \mathcal{G}$. If $G \in \mathcal{G}$, does it follow that $G^C$ is also?

$$\mathcal{X}_{R_p} = \mathcal{X}_{R_p \cap G} + \mathcal{X}_{R_p \cap G^C}$$

Hence

$$\mathcal{X}_{R_p} - \mathcal{X}_{R_p \cap G} = \mathcal{X}_{R_p \cap G^C}$$

Both of the functions on the left can be approximated in $L^2$ by the desired kind of functions and so the one on the right can also. It follows from Dynkin’s lemma that $\mathcal{G} \equiv \sigma(\mathcal{K})$ which is the product measurable sets. Thus if $U$ is any set in $\mathcal{B}^n$, it follows that $\mathcal{X}_U$ can be approximated in $L^2(T^n)$ with linear combinations of sets like $\mathcal{X}_{A_1 \times \cdots \times A_n}$. ■

Of course nothing is known about whether the sets $A_i$ are disjoint. Also it is not known whether these linear combinations of these functions equals 0 if $t_i = t_j$. Thus there is something which needs to be proved.

Lemma 19.2.7 The functions in $\mathcal{E}_n$ mentioned above are dense in $L^2(T^n)$.

Proof: From Lemma 19.2.6, it suffices to show that $\mathcal{X}_{A_1 \times \cdots \times A_n}$ can be approximated in $L^2(T^n)$ with functions in $\mathcal{E}_n$. This is where it will be important that the measure is sufficiently like Lebesgue measure. Let $\{B_k\}_{k=1}^m$ be a partition of $A_i$ such that $m(B_k) \leq \frac{2m(A_i)}{m}$. Let $\{B_k\}_{k=1}^m$ denote all these sets so $p = mn$. They are not necessarily disjoint. However, one can say that it is possible to choose $\varepsilon_i$ equal to either 0 or 1 such that

$$\mathcal{X}_{A_1 \times \cdots \times A_n} = \sum_{i} \varepsilon_i \mathcal{X}_{B_i^1 \times \cdots \times B_i^n}$$

where we can have $B_{ik} \subseteq A_k$. Let $J$ be those indices $i$ which involve a repeated set. That is some $B_{ij} = B_{ik}$ for some $j \neq k$. How many possibilities are there? There are no more than $C(n, 2) m, C(n, 2)$ possibilities for duplicates among the $A_k$ and then there are $m$ sets in the partition of $A_k$.

$$\int_T \cdots \int_T \left( \sum_{i \in J} \varepsilon_i \mathcal{X}_{B_1^i \times \cdots \times B_n^i} \right)^2 dt \cdots dt$$

$$= \int_T \cdots \int_T C(n, 2) m \mathcal{X}_{B_1^i \times \cdots \times B_n^i} dt \cdots dt$$

$$\leq C(n, 2) m \prod_{k=1}^m m(B_{ik})$$
The mixed terms are 0 because for a fixed \( k, \{B_{i_k}\}_{i=1}^m \) are disjoint. Now from the description of these, \( m (B_{i_k}) m < m (A_k) \) and so

\[
\int_T \cdots \int_T \left( \sum_{i \in J} e_i X_{B_{i_1} \times \cdots \times B_{i_n}} \right)^2 \, dt \cdots dt
\]

\[
\leq C (n, 2) m \prod_{k=1}^n \frac{2 m (A_k)}{m} = C (n, 2) m \prod_{k=1}^n m (A_k)
\]

which clearly converges to 0 as \( m \to \infty \) provided that \( n \geq 2 \). In case \( n = 1 \), all you have to do is approximate \( X_A \) from something in \( E_1 \) and of course you just use \( X_A \). ■

Let \( f, g \in E_1 \). Then from Lemma 19.2.7,

\[
E \left( (I_n (f - g))^2 \right) = n! \| f - g \|_{L^2(T^n)}^2
\]

\[
\| f \|_{L^2(T^n)} = \left( \int_T \cdots \int_T |f(t)|^2 \, dt \right)^{1/2}
\]

\[
= \left( \int_T \cdots \int_T \frac{1}{n!} \sum_{\sigma} f(t_{\sigma(1)}, \cdots, t_{\sigma(n)})^2 \, dt \right)^{1/2}
\]

\[
\leq \frac{1}{n!} \sum_{\sigma} \left( \int_T \cdots \int_T |f(t_{\sigma(1)}, \cdots, t_{\sigma(n)})|^2 \, dt \right)^{1/2}
\]

\[
= \frac{1}{n!} \sum_{\sigma} \| f \|_{L^2(T^n)} = \| f \|_{L^2(T^n)}
\]

Therefore,

\[
E \left( (I_n (f - g))^2 \right) = n! \| f - g \|_{L^2(T^n)}^2 \leq n! \| f - g \|_{L^2(T^n)}^2.
\]

The following theorem comes right away from this and Lemma 19.2.7.

**Theorem 19.2.8** The integral \( I_n \) defined on \( E_n \) extends uniquely to an integral \( I_n \) defined on \( L^2(T^n) \). This integral satisfies

\[
I_n (f) \in L^2(\Omega)
\]

Also

\[
E \left( I_n (f) I_n (g) \right) = \langle f, g \rangle_{L^2(T^n)}
\]

**Proof:** This follows right away from the density of \( E_n \) in \( L^2(T^n) \) and the inequality 19.2.7. ■

Obviously one wonders whether linear combinations \( \sum_{i} c_i I_n (f_i) \) are dense in \( L^2(\Omega) \). It looks like the important thing to notice is that for \( f \in E_n, I_n (f) \) is a polynomial in \( W(A_{i_k}) \equiv W(\mathcal{X}_{A_{i_k}}) \). Recall the corollary above, Corollary 19.2.6.

**Corollary 19.2.9** Let \( P_n^0 \) denote all polynomials of the form

\[
p(W(h_1), \cdots, W(h_k)), \quad \text{degree of } p \leq n, \text{ some } h_1, \cdots, h_k
\]

Also let \( \mathcal{P}_n \) denote the closure in \( L^2(\Omega, \mathcal{F}, P) \) of \( P_n^0 \). Then

\[
\mathcal{P}_n = \bigoplus_{i=0}^n \mathcal{H}_i
\]

Consider \( \cup_{p \leq n} \{ I_p (f) : f \in E_p \} \). This is a subset of \( P_n^0 \) and so it is a subset of \( \bigoplus_{i=0}^n \mathcal{H}_i \). Now for \( h \in L^2(T^n) \), it was shown above that there exists a sequence \( g_k \to h \) in \( L^2(T^n) \) where each \( h_k \in E_n \). Then \( I_n (g_k) \to I_n (h) \). In particular, if \( h \in L^2(T) \equiv H \), then there is a sequence \( g_k \in E_1 \) such that \( g_k \to h \) in \( L^2(T) \). Then clearly

\[
E \left( |W(g_k) - W(h)|^2 \right) = E \left( |W(g_k - h)|^2 \right) \to 0
\]
and so each polynomial \( p(W(h_1), \cdots, W(h_k)) \) can be approximated in \( L^2(\Omega) \) by one which is of the form
\[
p(W(g_1), \cdots, W(g_k))
\]
where each \( g_j \in \mathcal{E}_1 \). Corresponding to each \( g_j \) there is a list of disjoint sets. Now consider the union of all the sets just described and let \( \{A_k\} \) be a partition of this union such that the \( A_k \) are pairwise disjoint and for each \( j \), every set corresponding to \( g_j \) is partitioned by a subset of the \( \{A_k\} \). Thus
\[
g_j = \sum_i c_i \chi_{B_i} = \sum_i c_i \sum_{s=1}^{m_j} \chi_{A^s_i}
\]
where \( B_i \) is partitioned by the \( A^s_i \). Then consider \( p(g_1, \cdots, g_k) \). Then the terms of degree \( m \) are of the form
\[
p_m \equiv \sum_i c_i \chi_{A_{1i} \times \cdots \times A_{im}}
\]
(19.2.17)
where the \( A_{ik} \) come from the list of disjoint sets \( \{A_k\} \). The terms of degree \( m \) in \( p(W(g_1), \cdots, W(g_k)) \) are also of the form
\[
p_m(W(g_1), \cdots, W(g_k)) \equiv \sum_i c_i \prod_k W(A_{ik})
\]
The problem is that \( \mathcal{E}_m \) is not in \( \mathcal{E}_m \) because it is not known whether \( c_i = 0 \) if two indices are repeated. However, as explained in the proof of Lemma \( \mathcal{E}_m \) there is a further partition such that the contribution of those terms corresponding to \( I \) in which two indices are repeated can be made as small as desired. Therefore, the terms of order \( m \) are approximated in \( L^2(T^n) \) by \( g_m \in \mathcal{E}_m \). Assume this approximation is good enough that, from the estimates given above in Lemma \( \mathcal{E}_m \),
\[
E \left( |I_m(g_m) - p_m(W(g_1), \cdots, W(g_k))|^2 \right)^{1/2} < \frac{\varepsilon}{n+1}
\]
Thus, taking a succession of partitions if necessary,
\[
E \left( \left| p(W(g_1), \cdots, W(g_k)) - \sum_{m=0}^{n} I_m(g_m) \right|^2 \right)^{1/2} \leq \sum_{m=1}^{n} E \left( |I_m(g_m) - p_m(W(g_1), \cdots, W(g_k))|^2 \right)^{1/2} < \sum_{m=1}^{n} \frac{\varepsilon}{n+1} < \varepsilon.
\]
This has proved the following lemma.

**Lemma 19.2.10** Let \( n \) be given. Then \( \cup_{p \leq n} \{I_p(f) : f \in \mathcal{E}_p\} \) is dense in \( \mathcal{P}_n = \oplus_{i=0}^{n} \mathcal{H}_i \). Consequently, every \( f \in L^2(\Omega, \mathcal{F}) \) may be written as an infinite sum
\[
f = \sum_{k=1}^{\infty} c_k I_k(g_k)
\]
where \( g_k \in \mathcal{E}_k \) and it can also be assumed that \( g_k \) is symmetric.

**Proof:** It only remains to verify that \( g_k \) can be symmetric. However, this is obvious because if \( g_k \) is replaced with \( \bar{g}_k \), the integral \( I_k \) is unchanged. \( \blacksquare \)

### 19.3 The Skorokhod Integral

This integral allows for one to obtain a stochastic integral of functions which are not adapted. It is a generalization of the Itô integral. There is also a strange sort of derivative which can be defined and the two are related in a natural way.
19.3.1 The Derivative

Let $F: \mathbb{R}^n \to \mathbb{R}$ be smooth and have polynomial growth. Then consider

$$F(W(h_1), \ldots, W(h_n))$$

where $W$ is defined above. Recall that $h \in H$ a separable real Hilbert space and $W(h) \in L^2(\Omega, \mathcal{F}, P)$ where $\mathcal{F} = \sigma(W(h), h \in H)$. Also $(W(h_1), \cdots, W(h_n))$ is multivariate normal and $E(W(g)W(h)) = (h,g)_H$.

**Definition 19.3.1** In the above situation,

$$DF \equiv \sum_{k=1}^{n} D_{k} F(W(h_1), \cdots, W(h_n)) h_k$$

Thus from Lemma [16.5.7], $F$, $D_kF$ are in $L^p(\Omega)$ and so $DF$ is in $L^p(\Omega; H)$ for every $p$.

First it is good to consider whether $DF$ is well defined.

**Lemma 19.3.2** The derivative is well defined. Also, if $F(W(h_1), \cdots, W(h_n)) = 0$ for $\{h_1, \cdots, h_n\}$ independent, then for all $x, F(x) = 0$.

**Proof:** Suppose

$$F(W(h_1), \cdots, W(h_n)) = 0.$$ 

Is it true that $DF = 0$? Let $\lambda$ be the distribution measure of $(W(h_1), \cdots, W(h_n)) \equiv W(h)$. Then the above requires that for any ball $B$ in $\mathbb{R}^n$,

$$E\left(\chi_B(W(h))F^2(W(h))\right) = \int_B F^2(x)\, d\lambda(x) = 0$$

If $\{h_1, \cdots, h_n\}$ is independent, then $\lambda$ has a normal density function and $\lambda \ll m_n$ and so $F^2(x) = 0$ for a.e. $x$. Since $F$ is smooth, this means that $F = 0$ everywhere. Hence $D_kF = 0$ and so $DF = 0$. Thus the case where the $h_i$ are independent is easy.

Next suppose without loss of generality that a basis for

$$\{h_1, \cdots, h_n\}$$

is

$$\{h_1, \cdots, h_r\}$$

where $r < n$. Say $h_k = \sum_{i=1}^{r} c_i^k h_i$ for $k > r$. Then

\[
0 = F\left(W(h_1), \cdots W(h_r), W\left(\sum_{i=1}^{r} c_i^{r+1} h_i\right) \cdots W\left(\sum_{i=1}^{n} c_i^{n} h_i\right)\right) = F\left(W(h_1), \cdots W(h_r), \sum_{j=1}^{r} c_j^{r+1} W(h_j) \cdots \sum_{j=1}^{r} c_j^{n} W(h_j)\right) = G(W(h_1), \cdots W(h_r))
\]

and so in terms of $\{h_1, \cdots, h_r\}$,

\[
DF = \sum_{i=1}^{r} (D_i F) h_i + \sum_{i=r+1}^{n} (D_i F) \sum_{j=1}^{r} c_j^i h_j = \sum_{j=1}^{r} (D_j F) h_j + \sum_{j=1}^{r} \sum_{i=r+1}^{n} (D_i F) c_j^i h_j = \sum_{j=1}^{r} \left(D_j F + \sum_{i=r+1}^{n} (D_i F) c_j^i \right) h_j
\]
Now it was just shown that $G(x)$ is identically 0 and so $D_jG = 0$, $j \leq r$. So what is $D_jG$? From the above, it equals

$$D_jF + \sum_{i=r+1}^{n} (D_iF)c_i^j = 0$$

Hence $DF = 0$. Now if $F(W(h_1), \ldots, W(h_n)) = G(W(k_1), \ldots, W(k_m))$, then $F - G = 0$ and so from what was just shown, $D(F - G) = DF - DG = 0$. Thus the derivative is well defined. □

**Lemma 19.3.3** Let $\mathcal{P}$ denote the set of all polynomials in $W(h)$ for $h \in H$. Then $\mathcal{P}$ is dense in $L^p(\Omega)$.

**Proof:** Let $g \in L^p'(\Omega)$ and suppose that for every $f \in D$, $\int_\Omega gfdP = 0$. Does it follow that $g = 0$? If so, then by the Riesz representation theorem, $\mathcal{P}$ is dense in $L^p(\Omega)$. From Lemma [Lemma 19.3.3](#), for a given $h$, there is a sequence of functions of $\mathcal{P}, \{f_n\}$ which converges to $e^{W(h)}$ in $L^p(\Omega)$. It follows that

$$\int_\Omega g e^{W(h)}dP = \lim_{n \to \infty} \int_\Omega g f_n dP = 0$$

Hence by Lemma [Lemma 19.3.3](#) it follows that $g = 0$. Hence $\mathcal{P}$ is dense in $L^p(\Omega)$. □

Let $D^{1,p}$ denote the closure in $L^p(\Omega)$ of functions in $\mathcal{P}$ with respect to the seminorm

$$\|f\|_{1,p} = \left(\|f\|_{L^p(\Omega)}^p + \|Df\|_{L^p(\Omega,H)}^p\right)^{1/p}$$

By this we mean the following. The above $\|f\|_{1,p}$ makes perfect sense for every $f \in \mathcal{P}$ and is algebraically like a norm. Thus it makes $\mathcal{P}$ into a normed linear space. $D^{1,p}$ is just the completion of this normed linear space. Then for $f \in D^{1,p}$, we define $Df \equiv \lim_{n \to \infty} Df_n$ in $L^p(\Omega,H)$ where $f_n \in \mathcal{P}$.

### 19.3.2 The Integral

The derivative has been defined above. Now here is the definition of the integral defined on functions in $L^p'(\Omega, H)$, possibly not all of them.

**Definition 19.3.4** We say a random variable $F$ is “smooth” if it is of the form $F(\omega) = F(W(h_1), \ldots, W(h_r))$ where $x \to F(x)$ is a smooth function of the real variables $x_i$. It has polynomial growth if

$$\frac{|F(x)|}{\left(1 + |x|^2\right)^m}$$

is bounded for some positive integer $m$. Let $u \in L^p(\Omega, H)$. Then $u \in D(\delta)$ if for all $F$ smooth having polynomial growth in the $W(h)$,

$$|E(DF, u)| \leq C(u) \|F\|_{L^p(\Omega)}$$

Then $\delta u \in L^p(\Omega)$ is defined by

$$E(DF, u) \equiv E(F\delta u)$$

Thus you have $\delta$ is the adjoint of $D$.

$$L^p(\Omega) \subseteq D(\delta) \subseteq L^p'(\Omega, H)$$

Next it is shown that there are functions in $D(\delta)$ by giving examples of them. It turns out that functions of the form $\sum_i F_i h_i$ where $F_i$ is smooth with polynomial growth are in $D(\delta)$. Consider

$$E(DG, F(W(h_1), \ldots, W(h_n))h)$$

where $G = G(W(k_1), \ldots, W(k_p))$ and for simplicity, $\|h\|_H = 1$.

Consider the vectors \{h, h_1, \ldots, h_n, k_1, \ldots, k_p\}. Starting with the left and moving toward the right, delete vectors which are dependent on the preceding vectors, obtaining a linearly independent set of vectors which includes $h$. Then
let \( \{ h, e_1, \ldots, e_q \} \) be an orthonormal basis having the same span as the original vectors \( \{ h, h_1, \ldots, h_n, k_1, \ldots, k_p \} \).

Then from the fact that \( W \) is linear, there are smooth functions having polynomial growth \( \tilde{G}, \tilde{F} \) such that

\[
\begin{align*}
G( W(h_1), \ldots, W(k_p) ) &= \tilde{G}( W(h), W(e_1), \ldots, W(e_q)) \\
F( W(h_1), \ldots, W(h_n) ) &= \tilde{F}( W(h), W(e_1), \ldots, W(e_q))
\end{align*}
\]

Note that \( h_i = \sum_{j=1}^q (h_i, e_j) e_j + (h_i, h) h \). Thus

\[
F( W(h_1), \ldots, W(h_n) ) = \\
\sum_{j=1}^q (h_1, e_j) W(e_j) + (h_1, h) W(h) + \cdots + \sum_{j=1}^q (h_n, e_j) W(e_j) + (h_n, h) W(h)
\]

and so, \( D_1 \tilde{F} \) is given by

\[
D_1 \tilde{F} = \sum_{i=1}^n D_i( F( W(h_1), \ldots, W(h_n)))(h_i, h)
\]

Then by Lemma

\[
E \langle DG, F( W(h_1), \ldots, W(h_n)) h \rangle = E \left( D \tilde{G}, \tilde{F} h \right)
\]

\[
= E \left( D_1 \left( \tilde{G} h + \sum_{k=1}^q D_k( \tilde{G} ) e_k , \tilde{F} h \right) \right) = E \left( D_1 \left( \tilde{G} \right) \tilde{F} \right)
\]

\[
= \frac{1}{(2\pi)^{q+1}} \int_{|\mathbf{x}|=1} \int_{\mathbb{R}^q} D_1 \tilde{G}(x) \tilde{F}(x) e^{-\frac{1}{2} |x|^2} dx_1 d\mathbf{x}
\]

\[
= \frac{-1}{(2\pi)^{q+1}} \int_{|\mathbf{x}|=1} \int_{\mathbb{R}^q} \tilde{G}(x) D_1 \left( \tilde{F}(x) e^{-\frac{1}{2} |x|^2} - \tilde{F}(x) x_1 e^{-\frac{1}{2} |x|^2} \right) dx_1 d\mathbf{x}
\]

\[
= E \left( F W(h) - D_1 \tilde{F} \right) \tilde{G} = E \left( FW(h) - \sum_{i=1}^n D_i( F)(h_i, h) \right) G
\]

Thus \( Fh \in D(\delta) \) and

\[
\delta(Fh) = FW(h) - \sum_{i=1}^n D_i( F)(h_i, h)
\]

Since \( \delta \) is an adjoint map, it is clearly linear. Hence, if \( h \) is arbitrary, \( h \neq 0 \) of course,

\[
\delta(Fh) = \|h\| \delta \left( F \left( \frac{W(h)}{\|h\|} \right) \right) = \|h\| F \frac{W(h)}{\|h\|} - \|h\| \sum_{i=1}^n D_i( F)(h_i, h) \frac{h}{\|h\|}
\]

\[
= FW(h) - \sum_{i=1}^n D_i( F)(h_i, h) H = FW(h) - \langle DF, h \rangle \quad (19.3.18)
\]

Note how this looks just like integration by parts. More generally,

\[
\delta \left( \sum_{j=1}^m F_j h_j \right) = \sum_{j=1}^m \delta( F_j h_j) = \sum_{j=1}^m F_j W(h_j) - \langle DF_j, h_j \rangle
\]

Are functions like \( \sum_{j=1}^m F_j h_j \) where \( F_j \) is a polynomial in variables of the form \( W(h) \) dense in \( L^p(\Omega, H) \)? It was shown earlier in Lemma that polynomial functions \( F \) in the \( W(h) \) are dense in \( L^p(\Omega) \) for any \( p \). Let \( s(\omega) = \sum_{k=1}^n h_k \mathcal{X}_{E_k} \) be a simple function. Then \( \mathcal{X}_{E_k} \) is clearly in \( L^p(\Omega) \) and so there exists \( E_k \) a polynomial in the \( W(h) \) which is as close as desired to \( \mathcal{X}_{E_k} \) in \( L^p \). Hence \( \sum_{k=1}^n h_k E_k \) is close to \( s \) in \( L^p(\Omega, H) \) and so since these simple functions are dense, it follows that these kinds of functions are indeed dense in \( L^p(\Omega, H) \), this for any \( p > 1 \). The above discussion is summarized in the following lemma.
19.3. THE SKOROKHOD INTEGRAL

**Lemma 19.3.5** Functions of the form $\sum_{k=1}^{n} F_k h_k$ where $F_k$ is a polynomial in $\mathcal{W}(h)$ ($F_j \in \mathcal{P}$) are dense in $L^p(\Omega, H)$ for any $p > 1$. Also each function of this form is in $D\delta$ and

$$
\delta \left( \sum_{j=1}^{m} F_j h_j \right) = \sum_{j=1}^{m} \delta (F_j h_j) = \sum_{j=1}^{m} F_j W (h_j) - \langle DF_j, h_j \rangle
$$

What does $D$ do to $\delta (F h)$? It is shown above that $\delta (F h) = F \mathcal{W} (h) - \langle DF, h \rangle$. Say $F = F (\mathcal{W}(h_1), \cdots, \mathcal{W}(h_n))$. Then when you do $D$ to $\delta (F h)$, you would get

$$
F h + \sum_{k=1}^{n} D_k (F) W (h) h_k - \sum_{k=1}^{n} \sum_{j=1}^{n} D_j (D_k (F)) h_j (h_k, h)
$$

In other words,

$$
F h + W (h) D (F) - D \langle DF, h \rangle
$$

Recall that $DG$ is well defined. This means that we can replace $\{h_1, \cdots, h_n, h\}$ with an orthonormal basis

$$
\{e_1, \cdots, e_p, h\}
$$

as in

$$
G (\mathcal{W}(h_1), \cdots, \mathcal{W}(h_n), \mathcal{W}(h)) = \hat{G} (\mathcal{W}(e_1), \cdots, \mathcal{W}(e_p), \mathcal{W}(h))
$$

where we assume $\|h\| = 1$ for simplicity. Thus the above equals

$$
D (\delta (F)) = D \left( \delta \left( \hat{F} \right) \right) = \hat{F} h + W (h) D \left( \hat{F} \right) - D \langle D \hat{F}, h \rangle
$$

Now consider $E \left( \delta (F h)^2 \right) = E \langle D (\delta (F)), F h \rangle$. Thus the following must be considered.

$$
E \langle \hat{F} h + W (h) D \left( \hat{F} \right) - D \langle D \hat{F}, h \rangle, \hat{F} h \rangle
$$

(19.3.19)

Consider the terms involved. The first term is just $E \left( \| \hat{F} h \|_{H}^2 \right) = E \left( \| F h \|_{H}^2 \right)$. Now consider the third term. It equals

$$
- E \left( D \left( D_{p+1} \left( \hat{F} \right) \right), \hat{F} h \right) = - E \left( D_{p+1}^2 \left( \hat{F} \right) \hat{F} \right)
$$

$$
= \frac{-1}{(\sqrt{2\pi})^{p+1}} \int_{\mathbb{R}^{p+1}} \int_{\mathbb{R}} D_{p+1}^2 \left( \hat{F} (x) \right) \hat{F} (x) e^{-\frac{1}{2} |x|^2} dx_{p+1} d\xi_{p+1}
$$

$$
= \frac{1}{(\sqrt{2\pi})^{p+1}} \int_{\mathbb{R}^{p+1}} \int_{\mathbb{R}} D_{p+1} \left( \hat{F} (x) \right) D_{p+1} \left( \hat{F} (x) \right) e^{-\frac{1}{2} |x|^2} dx_{p+1} d\xi_{p+1}
$$

$$
- \frac{1}{(\sqrt{2\pi})^{p+1}} \int_{\mathbb{R}^{p+1}} \int_{\mathbb{R}} D_{p+1} \left( \hat{F} (x) \right) \hat{F} (x) e^{-\frac{1}{2} |x|^2} dx_{p+1} d\xi_{p+1}
$$

$$
= E \left( \left( D_{p+1} \hat{F} \right)^2 \right) - E \left( W (h) D_{p+1} \left( \hat{F} \right) \hat{F} \right)
$$

$$
= E \left( \left( D_{p+1} \hat{F} \right)^2 \right) - E \left( W (h) D \left( \hat{F} \right), \hat{F} h \right)
$$

Hence $\mathcal{D}$ reduces to

$$
E \left( \left\| \hat{F} h \right\|^2 \right) + E \left( \left( D_{p+1} \hat{F} \right)^2 \right) = E \left( \left\| \hat{F} h \right\|^2 \right) + E \left( \left( D \left( \hat{F} \right), h \right)^2 \right)
$$

$$
= E \left( \left\| F h \right\|^2 \right) + E \left( \left( D (F), h \right)^2 \right)
$$
This assumed that \( \|h\| = 1 \). For arbitrary nonzero \( h \),

\[
E \left( \delta (Fh)^2 \right) = \|h\|^2 E \left( \delta \left( F \frac{h}{\|h\|} \right)^2 \right) \\
= \|h\|^2 \left( E \left( \|F \frac{h}{\|h\|}\|^2 \right) + E \left( \langle D(F), \frac{h}{\|h\|} \rangle \right)^2 \right) \\
= E \left( \|Fh\|^2 \right) + E \left( \langle D(F), h \rangle \right)^2
\]

Next consider a generalization, \( u = \sum_{j=1}^{m} F_j h_j \) where the \( \{h_j\} \) is an orthonormal set of vectors. Say \( F_j = F_j (W(k_1), \ldots, W(k_n)) \). Let \( \{h_1, \ldots, h_m, e_1, \ldots, e_p\} = \{g_i\}_{i=1}^{m+p} \) be an orthonormal basis for the span of all the \( h_j \) and \( k_i \). Thus \( g_i = h_i \) for \( i \leq m \). Then let

\[
F_j (W(k_1), \ldots, W(k_n)) = \hat{F}_j (W(h_1), \ldots, W(h_m), W(e_1), \ldots, W(e_p))
\]

The computations will be done with respect to this orthonormal set because it will be simpler. Also, the above argument using the density function for the normal distribution will be used without explicitly repeating it.

It is desired to consider \( E \left( \delta (u)^2 \right) \). Recall that

\[
D (\delta (Fh)) = Fh + W(h) D(F) - D(DF,h).
\]

Thus \( E \left( \delta (u)^2 \right) = \sum_{j,k=1}^{m} E \left( \delta \left( \hat{F}_j h_j \right) \delta \left( \hat{F}_k h_k \right) \right) = \sum_{j,k=1}^{m} E \left( \langle D \left( \delta (\hat{F}_j h_j) \right), (\hat{F}_k h_k) \rangle \right) \)

\[
\sum_{j,k=1}^{m} E \left( \langle \hat{F}_j h_j + W(h_j) D(\hat{F}_j) - D(D\hat{F}_j, h_j), (\hat{F}_k h_k) \rangle \right)
\]

Separating out the first term this is

\[
= E \left( \sum_{k=1}^{m} \|\hat{F}_k\|^2 \right) + \sum_{k,k} E \left( \langle W(h_j) D(\hat{F}_j), \hat{F}_k h_k \rangle \right) \\
- \sum_{j,k} E \left( D_k \left( D_j \hat{F}_j \right) \hat{F}_k \right)
\]

\[
= E \left( \sum_{k=1}^{m} \|\hat{F}_k\|^2 \right) + \sum_{k,k} E \left( \langle W(h_j) D(\hat{F}_j), \hat{F}_k h_k \rangle \right) \\
- \sum_{j,k} E \left( D_k \left( D_j \hat{F}_j \right) \hat{F}_k \right)
\]

\[
= E \left( \sum_{k=1}^{m} \|\hat{F}_k\|^2 \right) + \sum_{k,k} E \left( \langle W(h_j) D_k (\hat{F}_j), \hat{F}_k \rangle \right) \\
- \sum_{j,k} E \left( D_k \left( D_j \hat{F}_j \right) \hat{F}_k \right)
\]

(19.3.20)

By equality of mixed partial derivatives, the third term equals

\[
- \sum_{j,k} E \left( D_j \left( D_k \hat{F}_j \right) \hat{F}_k \right) = \sum_{j,k} E \left( \left( D_k \hat{F}_j \right) \left( D_j \hat{F}_k \right) \right) - \sum_{j,k} E \left( D_k \left( \hat{F}_j \right) \hat{F}_k W(h_j) \right)
\]
Therefore, \( \| F \|_H^2 \) reduces to

\[
E \left( \delta \left( \sum_{j=1}^m F_j h_j \right)^2 \right) = E \left( \sum_{k=1}^m \| F_k \|^2_H \right) + \sum_{j,k=1}^m E \left( \langle DF_j, h_k \rangle \langle DF_k, h_j \rangle \right)
\]

because the derivative is well defined. All of this assumes the \( h_k \) form an orthonormal set. Suppose these are just orthogonal but nonzero. Then

\[
E \left( \delta \left( \sum_{j=1}^m F_j h_j \right)^2 \right) = E \left( \sum_{j=1}^m \delta (F_j h_j) \right)^2 = E \left( \sum_{j,k=1}^m \delta (F_j h_j) \delta (F_k h_k) \right)
\]

and doing exactly the same steps but keeping the factor \( \| h_j \| \| h_k \| \) throughout, this yields

\[
E \left( \sum_{k=1}^m \| h_k \|^2 \| F_k \|^2_H \right) + \sum_{j,k=1}^m E \left( \| h_j \| \| h_k \| \langle DF_j, h_k \rangle \langle DF_k, h_j \rangle \right)
\]

It appears from the computations to be correct, but it does not look right. This is because the second term is not clearly nonnegative. It is the expectation of the trace of \( A^2 \) where \( A \) is the matrix whose \( jk^{th} \) entry is \( \langle DF_j, h_k \rangle \). One wonders whether the end result is nonnegative.

### 19.3.3 The Ito And Skorokhod Integrals

If you let \( H = L^2 (0, \infty; U) \) where \( U \) is a separable Hilbert space, and if \( f \in D (\delta) \), it is very natural to ask whether \( f \mathcal{A}_{(0,t)} \in D (\delta) \). This is not so. There is a counter example given in \((\text{[?]}\). However, this is true if you change the definition of the integral such that in the definition of \( \delta \), it is only necessary for

\[
\langle DF, G \rangle \leq C \| F \|_{L^p (\Omega)}
\]

where \( F \) is in \( \mathcal{P} \). When you see why this is so, it will be clear why it is not so for the definition given above.

**Lemma 19.3.6** Suppose the definition of the Skorokhod integral \( \delta \) is changed so that it is only necessary to have

\[
\langle DF, G \rangle \leq C \| F \|_{L^p (\Omega)}
\]

for all \( F \) in \( \mathcal{P} \). Then let \( H = L^2 (0, \infty; U) \) or \( L^2 ([0,T]; U) \) where \( U \) is a separable real Hilbert space. For this modified definition of the integral, if \( f \in D (\delta) \), it follows that \( f \mathcal{A}_{(0,t)} \in D (\delta) \).

**Proof:** The case \( L^2 (0, \infty; U) \) is considered here. The other case is similar. \( \delta \) will be defined on some things in \( L^2 (\Omega, L^2 (0, \infty; U), \mathcal{F}) \) where, as discussed earlier,

\[
\mathcal{F} = \sigma (W (h) : h \in H)
\]
Then if you have $f \in D(\delta)$ so $f \in L^2(\Omega, L^2(0, \infty; U))$, does it follow that $f \mathcal{X}_{[0,t]} \in D(\delta)$ also? Let $F$ be one of those polynomial functions of some $W(h)$. Assume first that $a_0$, the constant term is 0 and consider

$$ E \langle DF, f \mathcal{X}_{[0,t]} \rangle = E \left( \sum_k D_k(F) h_k, f \mathcal{X}_{[0,t]} \right) $$

Since $h_k \in H = L^2(0, \infty; U)$, so is $h_k \mathcal{X}_{[0,t]}$. Thus the above reduces to

$$ = \sum_k E \langle D_k(F) h_k, f \mathcal{X}_{[0,t]} \rangle = \sum_k E \left( \int_0^\infty D_k(F(W(h_1), \cdots W(h_n))) h_k \mathcal{X}_{[0,t]} f dt \right) $$

Since $F$ is just a polynomial and $W$ is linear and $\mathcal{X}^q_{[0,t]} = \mathcal{X}_{[0,t]}$, this equals

$$ \sum_k E \left( \int_0^\infty D_k(F(W(\mathcal{X}_{[0,t]} h_1), \cdots W(\mathcal{X}_{[0,t]} h_n))) h_k \mathcal{X}_{[0,t]} f dt \right) $$

Let $F_t = F(W(\mathcal{X}_{[0,t]} h_1), \cdots W(\mathcal{X}_{[0,t]} h_n))$ and so the above is nothing more than

$$ E \langle DF, f \mathcal{X}_{[0,t]} \rangle = E \langle DF_t, f \rangle $$

and since $f \in D(\delta)$,

$$ |E \langle DF, f \mathcal{X}_{[0,t]} \rangle| = |E \langle DF_t, f \rangle| \leq C(f) \|F_t\|_{L^2(\Omega)} $$

Also

$$ \|F_t\|_{L^2(\Omega)}^2 = \int_\Omega F(W(\mathcal{X}_{[0,t]} h_1), \cdots W(\mathcal{X}_{[0,t]} h_n))^2 dP $$

$$ = \int_\Omega \mathcal{X}_{[0,t]} F(W(h_1), \cdots W(h_n))^2 dP $$

$$ \leq \int_\Omega F(W(h_1), \cdots W(h_n))^2 dP $$

Thus for such $F$ which have zero constant term,

$$ |E \langle DF, f \mathcal{X}_{[0,t]} \rangle| \leq C \|F\|_{L^2(\Omega)} $$

Now what if $F$ is a constant $a$? In this case, $DF = Da = 0$

$$ |E \langle Da, f \mathcal{X}_{[0,t]} \rangle| = 0 \leq \|a\|_{L^2(\Omega)} $$

It follows that $\mathcal{X}_{[0,t]} f \in D(\delta)$ whenever $f$ is.

Note how it was essential in this argument to have $F$ be a polynomial or perhaps more generally an analytic function. However, in the definition of the Skorokhod integral, one must test with functions $F$ which are smooth and have polynomial growth. In particular, this would include functions which are infinitely differentiable with compact support, none of which have valid power series.

How does the Skorokhod integral relate to the Ito integral? What about elementary functions and so forth? Let $0 = t_0 < t_1 < \cdots < t_n = T$. Consider

$$ \sum_{k=0}^{n-1} F_k \mathcal{X}_{(t_k, t_{k+1})} $$

As shown above, this is one of the things in $D(\delta)$.

$$ \delta \left( \mathcal{X}_{(0,t)} \sum_{k=0}^{n-1} F_k \mathcal{X}_{(t_k, t_{k+1})} \right) = \delta \left( \sum_{k=0}^{n-1} F_k \mathcal{X}_{[t_k \land t, t \land t_{k+1}]} \right) $$

$$ = \sum_{k=0}^{n-1} F_k W(\mathcal{X}_{[t_k \land t, t \land t_{k+1}]}) - \langle DF_k, \mathcal{X}_{[t_k \land t, t \land t_{k+1}]} \rangle $$
19.3. THE SKOROKHOD INTEGRAL

\[
= \sum_{k=0}^{n-1} F_k \left( W \left( X_{(0,t \wedge t_k)} \right) - W \left( X_{(0,t \wedge t_k)} \right) \right) - \left\langle DF_k, X_{[t_k,t \wedge t_{k+1}]} \right\rangle_H
\]

In terms of the Wiener process, this is of the form

\[
= \sum_{k=0}^{n-1} F_k \left( W \left( t \wedge t_{k+1} \right) - W \left( t \wedge t_k \right) \right) - \left\langle DF_k, X_{[0,t \wedge t_{k+1}]} - X_{[0,t \wedge t_k]} \right\rangle_H
\]

What if

\[
F_k = F_k \left( W \left( X_{[0,t;h]} \right) , \cdots , W \left( X_{[0,t;h]} \right) \right) ?
\]

Let \( F_k \equiv \sigma \left( W \left( X_{[0,t;h]} : h \in H \right) \right) \). Then this is clearly a filtration. If \( F_k \) is as just described, then \( F_k \) is \( F_k \) adapted.

\[
\left\langle DF_k, X_{[0,t \wedge t_{k+1}]} - X_{[0,t \wedge t_k]} \right\rangle = \int_0^\infty \sum_s D_s (F_k) X_{(0,t_k)h_s} X_{(t \wedge t_{k+1})} = 0
\]

because the intervals are disjoint. In this case, the troublesome term at the end vanishes and you are left with

\[
\sum_{k=0}^{n-1} F_k \left( W \left( t \wedge t_{k+1} \right) - W \left( t \wedge t_k \right) \right)
\]

which is similar to the usual definition for the Ito integral.

What if \( F \in L^2 \left( \Omega \times [0,T] \right) \) and is progressively measurable. Does it have a Skorokhod integral, and if so, is it the same as the Ito integral? Recall the following useful lemma. It is Lemma 19.3.7 on Page 473.

**Lemma 19.3.7** Let \( \Phi : [0,T] \times \Omega \to E \), be \( B \left( [0,T] \right) \times F \) measurable and suppose

\[
\Phi \in K \equiv L^p \left( [0,T] \times \Omega ; E \right) , \quad p \geq 1
\]

Then there exists a sequence of nested partitions, \( \mathcal{P}_k \subseteq \mathcal{P}_{k+1} \),

\[
\mathcal{P}_k \equiv \{ t_0^k , \cdots , t_{m_k}^k \}
\]

such that the step functions given by

\[
\Phi_k^L (t) = \sum_{j=1}^{m_k} \Phi \left( t_j^k \right) X_{[t_{j-1}^k,t_j^k]} (t)
\]

\[
\Phi_k^R (t) = \sum_{j=1}^{m_k} \Phi \left( t_{j-1}^k \right) X_{[t_{j-1}^k,t_j^k]} (t)
\]

both converge to \( \Phi \) in \( K \) as \( k \to \infty \) and

\[
\lim_{k \to \infty} \max \left\{ |t_j^k - t_{j+1}^k| : j \in \{ 0, \cdots , m_k \} \right\} = 0.
\]

Also, each \( \Phi \left( t_j^k \right) , \Phi \left( t_{j-1}^k \right) \) is in \( L^p \left( \Omega ; E \right) \). One can also assume that \( \Phi (0) = 0 \). The mesh points \( \{ t_j^k \}_{j=0}^{m_k} \) can be chosen to miss a given set of measure zero. In addition to this, we can assume that

\[
|t_j^k - t_{j-1}^k| = 2^{-n_k}
\]

except for the case where \( j = 1 \) or \( j = m_{n_k} \) when this is so, you could have \( |t_j^k - t_{j-1}^k| < 2^{-n_k} \).

**Theorem 19.3.8** Let \( F \in L^2 \left( \Omega \times [0,T] \right) \) and is progressively measurable. Then it has a Skorokhod integral which coincides with the Ito integral.

**Proof:** From Lemma 19.3.7, there is a sequence of left step functions denoted here as \( \{ F_k^L \}_{k=1}^\infty \) which converges to \( F \) in \( L^2 \left( \Omega \times [0,T] \right) \) where \( F_k^L \left( t_j^k \right) = F \left( t_j^k \right) \). We can take a subsequence if necessary and assume

\[
\left\| F_k^L - F \right\|_{L^2 ([0,T] \times \Omega )} < 2^{-k}
\]
Here the \( \{ t^k_j \} \) are mesh points corresponding to the \( k^{th} \) partition described above. Thus each \( F_k (t^k_j) \) is in \( L^2 (\Omega) \). By Lemma 19.3.3, there exists a random variable \( G^k_j (t^k_j) \) which is a polynomial function of some \( W (h) \) for \( h \in L^2 (0, t^k_j) \) which can approximate \( F_k (t^k_j) \) as closely as desired in \( L^2 (\Omega) \). Then choosing these sufficiently close, it can be assumed that the step functions

\[
G^k_j \equiv \sum_{j=0}^{m_k-1} G^j_k (t^k_j) \chi(t^k_j, t^k_{j+1}]
\]

also converge in \( L^2 (\Omega \times [0, T]) \) to \( F \). Of course, each of these last step functions are in \( D (\delta) \).

The idea is to show that \( \delta (G^k_k) \) is Cauchy in \( L^2 (\Omega) \) as \( k \to \infty \) and then use the fact that, since \( \delta \) is an adjoint, it must be a closed operator. This will show that \( F \in L^2 (\Omega \times [0, T]) \), considered as a subspace of \( L^2 (\Omega; L^2 (0, \infty, \mathbb{R})) \), is in \( D (\delta) \) and \( \delta (F) \) is equal to the above limit. Using \( C_44 \) which comes from the fact that the functions are adapted to the given filtration,

\[
|| \delta (X_{[0,T]} G^k_k) - \delta (X_{[0,T]} G^k_{k+1}) ||^2_{L^2 (\Omega)} = E \left( \sum_{j=0}^{m_k-1} (G^j_k (t^k_{j+1}) - G^j_{k+1} (t^k_{j+1})) (W (t^k_{j+1}) - W (t^k_j)) \right)^2
\]

Consider a mixed term. To save on space, let \( \Delta_j = G^j_k (t^k_{j+1}) - G^j_{k+1} (t^k_{j+1}) \) and say \( i < j \). Then

\[
E ( (\Delta_j) (\Delta_i) (W (t^k_{j+1}) - W (t^k_j)) (W (t^k_{i+1}) - W (t^k_i))) = 0
\]

By independence of the increments for \( W \), this is

\[
E ( (W (t^k_{j+1}) - W (t^k_j)) (W (t^k_{i+1}) - W (t^k_i))) = 0
\]

and so the above reduces to

\[
\sum_{j=0}^{m_k-1} E \left( \Delta_j^2 (W (t^k_{j+1}) - W (t^k_j))^2 \right) = \sum_{j=0}^{m_{k+1}-1} E (\Delta_j^2) E \left( (W (t^k_{j+1}) - W (t^k_j))^2 \right)
\]

\[
= \sum_{j=0}^{m_{k+1}-1} E \left( (G^j_k (t^k_{j+1}) - G^j_{k+1} (t^k_{j+1}))^2 \right) (t^k_{j+1} - t^k_j)
\]

\[
= E \left( \int_0^T (G^k_k - G^k_{k+1})^2 dt \right) \leq 2 \left( E \int_0^T (G^k_k - F)^2 dt + E \int_0^T (F - G^k_{k+1})^2 dt \right)
\]

which is given to converge to 0 as \( k \to \infty \). It follows that \( X_{[0,T]} G^k_k \to X_{[0,T]} F \) in \( L^2 (\Omega; L^2 (0, \infty, \mathbb{R})) \) by construction and \( \delta (X_{[0,T]} G^k_k) \) is a Cauchy sequence in \( L^2 (\Omega) \). Therefore, it converges to something in \( L^2 (\Omega) \) and since \( \delta \) is a closed operator, that which it converges to is \( \delta (F) \).

However, by the definition of the Ito integral, \( \delta (X_{[0,T]} G^k_k) \) also converges to the Ito integral \( \int_0^T F dW \).

It follows that the Skorokhod integral is more general than the Ito integral but it gives the Ito integral in the special case where the function is adapted. This also shows that the progressively measurable functions in \( L^2 ([0,T] \times \Omega) \) are in \( D (\delta) \), but as shown above, there are many other functions which are not progressively measurable but which are still in \( D (\delta) \). Just consider, for example \( \sum_{k=1}^n F h_k \) where \( F \) is just a polynomial in \( W (h) \) for \( h \in L^2 (0, \infty; \mathbb{R}) \).
Chapter 20

The Easy Ito Formula

First recall where it is shown that for every $\alpha$

$E (|W (t) - W (s)|^\alpha) \leq C_\alpha |t - s|^\alpha/2$,

and so by Kolmogorov Čentsov continuity theorem

$|W (t) - W (s)| \leq C_\gamma |t - s|^{\gamma}$ (20.0.1)

for every $\gamma < 1/2$.

20.1 The Situation

The idea is as follows. You have a sufficiently smooth function $F : [0, T] \times H \rightarrow \mathbb{R}$ where $H$ is a separable Hilbert space. You also have the random variable

$X (t) = X_0 + \int_0^t \phi (s) ds + \int_0^t \Phi dW$

where $\Phi$ is progressively measurable and in $L^2 ([0, T] \times \Omega; \mathcal{L}_2 (Q^{1/2}U, H))$ where $Q : U \rightarrow U$ is a positive self adjoint operator. Also assume $X_0$ is $\mathcal{F}_0$ measurable with values in $H$. Recall the descriptive diagram.

Here the Wiener process is in $U_1$ and the filtration with respect to which $\Phi$ is progressively measurable is the usual filtration determined by this Wiener process. Then the Ito formula is about writing the random variable $F (t, X (t))$ in terms of various integrals and derivatives of $F$.

20.2 Assumptions And A Lemma

Assume $F : [0, T] \times H \times \Omega \rightarrow \mathbb{R}^1$ has continuous partial derivatives $F_1, F_X,$ and $F_{XX}$ which are uniformly continuous and bounded on bounded subsets of $[0, T] \times H$ independent of $\omega \in \Omega$. Also assume $F_{XX}$ is uniformly bounded and that $F_{XXX}$ exists. Let $\phi : [0, T] \times \Omega \rightarrow H$ be progressively measurable and Bochner integrable for each $\omega$. Assume $\Phi$ is progressively measurable, and is in $L^2 ([0, T] \times \Omega; \mathcal{L}_2 (Q^{1/2}U, H))$.

Now here is the important lemma which makes the Ito formula possible.
Lemma 20.2.1 Suppose \( \eta_j \) are real random variables \( E(\eta_j^2) < \infty \), such that \( \eta_k \) is measurable with respect to \( G_j \) for all \( j > k \) where \( \{G_k\} \) is increasing. Then

\[
E \left( \sum_{k=0}^{m-1} \eta_k - \sum_{k=0}^{m-1} E(\eta_k|G_k)^2 \right)^2 = E \left( \sum_{k=0}^{m-1} \eta_k^2 - E(\eta_k|G_k)^2 \right)^2
\]

(20.2.2)

Proof: First consider a mixed term \( i < k \).

\[
E((\eta_i - E(\eta_i|G_i))(\eta_k - E(\eta_k|G_k)))
\]

This equals

\[
E(\eta_i \eta_k) - E(\eta_i E(\eta_k|G_k)) - E(\eta_k E(\eta_i|G_i)) + E(E(\eta_i|G_i) E(\eta_k|G_k))
\]

\[
= E(\eta_i \eta_k) - E(E(\eta_i \eta_k|G_k)) - E(E(\eta_k \eta_i|G_i)) + E(E(\eta_k \eta_i|G_k))
\]

\[
= E(\eta_i \eta_k) - E(\eta_i \eta_k) - E(\eta_k \eta_i) + E(\eta_k \eta_i) = 0
\]

Thus \( 20.2.2 \) equals

\[
\sum_{k=0}^{m-1} E(\eta_k^2 - E(\eta_k|G_k)^2)
\]

which equals

\[
\sum_{k=0}^{m-1} E(\eta_k^2) - 2E(\eta_k E(\eta_k|G_k)) + E(E(\eta_k|G_k)^2)
\]

\[
= \sum_{k=0}^{m-1} E(\eta_k^2) - 2E(E(\eta_k E(\eta_k|G_k)|G_k) + E(E(\eta_k|G_k)^2)
\]

\[
= \sum_{k=0}^{m-1} E(\eta_k^2) - 2E(E(\eta_k|G_k) E(\eta_k|G_k)) + E(E(\eta_k|G_k)^2)
\]

\[
= \sum_{k=0}^{m-1} E(\eta_k^2) - 2E(\eta_k|G_k)^2 + E(E(\eta_k|G_k)^2)
\]

\[
= \sum_{k=0}^{m-1} E(\eta_k^2) - E(E(\eta_k|G_k)^2)
\]

20.3 A Special Case

To make it simpler, first consider the situation in which \( \Phi = \Phi_0 \) where \( \Phi_0 \) is \( F_0 \) measurable and has finitely many values in \( \mathcal{L}(U, H) \), and \( \phi = \phi_0 \) where \( \phi_0 \) is \( F_0 \) measurable and a simple function with values in \( H \). Thus

\[
X(t) = X_0 + \int_0^t \phi_0 ds + \int_0^t \Phi_0 dW
\]

Now let \( \{t^n_k\}_{k=0}^{m_n} \) denote the \( n \)th partition of \([0, T]\), referred to as \( P_n \) such that

\[
\lim_{n \to \infty} \max \left\{ |t^n_k - t^n_{k-1}|, k = 0, 1, 2, \ldots, m_n \right\} = \lim_{n \to \infty} ||P_n|| = 0.
\]

The superscript \( n \) will be suppressed to save notation. Then

\[
F(T, X(T)) - F(0, X_0) = \sum_{k=0}^{m_n-1} (F(t_{k+1}, X(t_{k+1})) - F(t_k, X(t_k)))
\]
Consider one of the terms in (20.3.4). This equals:
\[
\sum_{k=0}^{m_n-1} (F(t_{k+1}, X(t_{k+1})) - F(t_k, X(t_k))) + \sum_{k=0}^{m_n-1} (F(t_k, X(t_{k+1})) - F(t_k, X(t_k)))
\]
\[
+ \frac{1}{2} \sum_{k=0}^{m_n-1} (F_{XX}(t_k, X(t_k)) (X(t_{k+1}) - X(t_k)) - (X(t_{k+1}) - X(t_k)))_H
\]
\[
+ \sum_{k=0}^{m_n-1} O \left( |X(t_{k+1}) - X(t_k)|^2_H \right)
\]

Recall
\[
X(t) = X_0 + \int_0^t \phi_0 ds + \int_0^t \Phi_0 dW
\]

From the properties of the Wiener process in (20.3.1), the term in (20.3.4) converges to 0 as \( n \to \infty \) since these properties of the Wiener process imply \( X \) is Holder continuous with exponent 2/5.

Now consider the term of (20.3.5). All terms converge to 0 except
\[
\frac{1}{2} \sum_{k=0}^{m_n-1} \left( F_{XX}(t_k, X(t_k)) \int_{t_k}^{t_{k+1}} \Phi_0 dW, \int_{t_k}^{t_{k+1}} \Phi_0 dW \right)_H
\]

Consider one of the terms in (20.3.7). Let \( A \in \mathcal{F}_{t_k} \). By Corollary (17.5.5),
\[
\int_{A} \frac{1}{2} \left( F_{XX}(t_k, X(t_k)) \int_{t_k}^{t_{k+1}} \Phi_0 dW, \int_{t_k}^{t_{k+1}} \Phi_0 dW \right)_H dP
\]
\[
= \int_{A} \frac{1}{2} \left( \int_{t_k}^{t_{k+1}} F_{XX}(t_k, X(t_k)) \Phi_0 dW, \int_{t_k}^{t_{k+1}} \Phi_0 dW \right)_H dP
\]
By independence,
\[
= P(A) \frac{1}{2} \int_{\Omega} \left( \int_{t_k}^{t_{k+1}} F_{XX}(t_k, X(t_k)) \Phi_0 dW, \int_{t_k}^{t_{k+1}} \Phi_0 dW \right)_H dP
\]
By the Ito isometry results presented earlier,
\[
= \int_{\Omega} \mathcal{L}_{2}(A) \frac{1}{2} \int_{t_k}^{t_{k+1}} (F_{XX}(t_k, X(t_k)) \Phi_0, \Phi_0)_{L_2} \, ds \, dP
\]
\[
= \int_{A} \frac{1}{2} \int_{t_k}^{t_{k+1}} (F_{XX}(t_k, X(t_k)) \Phi_0, \Phi_0)_{L_2} \, ds \, dP
\]
\[
= \int_{A} \frac{1}{2} (F_{XX}(t_k, X(t_k)) \Phi_0, \Phi_0)_{L_2} (t_{k+1} - t_k) \, dP
\]
Since \( A \in \mathcal{F}_{t_k} \) was arbitrary,
\[
E \left( \frac{1}{2} \left( F_{XX}(t_k, X(t_k)) \int_{t_k}^{t_{k+1}} \Phi_0 dW, \int_{t_k}^{t_{k+1}} \Phi_0 dW \right)_H | \mathcal{F}_{t_k} \right)
\]
\[
= \frac{1}{2} (F_{XX}(t_k, X(t_k)) \Phi_0, \Phi_0)_{L_2} (t_{k+1} - t_k)
\]
CHAPTER 20. THE EASY ITO FORMULA

From what was just shown, and Lemma 20.3.1,

\[
E \left( \frac{1}{2} \sum_{k=0}^{m-1} (F_{XX}(t_k, X(t_k)) \Phi_0 \Delta W(t_k), \Phi_0 \Delta W(t_k))_H - \sum_{k=0}^{m-1} \frac{1}{2} (F_{XX}(t_k, X(t_k)) \Phi_0, \Phi_0)_{L^2}(t_{k+1} - t_k) \right)^2 \right) \right)
\]

\[
= \frac{1}{4} E \left( \sum_{k=0}^{m-1} (F_{XX}(t_k, X(t_k)) \Phi_0 \Delta W(t_k), \Phi_0 \Delta W(t_k))_H^2 - \sum_{k=0}^{m-1} (F_{XX}(t_k, X(t_k)) \Phi_0, \Phi_0)_{L^2}(t_{k+1} - t_k)^2 \right)
\]

Now \(F_{XX}\) is bounded and so there exists a constant \(M\) independent of \(k\) and \(n\),

\[
M \geq ||\Phi_0 F_{XX}(t_k, X(t_k)) \Phi_0||, |(F_{XX}(t_k, X(t_k)) \Phi_0, \Phi_0)_{L^2}|
\]

Hence the above is dominated by

\[
\leq \frac{1}{4} M^2 \sum_{k=0}^{m-1} E ||\Delta W(t_k)||_{L^4}^4 + \frac{1}{4} M^2 \sum_{k=0}^{m-1} (t_{k+1} - t_k)^2
\]

\[
\leq \frac{M^2}{4} \left( \sum_{k=0}^{m-1} (C_4 + 1) (t_{k+1} - t_k)^2 \right)
\]

which converges to 0 as \(n \to \infty\). Then from 20.3.8, and referring to 20.3.9,

\[
\lim_{n \to \infty} \frac{1}{2} \sum_{k=0}^{m-1} (F_{XX}(t_k, X(t_k)) (X(t_{k+1}) - X(t_k)), (X(t_{k+1}) - X(t_k)))_H
\]

\[
= \lim_{n \to \infty} \frac{1}{2} \sum_{k=0}^{m-1} \left( F_{XX}(t_k, X(t_k)) \int_{t_k}^{t_{k+1}} \Phi_0 dW, \int_{t_k}^{t_{k+1}} \Phi_0 dW \right)_H
\]

\[
= \lim_{n \to \infty} \frac{1}{2} \sum_{k=0}^{m-1} (F_{XX}(t_k, X(t_k)) \Phi_0, \Phi_0)_{L^2}(t_{k+1} - t_k)
\]

if this last limit exists in \(L^2(\Omega)\). However, since \(F_{XX}\) is bounded, this limit certainly exists for a.e. \(\omega\) and equals

\[
= \frac{1}{2} \int_0^T (F_{XX}(t, X(t)) \Phi_0, \Phi_0)_{L^2} dt,
\]

The limit also exists in \(L^2(\Omega)\) obviously, since \(F_{XX}\) is assumed bounded. Therefore, a subsequence of 20.3.9, still denoted as \(n\) must converge for a.e. \(\omega\) to the above integral as \(n \to \infty\).

Next consider 20.3.10. From Corollary 17.5.4, it equals

\[
\sum_{k=0}^{m-1} \int_{t_k}^{t_{k+1}} F_X(t_k, X(t_k)) \Phi_0 dW
\]

\[
= \int_0^T \left( \sum_{k=0}^{m-1} X_{t_k, t_{k+1}}(t) F_X(t_k, X(t_k)) \right) \Phi_0 dW
\]
which converges as \( n \to \infty \) to
\[
\int_0^T F_X (t, X (t)) \Phi_0 dW
\]
because
\[
\lim_{n \to \infty} \left( \sum_{k=0}^{m_n-1} X_{(t_k, t_{k+1})} (t) F_X (t_k, X (t_k)) \right) \Phi_0 = F_X (t, X (t)) \Phi_0
\]
in \( L^2 ([0, T] \times \Omega; L_2 (Q^{1/2} U, H)) \). Next consider the first on the right in 20.3.10. It equals
\[
\sum_{k=0}^{m_n-1} (F_X (t_k, X (t_k)) \phi_0 (t_{k+1} - t_k))
\]
and converges to
\[
\int_0^T F_X (t, X (t)) \phi_0 dt.
\]
Finally, it is obviously the case that 20.3.10 converges to
\[
\int_0^T F_1 (t, X (t)) dt
\]
This has shown
\[
F (T, X (T)) = F (0, X_0) + \int_0^T F_1 (t, X (t)) + F_X (t, X (t)) \phi_0 dt
\]
\[
+ \int_0^T F_X (t, X (t)) \Phi_0 dW + \frac{1}{2} \int_0^T (F_{XX} (t, X (t)) \Phi_0, \Phi_0)_{L_2 (Q^{1/2} U, H)} dt
\]
when
\[
X (t) = X_0 + \int_0^t \phi_0 ds + \int_0^t \Phi_0 dW,
\]
\( \phi_0, \Phi_0 \mathcal{F}_0 \) measurable as described above. This is the first version of the Ito formula.

### 20.4 The Case Of Elementary Functions

Of course there was nothing special about the interval \([0, T]\). It follows that for \([a, b] \subseteq [0, T] \), \( \Phi_a \in L (U_1, U) \) and \( \mathcal{F}_a \) measurable, having finitely many values, \( \phi_a \) a simple function which is \( \mathcal{F}_a \) measurable,
\[
X (t) = X (a) + \int_a^t \phi_a dt + \int_a^t \Phi_a dW
\]
\[
F (b, X (b)) = F (a, X (a)) + \int_a^b (F_1 (t, X (t)) + F_X (t, X (t)) \phi_a) dt
\]
\[
+ \int_a^b F_X (t, X (t)) \Phi_a dW + \frac{1}{2} \int_a^b (F_{XX} (t, X (t)) \Phi_a, \Phi_a)_{L_2 (Q^{1/2} U, H)} dt.
\]
Therefore, if \( \Phi \) is any elementary function, being a sum of functions like \( \Phi_a \mathcal{X}_{[a, b]} \), and \( \phi \) a similar sort of elementary function with
\[
X (t) = X_0 + \int_0^t \phi ds + \int_0^t \Phi dW,
\]
then
\[
F (T, X (T)) = F (0, X_0) + \int_0^T F_1 (t, X (t)) + F_X (t, X (t)) \phi (t) dt
\]
\[
+ \int_0^T F_X (t, X (t)) \Phi dW + \frac{1}{2} \int_0^T (F_{XX} (t, X (t)) \Phi, \Phi)_{L_2 (Q^{1/2} U, H)} dt
\]
(20.4.11)
This has proved the following lemma.

**Lemma 20.4.1** Let \( \Phi, \phi \) be elementary functions as described and let
\[
X (t) = X_0 + \int_0^t \phi (s) ds + \int_0^t \Phi dW
\]
Then 20.4.11 holds.
20.5 The Integrable Case

Now let $\Phi \in L^2 \left( [0, T] \times \Omega; \mathcal{L}_2 \left( \mathcal{F}^{1/2} U, H \right) \right), \phi \in L^1 \left( [0, T] \times \Omega; H \right)$ and be progressively measurable. Let $\phi$ be as above, and let

$$X(t) = X_0 + \int_0^t \phi(s) \, ds + \int_0^t \Phi \, dW$$

(20.5.12)

Suppose also the additional condition that for some $M$,

$$|X(t, \omega)| < M \text{ for all } (t, \omega) \in [0, T] \times \mathcal{N}, \ P(N) = 0.$$

Does it follow that $$\Phi_{n, \tau}$$ holds?

There exists a sequence of elementary functions $\{\Phi_n\}$ converging to $\Phi \circ J^{-1}$ in $L^2 \left( [0, T] \times \Omega; \mathcal{L}_2 \left( \mathcal{F}^{1/2} U, H \right) \right)$. Similarly let $\{\phi_n\}$ converge to $\phi$ in $L^1 \left( [0, T] \times \Omega; H \right)$ where $\phi_n$ is also an elementary function, $|\phi_n| \leq |\phi|$ at the mesh points. You could use that theorem about approximating with left and right step functions if desired, Lemma 17.3.1. Let

$$X_n(t) = X_0 + \int_0^t \phi_n(s) \, ds + \int_0^t \Phi_n \, dW.$$

Also let $\tau_n$ be the stopping times

$$\tau_n \equiv \inf \{ t > 0 : |X_n(t)| > M \}.$$

Since $X_n$ is continuous, this is a well defined stopping time. Thus

$$X_{\tau_n}^\tau(t) = X_0 + \int_0^\tau X_{0, \tau_n} \phi_n(t) \, dt + \int_0^\tau X_{0, \tau_n} \Phi_n \, dW$$

and as noted in the discussion of localization for elementary functions, $X_{0, \tau_n} \Phi_n$ is an elementary function.

**Claim:** $\lim_{n \to \infty} X_{0, \tau_n} = 1.$

**Proof of claim:** From maximal estimates as in the construction of the stochastic integral and the Borel Cantelli lemma, it follows that there exists a subsequence still denoted by $n$ and a set of measure zero $N$ such that for $\omega \notin N_1$,

$$\int_0^\tau \Phi_n \, dW \to \int_0^\tau \Phi \, dW$$

uniformly on $[0, T]$. Also one can show that off a set of measure zero, there is a subsequence still called $n$ such that

$$\int_0^\tau \phi_n(s) \, ds \to \int_0^\tau \phi(s) \, ds$$

uniformly on $[0, T]$. Here is why.

$$E \left( \left| \int_0^\tau \phi_n(s) \, ds - \int_0^\tau \phi(s) \, ds \right| \right) \leq \int_\Omega \int_0^T |\phi_n - \phi| \, dt \, dP$$

which is given to converge to 0. Thus

$$P \left( \max_{t \in [0, T]} \left| \int_0^t \phi_n(s) \, ds - \int_0^t \phi(s) \, ds \right| > \lambda \right) \leq P \left( \int_0^T |\phi_n(s) - \phi(s)| \, ds > \lambda \right)$$

$$\leq \frac{1}{\lambda} \int_\Omega \int_0^T |\phi_n(s) - \phi(s)| \, ds \, dP$$

Thus

$$P \left( \max_{t \in [0, T]} \left| \int_0^t \phi_n(s) \, ds - \int_0^t \phi(s) \, ds \right| > 2^{-k} \right) \leq 2^k \int_\Omega \int_0^T |\phi_n(s) - \phi(s)| \, ds \, dP$$

If $n > n_k$, the right side is less than $2^{-k}$. Use $\phi_{n_k}$. Then there exists a set of measure zero $N_2$ such that for $\omega \notin N_2$,

$$\left| \int_0^t \phi_n(s) \, ds - \int_0^t \phi(s) \, ds \right| \to 0$$
20.6. THE GENERAL STOCHASTICALLY INTEGRABLE CASE

uniformly. Hence, you can take a couple of subsequences and assert that there exists a subsequence still called \( n \) and a set of measure zero \( N \) such that \( X_n (t) \rightarrow X (t) \) uniformly on \([0, T]\) for each \( \omega \notin N \). Since \( |X (t, \omega)| < M \), it follows that for each \( \omega \notin N \), when \( n \) is large enough, \( \tau_n = \infty \) and this proves the claim.

From the claim, it follows that \( X_{[0, \tau_n]} \Phi_n \rightarrow \Phi \circ J^{-1} \) in \( L^2 (\Omega; \mathbb{L}_2 (Q^{1/2} U, H)) \) and \( X_{[0, \tau_n]} \phi_n \rightarrow \phi \) in \( L^1 ([0, T] \times \Omega; H) \). Thus you can replace \( \Phi_n \) in the above with \( X_{[0, \tau_n]} \Phi_n \) and \( \phi_n \) with \( X_{[0, \tau_n]} \phi_n \). Thus there exists a subsequence, still called \( n \) and a set of measure zero \( N \) such that for \( \omega \notin N \),

\[
\int_0^t X_{[0, \tau_n]} \Phi_n dW \rightarrow \int_0^t \Phi dW
\]

uniformly and

\[
\int_0^t X_{[0, \tau_n]} \phi_n ds \rightarrow \int_0^t \phi ds
\]

uniformly. Hence also \( X_{[n]} (t) \rightarrow X (t) \) uniformly on \([0, T]\) whenever \( \omega \notin N \). Of course \( |X_{[n]} (t)|_H \) has the advantage of being bounded by \( M \).

From the above,

\[
F (T, X_{[n]} (T)) = F (0, X_0) + \int_0^T F_t (t, X_{[n]} (t)) + F_X (t, X_{[n]} (t)) X_{[0, \tau_n]} \phi_n (t) dt
\]

\[+ \int_0^T F_X (t, X_{[n]} (t)) \Phi_n dW + \frac{1}{2} \int_0^T \left(F_X X (t, X_{[n]} (t)) X_{[0, \tau_n]} \Phi_n, X_{[0, \tau_n]} \Phi_n \right)_{L_2 (Q^{1/2} U, H)} dt\]

Then it is obvious that one can pass to the limit in each of the non stochastic integrals in the above. It is necessary to consider the other one.

From the above claim, \( X_{[0, \tau_n]} \Phi_n \rightarrow \Phi \circ J^{-1} \) in \( L^2 ([0, T] \times \Omega; \mathbb{L}_2 (JQ^{1/2} U, H)) \) and also, thanks to the stopping times \( \tau_n \), \( F_X (t, X_{[n]} (t)) \) is bounded and converges to \( F_X (t, X (t)) \). Hence the dominated convergence theorem applies, and letting \( n \rightarrow \infty \), the following is obtained for a.e. \( \omega \)

\[
F (T, X (T)) = F (0, X_0) + \int_0^T F_t (t, X (t)) + F_X (t, X (t)) \phi (t) dt
\]

\[+ \int_0^T F_X (t, X (t)) \Phi dW + \frac{1}{2} \int_0^T \left(F_X X (t, X (t)) \Phi, \Phi \right)_{L_2 (Q^{1/2} U, H)} dt\]  \hspace{1cm} (20.5.13)

This is the Itô formula in case that \( \Phi \in L^2 ([0, T] \times \Omega; \mathbb{L}_2 (Q^{1/2} U, H)) \) and \( |X| \) is bounded above by \( M \).

It is easy to remove this assumption on \( |X| \). Let \( X \) be given in 20.5.12. Let \( \tau_n \) be the stopping time

\[
\tau_n = \inf \{ t > 0 : |X| > n \}
\]

Then 20.5.13 holds for the stopped process \( X_{[\tau_n]} \) and \( \Phi \) and \( \phi \) replaced with \( \Phi X_{[0, \tau_n]} \) and \( \phi X_{[0, \tau_n]} \) respectively. Then let \( n \rightarrow \infty \) in this expression, using the continuity of \( X \) and the fact that \( \tau_n \rightarrow \infty \) to to recover 20.5.13 without the restriction on \( |X| \).

20.6. The General Stochastically Integrable Case

Now suppose that \( \Phi \) is only progressively measurable and stochastically integrable

\[
P \left( \int_0^T ||\Phi||^2_{L_2 (Q^{1/2} U, H)} dt < \infty \right) = 1.
\]

Also \( \phi \) is only progressively measurable and Bochner integrable in \( t \). Define a stopping time

\[
\tau (\omega) = \inf \left\{ t \geq 0 : |X (t, \omega)|_H + \int_0^t ||\Phi||^2 ds + \int_0^t |\phi| ds > C \right\}
\]

This is just the first hitting time of an open set so it is a stopping time. For \( t \leq \tau \), all of the above quantities must be no larger than \( C \). In particular, \( X_{[0, \tau]} \Phi \in L^2 ([0, T] \times \Omega; \mathbb{L}_2 (Q^{1/2} U, H)) \). Then

\[
X^\tau (t) = X_0 + \int_0^t X_{[0, \tau]} \phi ds + \int_0^t X_{[0, \tau]} \Phi dW
\]
and so \(4.39.4\) holds with \(X \rightarrow X', \Phi \rightarrow \chi'_{[0,T]} \Phi \) and \(\phi \rightarrow \chi'_{[0,T]} \phi \). Now simply let \(C \rightarrow \infty \) and exploit the continuity of \(X\) given by the formula \(4.39.4\) to obtain the validity of \(4.39.4\) without any reference to the stopping time. Of course arbitrary \(t\) can replace \(T\). This leads to the main result.

**Theorem 20.6.1** Let \(\Phi\) be a progressively measurable process having values in \(L_2(Q^{1/2}U, H)\) which is stochastically integrable in \([0, T]\) because

\[
P \left( \left\| \int_0^T |\Phi(t)|^2_{L^2(Q^{1/2}U, H)} dt < \infty \right\| \right) = 1
\]

and let \(\phi : [0, T] \times \Omega \rightarrow H\) be progressively measurable and Bochner integrable on \([0, T]\) for a.e. \(\omega\), and let \(X_0\) be \(\mathcal{F}_0\) measurable and \(H\) valued. Let

\[
X(t) = X_0 + \int_0^t \phi(s) ds + \int_0^t \Phi dW.
\]

Let \(F : [0, T] \times H \times \Omega \rightarrow \mathbb{R}^1\) be progressively measurable, have continuous partial derivatives \(F_1, F_X, F_{XX}\) which are uniformly continuous on bounded subsets of \([0, T] \times H\) independent of \(\omega \in \Omega\). Also assume \(F_{XX}\) is bounded and let \(F_{XXX}\) exist and be bounded. Then the following formula holds for a.e. \(\omega\).

\[
F(t, X(t)) = F(0, X_0) + \int_0^t F_X(\cdot, X(\cdot)) \Phi dW + \int_0^t F_t(s, X(s)) + F_X(s, X(s)) \phi(s) ds + \frac{1}{2} \int_0^t (F_{XX}(s, X(s)) \Phi, \Phi)_{L^2(Q^{1/2}U, H)} ds
\]

The dependence of \(F\) on \(\omega\) is suppressed.

That last term is interesting and can be written differently. Let \(\{g_j\}\) be an orthonormal basis for \(Q^{1/2}U\). Then this integrand equals

\[
\sum_{i=1}^L (F_{XX}(s, X(s)) \Phi g_i, \Phi g_i)_H = \sum_{i=1}^L (\Phi^* F_{XX}(s, X(s)) \Phi g_i, g_i)_{Q^{1/2}H}
\]

and we write this as

\[
\text{trace} (\Phi^* (s) F_{XX}(s, X(s)) \Phi (s)).
\]

A simple special case is where \(Q = I\) and then \(Q^{1/2}U = U\). Thus it is only required that \(\Phi\) have values in \(L_2(U, H)\).

### 20.7 Remembering The Formula

I find it almost impossible to remember this formula. Here is a way to do it. Recall that \(|\Delta W|^2\) is like \(\Delta t\). Therefore, in what follows, neglect all terms which are like \(dW dt, dt^2\), but keep terms which are \(dW, dt, dW^2\). Then you start with \(dX = \phi dt + \Phi dW\). Thus for \(F(t, X)\),

\[
dF = F_t dt + F_X dX + \frac{1}{2} (F_{XX} dX, dX)
\]

other terms from Taylor’s formula are neglected because they involve \(dt dW\) or \(dt^2\). Now the above equals

\[
dF = F_t dt + F_X (\phi dt + \Phi dW) + \frac{1}{2} (F_{XX} \Phi dW, \Phi dW)
\]

Since the \(dW\) occurs twice, in that inner product, you get a \(dt\) out of it. Hence you get

\[
dF = (F_t + F_X \phi) dt + \frac{1}{2} (F_{XX} \Phi, \Phi) dt + F_X \Phi dW
\]

Now place an \(\int_0^t\) in front of everything and you have the Ito formula.
20.8 An Interesting Formula

Suppose everything is real valued and $\phi$ is progressively measurable and in $L^2([0,T] \times \Omega)$. Let

$$X(t) = \int_0^t \phi dW - \frac{1}{2} \int_0^t \phi^2 ds$$

and consider $F(X) = e^X$. Then from the Ito formula,

$$dF = - \left( e^X \phi^2 \frac{1}{2} \right) dt + \frac{1}{2} e^X \phi^2 dt + e^X \phi dW$$

and then do an integral

$$e^{X(t)} - 1 = \int_0^t e^X \phi dW$$

Thus

$$e^{X(t)} = 1 + \int_0^t e^X(s) \phi dW$$

That expression on the right is obviously a local martingale and so the expression on the left is also. To see this, you can use a localizing sequence of stopping times which depend on the size of $X(t)$. This will work fine because $X(t)$ is continuous.

20.9 Some Representation Theorems

In this section is a very interesting representation theorem which comes from the Ito formula. In all of this, $W$ will be a $Q$ Wiener process having values in $\mathbb{R}^n$ for which $Q = I$. Recall that, letting

$$\mathcal{G}_t \equiv \sigma(W(s) : s \leq t)$$

the normal filtration determined by the Wiener process is given by

$$\mathcal{F}_t \equiv \cap_{s>t} \mathcal{G}_s$$

where $\mathcal{G}_s$ is the completion of $\mathcal{G}_t$. In this section, the theorems will all feature the smaller filtration $\mathcal{G}_t$, not the filtration $\mathcal{F}_t$. First here are some simple observations which tie this specialized material to what was presented earlier.

When you have $f$ an $\mathcal{G}_t$ adapted function in $L^2(\Omega; \mathbb{R}^n)$, you can consider $f^T \in L^2([0,T] \times \Omega; \mathcal{L}_2(Q^{1/2} \mathbb{R}^n, \mathbb{R}))$ as follows. Letting $\{g_i\}$ be an orthonormal basis for the subspace $Q^{1/2} \mathbb{R}^n$, the norm of $Q^{1/2} \mathbb{R}^n$,

$$\|f\|_{\mathcal{L}_2(Q^{1/2} \mathbb{R}^n, \mathbb{R})}^2 \equiv \sum_i (f^T g_i)^2 < \infty$$

For simplicity, let $Q = I$. Then you have the simple situation that

$$\|f^T\|_{\mathcal{L}_2(Q^{1/2} \mathbb{R}^n, \mathbb{R})} = \|f\|_{\mathbb{R}^n}$$

In what follows $W_t$ will be the $Q$ Wiener process on $\mathbb{R}^n$ where $Q = I$. Then the Ito isometry is nothing more than the following lemma.

**Lemma 20.9.1** Let $f$ be $\mathcal{F}_t$ adapted in the sense that every component is $\mathcal{F}_t$ adapted and $f \in L^2(\Omega; \mathbb{R}^n)$. Here $\mathcal{F}_t$ is the normal filtration coming from the Wiener process. Then

$$\left\| \int_0^T f(s)^T dW \right\|_{L^2(\Omega)} = \|f\|_{L^2([0,T] ; \mathbb{R}^n)}.$$
Lemma 20.9.2 Let $X \geq 0$ and measurable and integrable. Also define a finite measure $\nu$ on $B(\mathbb{R}^p)$ by

$$\nu(B) = \int_{\Omega} X \chi_B(Y) \, dP$$

Then

$$\int_{\Omega} g(Y) X \, dP = \int_{\mathbb{R}^p} g(y) \, d\nu(y)$$

where here $Y$ is a measurable function with values in $\mathbb{R}^p$ and $g \geq 0$ is Borel measurable. Formally, $X \, dP = d\nu$.

**Proof:** First say $X = \chi_D$ and replace $g(Y)$ with $\chi_{Y^{-1}(B)}$. Let

$$\mu(B) = \int_{\Omega} \chi_D \chi_B(Y) \, dP$$

Then

$$\int_{\Omega} \chi_D \chi_{Y^{-1}(B)} \, dP = P(D \cap Y^{-1}(B))$$

Thus

$$\int_{\Omega} \chi_D \chi_{Y^{-1}(B)} \, dP = \int_{\Omega} \chi_D \chi_B(Y) \, dP = \int_{\mathbb{R}^p} \chi_B(y) \, d\mu(y)$$

Now let $s_n(y) \uparrow g(y)$, and let $s_n(y) = \sum_{k=1}^{m} c_k \chi_{B_k}(y)$ where $B_k$ is a Borel set. Then

$$\int_{\mathbb{R}^p} s_n(y) \, d\mu(y) = \int_{\mathbb{R}^p} \sum_{k=1}^{m} c_k \chi_{B_k}(y) \, d\mu(y) = \sum_{k=1}^{m} c_k \int_{\mathbb{R}^p} \chi_{B_k}(y) \, d\mu(y)$$

$$= \sum_{k=1}^{m} c_k P(D \cap Y^{-1}(B_k))$$

$$\int_{\Omega} s_n(Y) \chi_D \, dP = \sum_{k=1}^{m} c_k \int_{\Omega} \chi_D \chi_{B_k}(Y) \, dP = \sum_{k=1}^{m} c_k P(D \cap Y^{-1}(B_k))$$

which is the same thing. Therefore,

$$\int_{\Omega} s_n(Y) \chi_D \, dP = \int_{\mathbb{R}^p} s_n(y) \, d\mu(y)$$

Now pass to a limit using the monotone convergence theorem to obtain

$$\int_{\Omega} g(Y) \chi_D \, dP = \int_{\mathbb{R}^p} g(y) \, d\mu(y)$$

Next replace $\chi_D$ with $\sum_{k=1}^{m} d_k \chi_{D_k}(\omega) \equiv s_n(\omega)$, a simple function. Then from what was just shown,

$$\int_{\Omega} g(Y) \sum_{k=1}^{m} d_k \chi_{D_k} \, dP = \sum_{k=1}^{m} d_k \int_{\Omega} g(Y) \chi_{D_k} \, dP$$

$$= \sum_{k=1}^{m} d_k \int_{\mathbb{R}^p} g(y) \, d\mu_k$$

where $\mu_k(B) = \int_{\Omega} \chi_{D_k} \chi_B(Y) \, dP$. Now let

$$\nu_n(B) = \int_{\Omega} \sum_{k=1}^{m} d_k \chi_{D_k} \chi_B(Y) = \int_{\Omega} s_n \chi_B(Y) \, dP$$
Then
\[ \nu_n(B) = \sum_{k=1}^{m} d_k \int_{\Omega} \mathcal{X}_{D_k} \mathcal{X}_B(Y) \, dP = \sum_{k=1}^{m} d_k \mu_k(B) \]

Hence
\[ \int_{\Omega} g(Y) s_n \, dP = \int_{\Omega} g(Y) \sum_{k=1}^{m} d_k \mathcal{X}_{D_k} \, dP = \sum_{k=1}^{m} d_k \int_{\mathbb{R}^p} g(y) \, d\mu_k \]
\[ = \int_{\mathbb{R}^p} g(y) \sum_{k=1}^{m} d_k d\mu_k = \int_{\mathbb{R}^p} g(y) \, d\nu_n \]

Then let \( s_n(\omega) \uparrow X(\omega) \). Clearly \( \nu_n \ll \nu \) and so by the Radon Nikodym theorem \( d\nu_n = h_n \, d\nu \). Then by the monotone convergence theorem, for any \( B \) Borel in \( \mathbb{R}^p \),
\[ \int_{B} h_n \, d\nu = \nu_n(B) \equiv \int_{\Omega} s_n(\omega) \mathcal{X}_B(Y(\omega)) \, dP \uparrow \int_{\Omega} X(\omega) \mathcal{X}_B(Y(\omega)) \, dP \equiv \nu(B) \]
Thus for each \( B \) Borel, \( 0 \leq h_n \leq 1 \) and
\[ \int_{B} h_n \, d\nu \to \nu(B) \]
and so \( h_n \uparrow 1 \) \( \nu \) a.e. Thus, from the above,
\[ \int_{\Omega} g(Y) s_n \, dP = \int_{\mathbb{R}^p} g(y) \, d\nu_n = \int_{\mathbb{R}^p} g(y) \, h_n(y) \, d\nu \]

It follows from the monotone convergence theorem that one can pass to a limit in the above and obtain
\[ \int_{\Omega} g(Y) X \, dP = \int_{\mathbb{R}^p} g(y) \, d\nu \]

Note that the same conclusion will hold if the functions are suitably integrable without any restriction on the sign. In particular, this will hold if \( g(y) \) is bounded. One just considers positive and negative parts of real and imaginary parts of \( g \) and applies the above lemma.

Let
\[ \mathcal{G}_t \equiv \sigma(W(s) : s \leq t) \]
thus the normal filtration for the Wiener process and the Ito integral and so forth is
\[ \mathcal{F}_t = \bigcap_{s > t} \mathcal{G}_s \]

**Lemma 20.9.3** Let \( h \) be a deterministic step function of the form
\[ h = \sum_{i=0}^{m-1} a_i \mathcal{X}_{(t_i, t_{i+1})}, \ t_m = t \]

Then for \( h \) of this form, linear combinations of functions of the form
\[ \exp \left( \int_{0}^{t} h^T dW - \frac{1}{2} \int_{0}^{t} h \cdot h d\tau \right) \]  \hspace{1cm} (20.9.14)
are dense in \( L^2(\Omega, \mathcal{G}_t, P) \) for each \( t \).

**Proof:** I will show in the process of the proof that functions of the form \( \mathcal{X}_{(t_i, t_{i+1})} \) are in \( L^2(\Omega, P) \). Let \( g \in L^2(\Omega, \mathcal{G}_t, P) \) be such that
\[ \int_{\Omega} g(\omega) \exp \left( \int_{0}^{t} h^T dW - \frac{1}{2} \int_{0}^{t} h \cdot h d\tau \right) \, dP \]
\[ = \exp \left( -\frac{1}{2} \int_{0}^{t} h \cdot h d\tau \right) \int_{\Omega} g(\omega) \exp \left( \int_{0}^{t} h^T dW \right) \, dP = 0 \]
for all such \( h \). It is required to show that whenever this happens for all such functions \( \exp \left( \int_0^t h^T dW - \frac{1}{2} \int_0^t h \cdot h dt \right) \) then \( g = 0 \).

Letting \( h \) be given as above, \( \int_0^t h^T dW \)

\[
= \sum_{i=0}^{m-1} a_i^T (W(t_{i+1}) - W(t_i)) \quad (20.9.15)
\]

\[
= \sum_{i=1}^m a_{i-1}^T W(t_i) - \sum_{i=0}^{m-1} a_i^T W(t_i)
\]

\[
= \sum_{i=1}^{m-1} (a_{i-1}^T - a_i^T) W(t_i) + a_0^T W(t_0) + a_{n-1}^T W(t_n). \quad (20.9.16)
\]

Also \( \text{20.9.16} \) shows \( \exp \left( \int_0^t h^T dW \right) \) is in \( L^2(\Omega, P) \). To see this recall the \( W(t_{i+1}) - W(t_i) \) are independent and the density of \( W(t_{i+1}) - W(t_i) \) is

\[
C(n, \Delta t_i) \exp \left( -\frac{1}{2} \frac{|x|^2}{(t_{i+1} - t_i)} \right), \quad \Delta t_i \equiv t_{i+1} - t_i,
\]

so

\[
\int_{\Omega} \left( \exp \left( \int_0^t h^T dW \right) \right)^2 dP = \int_{\Omega} \exp \left( 2 \int_0^t h^T dW \right) dP
\]

\[
= \int_{\Omega} \exp \left( \sum_{i=0}^{m-1} 2a_i^T (W(t_{i+1}) - W(t_i)) \right) dP
\]

\[
= \int_{\Omega} \prod_{i=0}^{m-1} \exp \left( 2a_i^T (W(t_{i+1}) - W(t_i)) \right) dP
\]

\[
= \prod_{i=0}^{m-1} \int_{\Omega} \exp \left( 2a_i^T (W(t_{i+1}) - W(t_i)) \right) dP
\]

\[
= \prod_{i=0}^{m-1} \int_{\mathbb{R}^n} C(n, \Delta t_i) \exp \left( 2a_i^T x \right) \exp \left( -\frac{1}{2} \frac{|x|^2}{2 \Delta t_i} \right) dx < \infty
\]

Choosing the \( a_i \) appropriately in \( \text{20.9.16} \) the formula in \( \text{20.9.16} \) is of the form

\[
\sum_{i=0}^m y_i^T W_{t_i},
\]

where \( y_i \) is an arbitrary vector in \( \mathbb{R}^n \). It follows that for all choices of \( y_j \in \mathbb{R}^n \),

\[
\int_{\Omega} g(\omega) \exp \left( \sum_{j=0}^m y_j^T W_{t_j}(\omega) \right) dP = 0.
\]

Now the mapping

\[
y = (y_0, \cdots, y_m) \rightarrow \int_{\Omega} g(\omega) \exp \left( \sum_{j=0}^m y_j^T W_{t_j}(\omega) \right) dP
\]

is analytic on \( \mathbb{C}^{(m+1)n} \) and equals zero on \( \mathbb{R}^{(m+1)n} \) so from standard complex variable theory, this analytic function must equal zero on \( \mathbb{C}^{(m+1)n} \), not just on \( \mathbb{R}^{(m+1)n} \). In particular, for all \( y = (y_0, \cdots, y_m) \in \mathbb{R}^{n(m+1)} \),

\[
\int_{\Omega} g(\omega) \exp \left( \sum_{j=0}^m iy_j^T W_{t_j}(\omega) \right) dP = 0. \quad (20.9.17)
\]
This left side equals

\[ \int_{\Omega} g_+ (\omega) \exp \left( \sum_{j=0}^{m} iy_j^T W_{t_j} (\omega) \right) dP - \int_{\Omega} g_- (\omega) \exp \left( \sum_{j=0}^{m} iy_j^T W_{t_j} (\omega) \right) dP \]

where \( g_+ \) and \( g_- \) are the positive and negative parts of \( g \). By the Lemma 20.9.2 and the observation at the end, this equals

\[ \int_{\mathbb{R}^{nm}} \exp \left( \sum_{j=0}^{m} iy_j^T x_j \right) d\nu_+ - \int_{\mathbb{R}^{nm}} \exp \left( \sum_{j=0}^{m} iy_j^T x_j \right) d\nu_- \]

where \( \nu_+ (B) = \int_{\Omega} g_+ (\omega) \chi_B (W_{t_1} (\omega), \ldots, W_{t_m} (\omega)) dP \) and \( \nu_- \) is defined similarly. Then letting \( \nu \) be the measure \( \nu_+ - \nu_- \), it follows that

\[ 0 = \int_{\mathbb{R}^{nm}} \exp \left( \sum_{j=0}^{m} iy_j^T x_j \right) d\nu (y) \]

and this just says that the inverse Fourier transform of \( \nu \) is 0. It follows that \( \nu = 0 \). Thus

\[ \int_{\Omega} g (\omega) \chi_B (W_{t_1} (\omega), \ldots, W_{t_m} (\omega)) dP = 0 \]

for every \( B \) Borel in \( \mathbb{R}^{nm} \) where

\[ W_m (\omega) \equiv (W_{t_1} (\omega), \ldots, W_{t_m} (\omega)) \]

Let \( K \) be the \( \pi \) system defined as \( W_m^{-1} (B) \) for \( B \) of the form \( \bigcap_{i=1}^{m} U_i \) where \( U_i \) is open in \( \mathbb{R}^n \), this for some \( m \) a positive integer. This is indeed a \( \pi \) system because it includes \( W_1^{-1} (\mathbb{R}^n) = \Omega \) and the empty set. Also it is closed with respect to intersections because, in the situation where each \( s_i \) is larger than every \( t_i \),

\[ (W_{t_1}, \ldots, W_{t_{m_1}})^{-1} \left( \bigcap_{i=1}^{m_1} U_i \right) \cap (W_{s_1}, \ldots, W_{s_{m_2}})^{-1} \left( \bigcap_{i=1}^{m_2} V_i \right) = \]

\[ (W_{t_1}, \ldots, W_{t_{m_1}}, W_{s_1}, \ldots, W_{s_{m_2}})^{-1} \left( \bigcap_{i=1}^{m_1} U_i \times \prod_{k=1}^{m_2} \mathbb{R}^n \right) \]

\[ \cap \left( (W_{t_1}, \ldots, W_{t_{m_1}}, W_{s_1}, \ldots, W_{s_{m_2}})^{-1} \left( \prod_{i=1}^{m_1} \mathbb{R}^n \times \prod_{k=1}^{m_2} V_i \right) \right) \]

\[ = (W_{t_1}, \ldots, W_{t_{m_1}}, W_{s_1}, \ldots, W_{s_{m_2}})^{-1} \left( \bigcap_{i=1}^{m_1} U_i \times \prod_{k=1}^{m_2} V_k \right) \]

In general, you would just make the obvious modification where you insert a copy of \( \mathbb{R}^n \) in the appropriate position after rearranging so that the indices are increasing. It was just shown that \( K \subseteq \mathcal{G} \) where

\[ \mathcal{G} \equiv \left\{ U \in \mathcal{G}_t : \int_{\Omega} g \chi_U dP = 0 \right\} \]

Now it is clear that \( \mathcal{G} \) is closed with respect to countable disjoint unions and complements. The case of complements goes as follows. \( \Omega \in \mathcal{K} \) and so if \( U \in \mathcal{G} \),

\[ \int_{\Omega} g \chi_{U^c} dP + \int_{\Omega} g \chi_U dP = \int_{\Omega} g dP \]

The last on the left and the integral on the right are both 0 so it follows that \( \int_{\Omega} g \chi_{U^c} dP = 0 \) also. It follows from Dynkin’s lemma that \( \mathcal{G} \supseteq \sigma (\mathcal{K}) \). Now \( \sigma (\mathcal{K}) \) is \( \sigma (W (u) : u \leq t) \equiv \mathcal{G}_t \). Hence, \( \mathcal{G} = \mathcal{G}_t \) and so \( g \) is in \( L^2 (\Omega, \mathcal{G}_t) \) and for every \( U \in \mathcal{G}_t \),

\[ \int_{\Omega} g \chi_U dP = 0 \]
which requires $g = 0$. Thus functions of the above form are indeed dense in $L^2(\Omega, \mathcal{G}_t)$. ■

Note that this involves $g$ being $\mathcal{G}_t$ measurable, not $\mathcal{F}_t$ measurable. It is not clear to me whether it suffices to assume only that $g$ is $\mathcal{F}_t$ measurable. If true, this above has not proved it. The problem is the argument at the end using Dynkin’s lemma to conclude that $g = 0$.

Why such a funny lemma? It is because of the following computation which depends on Itô’s formula. Let

$$X = \int_0^t h^T dW - \frac{1}{2} \int_0^t h \cdot h d\tau$$

and $g(x) = e^x$ and consider $g(X) = Y$. Recall the Ito formula. Formally,

$$dY = g'(X) dX + g''(X) (dX)^2$$

$$dY = g(X) \left( h^T dW - \frac{1}{2} |h|^2 dt \right)$$

$$dY = \frac{1}{2} g(X) \left( h^T dW - \frac{1}{2} |h|^2 dt \right) \left( h^T dW - \frac{1}{2} |h|^2 dt \right)$$

$$= Y \left( h^T dW - \frac{1}{2} |h|^2 dt \right) + \frac{1}{2} Y \left( (h^T dW) (h^T dW) - h^T dW |h|^2 dt + \frac{1}{4} |h|^2 dt^2 \right)$$

Then neglecting the terms of the form $dW dt$. $dt^2$ and so forth,

$$dY = Y h^T dW - \frac{1}{2} Y |h|^2 dt + \frac{1}{2} Y (h^T dW) (h^T dW)$$

Now the $dW$ occurs twice in the last term so it leads to a $dt$ and you get

$$dY = Y h^T dW - \frac{1}{2} Y |h|^2 dt + \frac{1}{2} (Y h^T, h^T) dt$$

$$dY = Y h^T dW - \frac{1}{2} Y |h|^2 dt + \frac{1}{2} Y |h|^2 dt$$

$$dY = Y h^T dW$$

Note that $\| h^T \|_{L_2(\Omega, \mathcal{G}_t, P)} = \sum_{k=1}^n (h^T e_k)^2 = |h|^2_{R^n}$. Place an $\int_0^t$ in place of both sides to obtain

$$Y(t) - Y(0) = \int_0^t Y h^T dW$$

$$Y(t) = 1 + \int_0^t Y h^T dW$$

(20.9.18)

Now here is the interesting part of this formula.

$$E \left( \int_0^t Y h^T dW \right) = 0$$

because the stochastic integral is a martingale and equals 0 at $t = 0$.

$$E \left( \int_0^t Y h^T dW \right) = E \left( E \left( \int_0^t Y h^T dW \mid \mathcal{F}_0 \right) \right) = 0$$

Thus

$$E(Y(t)) = 1$$

and for $Y$ one obtains

$$Y(t) = E(Y(t)) + \int_0^t Y h^T dW$$

$$= E(Y(t)) + \int_0^t f^T dW$$

where $f^T$ is adapted and square integrable. It is just $Y h^T$ where $h$ does not depend on $\omega$ and $Y$ is a function of an adapted function.

Does such a function $f$ exist for all $F \in L^2(\Omega, \mathcal{G}_t, P)$? The answer is yes and this is the content of the next theorem which is called the Itô representation theorem.
Theorem 20.9.4 Let \( F \in L^2(\Omega, \mathcal{G}_t, P) \). Then there exists a unique \( \mathcal{G}_t \) adapted \( f \in L^2(\Omega \times [0,t]; \mathbb{R}^n) \) such that

\[
F = E(F) + \int_0^t f(s,\omega)^T \, dW.
\]

Proof: By Lemma 21.9, the span of functions of the form

\[
\exp \left( \int_0^t h^T \, dW - \frac{1}{2} \int_0^t h \cdot h \, dt \right)
\]

where \( h \) is a vector valued deterministic step function of the sort described in this lemma, are dense in \( L^2(\Omega, \mathcal{G}_t, P) \). Given \( F \in L^2(\Omega, \mathcal{G}_t, P) \), \( \{G_k\}_{k=1}^{\infty} \) be functions in the subspace of linear combinations of the above functions which converge to \( F \) in \( L^2(\Omega, \mathcal{G}_t, P) \). For each of these functions there exists \( f_k \) an adapted step function such that

\[
G_k = E(G_k) + \int_0^t f_k(s,\omega)^T \, dW.
\]

Then from the Itô isometry, and the observation that \( E(G_k-G_l)^2 \to 0 \) as \( k,l \to \infty \) by the above definition of \( G_k \) in which the \( G_k \) converge to \( F \) in \( L^2(\Omega) \),

\[
0 = \lim_{k,l \to \infty} E \left( (G_k-G_l)^2 \right) = \lim_{k,l \to \infty} E \left( \left( E(G_k) + \int_0^t f_k(s,\omega)^T \, dW - \left( E(G_l) + \int_0^t f_l(s,\omega)^T \, dW \right) \right)^2 \right)
\]

\[
= \lim_{k,l \to \infty} \left\{ E(G_k-G_l)^2 + 2E(G_k-G_l) \int_\Omega \int_0^t (f_k-f_l)^T \, dW \, dP \right. \\
+ \left. \int_\Omega \left( \int_0^t (f_k-f_l)^T \, dW \right)^2 \, dP \right\}
\]

\[
= \lim_{k,l \to \infty} \left\{ E(G_k-G_l)^2 + \int_\Omega \left( \int_0^t (f_k-f_l)^T \, dW \right)^2 \, dP \right\} = \lim_{k,l \to \infty} \int_\Omega \left( \int_0^t (f_k-f_l)^T \, dW \right)^2 \, dP = \lim_{k,l \to \infty} \|f_k-f_l\|_{L^2(\Omega \times [0,T]; \mathbb{R}^n)}
\]

(20.9.19)

Going from the third to the fourth equations, is justified because

\[
\int_\Omega \int_0^t (f_k-f_l)^T \, dW \, dP = 0
\]

thanks to the fact that the Ito integral is a martingale which equals 0 at \( t = 0 \).

This shows \( \{f_k\}_{k=1}^{\infty} \) is a Cauchy sequence in \( L^2(\Omega \times [0,t]; \mathbb{R}^n, \mathcal{P}) \), where \( \mathcal{P} \) denotes the progressively measurable sets. It follows there exists a subsequence and \( f \in L^2(\Omega \times [0,t]; \mathbb{R}^n) \) such that \( f_k \) converges to \( f \) in \( L^2(\Omega \times [0,t]; \mathbb{R}^n, \mathcal{P}) \) with \( f \) being progressively measurable. Then by the Itô isometry and the equation

\[
G_k = E(G_k) + \int_0^t f_k(s,\omega)^T \, dW
\]

you can pass to the limit as \( k \to \infty \) and obtain

\[
F = E(F) + \int_0^t f(s,\omega)^T \, dW
\]
Now \( E(G_k) \to E(F) \). Consider the stochastic integrals. By the maximal estimate, Theorem [Kallenberg, and the Ito isometry,]

\[
P \left( \sup_{s \in [0,t]} \left| \int_0^s f_k(\cdot,\omega)^T dW - \int_0^s f(\cdot,\omega)^T dW \right| > \delta \right) < \frac{E \left( \left| \int_0^t f_k(\cdot,\omega)^T dW - \int_0^t f(\cdot,\omega)^T dW \right|^2 \right)}{\delta^2}
\]

From the above convergence result and an application of the Borel Cantelli lemma, there is a set of measure zero \( N \) and a subsequence, still denoted as \( f_k \), such that for \( \omega \notin N \), the convergence of the stochastic integrals for this subsequence is uniform. Thus for \( \omega \notin N \),

\[
F = E(F) + \int_0^t f(s,\omega)^T dW
\]

This proves the existence part of this theorem.

It remains to consider the uniqueness. Suppose then that

\[
F = E(F) + \int_0^T f(t,\omega)^T dW = E(F) + \int_0^T f_1(t,\omega)^T dW.
\]

Then

\[
\int_0^T f(t,\omega)^T dW = \int_0^T f_1(t,\omega)^T dW
\]

and so

\[
\int_0^T (f(t,\omega)^T - f_1(t,\omega)^T) dW = 0
\]

and by the Itô isometry,

\[
0 = \left\| \int_0^T (f(t,\omega)^T - f_1(t,\omega)^T) dW \right\|_{L^2(\Omega)} = \left\| f - f_1 \right\|_{L^2(\Omega \times [0,T];\mathbb{R}^n)}
\]

which proves uniqueness. \( \blacksquare \)

With the above major result, here is another interesting representation theorem. Recall that if you have an \( \mathcal{F}_t \) adapted function \( f \) and \( f \in L^2(\Omega \times [0,T];\mathbb{R}^n) \), then \( \int_0^T f^T dW \) is a martingale. The next theorem is sort of a converse. It starts with a \( \mathcal{G}_t \) martingale and represents it as an Itô integral. In this theorem, \( \mathcal{G}_t \) continues to be the filtration determined by \( n \) dimensional Wiener process.

**Theorem 20.9.5** Let \( M \) be an \( \mathcal{G}_t \) martingale and suppose \( M(t) \in L^2(\Omega) \) for all \( t \geq 0 \). Then there exists a unique stochastic process, \( g(s,\omega) \) such that \( g \) is \( \mathcal{G}_t \) adapted and in \( L^2(\Omega \times [0,t]) \) for each \( t > 0 \), and for all \( t \geq 0 \),

\[
M(t) = E(M(0)) + \int_0^t g^T dW
\]

**Proof:** First suppose \( f \) is an adapted function of the sort that \( g \) is. Then the following claim is the first step in the proof.

**Claim:** Let \( t_1 < t_2 \). Then

\[
E \left( \int_{t_1}^{t_2} f^T dW | \mathcal{G}_{t_1} \right) = 0
\]

**Proof of claim:** This follows from the fact that the Ito integral is a martingale adapted to \( \mathcal{G}_t \). Hence the above reduces to

\[
E \left( \int_0^{t_2} f^T dW - \int_0^{t_1} f^T dW | \mathcal{G}_{t_1} \right) = \int_0^{t_1} f^T dW - \int_0^{t_1} f^T dW = 0.
\]
Then there exists a Borel function, $g$ such that

$$M(t) = \mathbb{E}(M(0)) + \int_0^t f(s, \cdot)^T dW.$$ 

Now let $t_1 < t_2$. Then since $M$ is a martingale and so is the Ito integral,

$$M(t_1) = \mathbb{E}(M(t_2)|\mathcal{G}_{t_1}) = \mathbb{E}\left(\mathbb{E}(M(0)) + \int_0^{t_2} f(t_2^2(s, \cdot))^T dW|\mathcal{G}_{t_1}\right)$$

and so

$$0 = \int_0^{t_1} f(t_1^2(s, \cdot))^T dW - \int_0^{t_1} f(t_2^2(s, \cdot))^T dW$$

and so by the Itô isometry,

$$\|f(t_1) - f(t_2)\|_{L^2(\Omega \times [0, t_1]; \mathbb{R}^n)} = 0.$$ 

Letting $N \in \mathbb{N}$, it follows that

$$M(t) = \mathbb{E}(M(0)) + \int_0^t f^N(s, \cdot)^T dW$$

for all $t \leq N$. Let $g = f^N$ for $t \in [0, N]$. Then aside from a set of measure zero, this is well defined and for all $t \geq 0$

$$M(t) = \mathbb{E}(M(0)) + \int_0^t g(s, \cdot)^T dW \quad \blacksquare$$

Surely this is an incredible theorem. Note that it implies all the martingales adapted to $\mathcal{G}_t$ which are in $L^2$ for each $t$ must be continuous a.e. and are obtained from an Ito integral. Also, any such martingale satisfies $M(0) = \mathbb{E}(M(0))$. Isn’t that amazing? Also note that this featured $\mathbb{R}^n$ as where $W$ has its values and $n$ was arbitrary. One could have $n = 1$ if desired.

The above theorems can also be obtained from another approach. It involves showing that random variables of the form

$$\phi(W(t_1), \cdots, W(t_k))$$

are dense in $L^2(\Omega, \mathcal{G}_T)$. This theorem is interesting for its own sake and it involves interesting results discussed earlier. Recall the Doob Dynkin lemma, Lemma 10.2.4 on Page 101 which is listed here.

**Lemma 20.9.6** Suppose $X, Y_1, Y_2, \cdots, Y_k$ are random vectors, $X$ having values in $\mathbb{R}^n$ and $Y_j$ having values in $\mathbb{R}^{p_j}$ and

$$X, Y_j \in L^1(\Omega).$$

Suppose $X$ is $\sigma(Y_1, \cdots, Y_k)$ measurable. Thus

$$\{X^{-1}(E) : E \text{ Borel}\} \subseteq \left\{(Y_1, \cdots, Y_k)^{-1}(F) : F \text{ Borel in } \prod_{j=1}^k \mathbb{R}^{p_j}\right\}$$

Then there exists a Borel function, $g : \prod_{j=1}^k \mathbb{R}^{p_j} \to \mathbb{R}^n$ such that

$$X = g(Y).$$

Recall also the submartingale convergence theorem.
Theorem 20.9.7 (submartingale convergence theorem) Let 
\[ \{(X_t, S_t)\}_{t=1}^{\infty} \]
be a submartingale with \( K \equiv \sup E(|X_t|) < \infty \). Then there exists a random variable \( X \), such that \( E(|X|) \leq K \) and 
\[ \lim_{n \to \infty} X_n(\omega) = X(\omega) \text{ a.e.} \]

Recall 
\[ G_t = \sigma(W(u) - W(r) : 0 \leq r < u \leq t) \]

It suffices to consider only \( t \) and \( u, r \) in a countable dense subset of \( \mathbb{R} \) denoted as \( D \). This follows from continuity of the Wiener process. To see this, let \( 0 \leq r < u \leq t \) be open and \( U_n \uparrow U \) where each \( U_n \) is open and \( U_n \subseteq U_{n+1}, U_n U_m = U \). Then letting \( u_n \uparrow u \) and \( r_n \uparrow r, u_n r_n \) being in the countable dense set,

\[
(W(u) - W(r))^{-1}(U_n) \subseteq \bigcup_{i=1}^{\infty} (W(u_i) - W(r_i))^{-1}(U_n) \subseteq (W(u) - W(r))^{-1}(U_n)
\]

and so

\[
(W(u) - W(r))^{-1}(U) = \bigcup_{i=1}^{\infty} (W(u_i) - W(r_i))^{-1}(U_n)
\]

Now the set in the middle which has two countable unions and a countable intersection is in

\[
\sigma(W(u) - W(r) : 0 \leq r < u \leq t, r, u \in D)
\]

Thus in particular, one would get the same filtration from 
\[ G_t = \sigma(W(u) - W(r) : 0 \leq r < u \leq t, r, u \in D) \]

Since \( W(0) = 0 \), this is the same as 
\[ G_t = \sigma(W(u) : 0 \leq u \leq t, r \in D) \]

Lemma 20.9.8 Random variables of the form 
\[ \phi(W(t_1), \cdots, W(t_k)), \phi \in C_c^\infty(\mathbb{R}^k) \]
are dense in \( L^2(\Omega, G_T, P) \) where \( t_1 < t_2 \cdots < t_k \) is a finite increasing sequence of \( (\mathbb{Q} \cup \{T\}) \cap [0, T] \).

Proof: Let \( g \in L^2(\Omega, G_T, P) \). Also let \( \{t_j\}_{j=1}^{\infty} \) be the points of \( (\mathbb{Q} \cup \{T\}) \cap [0, T] \). Let 
\[ G_m \equiv \sigma(W(t_k) : k \leq m) \]

Thus the \( G_m \) are increasing but each is generated by finitely many \( W(t_k) \). Also as explained above,

\[
G_T = \sigma(W(u) : 0 \leq u \leq T, u \in (\mathbb{Q} \cup \{T\}) \cap [0, T])
\]

Now consider the martingale, 
\[ \{E(g_m | G_m)\}_{m=1}^{\infty} \]

where here
\[
g_M(\omega) \equiv \begin{cases} 
g(\omega) & \text{if } g(\omega) \in [-M, M] \\
M & \text{if } g(\omega) > M \\
-M & \text{if } g(\omega) < -M \end{cases}
\]

and \( M \) is chosen large enough that
\[
\|g - g_M\|_{L^2(\Omega)} < \varepsilon/4. \quad (20.9.20)
\]
Now the terms of this martingale are uniformly bounded by $M$ because
$$|E(g_M|\mathcal{G}_m)| \leq E(|g_M| |\mathcal{G}_m)| \leq E(M|\mathcal{G}_m) = M.$$ 

It follows the martingale is certainly bounded in $L^1$ and so the martingale convergence theorem stated above can be applied, and so there exists $f$ measurable in $\sigma(\mathcal{G}_m, m < \infty)$ such that $\lim_{m \to \infty} E(g_M|\mathcal{G}_m)(\omega) = f(\omega)$ a.e. Also $|f(\omega)| \leq M$ a.e. Since all functions are bounded, it follows that this convergence is also in $L^2(\Omega)$.

Now letting $A \in \sigma(\mathcal{G}_m, m < \infty)$, it follows from the dominated convergence theorem that
$$\int_A f dP = \lim_{m \to \infty} \int_A E(g_M|\mathcal{G}_m) dP = \int_A g_M dP.$$

Now letting $A \in \sigma(\mathcal{G}_m, m < \infty)$, it follows from the dominated convergence theorem that
$$\int_A f dP = \lim_{m \to \infty} \int_A E(g_M|\mathcal{G}_m) dP = \int_A g_M dP.$$ 

By the Doob Dynkin lemma listed above, there exists a Borel measurable $h : \mathbb{R}^{nm} \to \mathbb{R}$ such that
$$E(g_M|\mathcal{G}_m) = h(W_{t_1}, \cdots, W_{t_m}) \text{ a.e.}$$

Of course $h$ is not in $C^\infty_c(\mathbb{R}^{nm})$. Let $m$ be large enough that
$$||g_M - E(g_M|\mathcal{G}_m)||_L^2 = ||f - E(g_M|\mathcal{G}_m)||_L^2 < \frac{\varepsilon}{4}.$$ 

Let $\lambda(W_{t_1}, \cdots, W_{t_m})$ be the distribution measure of the random vector $(W_{t_1}, \cdots, W_{t_m})$. Thus $\lambda(W_{t_1}, \cdots, W_{t_m})$ is a Radon measure and so there exists $\phi \in C_c(\mathbb{R}^{nm})$ such that
$$\left( \int_\Omega |E(g_M|\mathcal{G}_m) - \phi(W_{t_1}, \cdots, W_{t_m})|^2 dP \right)^{1/2} = \left( \int_\Omega |h(W_{t_1}, \cdots, W_{t_m}) - \phi(W_{t_1}, \cdots, W_{t_m})|^2 dP \right)^{1/2} = \left( \int_{\mathbb{R}^{nm}} |h(x_1, \cdots, x_m) - \phi(x_1, \cdots, x_m)|^2 d\lambda(W_{t_1}, \cdots, W_{t_m}) \right)^{1/2} < \varepsilon/4.$$ 

By convolving with a mollifier, one can assume that $\phi \in C^\infty_c(\mathbb{R}^{nm})$ also. It follows from (20.9.21) and (20.9.22) that
$$||g - \phi(W_{t_1}, \cdots, W_{t_m})||_L^2 \leq ||g - g_M||_L^2 + ||g_M - E(g_M|\mathcal{G}_m)||_L^2 + ||E(g_M|\mathcal{G}_m) - \phi(W_{t_1}, \cdots, W_{t_m})||_L^2 \leq 3 \left( \frac{\varepsilon}{4} \right) < \varepsilon.$$ 

By convolving with a mollifier, one can assume that $\phi \in C^\infty_c(\mathbb{R}^{nm})$ also. It follows from (20.9.21) and (20.9.22) that
$$||g - \phi(W_{t_1}, \cdots, W_{t_m})||_L^2 \leq ||g - g_M||_L^2 + ||g_M - E(g_M|\mathcal{G}_m)||_L^2 + ||E(g_M|\mathcal{G}_m) - \phi(W_{t_1}, \cdots, W_{t_m})||_L^2 \leq 3 \left( \frac{\varepsilon}{4} \right) < \varepsilon.$$
Chapter 21

Gelfand Triples

Let $H$ be a separable real Hilbert space and let $V \subseteq H$ be a separable Banach space which is embedded continuously into $H$ and which is also dense in $H$. Then identifying $H$ and $H'$ you can write

$$V \subseteq H = H' \subseteq V'.$$

This is called a Gelfand triple. If $V$ is reflexive, you could conclude separability of $V$ from the separability of $H$. However, if $V$ is not reflexive, this might not happen. For example, you could take $V = L^\infty (0,1)$ and $H = L^2 (0,1)$.

**Proposition 21.0.9** Suppose $V$ is reflexive and a subset of $H$ a separable Hilbert space with the inclusion map continuous. Suppose also that $V$ is dense in $H$. Then identifying $H$ and $H'$, it follows that $H$ is dense in $V'$ and $V$ is separable.

**Proof:** If $H$ is not dense in $V'$, then by the Hahn Banach theorem, there exists $\phi^{**} \in V''$ such that $\phi^{**} (H) = 0$ but $\phi^{**} (\phi^*) \neq 0$ for some $\phi^* \in V' \setminus H$. Since $V$ is reflexive there exists $v \in V$ such that $\phi^{**} = Jv$ for $J$ the standard mapping from $V$ to $V''$. Thus

$$\phi^{**} (h) \equiv \langle h, v \rangle \equiv \langle v, h \rangle_H = 0 \quad \text{for all } h \in H.$$

Therefore, $v = 0$ and so $Jv = 0 = \phi^{**}$ which contradicts $\phi^{**} (\phi^*) \neq 0$. Therefore, $H$ is dense in $V'$. Now by Theorem [19, 19], which says separability of the dual space implies separability of the space, it follows $V$ is separable as claimed. This proves the proposition.

From now on, it is assumed $V$ and $V'$ are both separable and that $H$ is dense in $V'$. This is summarized in the following definition.

**Definition 21.0.10** $V,H,V'$ will be called a Gelfand triple if $V,V'$ are separable, $V \subseteq H$ with the inclusion map continuous, $H = H'$, and $H = H'$ is dense in $V'$.

What about the Borel sets on $V$ and $H$?

**Proposition 21.0.11** Denote by $\mathcal{B}(X)$ the Borel sets of $X$ where $X$ is any separable Banach space. Then

$$\mathcal{B}(X) = \sigma (X').$$

Here $\sigma (X')$ is the smallest $\sigma$ algebra such that each $\phi \in X'$ is measurable. Also in the context of the above definition, $\mathcal{B}(V) = \sigma (i^* H')$ because $H'$ is dense in $V'$. Here $i^*$ is the restriction to $V$ so that $i^* h (v) \equiv h (v) \equiv (h,v)_H$ for all $v \in V$ and $\sigma (i^* H')$ denotes the smallest $\sigma$ algebra such that $i^* h$ is measurable for each $h \in H'$.

**Proof:** By Lemma [19, 19] there exists a countable subset of the unit ball in $X'$

$$\{ \phi_n \}_{n=1}^\infty = D'$$

such that

$$||v||_X = \sup \{ |\phi (v)| : \phi \in D' \}.$$

Consider a closed ball $\overline{B}(v_0, r)$ in $X$. This equals

$$\left\{ v \in X : \sup_n |\phi_n (v) - \phi_n (v_0)| \leq r \right\} = \bigcap_{n=1}^\infty \phi_n^{-1}\left( \overline{B}(\phi_n (v_0), r) \right)$$

461
and this last set is in \( \sigma (D') \). Therefore, every closed ball is in \( \sigma (D') \) which implies every open ball is also in \( \sigma (D') \) since open balls are the countable union of closed balls. Since \( X \) is separable, it follows every open set is the countable union of balls and so every open set is in \( \sigma (D') \). It follows \( \mathcal{B}(X) \subseteq \sigma (D') \subseteq \sigma (X') \). On the other hand, every \( \phi \in X' \) is continuous and so it is Borel measurable. Hence \( (X')' \subseteq \mathcal{B}(X) \).

Now consider the last claim. From Lemma \ref{lem} and density of \( H' = H \) in \( V' \), it can be assumed \( D' \subseteq H = H' \). Therefore, from the first part of the argument
\[
\mathcal{B}(V) \subseteq \sigma (D') \subseteq \sigma (i^*H')
\]
Also each \( i^*h \) is continuous on \( V \) so in fact, equality holds in the above because \( \sigma (i^*H') \subseteq \mathcal{B}(V) \). This proves the proposition.

Next I want to verify that \( V \) is in \( \mathcal{B}(H) \). This will be true if \( V \) is reflexive. More generally, here is an interesting result.

**Proposition 21.0.12** Let \( X \subseteq Y \), \( X \) dense in \( Y \) and suppose \( X, Y \) are Banach spaces and that \( X \) is reflexive. Then \( X \in \mathcal{B}(Y) \).

**Proof:** Define the functional
\[
\phi \equiv \left\{ \begin{array}{ll}
\|x\|_X & \text{if } x \in X \\
\infty & \text{if } x \in Y \setminus X
\end{array} \right.
\]
Then \( \phi \) is lower semicontinuous on \( Y \). Here is why. Suppose \( (x, a) \notin \text{epi} (\phi) \) so that \( a < \phi (x) \). I need to verify this situation persists for \( (x, b) \) near \( (x, a) \). If this is not so, there exists \( x_n \to x \) and \( a_n \to a \) such that \( a_n \geq \phi (x_n) \). If \( \liminf_{n \to \infty} \phi (x_n) < \infty \), then there exists a subsequence still denoted by \( n \) such that \( \|x_n\|_X \) is bounded. Then by the Eberlein Smulian theorem, there exists a further subsequence such that \( x_n \) converges weakly in \( X \) to some \( z \). Now since \( X \) is dense in \( Y \) it follows \( Y' \) can be considered a subspace of \( X' \) and so for \( f \in Y' \)
\[
f (x_n) \to f (z) , \ f (x_n) \to f (x)
\]
and so \( f (z - x) = 0 \) for all \( f \in Y' \) which requires \( z = x \). Now \( x \to \|x\|_X \) is convex and lower semicontinuous on \( X \) so it follows from Corollary \ref{cor}
\[
a = \lim \inf_{n \to \infty} a_n \geq \lim \inf_{n \to \infty} \phi (x_n) \geq \phi (x) > a
\]
which is a contradiction. If \( \lim\inf_{n \to \infty} \phi (x_n) = \infty \), then
\[
\infty > a = \lim \inf_{n \to \infty} a_n = \infty
\]
another contradiction. Therefore, \( \text{epi} (\phi) \) is closed and so \( \phi \) is lower semicontinuous as claimed. Therefore,
\[
X = Y \setminus \left( \cap_{n=1}^{\infty} \phi^{-1} ((n, \infty)) \right)
\]
and since \( \phi \) is lower semicontinuous, each \( \phi^{-1} ((n, \infty)) \) is open. Hence \( X \) is a Borel subset of \( Y \). This proves the proposition.

### 21.1 An Unnatural Example

Recall Gelfand triples are of the form
\[
V \subseteq H \subseteq V'
\]
where \( H \) is a Hilbert space and \( V \) is a Banach space contained in \( H \) and each of the above inclusions is continuous and each space is dense in the next one. The standard example of a Gelfand triple is \( H^1_0 (D) \subseteq L^2 (D) \subseteq (H^1_0 (D))' \) with the convention that \( L^2 (D) \) is identified with its dual space. Thus for \( f \in L^2 (D) \), \( f \) is considered as something in \( (H^1_0 (D))' \) according to the rule
\[
(f, \phi) \equiv (f, \phi)_{L^2(D)}
\]
This is a very pleasant thing to contemplate and it is natural and transparent. However, there are other ways to come up with a Gelfand triple which are much more perverse. The following is an example of such a thing along with an application. See \[\text{[2]}\] and references given there.
First consider the following situation.

\[ X \overset{\theta}{\rightarrow} Y \]

where \( \theta \) is continuous, linear and one to one and \( X \) is a Banach space. Then \( \theta (X) \subseteq Y \) and you could define

\[ \|\theta x\|_{\theta(X)} \equiv \|x\|_X . \]

Then \( \theta (X) \) can be considered the same thing as \( X \) because \( \theta \) preserves distances and all algebraic properties. Thus people write \( X \subseteq Y \) to save space. In the above simple example, it is obvious what \( \theta \) is. This is because the things in \( H^0_0 \) and things in \( L^2 \) are both functions defined on \( D \) and we can simply take \( \theta \) to be the identity map. However, you might have \( H \) be the dual space of something. Thus it consists of bounded linear transformations defined on some Banach space. Then it becomes necessary to specify the manner in which vectors in \( V \) can be considered as vectors of \( H \).

Let \( \infty > p \geq 2 \). Then letting \( D \) be a bounded open set, \( H^1_0 (D) \) embeds continuously into \( L^{p'} (D) \). That is

\[ \|\phi\|_{L^{p'}} \leq C \|\phi\|_{H^1_0} . \]  \hspace{1cm} (21.1.1)

Here \( \frac{1}{p} + \frac{1}{p'} = 1 \). Also note that an equivalent inner product on \( H^1_0 (D) \) is

\[ (f,g)_{H^1_0} \equiv \int_D \nabla f \cdot \nabla g dx \]

Then with respect to this inner product, the Riesz map is given by \( -\Delta \).

\[ -\Delta : H^1_0 (D) \rightarrow (H^1_0 (D))' \]

Thus a typical vector of \( (H^1_0 (D))' \) is of the form \( -\Delta \phi \) where \( \phi \in H^1_0 (D) \) and the following hold.

\[ (\phi,\psi)_{H^1_0} \equiv \langle -\Delta \phi, \psi \rangle , \quad (\Delta \phi, -\Delta \psi)_{H^1_0} \equiv (\phi,\psi)_{H^1_0} = \langle -\Delta \psi, \phi \rangle \]

The following is about the Gelfand triple

\[ V = L^p (D) \subseteq (H^1_0)' \subseteq (L^p (D))' \]

**Lemma 21.1.1** It is possible to consider \( L^p (D) \equiv V \) as a dense subspace of \( (H^1_0)' \equiv H \) as follows. For \( f \in L^p (D) \) and \( \phi \in H^1_0 (D) \),

\[ \langle f, \phi \rangle \equiv \int_D f (x) \phi (x) dx \]

One can also consider \( H \equiv (H^1_0)' \) as a dense subspace of \( (L^p (D))' \equiv V' \) as follows. For \( -\Delta \phi \in H \) and \( f \in L^p (D) \),

\[ \langle -\Delta \phi, f \rangle \equiv (\Delta \phi,f)_H \equiv \langle f, \phi \rangle \]

\( -\Delta \) maps \( H^1_0 (D) \) to \( H \equiv (H^1_0)' \subseteq V' \). \( -\Delta \) can be extended to yield a map \( -\Delta_1 \) from \( L^{p'} (D) \) to \( V' \).

\[ H^1_0 (D) \overset{-\Delta}{\rightarrow} (H^1_0)' \]

\[ L^{p'} (D) = V \overset{-\Delta_1}{\rightarrow} V' \]

**Proof:** First of all, note that by (21.1.1)

\[ |\langle f, \phi \rangle| \leq \|f\|_{L^p} \|\phi\|_{L^{p'}} \leq C \|f\|_{L^p} \|\phi\|_{H^1_0} \]

and so it is certainly possible to consider \( L^p \subseteq H \equiv (H^1_0)' \) as just claimed. Now why can \( L^p (D) \) be considered dense in \( H \equiv (H^1_0)' \)? If it isn’t dense, then there exists \( \psi \in H^1_0 (D), \psi \neq 0 \) such that

\[ \langle -\Delta \psi, f \rangle_H = 0 \]
for all \( f \in L^p(D) \). However, the above would say that for all \( f \in L^p \),
\[
(-\Delta \psi, f)_H \equiv \langle f, \psi \rangle \equiv \int_D f \psi = 0
\]

But \( \psi \in L^{p'}(D) \) because \( H^1_0(D) \) embeds continuously into \( L^{p'}(D) \) and so the above holding for all \( f \in L^p(D) \) implies by the usual Riesz representation theorem that \( \psi = 0 \) contrary to the way \( \psi \) was chosen.

Now consider the next claim. For \(-\Delta \phi \in H \equiv (H^1_0(D))' \) and \( f \in L^p(D) \) and from the first part
\[
|\langle -\Delta \phi, f \rangle| \equiv |\langle -\Delta \phi, f \rangle_H| \equiv |\langle f, \phi \rangle| \leq C \|f\|_{L^p} \|\phi\|_{H^1_0(D)}
\]

Thus \(-\Delta \phi \in H\) can be considered in \((L^p(D))'\). Why should \( H \) be dense in \((L^p(D))'\)? If it is not dense, then there exists \( g^* \in (L^p(D))' \) which is not the limit of vectors of \( H \). Then since \( L^p(D) \) is reflexive, an application of the Hahn Banach theorem shows there exists \( f \in L^p(D) \) such that
\[
\langle g^*, f \rangle_{(L^p(D))', L^p(D)} \neq 0, \quad \langle -\Delta \phi, f \rangle_{(L^p(D))', L^p(D)} = 0
\]

for all \(-\Delta \phi \in H\). However, it was just shown \( H \) could be considered a subset of \((L^p(D))'\) in the manner described above. Therefore, the last equation in the above is of the form
\[
0 = (-\Delta \phi, f)_H = \langle f, \phi \rangle = \int_D f \phi \, dx
\]

and since this holds for all \( \phi \in H^1_0(D) \), it follows by density of \( H^1_0(D) \) in \( L^{p'}(D) \), that \( f = 0 \) and now this contradicts the inequality in (21.1.2).

Now \( \Delta \) is defined on \( H^1_0(D) \) and it delivers something in \((H^1_0(D))' \equiv H\). Of course \( H^1_0(D) \) is dense in \( L^{p'}(D) \). Can \( \Delta \) be extended to all of \( L^{p'}(D) \)? The answer is yes and it is more of the same given above. For \( \phi \in H^1_0(D), -\Delta \phi \in H \subseteq (L^p(D))' \). Then by the above, for \( \phi \in H^1_0(D) \) and \( f \in L^p(D) \),
\[
\langle -\Delta \phi, f \rangle \equiv \langle f, \phi \rangle \equiv \int_D f \phi \, dx
\]

and so \( -\Delta \) is a continuous linear mapping defined on a dense subspace \( H^1_0(D) \) of \( L^{p'}(D) \) and so this does indeed extend to a continuous linear map defined on all of \( L^{p'}(D) \) given by the formula
\[
\langle -\Delta g, f \rangle \equiv \int_D f g \, dx
\]

This proves the lemma.

Thus letting \( V \equiv L^p(D) \), and \( H \equiv (H^1_0(D))' \), it follows \( V \subseteq H \subseteq V' \) is a Gelfand triple with the understanding of what it means for one space to be included in another described above. To emphasize the above, for \(-\Delta \phi \in H, f \in L^p\),
\[
\langle -\Delta \phi, f \rangle \equiv (-\Delta \phi, f)_H \equiv \langle f, \phi \rangle \equiv \int_D f \phi \, dx
\]

More generally, for \( g \in L^{p'}(D), -\Delta g \in (L^p(D))' \) according to the rule
\[
\langle -\Delta g, f \rangle \equiv \int_D f g \, dx.
\]

With this example of a Gelfand triple, one can define a “porous medium operator” \( A : V \to V' \). Let \( \Psi \) be a real valued function defined on \( \mathbb{R} \) which satisfies
\[
\Psi \text{ is continuous} \quad (21.1.3)
\]
\[
(t - s) (\Psi (t) - \Psi (s)) \geq 0 \quad (21.1.4)
\]

There exists \( p \geq 2, p < \infty \) and \( \alpha \in (0, \infty) \) such that for all \( s \in \mathbb{R} \)
\[
s \Psi (s) \geq \alpha |s|^p - c \quad (21.1.5)
\]
There exist \( c_3, c_4 \in (0, \infty) \) such that for all \( s \in \mathbb{R} \)

\[
|\Psi(s)| \leq c_4 + c_3 |s|^{p-1}
\]

(21.1.6)

Note that (21.1.6) implies that if \( v \in L^p(D) \), then

\[
\int_D |\Psi(v)|^{p'} \, dx \leq C \int_D \left(1 + |v|^{p'(p-1)}\right) \, dx = C \int_D (1 + |v|^p) \, dx < \infty.
\]

Thus for \( v \in L^p(D) \), \( \Psi(v) \) is something you can do \( \Delta \) to and obtain something in \( V' \). The porous medium operator \( A : V \to V' \) is given as follows.

\[
\langle Av, w \rangle_{V', V} \equiv \langle \Delta \Psi(v), w \rangle_{V', V} \equiv -\int_D \Psi(v) \, w \, dx
\]

What are the properties of \( A \)?

\[
\langle A(u + \lambda v), w \rangle \equiv -\int_D \Psi(u + \lambda v) \, w \, dx
\]

and this is easily seen to be a continuous function of \( \lambda \) Thus \( A \) is Hemicontinuous.

\[
\langle A(u) - A(v), u - v \rangle \equiv -\int_D \Psi(u)(u - v) \, dx + \int_D \Psi(v)(u - v) \, dx \leq 0
\]

Thus \( -A \) is monotone. Also there is a coercivity estimate which is routine.

\[
\langle A(v), v \rangle \equiv -\int_D \Psi(v) \, v \leq \int_D c - \alpha |v|^p \, dx = C - \alpha \|v\|_{V'}^p
\]

This operator also has a boundedness estimate.

\[
\|A(v)\|_{V'} \equiv \sup_{\|w\|_V \leq 1} |\langle A(v), w \rangle| \equiv \sup_{\|w\|_V \leq 1} \left|\int_D \Psi(v) \, w \right|
\]

\[
\leq \sup_{\|w\|_V \leq 1} \left( \int_D \left(c_4 + c_3 |v|^{p-1}\right) \, w \, dx \right)
\]

\[
\leq \left( \int_D C (1 + |v|^p) \, dx \right)^{1/p'} \leq C + C \left( \int_D |v|^p \, dx \right)^{1/p'}
\]

\[
= C + C \|v\|_{V'}^{p'/p} = C + C \|v\|_{V'}^{p-1}.
\]

Since \( \Psi \) is continuous, it will also follow that \( A \) is \( \mathcal{B}(V) \) measurable. Consider

\[
u \to \langle Au, w \rangle \equiv -\int_D \Psi(u) \, w \, dx
\]

for fixed \( w \in V \). Suppose \( u_n \to u \) in \( V \) and fix \( w \in L^\infty(D) \subseteq V \). Then it follows from an easy argument using the Vitali convergence theorem and the fact that from the estimates above

\[
\Psi(u_n)w
\]

is uniformly integrable that

\[
u \to -\int_D \Psi(u) \, w \, dx
\]

is continuous. For general \( w \in L^p(D) \), let \( w_n \to w \) in \( L^p(D) \) where each \( w_n \) is in \( L^\infty(D) \). Then the function

\[
u \to -\int_D \Psi(u) \, w \, dx \equiv \langle Au, w \rangle
\]

(21.1.7)

is the limit of the continuous functions

\[
u \to -\int_D \Psi(u) \, w_n \, dx
\]

and so the function (21.1.7) is Borel measurable. Now by the Pettis theorem this shows \( A : V \to V' \) is \( \mathcal{B}(V) \) measurable.

This shows \( A \) is an example of an operator which satisfies some conditions which will be considered later.
21.2 Some Standard Techniques In Evolution Equations

In this section, several significant theorems are presented. Unless indicated otherwise, the measure will be Lebesgue measure. First here is a lemma.

Lemma 21.2.1 Suppose $g \in L^1([a,b];X)$ where $X$ is a Banach space. Then if $\int_a^b g(t) \, \phi(t) \, dt = 0$ for all $\phi \in C_c^\infty(a,b)$, then $g(t) = 0$ a.e.

Proof: Let $E$ be a measurable subset of $(a,b)$ and let $K \subseteq E \subseteq V \subseteq (a,b)$ where $K$ is compact, $V$ is open and $m(V \setminus K) < \varepsilon$. Let $K \prec h \prec V$ as in the proof of the Riesz representation theorem for positive linear functionals. Enlarging $K$ slightly and convolving with a mollifier, it can be assumed $h \in C_c^\infty(a,b)$. Then

\[
\left| \int_a^b X_{E}(t) g(t) \, dt \right| = \left| \int_a^b (X_E(t) - h(t)) g(t) \, dt \right| \\
\leq \int_a^b |X_E(t) - h(t)| \|g(t)\| \, dt \\
\leq \int_{V \setminus K} \|g(t)\| \, dt.
\]

Now let $K_n \subseteq E \subseteq V_n$ with $m(V_n \setminus K_n) < 2^{-n}$. Then from the above,

\[
\left| \int_a^b X_{E}(t) g(t) \, dt \right| \leq \int_a^b \|g(t)\| \, dt
\]

and the integrand of the last integral converges to 0 a.e. as $n \to \infty$ because $\sum_n m(V_n \setminus K_n) < \infty$. By the dominated convergence theorem, this last integral converges to 0. Therefore, whenever $E \subseteq (a,b)$,

\[
\int_a^b X_{E}(t) g(t) \, dt = 0.
\]

Since the endpoints have measure zero, it also follows that for any measurable $E$, the above equation holds.

Now $g \in L^1([a,b];X)$ and so it is measurable. Therefore, $g([a,b])$ is separable. Let $D$ be a countable dense subset and let $E$ denote the set of linear combinations of the form $\sum a_i d_i$ where $a_i$ is a rational point of $\mathbb{F}$ and $d_i \in D$. Thus $E$ is countable. Denote by $Y$ the closure of $E$ in $X$. Thus $Y$ is a separable closed subspace of $X$ which contains all the values of $g$.

Now let $S_n = g^{-1}(B(y_n, \|y_n\|/2))$ where $E = \{y_n\}_{n=1}^\infty$. Therefore, $\cup_n S_n = g^{-1}(X \setminus \{0\})$. This follows because if $x \in Y$ and $x \neq 0$, then in $B\left(x, \frac{\|x\|}{4}\right)$ there is a point of $E$, $y_n$. Therefore, $\|y_n\| > \frac{3}{4} \|x\|$ and so $\frac{\|y_n\|}{2} > \frac{3\|x\|}{8} > \frac{\|x\|}{4}$ so $x \in B(y_n, \|y_n\|/2)$. It follows that if each $S_n$ has measure zero, then $g(t) = 0$ for a.e. $t$. Suppose then that for some $n$, the set, $S_n$ has positive measure. Then from what was shown above,

\[
\|y_n\| = \left\| \frac{1}{m(S_n)} \int_{S_n} g(t) \, dt - y_n \right\| = \left\| \frac{1}{m(S_n)} \int_{S_n} g(t) \, dt - y_n \right\|
\]

and so $y_n = 0$ which implies $S_n = \emptyset$, a contradiction to $m(S_n) > 0$. This contradiction shows each $S_n$ has measure zero and so as just explained, $g(t) = 0$ a.e. □

Definition 21.2.2 For $f \in L^1(a,b;X)$, define an extension, $\overline{f}$ defined on

\[
[2a - b, 2b - a] = [a - (b - a), b + (b - a)]
\]

as follows.

\[
\overline{f}(t) = \begin{cases} 
  f(t) & \text{if } t \in [a,b] \\
  f(2a - t) & \text{if } t \in [2a - b, a] \\
  f(2b - t) & \text{if } t \in [b, 2b - a]
\end{cases}
\]
21.2. SOME STANDARD TECHNIQUES IN EVOLUTION EQUATIONS

Definition 21.2.3 Also if \( f \in L^p(a, b; X) \) and \( h > 0 \), define for \( t \in [a, b] \), \( f_h(t) = \overline{f}(t - h) \) for all \( h < b - a \). Thus the map \( f \to f_h \) is continuous and linear on \( L^p(a, b; X) \). It is continuous because

\[
\int_a^b ||f_h(t)||^p \, dt = \int_a^{a+h} ||f(2a - t + h)||^p \, dt + \int_{a+h}^{b-h} ||f(t)||^p \, dt \\
= \int_a^{a+h} ||f(t)||^p \, dt + \int_{a}^{b-h} ||f(t)||^p \, dt \leq 2 ||f||_p^p.
\]

The following lemma is on continuity of translation in \( L^p(a, b; X) \).

Lemma 21.2.4 Let \( \overline{f} \) be as defined in Definition 21.2.3. Then for \( f \in L^p(a, b; X) \) for \( p \in [1, \infty) \),

\[
\lim_{\delta \to 0} \int_a^b ||\overline{f}(t - \delta) - f(t)||^p_X \, dt = 0.
\]

Proof: Regarding the measure space as \((a, b)\) with Lebesgue measure, by regularity of the measure, there exists \( g \in C_c(a, b; X) \) such that \( ||f - g||_p < \varepsilon \). Here the norm is the norm in \( L^p(a, b; X) \). Therefore,

\[
||f_h - f||_p \leq ||f_h - g_h||_p + ||g_h - g||_p + ||g - f||_p \\
\leq (2^{1/p} + 1) ||f - g||_p + ||g_h - g||_p \\
< (2^{1/p} + 1) \varepsilon + \varepsilon
\]

whenever \( h \) is sufficiently small. This is because of the uniform continuity of \( g \). Therefore, since \( \varepsilon > 0 \) is arbitrary, this proves the lemma. \( \blacksquare \)

Definition 21.2.5 Let \( f \in L^1(a, b; X) \). Then the distributional derivative in the sense of \( X \) valued distributions is given by

\[
f'(\phi) \equiv - \int_a^b f(t) \phi'(t) \, dt
\]

Then \( f' \in L^1(a, b; X) \) if there exists \( h \in L^1(a, b; X) \) such that for all \( \phi \in C_c^\infty(a, b) \),

\[
f'(\phi) = \int_a^b h(t) \phi(t) \, dt.
\]

Then \( f' \) is defined to equal \( h \). Here \( f \) and \( f' \) are considered as vector valued distributions in the same way as was done for scalar valued functions.

Lemma 21.2.6 The above definition is well defined.

Proof: Suppose both \( h \) and \( g \) work in the definition. Then for all \( \phi \in C_c^\infty(a, b) \),

\[
\int_a^b (h(t) - g(t)) \phi(t) \, dt = 0.
\]

Therefore, by Lemma 21.2.3 \( h(t) - g(t) = 0 \) a.e. \( \blacksquare \)

The other thing to notice about this is the following lemma. It follows immediately from the definition.

Lemma 21.2.7 Suppose \( f, f' \in L^1(a, b; X) \). Then if \( [c, d] \subseteq [a, b] \), it follows that \( (f|_{[c, d]})' = f'|_{[c, d]} \). This notation means the restriction to \([c, d]\).

Recall that in the case of scalar valued functions, if you had both \( f \) and its weak derivative, \( f' \) in \( L^1(a, b) \), then you were able to conclude that \( f \) is almost everywhere equal to a continuous function, still denoted by \( f \) and

\[
f(t) = f(a) + \int_a^t f'(s) \, ds.
\]

In particular, you can define \( f(a) \) to be the initial value of this continuous function. It turns out that an identical theorem holds in this case. To begin with here is the same sort of lemma which was used earlier for the case of scalar valued functions. It says that if \( f' = 0 \) where the derivative is taken in the sense of \( X \) valued distributions, then \( f \) equals a constant.
Lemma 21.2.8 Suppose \( f \in L^1(a, b; X) \) and for all \( \phi \in C_c^\infty(a, b) \),

\[
\int_a^b f(t) \phi'(t) \, dt = 0.
\]

Then there exists a constant, \( a \in X \) such that \( f(t) = a \) a.e.

Proof: Let \( \phi \in C_c^\infty(a, b) \), \( \int_a^b \phi_0(x) \, dx = 1 \) and define for \( \phi \in C_c^\infty(a, b) \)

\[
\psi_\phi(x) = \int_a^x [\phi(t) - \left( \int_a^b \phi(y) \, dy \right) \phi_0(t)] \, dt
\]

Then \( \psi_\phi \in C_c^\infty(a, b) \) and \( \psi'_\phi = \phi - \left( \int_a^b \phi(y) \, dy \right) \phi_0 \). Then

\[
\int_a^b f(t) (\phi(t)) \, dt = \int_a^b f(t) \left( \psi'_\phi(t) + \left( \int_a^b \phi(y) \, dy \right) \phi_0(t) \right) \, dt
\]

\( = 0 \) by assumption

\[
= \int_a^b f(t) \psi'_\phi(t) \, dt + \left( \int_a^b \phi(y) \, dy \right) \int_a^b f(t) \phi_0(t) \, dt
\]

\[
= \left( \int_a^b \left( \int_a^t f(t) \phi_0(t) \, dt \right) \phi(y) \, dy \right).
\]

It follows that for all \( \phi \in C_c^\infty(a, b) \),

\[
\int_a^b \left( f(y) - \left( \int_a^b f(t) \phi_0(t) \, dt \right) \right) \phi(y) \, dy = 0
\]

and so by Lemma 21.2.11,

\[
f(y) - \left( \int_a^b f(t) \phi_0(t) \, dt \right) = 0 \text{ a.e. } y
\]

Theorem 21.2.9 Suppose \( f, f' \) both are in \( L^1(a, b; X) \) where the derivative is taken in the sense of \( X \) valued distributions. Then there exists a unique point of \( X \), denoted by \( f(a) \) such that the following formula holds a.e. \( t \).

\[
f(t) = f(a) + \int_a^t f'(s) \, ds
\]

Proof:

\[
\int_a^b \left( f(t) - \int_a^t f'(s) \, ds \right) \phi'(t) \, dt = \int_a^b f(t) \phi'(t) \, dt - \int_a^b \int_a^t f'(s) \phi'(t) \, ds \, dt.
\]

Now consider \( \int_a^b \int_a^t f'(s) \phi'(t) \, ds \, dt \). Let \( \Lambda \in X' \). Then it is routine from approximating \( f' \) with simple functions to verify

\[
\Lambda \left( \int_a^b \int_a^t f'(s) \phi'(t) \, ds \, dt \right) = \int_a^b \int_a^t \Lambda(f'(s)) \phi'(t) \, ds \, dt.
\]

Now the ordinary Fubini theorem can be applied to obtain

\[
= \int_a^b \int_s^b \Lambda(f'(s)) \phi'(t) \, dt \, ds = \Lambda \left( \int_a^b \int_s^b f'(s) \phi'(t) \, dt \, ds \right).
\]

Since \( X' \) separates the points of \( X \), it follows

\[
\int_a^b \int_a^t f'(s) \phi'(t) \, ds \, dt = \int_a^b \int_s^b f'(s) \phi'(t) \, dt \, ds.
\]
21.2. SOME STANDARD TECHNIQUES IN EVOLUTION EQUATIONS

Therefore,
\[
\int_a^b \left( f(t) - \int_a^t f'(s) \, ds \right) \phi'(t) \, dt = \left. \int_a^t f(t) \phi'(t) \, dt \right|_a^b - \int_a^b f'(s) \phi'(t) \, ds \, dt = \int_a^b f(t) \phi'(t) \, dt - \int_a^b f'(s) \int_s^b \phi'(t) \, dt \, ds = \int_a^b f(t) \phi'(t) \, dt + \int_a^b f'(s) \phi(s) \, ds = 0.
\]

Therefore, by Lemma 21.2.8, there exists a constant, denoted as \( f(a) \) such that
\[ f(t) - \int_a^t f'(s) \, ds = f(a) \quad \square \]

There is also a useful theorem about continuity of pointwise evaluation.

**Corollary 21.2.10** Let \( f, f' \in L^1(a, b; X) \) so that
\[ f(t) = f(0) + \int_0^t f'(s) \, ds \quad (21.2.8) \]

where in this formula, \( t \to f(t) \) is the continuous representative of \( f \). Then there exists a constant \( C \) such that for each \( t \in [a, b] \),
\[ \| f(t) \|_X \leq C \left( \| f \|_{L^1(a, b; X)} + \| f' \|_{L^1(a, b; X)} \right) \]

**Proof:** From the integral equation 21.2.8,
\[ f(t) = f(s) + \int_s^t f'(r) \, dr \]

\[ \| f(t) \|_X \leq \| f(s) \|_X + \left| \int_s^t \| f'(r) \|_X \, dr \right| \leq \| f(s) \|_X + \int_a^b \| f'(r) \|_X \, dr \]

and so, integrating both sides with respect to \( s \)
\[ (b - a) \| f(t) \|_X \leq \| f \|_{L^1(a, b; X)} + (b - a) \| f' \|_{L^1(a, b; X)} \]

and so
\[ \| f(t) \|_X \leq \left( \frac{1}{b - a} + 1 \right) \left( \| f \|_{L^1(a, b; X)} + \| f' \|_{L^1(a, b; X)} \right) \quad \square \]

Let \( \mathcal{X} \) be the space of functions \( f \in L^1(a, b; X) \) such that their weak derivatives \( f' \) are also in \( L^1(a, b; X) \). Then \( \mathcal{X} \) is a Banach space with norm given by
\[ \| f \|_X \equiv \| f \|_{L^1(a, b; X)} + \| f' \|_{L^1(a, b; X)} \]

This is because the map \( f \to f' \) is a closed map. If \( f_n \to f \) in \( L^1(a, b; X) \) and \( f'_n \to \xi \) in \( L^1(a, b; X) \), then for \( \phi \in C_c^\infty(a, b) \),
\[ \int_a^b \xi \phi \, dt = \lim_{n \to \infty} \int_a^b f'_n \phi \, dt = \lim_{n \to \infty} - \int_a^b f_n \phi' \, dt = - \int_a^b f \phi' \, dt \]

showing that \( \xi = f' \). Thus if you have a Cauchy sequence in \( \mathcal{X} \), \( \{ f_n \} \), then \( f_n \to f \) in \( L^1(a, b; X) \) and \( f'_n \to \xi \) in \( L^1(a, b; X) \) for some \( \xi \). Hence \( f' = \xi \).
Then the above corollary says that pointwise evaluation is continuous as a map from $\mathfrak{X}$ to $X$. This is clearly a linear map. Also the formula obtained shows that in fact, this is continuous into $C([a,b];X)$.

$$
\|f\|_{C([a,b];X)} = \sup_{t \in [a,b]} \|f(t)\|_X \leq C \left( \|f\|_{L^1(a,b;X)} + \|f'\|_{L^1(a,b;X)} \right) = C \|f\|_X.
$$

Now let $\theta : \mathfrak{X} \to C([a,b];X)$ be given by $\theta f(t) \equiv f(t)$ where $f(t) = f(0) + \int_0^t f'(s) \, ds$, $f$ being the continuous representative of $f$. Then $\theta$ is continuous and linear. If $\theta f \equiv f(t)$ so that it is pointwise evaluation at $t$, then this $\theta_t$ is also continuous and linear. Suppose $X$ is also reflexive. It follows that if you have a sequence in $X \{f_n\}$ which is converging weakly to $f \in \mathfrak{X}$, then you would also have $\theta f_n \to \theta f$ weakly in $X$. If this is not so, then since $X$ is reflexive, there is a subsequence, still denoted as $f_n$ such that $\theta f_n \to \xi \neq f(t)$. However, this says that $(f,\xi)$ is in the weak closure of the graph of $\theta_t$. Since this graph is strongly closed and convex, it is also weakly closed and hence $\xi = \theta_t f \equiv f(t)$, a contradiction. This proves the following nice corollary.

**Corollary 21.2.11** Suppose $f_n \to f$ weakly in $\mathfrak{X}$ where we assume also that $X$ is reflexive. Then $f_n(t) \to f(t)$ weakly in $X$.

The integration by parts formula is also important.

**Corollary 21.2.12** Suppose $f, f' \in L^1(a,b;X)$ and suppose $\phi \in C^1([a,b])$. Then the following integration by parts formula holds.

$$
\int_a^b f(t) \phi'(t) \, dt = f(b) \phi(b) - f(a) \phi(a) - \int_a^b f'(t) \phi(t) \, dt.
$$

**Proof:** From Theorem 21.2.4

$$
\int_a^b f(t) \phi'(t) \, dt = \int_a^b \left( f(a) + \int_a^t f'(s) \, ds \right) \phi'(t) \, dt
= f(a) (\phi(b) - \phi(a)) + \int_a^b \int_a^t f'(s) \, ds \phi'(t) \, dt
= f(a) (\phi(b) - \phi(a)) + \int_a^b f'(s) \int_s^b \phi'(t) \, dt \, ds
= f(a) (\phi(b) - \phi(a)) + \int_a^b f'(s) (\phi(b) - \phi(s)) \, ds
= f(a) (\phi(b) - \phi(a)) - \int_a^b f'(s) \phi(s) \, ds + (f(b) - f(a)) \phi(b)
= f(b) \phi(b) - f(a) \phi(a) - \int_a^b f'(s) \phi(s) \, ds.
$$

The interchange in order of integration is justified as in the proof of Theorem 21.2.4.

With this integration by parts formula, the following interesting lemma is obtained. This lemma shows why it was appropriate to define $\overline{\mathcal{L}}$ as in Definition 21.2.2.

**Lemma 21.2.13** Let $\overline{\mathcal{L}}$ be given in Definition 21.2.2 and suppose $f, f' \in L^1(a,b;X)$. Then $\overline{\mathcal{L}}, \overline{\mathcal{L}}' \in L^1(2a-b, 2b-a;X)$ also and

$$
\overline{\mathcal{L}}'(t) \equiv \begin{cases} 
  f'(t) & \text{if } t \in [a,b] \\
  -f'(2a-t) & \text{if } t \in [2a-b,a] \\
  -f'(2b-t) & \text{if } t \in [b,2b-a]
\end{cases}
$$

(21.2.9)

**Proof:** It is clear from the definition of $\overline{\mathcal{L}}$ that $\overline{\mathcal{L}} \in L^1(2a-b, 2b-a;X)$ and that in fact

$$
\|\overline{\mathcal{L}}\|_{L^1(2a-b,2b-a;X)} \leq 3 \|f\|_{L^1(a,b;X)}. \tag{21.2.10}
$$
Let \( \phi \in C_c^\infty (2a - b, 2b - a) \). Then from the integration by parts formula,

\[
\int_{2a-b}^{2b-a} \overline{f}(t) \phi'(t) \, dt = \int_{a}^{b} f(t) \phi'(t) \, dt + \int_{a}^{b} f(2b - t) \phi'(t) \, dt + \int_{a}^{a} f(2a - t) \phi'(t) \, dt
\]

\[
= \int_{a}^{b} f(t) \phi'(t) \, dt + \int_{a}^{b} f(u) \phi'(2b - u) \, du + \int_{a}^{b} f(u) \phi'(2a - u) \, du
\]

\[
= f(b) \phi(b) - f(a) \phi(a) - \int_{a}^{b} f'(t) \phi(t) \, dt - f(b) \phi(b) + f(a) \phi(2b - a)
\]

\[
+ \int_{a}^{b} f'(u) \phi(2b - u) \, du - f(b) \phi(2a - b)
\]

\[
+ f(a) \phi(a) + \int_{a}^{b} f'(u) \phi(2a - u) \, du
\]

\[
= - \int_{a}^{b} f'(t) \phi(t) \, dt + \int_{a}^{b} f'(u) \phi(2b - u) \, du + \int_{a}^{b} f'(u) \phi(2a - u) \, du
\]

\[
= - \int_{a}^{b} f'(t) \phi(t) \, dt - \int_{b}^{2b-a} f'(2b - t) \phi(t) \, dt - \int_{2a-b}^{a} f'(2a - t) \phi(t) \, dt
\]

where \( \overline{f}(t) \) is given in 21.2.14. □

**Definition 21.2.14** Let \( V \) be a Banach space and let \( H \) be a Hilbert space. (Typically \( H = L^2(\Omega) \)) Suppose \( V \subseteq H \) is dense in \( H \) meaning that the closure in \( H \) of \( V \) gives \( H \). Then it is often the case that \( H \) is identified with its dual space, and then because of the density of \( V \) in \( H \), it is possible to write

\[
V \subseteq H = H' \subseteq V'
\]

When this is done, \( H \) is called a pivot space. Another notation which is often used is \( (f, g) \) to denote \( f(g) \) for \( f \in V' \) and \( g \in V \). This may also be written as \( (f, g)_{V',V} \). Another term is that \( V \subseteq H = H' \subseteq V' \) is called a Gelfand triple.

The next theorem is an example of a trace theorem. In this theorem, \( f \in L^p(0, T; V) \) while \( f' \in L^p(0, T; V') \). It makes no sense to consider the initial values of \( f \) in \( V \) because it is not even continuous with values in \( V \). However, because of the derivative of \( f \) it will turn out that \( f \) is continuous with values in a larger space and so it makes sense to consider initial values of \( f \) in this other space. This other space is called a trace space.

**Theorem 21.2.15** Let \( V \) and \( H \) be a Banach space and Hilbert space as described in Definition 21.2.14. Suppose \( f \in L^p(0, T; V) \) and \( f' \in L^p(0, T; V') \). Then \( f \) is a.e. equal to a continuous function mapping \([0, T]\) to \( H \). Furthermore, there exists \( f(0) \in H \) such that

\[
\frac{1}{2} |f(t)|_H^2 - \frac{1}{2} |f(0)|_H^2 = \int_0^t \langle f'(s), f(s) \rangle \, ds,
\]

(21.2.11)

and for all \( t \in [0, T] \),

\[
\int_0^t f'(s) \, ds \in H,
\]

(21.2.12)

and for a.e. \( t \in [0, T] \),

\[
f(t) = f(0) + \int_0^t f'(s) \, ds \text{ in } H,
\]

(21.2.13)

Here \( f' \) is being taken in the sense of \( V' \) valued distributions and \( \frac{1}{p} + \frac{1}{p'} = 1 \) and \( p \geq 2 \).
Recall in Section 21.2.13 yields that \( \Psi \) is a mollifier having support in \((-1/n, 1/n)\). Then by Minkowski’s inequality
\[
\left\| f_n - \tilde{f} \right\|_{L^p(\mathbb{R}; V)} = \left( \int_{\mathbb{R}} \left\| \int_{-1/n}^{1/n} (\tilde{f}(t) - \tilde{f}(t-s)) \phi_n(s) \, ds \right\|_V^p \, dt \right)^{1/p}
\]
\[
= \left( \int_{\mathbb{R}} \left( \int_{-1/n}^{1/n} \left\| \tilde{f}(t) - \tilde{f}(t-s) \right\|_V \phi_n(s) \, ds \right)^p \, dt \right)^{1/p}
\]
\[
\leq \int_{-1/n}^{1/n} \phi_n(s) \left( \int_{\mathbb{R}} \left\| \tilde{f}(t) - \tilde{f}(t-s) \right\|_V^p \, dt \right)^{1/p} \, ds
\]
\[
\leq \int_{-1/n}^{1/n} \phi_n(s) \varepsilon \, ds = \varepsilon
\]
provided \( n \) is large enough. This follows from continuity of translation in \( L^p \) with Lebesgue measure. Since \( \varepsilon > 0 \) is arbitrary, it follows \( f_n \to \tilde{f} \) in \( L^p(\mathbb{R}; V) \). Similarly, \( f_n \to f \) in \( L^2(\mathbb{R}; H) \). This follows because \( p \geq 2 \) and the norm in \( V \) and norm in \( H \) are related by \( \|x\|_H \leq C \|x\|_V \) for some constant, \( C \). Now
\[
\tilde{f}(t) = \begin{cases} 
\Psi(t) f(t) & \text{if } t \in [0, T], \\
\Psi(t) f(2T - t) & \text{if } t \in [T, 2T], \\
\Psi(t) f(-t) & \text{if } t \in [0, T], \\
0 & \text{if } t \notin [-T, 2T].
\end{cases}
\]

An easy modification of the argument of Lemma 21.4 yields
\[
\tilde{f}'(t) = \begin{cases} 
\Psi'(t) f(t) + \Psi(t) f'(t) & \text{if } t \in [0, T], \\
\Psi'(t) f(2T - t) - \Psi(t) f'(2T - t) & \text{if } t \in [T, 2T], \\
\Psi'(t) f(-t) - \Psi(t) f'(-t) & \text{if } t \in [-T, 0], \\
0 & \text{if } t \notin [-T, 2T].
\end{cases}
\]
Recall
\[
f_n(t) = \int_{-1/n}^{1/n} \tilde{f}(t-s) \phi_n(s) \, ds = \int_{\mathbb{R}} \tilde{f}(t-s) \phi_n(s) \, ds
\]
\[
= \int_{\mathbb{R}} \tilde{f}(s) \phi_n(t-s) \, ds.
\]
Therefore,
\[
f_n'(t) = \int_{\mathbb{R}} \tilde{f}(s) \phi'_n(t-s) \, ds = \int_{-T - \frac{1}{n}}^{2T + \frac{1}{n}} \tilde{f}(s) \phi'_n(t-s) \, ds
\]
\[
= \int_{-T - \frac{1}{n}}^{2T + \frac{1}{n}} \tilde{f}'(s) \phi_n(t-s) \, ds = \int_{\mathbb{R}} \tilde{f}'(s) \phi_n(t-s) \, ds
\]
\[
= \int_{-1/n}^{1/n} \tilde{f}'(t-s) \phi_n(s) \, ds.
\]
and it follows from the first line above that $f'_n$ is continuous with values in $V$ for all $t \in \mathbb{R}$. Also note that both $f'_n$ and $f_n$ equal zero if $t \not\in [-T, 2T]$ whenever $n$ is large enough. Exactly similar reasoning to the above shows that $f'_n \to \hat{f}'$ in $L^{p'}(\mathbb{R}; V')$.

Now let $\phi \in C_0^\infty(0, T)$.

$$
\int_R |f_n(t)|^2_H \phi'(t) \, dt = \int_R (f_n(t), f_n(t))_H \phi'(t) \, dt = -\int_R 2(f'_n(t), f_n(t)) \phi(t) \, dt
$$

(21.2.15)

Now

$$
\int_R \left| \int_R (f'_n(t), f_n(t)) \phi(t) \, dt - \int_R (f'(t), f(t)) \phi(t) \, dt \right| \leq \int_R \left( |(f'_n(t) - f'(t), f_n(t))| + |(f'(t), f_n(t) - f(t))| \right) \phi(t) \, dt.
$$

From the first part of this proof which showed that $f_n \to \hat{f}$ in $L^p(\mathbb{R}; V)$ and $f'_n \to \hat{f}'$ in $L^{p'}(\mathbb{R}; V')$, an application of Holder’s inequality shows the above converges to 0 as $n \to \infty$. Therefore, passing to the limit as $n \to \infty$ in the

$$
\int_R |\hat{f}'(t)|^2_H \phi'(t) \, dt = -\int_R 2 \langle \hat{f}'(t), \hat{f}(t) \rangle \phi(t) \, dt
$$

which shows $t \to |\hat{f}'(t)|^2_H$ equals a continuous function a.e. and it also has a weak derivative equal to $2 \langle \hat{f}'', \hat{f}' \rangle$.

It remains to verify that $\hat{f}$ is continuous on $[0, T]$. Of course $\hat{f} = f$ on this interval. Let $N$ be large enough that $f_n(-T) = 0$ for all $n > N$. Then for $m, n > N$ and $t \in [-T, 2T]$,

$$
|f_n(t) - f_m(t)|^2_H = 2 \int_{-T}^t (f'_n(s) - f'_m(s), f_n(s) - f_m(s)) \, ds
$$

$$
= 2 \int_{-T}^t (f'_n(s), f_n(s) - f_m(s))_{V', V} \, ds
$$

$$
\leq 2 \int_R |f'_n(s) - f'_m(s)||_{V'} ||f_n(s) - f_m(s)||_V \, ds
$$

$$
\leq 2 \big| f_n - f_m \big|_{L^{p'}(\mathbb{R}; V')} \big| f_n - f_m \big|_{L^p(\mathbb{R}; V)}
$$

which shows from the above that $\{f_n\}$ is uniformly Cauchy on $[-T, 2T]$ with values in $H$. Therefore, there exists $g$ a continuous function defined on $[-T, 2T]$ having values in $H$ such that

$$
\lim_{n \to \infty} \max \{|f_n(t) - g(t)|_H; t \in [-T, 2T]\} = 0.
$$

However, $g = \hat{f}$ a.e. because $f_n$ converges to $f$ in $L^p(0, T; V')$. Therefore, taking a subsequence, the convergence is a.e. It follows from the fact that $V \subseteq H = H' \subseteq V'$ and Theorem 21.2.1, there exists $f(0) \in V'$ such that for a.e. $t$,

$$
f(t) = f(0) + \int_0^t f'(s) \, ds \text{ in } V'
$$

Now $g = f$ a.e. and $g$ is continuous with values in $H$ hence continuous with values in $V'$ and so

$$
g(t) = f(0) + \int_0^t f'(s) \, ds \text{ in } V'
$$

for all $t$. Since $g$ is continuous with values in $H$ it is continuous with values in $V'$. Taking the limit as $t \downarrow 0$ in the above, $g(a) = \lim_{t \to a^+} g(t) = f(0)$, showing that $f(0) \in H$. Therefore, for a.e. $t$,

$$
f(t) = f(0) + \int_0^t f'(s) \, ds \text{ in } H, \quad \int_0^t f'(s) \, ds \in H.\blacksquare
$$

Note that if $f \in L^p(0, T; V)$ and $f' \in L^{p'}(0, T; V')$, then you can consider the initial value of $f$ and it will be in $H$. What if you start with something in $H$? Is it an initial condition for a function $f \in L^p(0, T; V)$ such that $f' \in L^{p'}(0, T; V')$? This is worth thinking about. If it is not so, what is the space of initial values? How can you give this space a norm? What are its properties? It turns out that if $V$ is a closed subspace of the Sobolev space, $W^{1,p}(\Omega)$ which contains $W^{1,p}_0(\Omega)$ for $p \geq 2$ and $H = L^2(\Omega)$ the answer to the above question is yes. Not surprisingly, there are many generalizations of the above ideas.
21.3 An Important Formula

It is not necessary to have \( p > 2 \) in order to do the sort of thing just described. Here is a major result which will have a much more difficult stochastic version presented later. First is an approximation theorem of Doob. See Lemma 21.3.1.

Lemma 21.3.1 Let \( \Phi : [0, T] \to E \), be Lebesgue measurable and suppose

\[
\Phi \in K = L^p([0, T]; E), \quad p \geq 1
\]

Then there exists a sequence of nested partitions, \( P_k \subseteq P_{k+1} \),

\[
P_k = \{ t^k_0, \ldots, t^k_{m_k} \}
\]

such that the step functions given by

\[
\Phi^k(t) = \sum_{j=1}^{m_k} \Phi(t_j^k) \chi_{[t_{j-1}^k, t_j^k)}(t)
\]

both converge to \( \Phi \) in \( K \) as \( k \to \infty \) and

\[
\lim_{k \to \infty} \max \left\{ |t_j^k - t_{j+1}^k| : j \in \{0, \ldots, m_k\} \right\} = 0.
\]

In the formulas, define \( \Phi(0) = 0 \). The mesh points \( \{t^k_j\}_{j=0}^{m_k} \) can be chosen to miss a given set of measure zero.

Note that it would make no difference in terms of the conclusion of this lemma if you defined

\[
\Phi^k(t) = \sum_{j=1}^{m_k} \Phi(t_{j-1}) \chi_{[t_{j-1}^k, t_j^k)}(t)
\]

because the modified function equals the one given above off a countable subset of \([0, T]\), the union of the mesh points.

Proof: For \( t \in \mathbb{R} \) let \( \gamma_n(t) = k/2^n \), \( \delta_n(t) = (k+1)/2^n \), where \( t \in (k/2^n, (k+1)/2^n] \), and \( 2^{-n} < T/4 \). Also suppose \( \Phi \) is defined to equal 0 on \([0, T] \times \Omega \). There exists a set of measure zero \( N \) such that for \( \omega \notin N, t \to ||\Phi(t, \omega)|| \) is in \( L^p(\mathbb{R}) \). Therefore by continuity of translation, as \( n \to \infty \) it follows that for \( \omega \notin N, t \in [0, T] \),

\[
\int_{\mathbb{R}} ||\Phi(\gamma_n(t) + s) - \Phi(t + s)||_E^p \, ds \to 0
\]

The above is dominated by

\[
\int_{\mathbb{R}} 2^{p-1} (||\Phi(s)||^p + ||\Phi(s)||^p) \chi_{[-2T, 2T]}(s) \, ds
\]

\[
= \int_{-2T}^{2T} 2^{p-1} (||\Phi(s)||^p + ||\Phi(s)||^p) \, ds < \infty
\]

Consider

\[
\int_{-2T}^{2T} \left( \int_{\mathbb{R}} ||\Phi(\gamma_n(t) + s) - \Phi(t + s)||_E^p \, ds \right) \, dt
\]

By the dominated convergence theorem, this converges to 0 as \( n \to \infty \). Now Fubini. This yields

\[
\int_{\mathbb{R}} \int_{-2T}^{2T} ||\Phi(\gamma_n(t) + s) - \Phi(t + s)||_E^p \, dt \, ds
\]

Change the variables on the inside.

\[
\int_{\mathbb{R}} \int_{-2T+s}^{2T+s} ||\Phi(\gamma_n(t - s) + s) - \Phi(t)||_E^p \, dt \, ds
\]
Now by definition, \( \Phi(t) \) vanishes if \( t \notin [0,T] \), thus the above reduces to
\[
\int_{\mathbb{R}} \int_0^T \| \Phi(\gamma_n(t-s)+s) - \Phi(t) \|^p_E \, dt \, ds
+ \int_{\mathbb{R}} \int_{-2T+s}^{2T+s} \mathcal{X}_{[0,T]^c \mid [0,T]} \| \Phi(\gamma_n(t-s)+s) - \Phi(t) \|^p_E \, dt \, ds
\]
\[
= \int_{\mathbb{R}} \int_0^T \| \Phi(\gamma_n(t-s)+s) - \Phi(t) \|^p_E \, dt \, ds
+ \int_{\mathbb{R}} \int_{-2T+s}^{2T+s} \mathcal{X}_{[0,T]^c \mid [0,T]} \| \Phi(\gamma_n(t-s)+s) - \Phi(t) \|^p_E \, dt \, ds
\]
Also by definition, \( \gamma_n(t-s)+s \) is within \( 2^{-n} \) of \( t \) and so the integrand in the integral on the right equals 0 unless \( t \in [-2^{-n} - T, T + 2^{-n}] \subseteq [-2T, 2T] \). Thus the above reduces to
\[
\int_{\mathbb{R}} \int_{-2T}^{2T} \| \Phi(\gamma_n(t-s)+s) - \Phi(t) \|^p_E \, dt \, ds.
\]
This converges to 0 as \( n \to \infty \) as was shown above. Therefore,
\[
\int_0^T \int_0^T \| \Phi(\gamma_n(t-s)+s) - \Phi(t) \|^p_E \, dt \, ds
\]
also converges to 0 as \( n \to \infty \). The only problem is that \( \gamma_n(t-s)+s \geq t - 2^{-n} \) and so \( \gamma_n(t-s)+s \) could be less than 0 for \( t \in [0, 2^{-n}] \). Since this is an interval whose measure converges to 0 it follows
\[
\int_0^T \int_0^T \| \Phi(\gamma_n(t-s)+s)^+ - \Phi(t) \|^p_E \, dt \, ds
\]
converges to 0 as \( n \to \infty \). Let
\[
m_n(s) = \int_0^T \| \Phi(\gamma_n(t-s)+s)^+ - \Phi(t) \|^p_E \, dt
\]
Then letting \( \mu \) denote Lebesgue measure,
\[
\mu([m_n(s) > \lambda]) \leq \frac{1}{\lambda} \int_0^T m_n(s) \, ds.
\]
It follows there exists a subsequence \( n_k \) such that
\[
\mu \left( \left[ m_{n_k}(s) > \frac{1}{k} \right] \right) < 2^{-k}
\]
Hence by the Borel Cantelli lemma, there exists a set of measure zero \( N \) such that for \( s \notin N \),
\[
m_{n_k}(s) \leq 1/k
\]
for all \( k \) sufficiently large. Pick such an \( s \). Then consider \( t \to \Phi(\gamma_{n_k}(t-s)+s)^+ \). For \( n_k, t \to (\gamma_{n_k}(t-s)+s)^+ \) has jumps at points of the form \( 0, s + 12^{-n} \) where \( l \) is an integer. Thus \( \mathcal{P}_{n_k} \) consists of points of \( [0,T] \) which are of this form and these partitions are nested. Define \( \Phi_k(t) \equiv \Phi(\gamma_{n_k}(t-s)+s)^+ \). Now suppose \( N_1 \) is a set of measure zero. Can \( s \) be chosen such that all jumps for all partitions occur off \( N_1 \)? Let \((a, b)\) be an interval contained in \([0, T]\). Let \( S_j \) be the points of \((a, b)\) which are translations of the measure zero set \( N_1 \) by \( t_j \) for some \( j \). Thus \( S_j \) has measure 0. Now pick \( s \in (a, b) \setminus \cup_j S_j \). To get the other sequence of step functions, the right step functions, just use a similar argument with \( \delta_n \) in place of \( \gamma_n \). Just apply the argument to a subsequence of \( n_k \) so that the same \( s \) can hold for both. \( \blacksquare \)
**Theorem 21.3.2** Let $V \subseteq H = H' \subseteq V'$ be a Gelfand triple and suppose $Y \in L^p (0, T; V') \equiv K'$ and

$$X (t) = X_0 + \int_0^t Y (s) \, ds \quad \text{in} \quad V'$$

(21.3.16)

where $X_0 \in H$, and it is known that $X \in L^p (0, T, V) \equiv K$ for $p > 1$. Then $t \to X (t)$ is in $C ([0, T], H)$ and also

$$\frac{1}{2} |X (t)|_H^2 = \frac{1}{2} |X_0|_H^2 + \int_0^t \langle Y (s), X (s) \rangle \, ds$$

**Proof:** By Lemma 17.3.1, there exists a sequence of uniform partitions $\{t_k^n\}_{k=0}^{m_n} = \mathcal{P}_n, \mathcal{P}_n \subseteq \mathcal{P}_{n+1}$, of $[0, T]$ such that the step functions

$$\sum_{k=0}^{m_n-1} X (t_k^n) \mathcal{X}_{(t_k^n, t_{k+1}^n)} (t) = X^l (t)$$

$$\sum_{k=0}^{m_n-1} X (t_{k+1}^n) \mathcal{X}_{(t_k^n, t_{k+1}^n)} (t) = X^r (t)$$

close to $X$ in $K$ and in $L^2 ([0, T], H)$.

**Lemma 21.3.3** Let $s < t$. Then for $X, Y$ satisfying 21.3.16

$$|X (t)|^2 = |X (s)|^2 + 2 \int_s^t \langle Y (u), X (t) \rangle \, du - |X (t) - X (s)|^2$$ 

(21.3.17)

**Proof:** It follows from the following computations

$$X (t) - X (s) = \int_s^t Y (u) \, du$$

$$- |X (t) - X (s)|^2 = - |X (t)|^2 + 2 \langle X (t), X (s) \rangle - |X (s)|^2$$

$$= - |X (t)|^2 + 2 \left( \langle X (t), X (t) - \int_s^t Y (u) \, du \rangle - |X (s)|^2 \right)$$

$$= - |X (t)|^2 + 2 |X (t)|^2 - 2 \left( \int_s^t \langle Y (u), X (t) \rangle \, du - |X (s)|^2 \right)$$

Hence

$$|X (t)|^2 = |X (s)|^2 + 2 \int_s^t \langle Y (u), X (t) \rangle \, du - |X (t) - X (s)|^2$$

**Lemma 21.3.4** In the above situation,

$$\sup_{t \in [0, T]} |X (t)|_H \leq C \left( \|Y\|_{K'}, \|X\|_K \right)$$

Also, $t \to X (t)$ is weakly continuous with values in $H$.

**Proof:** From the above formula applied to the $k^{th}$ partition of $[0, T]$ described above,

$$|X (t_m)|^2 - |X_0|^2 = \sum_{j=0}^{m-1} |X (t_{j+1})|^2 - |X (t_j)|^2$$

$$\quad = \sum_{j=0}^{m-1} 2 \int_{t_j}^{t_{j+1}} \langle Y (u), X (t_j) \rangle \, du - |X (t_{j+1}) - X (t_j)|_H^2$$

$$\quad = \sum_{j=0}^{m-1} 2 \int_{t_j}^{t_{j+1}} \langle Y (u), X^r (u) \rangle \, du - |X (t_{j+1}) - X (t_j)|_H^2$$
Thus, discarding the negative terms and denoting by $\mathcal{P}_k$ the $k^{th}$ of these partitions,

$$\sup_{t_j \in \mathcal{P}_k} |X(t_j)|^2_H \leq |X_0|^2 + 2 \int_0^T \|Y(u), X^+_k(u)\|_{V'} \, du$$

$$\leq |X_0|^2 + 2 \int_0^T \|Y(u)\|_{V'}, \|X^+_k(u)\|_{V'} \, du$$

$$\leq |X_0|^2 + 2 \left( \int_0^T \|Y(u)\|_{V'}^p \, du \right)^{1/p} \left( \int_0^T \|X^+_k(u)\|_{V'}^p \, du \right)^{1/p} \leq C(\|Y\|_{K'}, \|X\|_K)$$

because these partitions are chosen such that

$$\lim_{k \to \infty} \left( \int_0^T \|X^+_k(u)\|_{V'}^p \, du \right)^{1/p} = \left( \int_0^T \|X(u)\|_{V'}^p \, du \right)^{1/p}$$

and so these are bounded. This has shown that for the dense subset of $[0,T], D \equiv \cup_k \mathcal{P}_k$, $\sup_{t \in D} |X(t)| < C(\|Y\|_{K'}, \|X\|_K)$

Now let $\{g_k\}^\infty_{k=1}$ be linearly independent vectors of $V$ whose span is dense in $V$. This is possible because $V$ is separable. Then let $\{e_j\}^\infty_{j=1}$ be an orthonormal basis for $H$ such that $e_k \in \text{span}(g_1, \ldots, g_k)$ and each $g_k \in \text{span}(e_1, \ldots, e_k)$. This is done with the Gram Schmidt process. Then it follows that span $\{e_k\}^\infty_{k=1}$ is dense in $V$. I claim

$$|y|^2_H = \sum_{j=1}^\infty |\langle y, e_j \rangle|^2.$$ 

This is certainly true if $y \in H$ because

$$\langle y, e_j \rangle = \langle y, e_j \rangle_H$$

If $y \notin H$, then the series must diverge since otherwise, you could consider the infinite sum

$$\sum_{j=1}^\infty \langle y, e_j \rangle e_j \in H$$

because

$$\left| \sum_{j=p}^q \langle y, e_j \rangle e_j \right|^2 = \sum_{j=p}^q |\langle y, e_j \rangle|^2 \to 0 \text{ as } p, q \to \infty.$$ 

Letting $z = \sum_{j=1}^\infty \langle y, e_j \rangle e_j$, it follows that $\langle y, e_j \rangle$ is the $j^{th}$ Fourier coefficient of $z$ and that

$$\langle z - y, v \rangle = 0$$

for all $v \in \text{span}(\{e_k\}^\infty_{k=1})$ which is dense in $V$. Therefore, $z = y$ in $V'$ and so $y \in H$.

It follows

$$|X(t)|^2 = \sup_n \sum_{j=1}^n |\langle X(t), e_j \rangle|^2$$

which is just the sup of continuous functions of $t$. Therefore, $t \to |X(t)|^2$ is lower semicontinuous. It follows that for any $t$, letting $t_j \to t$ for $t_j \in D$,

$$|X(t)|^2 \leq \liminf_{j \to \infty} |X(t_j)|^2 \leq C(\|Y\|_{K'}, \|X\|_K)$$

This proves the first claim of the lemma.

Consider now the claim that $t \to X(t)$ is weakly continuous. Letting $v \in V$,

$$\lim_{t \to s} (X(t), v) = \lim_{t \to s} \langle X(t), v \rangle = \langle X(s), v \rangle = (X(s), v)$$
Since it was shown that \(|X(t)|\) is bounded independent of \(t\), and since \(V\) is dense in \(H\), the claim follows.  

Now

\[
- \sum_{j=0}^{m-1} |X(t_{j+1}) - X(t_j)|_H^2 = |X(t_m)|^2 - |X_0|^2 - \sum_{j=0}^{m-1} 2 \int_{t_j}^{t_{j+1}} \langle Y(u), X_k^r(u) \rangle \, du
\]

\[
= |X(t_m)|^2 - |X_0|^2 - 2 \int_0^{t_m} \langle Y(u), X_k^r(u) \rangle \, du
\]

Thus, since the partitions are nested, eventually \(|X(t_m)|^2\) is constant for all \(k\) large enough and the integral term converges to

\[
\int_0^{t_m} \langle Y(u), X(u) \rangle \, du
\]

It follows that the term on the left does converge to something. It just remains to consider what it does converge to. However, from the equation solved by \(X\),

\[
X(t_{j+1}) - X(t_j) = \int_{t_j}^{t_{j+1}} Y(u) \, du
\]

Therefore, this term is dominated by an expression of the form

\[
\sum_{j=0}^{m-1} \left( \int_{t_j}^{t_{j+1}} Y(u) \, du, X(t_{j+1}) - X(t_j) \right) = \sum_{j=0}^{m-1} \left( \int_{t_j}^{t_{j+1}} Y(u) \, du, X(t_{j+1}) - X(t_j) \right)
\]

\[
= \sum_{j=0}^{m-1} \int_{t_j}^{t_{j+1}} \langle Y(u), X(t_{j+1}) - X(t_j) \rangle \, du
\]

\[
= \int_0^T \langle Y(u), X^r(u) \rangle \, du - \int_0^T \langle Y(u), X^l(u) \rangle \, du
\]

However, both \(X^r\) and \(X^l\) converge to \(X\) in \(K = L^p(0,T,V)\). Therefore, this term must converge to 0. Passing to a limit, it follows that for all \(t \in D\), the desired formula holds. Thus, for such \(t\),

\[
|X(t)|^2 = |X_0|^2 + 2 \int_0^t \langle Y(u), X(u) \rangle \, du
\]

It remains to verify that this holds for all \(t\). Let \(t \notin D\) and let \(t(k) \in \mathcal{P}_k\) be the largest point of \(\mathcal{P}_k\) which is less than \(t\). Suppose \(t(m) \leq t(k)\) so that \(m \leq k\). Then

\[
X(t(m)) = X_0 + \int_0^{t(m)} Y(s) \, ds,
\]

a similar formula for \(X(t(k))\). Thus for \(t > t(m)\),

\[
X(t) - X(t(m)) = \int_{t(m)}^t Y(s) \, ds
\]

which is the same sort of thing already looked at except that it starts at \(t(m)\) rather than at 0 and \(X_0 = 0\). Therefore,

\[
|X(t(k)) - X(t(m))|^2 = 2 \int_{t(m)}^{t(k)} \langle Y(s), X(s) - X(t(m)) \rangle \, ds
\]

Thus, for \(m \leq k\)

\[
\lim_{m,k \to \infty} |X(t(k)) - X(t(m))|^2 = 0
\]
Hence \( \{ X ( t ( k ) ) \}_{k=1}^{\infty} \) is a convergent sequence in \( H \). Does it converge to \( X ( t ) \)? Let \( \xi ( t ) \in H \) be what it does converge to. Let \( v \in V \). Then

\[
( \xi ( t ), v ) = \lim_{k \to \infty} ( X ( t ( k ) ), v ) = \lim_{k \to \infty} ( X ( t ( k ) ), v ) = ( X ( t ), v ) = ( X ( t ), v )
\]

because it is known that \( t \to X ( t ) \) is continuous into \( V' \) and it is also known that \( X ( t ) \in H \) and that the \( X ( t ) \) for \( t \in [0, T] \) are uniformly bounded. Therefore, since \( V \) is dense in \( H \), it follows that \( \xi ( t ) = X ( t ) \).

Now for every \( t \in D \), it was shown above that

\[
| X ( t ) |^2 = | X_0 |^2 + 2 \int_0^t ( Y ( s ), X ( s ) ) \, ds
\]

Thus, using what was just shown, if \( t \notin D \) and \( t_k \to t \),

\[
| X ( t ) |^2 = \lim_{k \to \infty} | X ( t_k ) |^2 = \lim_{k \to \infty} \left( | X_0 |^2 + 2 \int_0^{t_k} ( Y ( s ), X ( s ) ) \, ds \right) = | X_0 |^2 + 2 \int_0^t ( Y ( s ), X ( s ) ) \, ds
\]

which proves the desired formula. From this it follows right away that \( t \to X ( t ) \) is continuous into \( H \) because it was just shown that \( t \to | X ( t ) | \) is continuous and \( t \to X ( t ) \) is weakly continuous. Since Hilbert space is uniformly convex, this implies the \( t \to X ( t ) \) is continuous. To see this in the special case of Hilbert space,

\[
| X ( t ) - X ( s ) |^2 = | X ( t ) |^2 - 2 ( X ( s ), X ( t ) ) + | X ( s ) |^2
\]

Then \( \lim_{t \to s} \left( | X ( t ) |^2 - 2 ( X ( s ), X ( t ) ) + | X ( s ) |^2 \right) = 0 \) by weak convergence of \( X ( t ) \) to \( X ( s ) \) and the convergence of \( | X ( t ) |^2 \) to \( | X ( s ) |^2 \). □

## 21.4 The Implicit Case

The above theorem can be generalized to the case where the formula is of the form

\[
BX ( t ) = BX_0 + \int_0^t Y ( s ) \, ds
\]

This involves an operator \( B \in \mathcal{L}(W, W') \) and \( B \) satisfies

\[
\langle Bx, x \rangle \geq 0, \quad \langle Bx, y \rangle = \langle By, x \rangle
\]

for

\[
V \subseteq W, W' \subseteq V'
\]

Where \( V \) is dense in the Banach space \( W \). Before giving the theorem, here is a technical lemma. First is one which is not so technical.

**Lemma 21.4.1** Let \( V \) be a separable Banach space. Then there exists \( \{ g_k \}_{k=1}^{\infty} \) which are linearly independent and whose span is dense in \( V \).

**Proof:** Let \( \{ f_k \} \) be a countable dense subset. Delete \( f_{k_1} \) such that \( k_1 \) is the first index such that \( f_k \) is in the span of the other vectors. That is, it is the first which is a finite linear combination of the others. If no such vector exists, then you have what is wanted. Next delete \( f_{k_2} \) where \( k_2 \) is the next for which \( f_k \) is a linear combination of the others. Continue. The remaining vectors must be linearly independent. If not, there would be a first which is a linear combination of the others. Say \( f_m \). But the process would have eliminated it at the \( m^{th} \) step. □

**Lemma 21.4.2** Suppose \( V, W \) are separable Banach spaces such that \( V \) is dense in \( W \) and \( B \in \mathcal{L}(W, W') \) satisfies

\[
\langle Bx, x \rangle \geq 0, \quad \langle Bx, y \rangle = \langle By, x \rangle, \quad B \neq 0.
\]

Then there exists a countable set \( \{ e_i \} \) of vectors in \( V \) such that

\[
\langle Be_i, e_j \rangle = \delta_{ij}
\]
and for each \( x \in W \),

\[
\langle Bx, x \rangle = \sum_{i=1}^{\infty} |\langle Bx, e_i \rangle|^2,
\]
and also

\[
Bx = \sum_{i=1}^{\infty} \langle Bx, e_i \rangle Be_i,
\]

the series converging in \( W' \). If \( B = B(\omega) \) and \( B \) is \( \mathcal{F} \) measurable into \( \mathcal{L}(W, W') \) and if the \( e_i = e_i(\omega) \) are as described above, then these \( e_i \) are measurable into \( V \). Therefore, if \( t \to B(t, \omega) \) is \( C^1([0, T], \mathcal{L}(W, W')) \) and if for each \( w \in W \),

\[
\langle B'(t, \omega) w, w \rangle \leq k_{w, \omega}(t) \langle B(t, \omega) w, w \rangle
\]

Where \( k_{w, \omega} \in L^1([0, T]) \), then the vectors \( e_i(t) \) can be chosen to also be right continuous functions of \( t \).

In the case of dependence on \( t \), the extra condition is trivial if \( \langle B(t, \omega) x, x \rangle \geq \delta \|w\|^2_W \) for example. This includes the usual case of evolution equations where \( W = H = H' = W' \). It also includes the case where \( B \) does not depend on \( t \).

**Proof:** Let \( \{g_k\}_{k=1}^{\infty} \) be linearly independent vectors of \( V \) whose span is dense in \( V \). This is possible because \( V \) is separable. Thus, their span is also dense in \( W \). Let \( n_1 \) be the first index such that \( \langle Bg_{n_1}, g_{n_1} \rangle \neq 0 \).

**Claim:** If there is no such index, then \( B = 0 \).

**Proof of claim:** First note that if there is no such first index, then if \( x = \sum_{i=1}^{k} a_i g_i \)

\[
|\langle Bx, x \rangle| = \left| \sum_{i \neq j} a_i a_j \langle Bg_i, g_j \rangle \right| \leq \sum_{i \neq j} |a_i| |a_j| |\langle Bg_i, g_j \rangle| \leq \sum_{i \neq j} |a_i| |a_j| \langle Bg_i, g_j \rangle^{1/2} \langle Bg_j, g_j \rangle^{1/2} = 0
\]

Therefore, if \( x \) is given, you could take \( x_k \) in the span of \( \{g_1, \ldots, g_k\} \) such that \( \|x_k - x\|_W \to 0 \). Then

\[
|\langle Bx, y \rangle| = \lim_{k \to \infty} |\langle Bx_k, y \rangle| \leq \lim_{k \to \infty} \langle Bx_k, x_k \rangle^{1/2} \langle By, y \rangle^{1/2} = 0
\]

because \( \langle Bx_k, x_k \rangle \) is zero by what was just shown. Hence the conclusion of the lemma is trivially true. Just pick \( e_1 = g_1 \) and let \( \{e_1\} \) be your set of vectors.

Thus assume there is such a first index. Let

\[
e_1 \equiv \frac{g_{n_1}}{\langle Bg_{n_1}, g_{n_1} \rangle^{1/2}}
\]

Then \( \langle Be_1, e_1 \rangle = 1 \). Now if you have constructed \( e_j \) for \( j \leq k \),

\[
e_j \in \text{span} \{g_1, \ldots, g_k\}, \quad \langle Be_i, e_j \rangle = \delta_{ij},
\]

\( g_{n_{j+1}} \) being the first in the list \( \{g_j\} \) for which

\[
\left( Bg_{n_{j+1}} - \sum_{i=1}^{j} \langle Bg_{n_{j+1}}, e_i \rangle Be_i, g_{n_{j+1}} - \sum_{i=1}^{j} \langle Bg_{n_{j}}, e_i \rangle e_i \right) \neq 0,
\]

and

\[
\text{span} \{g_1, \ldots, g_k\} = \text{span} \{e_1, \ldots, e_k\},
\]

let \( g_{n_{k+1}} \) be such that \( g_{n_{k+1}} \) is the first in the list \( \{g_n\} \) \( n_{k+1} > n_k \) such that

\[
\left( Bg_{n_{k+1}} - \sum_{i=1}^{k} \langle Bg_{n_{k+1}}, e_i \rangle Be_i, g_{n_{k+1}} - \sum_{i=1}^{k} \langle Bg_{n_{k+1}}, e_i \rangle e_i \right) \neq 0
\]

Note the difference between this and the Gram Schmidt process. Here you don’t necessarily use all of the \( g_k \) due to the possible degeneracy of \( B \).
Claim: If there is no such first \( g_{n_0+1} \), then \( B(\langle e_1, \cdots, e_k \rangle) = BW \) so in this case, \( \{Be_i\}_{i=1}^k \) is actually a basis for \( BW \).

Proof: To see this, note that if \( p \in (n_j, n_{j+1}) \), then by assumption,

\[
\left< B \left( g_p - \sum_{i=1}^{j} \langle Bg_p, e_i \rangle e_i \right) , g_p - \sum_{i=1}^{j} \langle Bg_p, e_i \rangle e_i \right> = 0
\]

Therefore,

\[Bg_p = \sum_{i=1}^{j} \langle Bg_p, e_i \rangle Be_i\]

Also, by assumption, if \( p > n_k\)
\[
\left< B \left( g_p - \sum_{i=1}^{k} \langle Bg_p, e_i \rangle e_i \right) , g_p - \sum_{i=1}^{k} \langle Bg_p, e_i \rangle e_i \right> = 0
\]

so

\[Bg_p = \sum_{i=1}^{k} \langle Bg_p, e_i \rangle Be_i\]

which shows that \( \text{span} \left( \{Bg_j \}_{j=1}^{\infty} \right) \subseteq \text{span} \left( \{Be_i \}_{i=1}^{k} \right) \). Hence if \( x \in W \), then letting \( x_r \in \text{span} \left( \{g_j \}_{j=1}^{\infty} \right) \) with \( x_r \to x \) in \( W \), it follows

\[Bx_r = \sum_{i=1}^{k} a_i Be_i = \sum_{i=1}^{k} \langle Bx_r, e_i \rangle Be_i\]

Then passing to a limit, you get

\[Bx = \sum_{i=1}^{k} \langle Bx, e_i \rangle Be_i\]

Thus \( \{Be_i\}_{i=1}^{k} \) is a basis for \( BW \). This proves the claim.

If this happens, the process being described stops. You have found what is desired which has only finitely many vectors involved.

If the process does not stop, let

\[e_{k+1} = \frac{g_{n_{k+1}} - \sum_{i=1}^{k} \langle Bg_{n_{k+1}}, e_i \rangle e_i}{\left< B \left( g_{n_{k+1}} - \sum_{i=1}^{k} \langle Bg_{n_{k+1}}, e_i \rangle e_i \right) , g_{n_{k+1}} - \sum_{i=1}^{k} \langle Bg_{n_{k+1}}, e_i \rangle e_i \right>^{1/2}}\]

Thus, as in the usual argument for the Gram Schmidt process, \( \langle Be_i, e_j \rangle = \delta_{ij} \) for \( i, j \leq k + 1 \). This is already known for \( i, j \leq k \). Letting \( l \leq k \), and using the orthogonality already shown,

\[
\langle Be_{k+1}, e_l \rangle = C \left< B \left( g_{n_{k+1}} - \sum_{i=1}^{k} \langle Bg_{n_{k+1}}, e_i \rangle e_i \right) , e_l \right> = C \left( \langle Bg_{k+1}, e_l \rangle - \langle Bg_{n_{k+1}}, e_l \rangle \right) = 0
\]

Consider

\[
\left< Bg_p - B \left( \sum_{i=1}^{k} \langle Bg_p, e_i \rangle e_i \right) , g_p - \sum_{i=1}^{k} \langle Bg_p, e_i \rangle e_i \right>
\]
If \( p \in (n_k, n_{k+1}) \), then the above equals zero which implies
\[
Bg_p = \sum_{i=1}^{k} \langle Bg_p, e_i \rangle Be_i
\]

On the other hand, suppose \( g_p = g_{n_{k+1}} \) for some \( n_{k+1} \) and so, from the construction, \( g_{n_{k+1}} = g_p \in \text{span} (e_1, \cdots, e_{k+1}) \) and therefore,
\[
g_p = \sum_{j=1}^{k+1} a_j e_j
\]

which requires easily that
\[
Bg_p = \sum_{i=1}^{k+1} \langle Bg_p, e_i \rangle Be_i,
\]
the above holding for all \( k \) large enough. To see this last claim, note that the coefficients of \( Bg = \sum_{j=1}^{m} a_j Be_j \) are required to be \( a_j = \langle Bg, e_j \rangle \) and from the construction, \( \langle Be_i, e_j \rangle = \delta_{ij} \). Thus if the upper limit is increased beyond what is needed, the new terms are all zero. It follows that for any \( x \in \text{span} (\{g_k\}_{k=1}^{\infty}) \), (finite linear combination of vectors in \( \{g_k\}_{k=1}^{\infty} \))
\[
Bx = \sum_{i=1}^{\infty} \langle Bx, e_i \rangle Be_i
\]
because for all \( k \) large enough,
\[
Bx = \sum_{i=1}^{k} \langle Bx, e_i \rangle Be_i
\]

Also note that for such \( x \in \text{span} (\{g_j\}_{j=1}^{\infty}) \),
\[
\langle Bx, x \rangle = \sum_{i=1}^{k} \langle Bx, e_i \rangle Be_i, x = \sum_{i=1}^{k} \langle Bx, e_i \rangle \langle Bx, e_i \rangle
\]
\[
= \sum_{i=1}^{k} |\langle Bx, e_i \rangle|^2 = \sum_{i=1}^{\infty} |\langle Bx, e_i \rangle|^2
\]
Now for \( x \) arbitrary, let \( x_k \to x \) in \( W \) where \( x_k \in \text{span} (\{g_k\}_{k=1}^{\infty}) \). Then by Fatou’s lemma,
\[
\sum_{i=1}^{\infty} |\langle Bx, e_i \rangle|^2 \leq \liminf_{k \to \infty} \sum_{i=1}^{\infty} |\langle Bx_k, e_i \rangle|^2
\]
\[
= \liminf_{k \to \infty} \langle Bx_k, x_k \rangle = \langle Bx, x \rangle
\]
\[
\leq \|Bx\|_W \|x\|_W \leq \|B\| \|x\|_W^2
\]
Thus the series on the left converges. Then also, from the above inequality,
\[
\left| \sum_{i=p}^{q} \langle Bx, e_i \rangle Be_i, y \right| \leq \sum_{i=p}^{q} |\langle Bx, e_i \rangle| |\langle Be_i, y \rangle|
\]
\[
\leq \left( \sum_{i=p}^{q} |\langle Bx, e_i \rangle|^2 \right)^{1/2} \left( \sum_{i=p}^{q} |\langle By, e_i \rangle|^2 \right)^{1/2}
\]
\[
\leq \left( \sum_{i=p}^{q} |\langle Bx, e_i \rangle|^2 \right)^{1/2} \left( \sum_{i=1}^{\infty} |\langle By, e_i \rangle|^2 \right)^{1/2}
\]
21.4. THE IMPLICIT CASE

By (21.4.18),

\[ \leq \left( \sum_{i=p}^{q} |\langle Bx, e_i \rangle|^2 \right)^{1/2} \left( \|B\| \|y\|_W \right)^{1/2} \leq \left( \sum_{i=p}^{q} |\langle Bx, e_i \rangle|^2 \right)^{1/2} \|B\|^{1/2} \|y\|_W \]

It follows that

\[ \sum_{i=1}^{\infty} \langle Bx, e_i \rangle \cdot Be_i \] (21.4.20)

converges in \( W' \) because it was just shown that

\[ \left\| \sum_{i=p}^{q} \langle Bx, e_i \rangle \cdot Be_i \right\|_{W'} \leq \left( \sum_{i=p}^{q} |\langle Bx, e_i \rangle|^2 \right)^{1/2} \|B\|^{1/2} \]

and it was shown above that \( \sum_{i=1}^{\infty} |\langle Bx, e_i \rangle|^2 < \infty \), so the partial sums of the series are a Cauchy sequence in \( W' \). Also, the above estimate shows that for \( \|y\| = 1 \),

\[ \left\langle \sum_{i=1}^{\infty} \langle Bx, e_i \rangle \cdot Be_i, y \right\rangle \leq \left( \sum_{i=1}^{\infty} |\langle Bx, e_i \rangle|^2 \right)^{1/2} \left( \sum_{i=1}^{\infty} |\langle Bx, e_i \rangle|^2 \right)^{1/2} \|B\|^{1/2} \]

and so

\[ \left\| \sum_{i=1}^{\infty} \langle Bx, e_i \rangle \cdot Be_i \right\|_{W'} \leq \left( \sum_{i=1}^{\infty} |\langle Bx, e_i \rangle|^2 \right)^{1/2} \|B\|^{1/2} \] (21.4.21)

Now for \( x \) arbitrary, let \( x_k \in \text{span} \left\{ g_i \right\}_{i=1}^{\infty} \) and \( x_k \to x \) in \( W \). Then for a fixed \( k \) large enough,

\[ \left\| Bx - \sum_{i=1}^{\infty} \langle Bx, e_i \rangle \cdot Be_i \right\| \leq \left\| Bx - Bx_k \right\| \]

\[ + \left\| Bx_k - \sum_{i=1}^{\infty} \langle Bx_k, e_i \rangle \cdot Be_i \right\| \]

\[ + \left\| \sum_{i=1}^{\infty} \langle Bx_k, e_i \rangle \cdot Be_i - \sum_{i=1}^{\infty} \langle Bx, e_i \rangle \cdot Be_i \right\| \]

\[ \leq \epsilon + \left\| \sum_{i=1}^{\infty} \langle B(x_k - x), e_i \rangle \cdot Be_i \right\| \]

the term

\[ \left\| Bx_k - \sum_{i=1}^{\infty} \langle Bx_k, e_i \rangle \cdot Be_i \right\| \]

equaling 0 by (21.4.18). From (21.4.21) and (21.4.19),

\[ \leq \epsilon + \|B\|^{1/2} \left( \sum_{i=1}^{\infty} |\langle B(x_k - x), e_i \rangle|^2 \right)^{1/2} \]

\[ \leq \epsilon + \|B\|^{1/2} \langle B(x_k - x), x_k - x \rangle^{1/2} < 2\epsilon \]

whenever \( k \) is large enough, the second inequality being implied by (21.4.18). Therefore,

\[ Bx = \sum_{i=1}^{\infty} \langle Bx, e_i \rangle \cdot Be_i \]

in \( W' \). It follows that

\[ \langle Bx, x \rangle = \lim_{k \to \infty} \left( \sum_{i=1}^{k} \langle Bx, e_i \rangle \cdot Be_i, x \right) = \lim_{k \to \infty} \sum_{i=1}^{k} |\langle Bx, e_i \rangle|^2 \equiv \sum_{i=1}^{\infty} |\langle Bx, e_i \rangle|^2 \]
Now consider the measurability assertion on the \( e_i \). Consider first \( e_1 \). Begin by considering \( n_1 (\omega) \)
\[
E^1_k \equiv \{ \omega : \langle B (\omega) \, g_k, g_k \rangle \neq 0 \} \cap \cap_{j<k} \{ \omega : \langle B (\omega) \, g_j, g_j \rangle = 0 \}
\]
As explained above, \( B (\omega) = 0 \), if and only if \( E^1_k = \emptyset \) for all \( k \). Also note that these \( E^1_k \) are disjoint and \( \mathcal{F} \) measurable. Then
\[
n_1 (\omega) = \begin{cases} 
1 & \text{if } \omega \notin \cup_k E^1_k \\
0 & \text{if } \omega \in E^1_k 
\end{cases}
\]
Thus \( n_1 (\omega) \) is clearly measurable because it is constant on measurable sets. Then from the algorithm,
\[
e_1 (\omega) \equiv \mathcal{X}_{\cup_k E^1_k} (\omega) \frac{g_{n_1 (\omega)}}{\langle B g_{n_1 (\omega)}, g_{n_1 (\omega)} \rangle^{1/2}}
\]
Thus \( e_1 (\omega) = 0 \) if \( \omega \notin \cup_k E^1_k \). Also \( e_1 (\omega) \) is measurable because \( \omega \to n_1 (\omega) \) is measurable. Thus \( e_1 \) has constant values on measurable sets. So suppose \( n_1 (\omega) \) is measurable for \( i \leq m \). Then define \( E^{m+1}_p \equiv \)
\[
\{ \omega : \langle B g_p - \sum_{i=1}^{m} \langle B g_p, e_i \rangle B e_i, g_p - \sum_{i=1}^{m} \langle B g_p, e_i \rangle e_i \rangle \neq 0 \} \cap \{ \omega : n_m (\omega) < p \}
\]
As earlier, these sets \( \{ E^{m+1}_p \}_{p=1}^{\infty} \) are disjoint and measurable. As before, let \( n_{m+1} (\omega) = p \) where \( \omega \in E^{m+1}_p \). Then from the algorithm, \( e_{m+1} (\omega) \equiv \)
\[
\mathcal{X}_{\cup_p E^{m+1}_p} (\omega) \frac{g_{n_{m+1} (\omega)} - \sum_{i=1}^{m} \langle B g_{n_{m+1} (\omega)}, e_i \rangle e_i}{D_m}
\]
where
\[
D_m = \langle B \left( g_{n_{m+1} (\omega)} - \sum_{i=1}^{m} \langle B g_{n_{m+1} (\omega)}, e_i \rangle e_i \right), g_{n_{m+1} (\omega)} - \sum_{i=1}^{m} \langle B g_{n_{m+1} (\omega)}, e_i \rangle e_i \rangle^{1/2}
\]
Thus the \( e_k (\omega) \) are all measurable into \( W \) thanks to the algorithm. However, they all have values in \( V \). Thus if \( \phi \in V' \), let \( \phi_n \to \phi \) in \( V' \) where \( \phi_n \in W' \).
\[
\langle \phi, e_k (\omega) \rangle_{V', V} = \lim_{n \to \infty} \langle \phi, e_k (\omega) \rangle_{V', V} = \lim_{n \to \infty} \langle \phi, e_k (\omega) \rangle_{W', W}
\]
which is the limit of measurable functions. By the Pettis theorem, this shows \( e_k \) is measurable into \( V \) also.

To verify the assertion on right continuity, the same kind of argument holds. We suppress the dependence on \( \omega \). Consider first \( e_1 \). Begin by considering \( n_1 (t) \)
\[
E^1_k \equiv \{ t : \langle B (t) \, g_k, g_k \rangle \neq 0 \} \cap \cap_{j<k} \{ t : \langle B (t) \, g_j, g_j \rangle = 0 \}
\]
As explained above, \( B (t) = 0 \), if and only if \( E^1_k = \emptyset \) for all \( k \). Also note that these \( E^1_k \) are disjoint. Then
\[
n_1 (t) = \begin{cases} 
1 & \text{if } t \notin \cup_k E^1_k \\
0 & \text{if } t \in E^1_k 
\end{cases}
\]
If \( t \in E^1_k \), then from the definition, \( \langle B (t) \, g_k, g_k \rangle \neq 0 \) and \( k \) is the first index for which this is nonzero. Let \( t_l \downarrow t \). Then by continuity, for all \( l \) large enough, \( \langle B (t_l) \, g_k, g_k \rangle \neq 0 \). What of \( \langle B (t_l) \, g_j, g_j \rangle \) for \( j < k \)? By assumption,
\[
\langle B' (t) \, g_j, g_j \rangle \leq k_{g_j} (t) \langle B (t) \, g_j, g_j \rangle
\]
and so, letting \( K_{g_j} (t) = \int_0^t k_{g_j} (s) \, ds \),
\[
\frac{d}{dt} \left( e^{-K_{g_j} (t)} \langle B (t) \, g_j, g_j \rangle \right) \leq 0
\]
\[
e^{-K_{g_j} (t_l)} \langle B (t_l) \, g_j, g_j \rangle \leq e^{-K_{g_j} (t)} \langle B (t) \, g_j, g_j \rangle = 0
\]
Thus one obtains right continuity of \( t \to n_1(t) \) and for \( E_k^1 \), there is an interval \([t, t + \delta]\) \( \subseteq E_k^1 \). From the algorithm,
\[
e_1(t) \equiv \mathcal{X}_{\cup_k E_k^1}(t) \frac{g_{n_1(t)}}{\langle Bg_{n_1(t)}, g_{n_1(t)} \rangle^{1/2}}
\]

Thus \( e_1(t) = 0 \) if \( t \notin \cup_k E_k^1 \). Also \( e_1(t) \) is right continuous because \( t \to n_1(t) \) is. Thus \( e_1 \) has constant values on a small interval starting at \( t \). But what about \( t \notin \cup_k E_k^1 \)? Why should it be right continuous there? If there is such a \( t \), then as explained above, \( B(t) = 0 \). Then if \( s \) is arbitrary, \( s > t \) and \( x \in W \),
\[
\langle B'(s)x, x \rangle \leq k_x \langle B(s)x, x \rangle
\]
and so as above,
\[
e^{-K_x(s)} \langle B(s)x, x \rangle \leq 0
\]

Thus this case reduces to having \( B(s) = 0 \) for all \( s \geq t \) and there is nothing to prove. You have \( n_1(s) = 1 \) and \( e_1(s) = 0 \) for all \( s \geq t \).

Suppose \( t \to n_i(t) \) is right continuous for \( i \leq m \) and that \( e_i \) is also. Then define \( E_p^{m+1} \equiv \)
\[
\left\{ t : \left\langle B g_p - \sum_{i=1}^{m} \langle B g_p, e_i \rangle B e_i, g_p - \sum_{i=1}^{m} \langle B g_p, e_i \rangle e_i \right\rangle \right\} \cap \{ t : n_m(t) < p \}
\]
\[
\cap \{ t : \langle B g_p - \sum_{i=1}^{m} \langle B g_p, e_i \rangle B e_i, g_p - \sum_{i=1}^{m} \langle B g_p, e_i \rangle e_i \rangle \}
\]
As earlier, these sets \( \{ E_p^{m+1} \right\}_{p=1}^{\infty} \) are disjoint. As before, let \( n_{m+1}(t) = p \) where \( t \in E_p^{m+1} \). Then by similar reasoning to the above, for small \( \delta, [t, t + \delta] \in E_p^{m+1} \) and \( n_{m+1}(s) = p \) for \( s \in [t, t + \delta] \). Then from the algorithm, \( e_{m+1}(t) \equiv \)
\[
\mathcal{X}_{\cup_p E_p^{m+1}}(t) \frac{g_{n_{m+1}(t)} - \sum_{i=1}^{m} \langle B g_{n_{m+1}(t)}, e_i(t) \rangle e_i(t)}{D_m}
\]
where \( D_m = \)
\[
\left\langle B \left( \frac{g_{n_{m+1}(t)} - \sum_{i=1}^{m} \langle B g_{n_{m+1}(t)}, e_i(t) \rangle e_i(t)}{g_{n_{m+1}(t)} - \sum_{i=1}^{m} \langle B g_{n_{m+1}(t)}, e_i(t) \rangle e_i(t)} \right) , \frac{1/2}{g_{n_{m+1}(t)} - \sum_{i=1}^{m} \langle B g_{n_{m+1}(t)}, e_i(t) \rangle e_i(t)} \right\rangle^{1/2}
\]
and so is right continuous. What of \( t \notin \cup_p E_p^{m+1} \)? In this case, the process has terminated and what is desired has been found. ■

Then the main result in this section is the following integration by parts theorem.

**Theorem 21.4.3** Let \( V \subseteq W, W' \subseteq V' \) be separable Banach spaces, and let \( Y \in L^{p'}(0, T; V') \) and
\[
Bu(t) = Bu_0 + \int_0^t Y(s) ds \text{ in } V', \ u_0 \in W, \ Bu(t) = B(u(t)) \text{ for a.e. } t \quad (21.4.22)
\]
As indicated, \( Bu \) is the name of a function satisfying the above equation which satisfies \( Bu(t) = B(u(t)) \) for a.e. \( t \). Thus \( Y = (Bu)' \) as a weak derivative in the sense of \( V' \) valued distributions. It is known that \( u \in L^p(0, T, V) \) for \( p > 1 \). Then \( t \to Bu(t) \) is continuous into \( W' \) for \( t \) off a set of measure zero \( N \) and also there exists a continuous function \( t \to \langle Bu, u(t) \rangle \) such that for all \( t \notin N \), \( \langle Bu, u(t) \rangle = \langle B(u(t)), u(t) \rangle \), \( Bu(t) = B(u(t)) \), and for all \( t \),
\[
\frac{1}{2} \langle Bu, u \rangle(t) = \frac{1}{2} \langle Bu_0, u_0 \rangle + \int_0^t \langle Y(s), u(s) \rangle ds
\]

**Proof:** By Lemma 21.4.3, there exists a sequence of partitions \( \{ t^n_k \}_{k=0}^{\infty} = P_n, P_n \subseteq P_{n+1} \) of \([0, T]\) such that the lengths of the sub intervals converge uniformly to 0 as \( n \to \infty \) and the step functions
\[
\sum_{k=0}^{m-n-1} u(t^n_k) \mathcal{X}_{(t^n_k, t^n_{k+1})}(t) = u^l(t)
\]
\[
\sum_{k=0}^{m-n-1} u(t^n_{k+1}) \mathcal{X}_{(t^n_k, t^n_{k+1})}(t) = u^r(t)
\]
converge to \( u \) in \( L^p(0, T; V) \equiv K \). We assume that all of these partition points have empty intersection with the set of measure zero where \( Bu(t) \neq B(u(t)) \). Thus, at every partition point, \( Bu(t_k) = B(u(t_k)) \). As just mentioned, \( L^p(0, T; V) \equiv K \), \( L^{p'}(0, T; V') = K' \).
Lemma 21.4.4 Let \( s < t \). Then for \( u, Y \) satisfying

\[
\langle Bu(t), u(t) \rangle = \langle Bu(s), u(s) \rangle + 2 \int_s^t \langle Y(r), u(t) \rangle \, dr - \langle Bu(t) - Bu(s), u(t) - u(s) \rangle
\]

(21.4.23)

**Proof:** It follows from the following computations

\[
Bu(t) - Bu(s) = \int_s^t Y(r) \, dr
\]

and so

\[
2 \int_s^t \langle Y(r), u(t) \rangle \, dr - \langle Bu(t) - Bu(s), u(t) - u(s) \rangle
\]

\[
= 2 \left( \int_s^t Y(r) \, dr, u(t) \right) - \langle Bu(t) - Bu(s), u(t) - u(s) \rangle
\]

\[
= 2 \langle Bu(t) - Bu(s), u(t) \rangle - \langle Bu(t) - Bu(s), u(t) - u(s) \rangle
\]

\[
= 2 \langle Bu(t), u(t) \rangle - 2 \langle Bu(s), u(t) \rangle - \langle Bu(t), u(t) \rangle
\]

\[
+ 2 \langle Bu(s), u(t) \rangle - \langle Bu(s), u(s) \rangle
\]

\[
= \langle Bu(t), u(t) \rangle - \langle Bu(s), u(s) \rangle
\]

Thus

\[
\langle Bu(t), u(t) \rangle - \langle Bu(s), u(s) \rangle
\]

\[
= 2 \int_s^t \langle Y(r), u(t) \rangle \, dr - \langle Bu(t) - Bu(s), u(t) - u(s) \rangle
\]

Lemma 21.4.5 In the above situation,

\[
\sup_{t \in N_C} \langle Bu(t), u(t) \rangle \leq C (\|Y\|_{K'}, \|u\|_K)
\]

Also, \( t \rightarrow Bu(t) \) is weakly continuous with values in \( W' \) on \( N_C \) where \( N \) is the set of measure zero where \( Bu(t) \neq B(u(t)) \).

**Proof:** From the above formula of Lemma 21.4.4 applied to the \( k^{th} \) partition of \( [0, T] \) described above,

\[
\langle Bu(t_m), u(t_m) \rangle - \langle Bu_0, u_0 \rangle = \sum_{j=0}^{m-1} \langle Bu(t_{j+1}), u(t_{j+1}) \rangle - \langle Bu(t_j), u(t_j) \rangle
\]

\[
= \sum_{j=0}^{m-1} 2 \int_{t_j}^{t_{j+1}} \langle Y(r), u(t_{j+1}) \rangle \, dr - \langle B(u(t_{j+1}) - u(t_j)), u(t_{j+1}) - u(t_j) \rangle
\]

\[
= \sum_{j=0}^{m-1} 2 \int_{t_j}^{t_{j+1}} \langle Y(r), u_k^r (r) \rangle \, dr - \langle B(u(t_{j+1}) - u(t_j)), u(t_{j+1}) - u(t_j) \rangle
\]

Thus, discarding the negative terms and denoting by \( \mathcal{P}_k \) the \( k^{th} \) of these partitions,

\[
\sup_{t_j \in \mathcal{P}_k} \langle Bu(t_j), u(t_j) \rangle \leq \langle Bu_0, u_0 \rangle + 2 \int_0^T \|Y(r), u_k^r (r)\| \, dr
\]

\[
\leq \langle Bu_0, u_0 \rangle + 2 \int_0^T \|Y(r)\|_{V'}, \|u_k^r (r)\|_V \, dr
\]
\[
\leq \langle Bu_0, u_0 \rangle + 2 \left( \int_0^T \|Y(r)\|_{V'}^p \, dr \right)^{1/p'} \left( \int_0^T \|u_k^r(r)\|_V^p \, dr \right)^{1/p} \\
\leq C (\|Y\|_{K'}, \|u\|_K)
\]

because these partitions are chosen such that

\[
\lim_{k \to \infty} \left( \int_0^T \|u_k^r(r)\|_V^p \right)^{1/p} = \left( \int_0^T \|u(r)\|_V^p \right)^{1/p}
\]

and so these are bounded. This has shown that for the dense subset of \([0, T], D \equiv \cup_k \mathcal{P}_k,\)

\[
\sup_{t \in D} \langle Bu(t), u(t) \rangle < C (\|Y\|_{K'}, \|u\|_K)
\]

From Lemma 21.4.2 above, there exists \(\{e_i\} \subseteq V\) such that \(\langle Be_i, e_j \rangle = \delta_{ij}\) and for \(t \notin N,\)

\[
\langle Bu(t), u(t) \rangle = \sum_{k=1}^\infty |\langle Bu(t), e_i \rangle|^2 = \sup_m \sum_{k=1}^m |\langle Bu(t), e_i \rangle|^2
\]

Thus, if \(s_n \to t, s_n \in D,\) Fatou’s lemma implies

\[
\langle Bu, u \rangle(t) = \langle B(u(t)), u(t) \rangle = \sum_{k=1}^\infty |\langle Bu(t), e_i \rangle|^2
\]

\[
\leq \lim \inf_{n \to \infty} \sum_{k=1}^\infty |\langle Bu(s_n), e_i \rangle|^2 \leq C (\|Y\|_{K'}, \|u\|_K)
\]

and so

\[
\sup_{t \in N^C} \langle Bu, u \rangle(t) = \sup_{t \in N^C} \langle B(u(t)), u(t) \rangle \leq C (\|Y\|_{K'}, \|u\|_K)
\]

It only remains to verify the claim about weak continuity.

Consider now the claim that \(t \to Bu(t)\) is weakly continuous on \(N^C.\) Letting \(v \in V, s \in N^C,\)

\[
\lim_{t \to s} \langle Bu(t), v \rangle = \langle Bu(s), v \rangle = \langle Bu(s), v \rangle
\]

(21.4.24)

The limit follows from the formula which implies \(t \to Bu(t)\) is continuous into \(V'.\) Now for \(t \in N^C,\)

\[
\|Bu(t)\| = \sup_{\|v\| \leq 1} |\langle Bu(t), v \rangle| \leq \langle Bv, v \rangle^{1/2} \langle Bu(t), u(t) \rangle^{1/2}
\]

which was shown to be bounded for \(t, s \in N^C.\) Let \(w \in W.\) Then

\[
|\langle Bu(t), w \rangle - \langle Bu(s), w \rangle| \leq \|Bu(t) - Bu(s), w - v\| + |\langle Bu(t) - Bu(s), v \rangle|
\]

Then the first term is less than \(\varepsilon\) if \(v\) is close enough to \(w\) and the second converges to 0 so holds for all \(v \in W\) and so this shows the weak continuity on \(N^C.\) □

Now pick \(t \in D,\) the union of all the mesh points. Then for all \(k \) large enough, \(t \in \mathcal{P}_k.\) Say \(t = t_m.\) From Lemma 21.4.2,

\[
- \sum_{j=0}^{m-1} \langle B(u(t_{j+1}) - u(t_j)), u(t_{j+1}) - u(t_j) \rangle = \\
\langle Bu(t_m), u(t_m) \rangle - \langle Bu_0, u_0 \rangle - 2 \sum_{j=0}^{m-1} \int_{t_j}^{t_{j+1}} \langle Y(r), u_k^r(r) \rangle \, dr
\]

Thus, \(\langle Bu(t_m), u(t_m) \rangle\) is constant for all \(k \) large enough and the integral term converges to

\[
\int_0^{t_m} \langle Y(r), u(r) \rangle \, dr
\]
It follows that the term on the left does converge to something as \( k \to \infty \). It just remains to consider what it does converge to. However, from the equation solved by \( u \),

\[
Bu(t_{j+1}) - Bu(t_j) = \int_{t_j}^{t_{j+1}} Y(r) \, dr
\]

Therefore, this term is dominated by an expression of the form

\[
\left| \sum_{j=0}^{m_k-1} \int_{t_j}^{t_{j+1}} \langle Y(r), u(t_{j+1}) - u(t_j) \rangle \right| = \left| \sum_{j=0}^{m_k-1} \int_{t_j}^{t_{j+1}} \langle Y(r), u(t_{j+1}) - u(t_j) \rangle \right|
\]

\[
= \sum_{j=0}^{m_k-1} \int_{t_j}^{t_{j+1}} \langle Y(r), u(t_{j+1}) - u(t_j) \rangle - \sum_{j=0}^{m_k-1} \int_{t_j}^{t_{j+1}} \langle Y(r), u(t_j) \rangle \right| = \int_{0}^{t_m} \langle Y(r), u^r(r) \rangle \, dr - \int_{0}^{t_m} \langle Y(r), u^l(r) \rangle \, dr
\]

\[
\leq \int_{0}^{T} |\langle Y(r), u^r(r) - u^l(r) \rangle| \, dr
\]

However, both \( u^r \) and \( u^l \) converge to \( u \) in \( K = L^p(0, T, V) \). Therefore, this term must converge to 0. Passing to a limit, it follows that for all \( t \in D \), the desired formula holds. Thus, for such \( t \in D \),

\[
\langle Bu(t), u(t) \rangle = \langle Bu_0, u_0 \rangle + 2 \int_{0}^{t} \langle Y(r), u(r) \rangle \, dr
\]

It remains to verify that this holds for all \( t \notin N \). Let \( t \in N^C \setminus D \) and let \( t(k) \in P_k \) be the largest point of \( P_k \) which is less than \( t \). Suppose \( t(m) < t(k) \) so that \( m \leq k \). Then

\[
Bu(t(m)) = Bu_0 + \int_{0}^{t(m)} Y(s) \, ds
\]

a similar formula for \( u(t(k)) \). Thus for \( t > t(m) \),

\[
Bu(t) - Bu(t(m)) = \int_{t(m)}^{t} Y(s) \, ds
\]

which is the same sort of thing already looked at except that it starts at \( t(m) \) rather than at 0 and \( u_0 = 0 \). Therefore,

\[
\langle B(u(t(k)) - u(t(m))), u(t(k)) - u(t(m)) \rangle = 2 \int_{t(m)}^{t(k)} \langle Y(s), u(s) - u(t(m)) \rangle \, ds
\]

Thus, for \( m \leq k \)

\[
\lim_{m,k \to \infty} \langle B(u(t(k)) - u(t(m))), u(t(k)) - u(t(m)) \rangle = 0 \quad (21.4.25)
\]

Hence \( \{Bu(t(k))\}_{k=1}^{\infty} \) is a convergent sequence in \( W' \) because

\[
|\langle B(u(t(k)) - u(t(m))), y \rangle| \leq \langle B(u(t(k)) - u(t(m))), u(t(k)) - u(t(m)) \rangle^{1/2} \langle By, y \rangle^{1/2} \leq \langle B(u(t(k)) - u(t(m))), u(t(k)) - u(t(m)) \rangle^{1/2} ||B||^{1/2} ||y||_W
\]

Does it converge to \( Bu(t) \)? Let \( \xi(t) \in W' \) be what it does converge to. Let \( v \in V \). Then

\[
\langle \xi(t), v \rangle = \lim_{k \to \infty} \langle Bu(t(k)), v \rangle = \lim_{k \to \infty} \langle Bu(t(k)), v \rangle = \langle Bu(t), v \rangle
\]
because it is known that \( t \to Bu(t) \) is continuous into \( V' \). It is also known that for \( t \in N^C \), \( Bu(t) \in W' \subseteq V' \) and that the \( Bu(t) \) for \( t \in N^C \) are uniformly bounded in \( W' \). Therefore, since \( V \) is dense in \( W \), it follows that \( \xi(t) = Bu(t) \).

Now for every \( t \in D \), it was shown above that

\[
\langle Bu(t), u(t) \rangle = \langle Bu_0, u_0 \rangle + 2 \int_0^t \langle Y(r), u(r) \rangle \, dr
\]

Also it was just shown that \( Bu(t(k)) \to Bu(t) \) for \( t \notin N \). Then for \( t \notin N \)

\[
|\langle Bu(t(k)), u(t(k)) \rangle - \langle Bu(t), u(t) \rangle| \leq |\langle Bu(t(k)), u(t(k)) - u(t) \rangle| + |\langle Bu(t(k)) - Bu(t), u(t) \rangle|
\]

Then the second term converges to 0. The first equals

\[
|\langle Bu(t(k)) - Bu(t), u(t(k)) \rangle| \leq \langle B(u(t(k)) - u(t)), u(t(k)) - u(t) \rangle^{1/2} \langle Bu(t(k)), u(t(k)) \rangle^{1/2}
\]

From the above, this is dominated by an expression of the form

\[
\langle B(u(t(k)) - u(t)), u(t(k)) - u(t) \rangle^{1/2} C
\]

Then using the lower semicontinuity of \( t \to \langle B(u(t(k)) - u(t)), u(t(k)) - u(t) \rangle \) on \( N^C \) which follows from the above, this is no larger than

\[
\lim_{m \to \infty} \inf \langle B(u(t(k)) - u(t(m))), u(t(k)) - u(t(m)) \rangle^{1/2} C \leq \varepsilon
\]

provided \( k \) is large enough. This follows from \( \bigcup \bigcup \bigcup \). Since \( \varepsilon \) is arbitrary, it follows that

\[
\lim_{k \to \infty} |\langle Bu(t(k)), u(t(k)) \rangle - \langle Bu(t), u(t) \rangle| = 0
\]

Then from the formula,

\[
\langle Bu(t), u(t) \rangle = \langle Bu_0, u_0 \rangle + 2 \int_0^t \langle Y(r), u(r) \rangle \, dr
\]

valid for \( t \in D \), it follows that the same formula holds for all \( t \notin N \). Then define \( \langle Bu, u \rangle(t) \) to equal \( \langle Bu(t), u(t) \rangle \) off \( N \) and the right side for \( t \in N \). Thus \( t \to \langle Bu, u \rangle(t) \) is continuous and for all \( t \in [0, T] \),

\[
\langle Bu, u \rangle(t) = \langle Bu_0, u_0 \rangle + 2 \int_0^t \langle Y(r), u(r) \rangle \, dr
\]

Also recall that \( t \to Bu(t) \) was shown to be weakly continuous into \( W' \) on \( N^C \). Then for \( t, s \in N^C \),

\[
\langle B(u(t) - u(s)), u(t) - u(s) \rangle = \langle Bu(t), u(t) \rangle - 2 \langle Bu(t), u(s) \rangle + \langle Bu(s), u(s) \rangle
\]

From this, it follows that \( t \to Bu(t) \) is continuous into \( W' \) on \( N^C \) because \( \lim_{u \to s} \) of the right side gives 0 and so the same is true of the left. Hence,

\[
|\langle B(u(t) - u(s)), y \rangle| \leq \langle By, y \rangle^{1/2} \langle B(u(t) - u(s)), u(t) - u(s) \rangle^{1/2}
\leq \|B\|^{1/2} \langle B(u(t) - u(s)), u(t) - u(s) \rangle^{1/2} \|y\|
\]

so

\[
\|B(u(t) - u(s))\|_{W'} \leq \|B\|^{1/2} \langle B(u(t) - u(s)), u(t) - u(s) \rangle^{1/2}
\]

which converges to 0 as \( t \to s \). ■

Consider the case that \( t \to B(u(t)) \) has a weak derivative, denoted as \( (Bu)'(t) \) which is in \( L^p(0; T; V') \). Then as shown above, there is a continuous function, denoted as \( Bu(t) \) which equals \( B(u(t)) \) for a.e. \( t \) and

\[
Bu(t) = Bu(0) + \int_0^t (Bu)'(s) \, ds
\]

Then the above theorem applies. Then one obtains the following corollary.
Corollary 21.4.6 Let $V \subseteq W, W' \subseteq V'$ be separable Banach spaces, and $B \in \mathcal{L}(W, W')$ is nonnegative and self adjoint. Also suppose $t \to B(u(t))$ has a weak derivative $(Bu)' \in L^p(0, T; V')$ for $u \in L^p(0, T; V)$. Then there is a continuous function denoted as $Bu(t)$ which equals $B(u(t))$ a.e. $t$. Say for $t \notin N$. Suppose $Bu(0) = Bu_0, u_0 \in W$. Then

$$Bu(t) = Bu_0 + \int_0^t (Bu)'(s) \, ds \text{ in } V'$$

(21.4.26)

Then $t \to Bu(t)$ is in $C(N^C, W')$ and also for such $t$,

$$\frac{1}{2} \langle Bu(t), u(t) \rangle = \frac{1}{2} \langle Bu_0, u_0 \rangle + \int_0^t \langle (Bu)'(s), u(s) \rangle \, ds$$

There exists a continuous function $t \to \langle Bu, u \rangle(t)$ which equals the right side of the above for all $t$ and equals $\langle Bu(t), u(t) \rangle(0)$ a.e. Then

$$\sup_{t \in [0, T]} \langle Bu, u \rangle(t) \leq C (\|Y\|_{K'}, \|u\|_K)$$

This also makes it easy to verify continuity of pointwise evaluation of $Bu$. Let $Lu = (Bu)'$.

$$u = D(L) \equiv \left\{ u \in L^p(0, T; V) : Lu \in L^p(0, T, V') \right\}$$

$$\|u\|_X \equiv \|u\|_{L^p(0, T; V)} + \|Lu\|_{L^p(0, T, V')}$$

Since $L$ is closed, this $X$ is a Banach space. Then the following theorem is obtained.

Theorem 21.4.7 In the above corollary, the map $u \to Bu(t)$ is continuous as a map from $X$ to $V'$. Also if $Y$ denotes those $f \in L^p([0, T]; V')$ for which $f' \in L^p([0, T]; V')$, so that $f$ has a representative such that $f(t) = f(0) + \int_0^t f'(s) \, ds$, then if $\|f\|_Y \equiv \|f\|_{L^p([0, T]; V')} + \|f'\|_{L^p([0, T]; V')}$ the map $f \to f(t)$ is continuous.

Proof: First, why is $u \to Bu(0)$ continuous? Say $u, v \in X$ and say $p \geq 2$ first.

$$Bu(t) - Bv(t) = Bu(0) - Bv(0) + \int_0^t (Bu)'(s) - (Bv)'(s) \, ds$$

and so

$$\left( \int_0^T \|Bu(0) - Bv(0)\|_{V'}^p, dt \right)^{1/p'} \leq \left( \int_0^T \|Bu(t) - Bv(t)\|_{V'}^p, dt \right)^{1/p'}$$

and so

$$\|Bu(0) - Bv(0)\|_{V'}, T^{1/p'} \leq \left( \|B\| \|u - v\|_{L^{p'}([0, T]; V')} + T^{1/p'} \| (Bu)' - (Bv)' \|_{L^{p'}([0, T]; V')} \right)$$

$$\leq C (\|B\|, T) \|u - v\|_X$$

Thus $u \to Bu(0)$ is continuous into $V'$. If $p < 2$, then you do something similar.

$$\left( \int_0^T \|Bu(0) - Bv(0)\|_{V'}^p, dt \right)^{1/p} \leq \left( \int_0^T \|Bu(t) - Bv(t)\|_{V'}^p, dt \right)^{1/p}$$

and so

$$\|Bu(0) - Bv(0)\|_{V'}, T^{1/p} \leq \|B\| \|u - v\|_{L^p} + C(T) \| (Bu)' - (Bv)' \|_{L^{p'}([0, T]; V')}$$

$$\leq C (\|B\|, T) \|u - v\|_X.$$

However, one could just as easily have done this for an arbitrary $s < T$ by repeating the argument for

$$Bu(t) = Bu(s) + \int_s^t (Bu)'(r) \, dr$$

Thus this mapping is certainly continuous into $V'$. The last assertion is similar. ■
21.5 The Implicit Case, \( B = B(t) \)

The above theorem can be generalized to the case where the formula is of the form

\[ BX(t) = BX_0 + \int_0^t Y(s) \, ds \]

This involves an operator \( B(t) \in \mathcal{L}(W,W') \) and \( B(t) \) satisfies

\[ \langle B(t)x,x \rangle \geq 0, \quad \langle B(t)x,y \rangle = \langle B(t)y,x \rangle \]

for

\[ V \subseteq W, W' \subseteq V' \]

Where we assume \( t \to B(t) \) is in \( C^1([0,T]; \mathcal{L}(W,W')) \) and \( V \) is dense in the Banach space \( W \).

Then the main result in this section is the following integration by parts theorem.

**Theorem 21.5.1** Let \( V \subseteq W, W' \subseteq V' \) be separable Banach spaces, and let \( Y \in L^p(0,T;W') \) and

\[ Bu(t) = Bu_0 + \int_0^t Y(s) \, ds \quad \text{in} \quad V', \quad u_0 \in W, \quad Bu(t) = B(t)(u(t)) \quad \text{for a.e.} \quad t \quad (21.5.27) \]

As indicated, \( Bu \) is the name of a function satisfying the above equation which satisfies \( Bu(t) = B(t)(u(t)) \) for a.e. \( t \). Thus \( Y = (Bu)' \) as a weak derivative in the sense of \( V' \)-valued distributions. Suppose that \( u \in L^p([0,T], V) \) and \( (s,t) \to B'(s)u(t) \) is bounded in \( V' \) in case \( p < 2 \). (If \( B(t) \) is constant in \( t \) this is obvious.) In the case where \( p \geq 2 \), it is enough to assume \( B' \in C^1([0,T]; \mathcal{L}(W,W')) \). Then \( t \to Bu(t) \) is continuous into \( W' \) for \( t \) off a set of measure zero \( N \) and also there exists a continuous function \( t \to (Bu,u)(t) \) such that for all \( t \notin N \), \( (Bu,u)(t) = \langle Bu(t),u(t) \rangle \), \( Bu(t) = B(u(t)) \), and for all \( t \),

\[ \frac{1}{2} \langle Bu,u \rangle(t) + \frac{1}{2} \int_0^t \langle Bu,u \rangle(s) \, ds = \frac{1}{2} \langle Bu_0,u_0 \rangle + \int_0^t \langle Y(s),u(s) \rangle \, ds \]

**Proof:** By Lemma 21.5.4, there exists a sequence of partitions \( \{t^n_k\}_{k=0}^{m_n} = \mathcal{P}_n, \mathcal{P}_n \subseteq \mathcal{P}_{n+1} \), of \( [0,T] \) such that the lengths of the sub intervals converge uniformly to 0 as \( n \to \infty \) and the step functions

\[ \sum_{k=0}^{m_n-1} u(t^n_k) \mathcal{X}_{[t^n_k,t^n_{k+1}]}(t) = u^n(t) \]

\[ \sum_{k=0}^{m_n-1} u(t^n_{k+1}) \mathcal{X}_{[t^n_k,t^n_{k+1}]}(t) = u^n_\prime(t) \]

converge to \( u \) in \( L^p(0,T;V) \equiv K \). We assume that all of these partition points have empty intersection with the set of measure zero where \( Bu(t) \neq B(t)(u(t)) \). Thus, at every partition point, \( Bu(t_k) = B(t_k)(u(t_k)) \). As just mentioned, \( L^p(0,T;V) \equiv K \), \( L^p(0,T;V') = K' \).

Taking a subsequence, we can have

\[ \|u^n_\prime - u\|_K + \|u^n - u\|_K + \|Bu^n_\prime - Bu\|_K + \|Bu^n - Bu\|_K \]

\[ + \|B'u^n_\prime - B'u\|_{L^2([0,T];W')} + \|B'u^n - B'u\|_{L^2([0,T];W')} < 2^{-n} \quad (21.5.28) \]

and so, we can assume that \( a.e. \) convergence also takes place for \( Bu^n_\prime, Bu^n, B'u^n_\prime, B'u^n, u^n_\prime, u^n \).

Is \( Bu(0) = B(0)u_0 \)? The integral equation gives this it seems. To save notation, \( B(0)u_0 \) will be written as \( Bu_0 \). This is not inconsistent because \( t \to B(t)u_0 \) is continuous and its value at 0 is \( B(0)u_0 \).

**Lemma 21.5.2** Let \( s < t \). Then for \( u,Y \) satisfying \( 21.5.27 \)

\[ \langle Bu(t),u(t) \rangle - \langle Bu(s),u(s) \rangle + \langle (B(t) - B(s))u(s),u(t) \rangle + \langle (B(t) - B(s))u(s),u(t) \rangle = 2 \int_s^t \langle Y(r),u(t) \rangle \, dr \]

\[ - \langle B(t)u(t) - B(t)u(s),u(t) - u(s) \rangle \]

\[ (21.5.29) \]
Proof: It follows from the following computations

$$B(t)u(t) - B(s)u(s) = \int_s^t Y(r) \, dr$$

and so

$$2 \int_s^t \langle Y(r) , u(t) \rangle \, dr - \langle B(t)u(t) - B(s)u(s), u(t) - u(s) \rangle$$

$$= 2 \left( \int_s^t Y(r) \, dr , u(t) \right) - \langle B(t)u(t) - B(s)u(s), u(t) - u(s) \rangle$$

$$= 2 \langle B(t)u(t) - B(s)u(s), u(t) \rangle - \langle B(t)u(t) - B(s)u(s), u(t) - u(s) \rangle$$

$$= 2 \langle B(t)u(t), u(t) \rangle - 2 \langle B(s)u(s), u(t) \rangle - \langle B(t)u(t), u(t) \rangle + \langle B(t)u(t), u(s) \rangle + \langle B(s)u(s), u(t) \rangle - \langle B(s)u(s), u(s) \rangle$$

$$= \langle B(t)u(t), u(t) \rangle - \langle B(s)u(s), u(s) \rangle + \langle B(t)u(t) - B(s)u(s), u(t) \rangle$$

Thus

$$\langle Bu(t), u(t) \rangle - \langle Bu(s), u(s) \rangle + \langle (B(t) - B(s))u(s), u(t) \rangle$$

$$= 2 \int_s^t \langle Y(r), u(t) \rangle \, dr - \langle B(t)u(t) - B(s)u(s), u(t) - u(s) \rangle$$

Now consider the last term. It equals

$$\langle B(t)u(t) - (B(s) - B(t) + B(t))u(s), u(t) - u(s) \rangle$$

$$= \langle B(t)u(t) - (B(s) - B(t))u(s) + B(t)u(s), u(t) - u(s) \rangle$$

$$= \langle B(t)u(t) - B(t)u(s), u(t) - u(s) \rangle + \langle (B(t) - B(s))u(s), u(t) - u(s) \rangle$$

It follows that

$$\langle Bu(t), u(t) \rangle - \langle Bu(s), u(s) \rangle + \langle (B(t) - B(s))u(s), u(t) \rangle$$

$$+ \langle (B(t) - B(s))u(s), u(t) - u(s) \rangle$$

$$= 2 \int_s^t \langle Y(r), u(t) \rangle \, dr - \langle B(t)u(t) - B(t)u(s), u(t) - u(s) \rangle$$

Of course this computation is under the assumption that neither $s$, $t$ are in the exceptional set off which $B(t)u(t) = Bu(t)$. In case $s = 0$ the same formula holds except you need to replace $u(s)$ with $u_0$ and $Bu(s)$ with $Bu(0)$. ■

It is good to emphasize part of the above.

$$\langle B(t)u(t) - B(t)u(s), u(t) - u(s) \rangle - \langle B(t)u(t) - B(s)u(s), u(t) - u(s) \rangle$$

$$= \langle (B(s) - B(t))u(s), u(t) - u(s) \rangle$$

Lemma 21.5.3 Let the partitions $\mathcal{P}_k$ be as above such that $2 \mathcal{P}_k \geq \mathcal{P}_k = \{ t^k_j \}_{j=0}^{m_k}$. Then for any $m \leq m_k$,

$$\sum_{j=0}^{m-1} \langle B(t^k_{j+1})u(t^k_{j+1}) - B(t^k_{j+1})u(t^k_j), u(t^k_{j+1}) - u(t^k_j) \rangle$$

$$\sum_{j=0}^{m-1} \langle B(t^k_{j+1})u(t^k_{j+1}) - B(t^k_j)u(t^k_j), u(t^k_{j+1}) - u(t^k_j) \rangle = \varepsilon^m(k)$$

where $\lim_{k \to \infty} \varepsilon^m(k) = 0$. Here

$$\varepsilon^m(k) = \sum_{j=0}^{m-1} \langle (B(t^k_j) - B(t^k_{j+1}))u(t^k_j), u(t^k_{j+1}) - u(t^k_j) \rangle$$
21.5. **THE IMPLICIT CASE, \( B = B(T) \)**

**Proof:** From the above lemma, the absolute value of the left side is no larger than

\[
\sum_{j=0}^{m-1} |((B(t^k_j) - B(t^k_{j+1})) u(t^k_j), u(t^k_{j+1}) - u(t^k_j))|
\]

\[
\leq \sum_{j=0}^{m-1} \int_{t^k_j}^{t^k_{j+1}} \|B'(s) u(t^k_j)\|_{W'} d\tau \|u(t^k_{j+1}) - u(t^k_j)\|_W
\]

(21.5.30)

In case \( p \geq 2 \) then for \( C = \max_{s} \|B'(s)\|_{\mathcal{L}(W,W')} \),

\[
\leq C \sum_{j=0}^{m-1} \int_{t^k_j}^{t^k_{j+1}} \|u^r(s)\|_W \|u^r_k(s) - u^r_k(s)\|_W d\tau
\]

\[
= C \sum_{j=0}^{m-1} \int_{t^k_j}^{t^k_{j+1}} \chi[t^k_j, t^k_{j+1}] (s) \|u^r_k(s) - u^r_k(s)\|_W d\tau
\]

\[
= C \int_{t^k_0}^{t^k_m} \|u^r(s)\|_W \|u^r_k(s) - u^r_k(s)\|_W d\tau
\]

\[
\leq C \|u^r_k\|_{L^p([0,T],V)} \|u^r_k(s) - u^r_k(s)\|_W
\]

\[
\leq \hat{C} (2) 2^{-k}
\]

by \[4\text{.7.25} \]. In case \( p < 2 \), then from assumption and \[4\text{.7.31} \], the absolute value of the left side is no larger than

\[
\sum_{j=0}^{m-1} C (t^k_{j+1} - t^k_j) \|u(t^k_{j+1}) - u(t^k_j)\|_W
\]

\[
= C \sum_{j=0}^{m-1} \int_{t^k_j}^{t^k_{j+1}} \chi[t^k_j, t^k_{j+1}] (s) \|u^r_k(s) - u^r_k(s)\|_W d\tau
\]

\[
= C \int_{t^k_0}^{t^k_m} \|u^r_k(s) - u^r_k(s)\|_W
\]

which converges to 0 as \( k \to \infty \) thanks to \[4\text{.7.28} \].

**Lemma 21.5.4** In the above situation,

\[
\sup_{t \in \mathcal{N}^C} \langle Bu(t), u(t) \rangle + \int_0^T \langle B'u, u \rangle ds \leq C (\|Y\|_{K'}, \|u\|_K)
\]

Also, \( t \to Bu(t) \) is weakly continuous with values in \( W' \) on \( \mathcal{N}^C \) where \( N \) is a set of measure zero including the set where \( Bu(t) \neq B(t)(u(t)) \).

**Proof:** From the above formula of Lemma \[4\text{.5.2} \] applied to the \( k \)th partition of \([0,T] \) described above,

\[
\langle Bu(t_m), u(t_m) \rangle - \langle Bu_0, u_0 \rangle + \sum_{j=0}^{m-1} ((B(t_{j+1}) - B(t_j)) u(t_j), u(t_{j+1})
\]

\[
+ \sum_{j=0}^{m-1} (B(t_{j+1}) - B(t_j)) u(t_j), u(t_{j+1}) - u(t_j)
\]

\[
= \sum_{j=0}^{m-1} 2 \int_{t_j}^{t_{j+1}} \langle Y(r), u(t_{j+1}) \rangle dr - \langle B(t_{j+1}) u(t_{j+1}) - B(t_j) u(t_j), u(t_{j+1}) - u(t_j) \rangle
\]

(21.5.31)
Consider the third term on the left,

\[
\sum_{j=0}^{m-1} \langle (B (t_{j+1}^n) - B (t_j^n)) u (t_j^n), u (t_{j+1}^n) \rangle
\]

\[
= \int_0^t \left\langle \sum_{j=0}^{m-1} K(t_j^n, t_{j+1}^n) (t) \frac{B (t_{j+1}^n) - B (t_j^n)}{t_{j+1}^n - t_j^n} u_n^t (t), u_n^t (t) \right\rangle dt
\]

Using a simple approximate identity argument and the assumption that \( t \to B (t) \) is in \( C^1 ([0, T], \mathcal{L} (W, W')) \),

\[
\sum_{j=0}^{m-1} K(t_j^n, t_{j+1}^n) (t) \frac{B (t_{j+1}^n) - B (t_j^n)}{t_{j+1}^n - t_j^n} \to B' (t)
\]

uniformly on \((0, T)\). Then \( \sum_{j=0}^{m-1} K(t_j^n, t_{j+1}^n) (t) \frac{B (t_{j+1}') - B (t_j')}{{t_{j+1}'} - t_j'} u_n' \to B' u \) strongly in \( L^2 ([0, T], W') \) while \( u_n' \to u \) strongly in \( L^2 ([0, T]; W) \). It follows that the third term on the left in (21.5.31) is

\[
\varepsilon (k) + 2 \int_0^T \langle B'u, u \rangle ds, \quad \varepsilon (k) \to 0.
\]

whenever \( n \) is sufficiently large. Also, \( T \) could be replaced with \( t_j \) for any of the mesh points.

Next consider the term labelled *. From Lemma 21.5.5, it is of the form \( \varepsilon^m (k) \) where \( \lim_{k \to \infty} \varepsilon^m (k) = 0 \). Thus (21.5.31) reduces to

\[
\langle Bu (t_m), u (t_m) \rangle - \langle Bu_0, u_0 \rangle + \int_0^{t_m} \langle B'u, u \rangle ds = \sum_{j=0}^{m-1} 2 \int_{t_j}^{t_{j+1}} \langle Y (r), u_k^r (r) \rangle dr
\]

\[
- \sum_{j=0}^{m-1} \langle B (t_{j+1}) u (t_{j+1}) - B (t_j) u (t_j), u (t_{j+1}) - u (t_j) \rangle + \varepsilon (k) \tag{21.5.32}
\]

where \( t_m \in \mathcal{P}_k \).

Thus, discarding the negative terms which occur at the end and denoting by \( \mathcal{P}_k \) the \( k^{th} \) of these partitions,

\[
\sup_{t_j \in \mathcal{P}_k} \langle Bu (t_j), u (t_j) \rangle + \int_0^T \langle B'u, u \rangle ds \leq \langle Bu_0, u_0 \rangle + 2 \int_0^T \| \langle Y (r), u_k^r (r) \rangle \|_V dr + \varepsilon
\]

\[
\leq \langle Bu_0, u_0 \rangle + 2 \int_0^T \| Y (r) \|_V \| u_k^r (r) \|_V dr + \varepsilon
\]

\[
\leq \langle Bu_0, u_0 \rangle + 2 \left( \int_0^T \| Y (r) \|_V^p \| u_k^r (r) \|_V^p \right)^{1/p} \left( \int_0^T \| u_k^r (r) \|_V^p \right)^{1/p} + \varepsilon
\]

\[
\leq C (\| Y \|_{K'}, \| u \|_K) + \varepsilon
\]

whenever \( k \) is large enough because these partitions are chosen such that

\[
\lim_{k \to \infty} \left( \int_0^T \| u_k^r (r) \|_V^p \right)^{1/p} = \left( \int_0^T \| u (r) \|_V^p \right)^{1/p}
\]

and so these are bounded. This has shown that for the dense subset of \([0, T], D \equiv \cup_P \mathcal{P}_k \),

\[
\sup_{t \in D} \langle B (t) u (t), u (t) \rangle + \int_0^T \langle B'u, u \rangle ds < C (\| Y \|_{K'}, \| u \|_K) + \varepsilon
\]

However, \( \varepsilon \) was arbitrary and the partitions are nested. Hence the above holds for all \( \varepsilon \) and so

\[
\sup_{t \in D} \langle B (t) u (t), u (t) \rangle + \int_0^T \langle B'u, u \rangle ds < C (\| Y \|_{K'}, \| u \|_K)
\]
By and the integral equation, there is a set of measure zero including all the earlier sets of measure zero \( N \) such that for \( t \notin N, u_n^x(t), u_n^r(t) \to u(t) \) pointwise in \( V \). Also, \( B(t) u_n^x(t) \to Bu(t) \) in \( V' \). This last can be obtained from the integral equation solved. \( t \to Bu(t) \) is continuous into \( V' \). Then let \( t \notin N \). We have \( u_n^r(t) \to u(t) \) in \( V \). Now \( B(t) u_n^x(t) = B(t) u(s_n) \) where \( s_n \in D \) and \( s_n \to t \). Then \( Bu(t) = B(t) u(t) \) and

\[
\| B(s_n) u(s_n) - B(t) u(t) \|_{V'} \leq \| B(s_n) - B(t) \|_{V'} \| u(s_n) \|_{V} + \| B(t) u(s_n) - u(t) \|_{V},
\]

\[
\leq C_t \| B(s_n) - B(t) \|_{V'} + C \| u(s_n) - u(t) \|_{V},
\]

where \( C_t \) is a constant which comes because \( u(s_n) \to u(t) \) in \( V \) and so is bounded. The constant \( C \) is just \( \max_{t \in [0,T]} \| B(t) \|_{V'} \). Then, since the two terms on the right converge to 0 as \( n \to \infty \), it follows that as \( s_n \to t, B(s_n) u(s_n) \to B(t) u(t) = Bu(t) \) in \( V' \) while \( u(s_n) \to u(t) \) in \( V \). It follows that for \( t \notin N \),

\[
\langle Bu(t), u(t) \rangle + \int_0^T \langle B'u, u \rangle \, ds = \lim_{n \to \infty} \langle Bu(s_n), u(s_n) \rangle + \int_0^T \langle B'u, u \rangle \, ds \leq C (\| Y \|_{K'}, \| u \|_K)
\]

Hence,

\[
\sup_{t \notin N} \langle Bu(t), u(t) \rangle + \int_0^T \langle B'u, u \rangle \, ds \leq C (\| Y \|_{K'}, \| u \|_K)
\]

It only remains to verify the claim about weak continuity.

Consider now the claim that \( t \to Bu(t) \) is weakly continuous on \( N^C \). Letting \( v \in V, s \in N^C \),

\[
\lim_{t \to s} \langle Bu(t), v \rangle = \langle Bu(s), v \rangle = \langle Bu(s), v \rangle
\]

(21.5.33)

The limit follows from the formula which implies \( t \to Bu(t) \) is continuous into \( V' \). Now for \( t \in N^C \),

\[
\| Bu(t) \|_{V'} = \sup_{\| v \|_{V'} \leq 1} |\langle Bu(t), v \rangle| \leq \langle Bu, v \rangle^{\frac{1}{2}} \langle Bu(t), u(t) \rangle^{\frac{1}{2}}
\]

\[
\leq \left( C (\| Y \|_{K'}, \| u \|_K) - \int_0^T \langle B'u, u \rangle \, ds \right) \sup_{t \notin N} \| Bu(t) \|_{V'} \leq \left( C (\| Y \|_{K'}, \| u \|_K) - \int_0^T \langle B'u, u \rangle \, ds \right)
\]

Now let \( w \in W \). Then

\[
|\langle Bu(t), w \rangle - \langle Bu(s), w \rangle| \leq |\langle Bu(t) - Bu(s), w-v \rangle| + |\langle Bu(t) - Bu(s), v \rangle_{V', V}|
\]

Then the first term is less than \( \varepsilon \) if \( v \) is close enough to \( w \) and the second converges to 0 by continuity of \( t \to Bu(t) \) which comes from the integral equation, so holds for all \( v \in W \) and so this shows the weak continuity of \( t \to Bu(t) \) on \( N^C \).

Now pick \( t \in D \), the union of all the mesh points. Then for all \( k \) large enough, \( t \in P_k \). Say \( t = t_m \). From

\[
\langle Bu(t_m), u(t_m) \rangle - \langle Bu_0, u_0 \rangle + \int_0^{t_m} \langle B'u, u \rangle \, ds = \sum_{j=0}^{m-1} 2 \int_{t_j}^{t_{j+1}} \langle Y(r), u^r(r) \rangle \, dr + \varepsilon(k)
\]

\[
- \sum_{j=0}^{m-1} \langle B(t_{j+1}) u(t_{j+1}) - B(t_{j+1}) u(t_j), u(t_{j+1}) - u(t_j) \rangle
\]

(21.5.34)

where \( \varepsilon(k) \to 0 \). By Lemma, you can modify \( \varepsilon(k) \) and write this in the form

\[
\langle Bu(t_m), u(t_m) \rangle - \langle Bu_0, u_0 \rangle + \int_0^{t_m} \langle B'u, u \rangle \, ds = \sum_{j=0}^{m-1} 2 \int_{t_j}^{t_{j+1}} \langle Y(r), u^r(r) \rangle \, dr + \varepsilon(k)
\]

\[
- \sum_{j=0}^{m-1} \langle B(t_{j+1}) u(t_{j+1}) - B(t_{j+1}) u(t_j), u(t_{j+1}) - u(t_j) \rangle
\]

(21.5.35)
Thus, \( \langle Bu(t_m), u(t_m) \rangle \) is constant for all \( k \) large enough and the integral term on the right converges as \( k \to \infty \) to

\[
\int_0^{t_m} \langle Y(r), u(r) \rangle \, dr
\]

It follows that the last term on the right does converge to something as \( k \to \infty \). It just remains to consider what it does converge to. However, from the equation solved by \( u \),

\[
Bu(t_{j+1}) - Bu(t_j) = \int_{t_j}^{t_{j+1}} Y(r) \, dr
\]

Therefore, this term is dominated by an expression of the form

\[
\begin{align*}
&\left| \sum_{j=0}^{m-1} \left( \int_{t_j}^{t_{j+1}} Y(r) \, dr, u(t_{j+1}) - u(t_j) \right) \right| \\
= &\left| \sum_{j=0}^{m-1} \int_{t_j}^{t_{j+1}} \langle Y(r), u(t_{j+1}) - u(t_j) \rangle \, dr \right|
\end{align*}
\]

\[
\begin{align*}
= &\left| \sum_{j=0}^{m-1} \int_{t_j}^{t_{j+1}} \langle Y(r), u(t_{j+1}) \rangle - \sum_{j=0}^{m-1} \int_{t_j}^{t_{j+1}} \langle Y(r), u(t_j) \rangle \right| \\
= &\left| \int_0^{t_m} \langle Y(r), u^r(r) \rangle \, dr - \int_0^{t_m} \langle Y(r), u^l(r) \rangle \, dr \right| \\
\leq &\int_0^T \left| \langle Y(r), u^r(r) - u^l(r) \rangle \right| \, dr
\end{align*}
\]

However, both \( u^r \) and \( u^l \) converge to \( u \) in \( K = L^p(0, T, V) \). Therefore, this term must converge to 0. Passing to a limit, it follows that for all \( t \in D \), the desired formula holds. Thus, for such \( t \in D \),

\[
\langle Bu(t), u(t) \rangle + \int_0^t \langle B'u, u \rangle \, dr = \langle Bu_0, u_0 \rangle + 2 \int_0^t \langle Y(r), u(r) \rangle \, dr
\]

It remains to verify that this holds for all \( t \notin N \). Let \( t \in N^C \setminus D \) and let \( t(k) \in \mathcal{P}_k \) be the largest point of \( \mathcal{P}_k \) which is less than \( t \). Suppose \( t(m) \leq t(k) \) so that \( m \leq k \). Then

\[
Bu(t(m)) = Bu_0 + \int_0^{t(m)} Y(s) \, ds,
\]
a similar formula for \( u(t(k)) \). Thus for \( t > t(m) \),

\[
Bu(t) - Bu(t(m)) = \int_{t(m)}^t Y(s) \, ds
\]

which is the same sort of thing already looked at except that it starts at \( t(m) \) rather than at 0 and \( u_0 = 0 \). Therefore,

\[
\begin{align*}
&\langle B(u(t(k)) - u(t(m))), u(t(k)) - u(t(m)) \rangle \\
+ &\int_{t(m)}^{t(k)} \langle B'(s)(u(s) - u(t(m))), u(s) - u(t(m)) \rangle \\
&= 2 \int_{t(m)}^{t(k)} \langle Y(s), u(s) - u(t(m)) \rangle \, ds
\end{align*}
\]

Thus, for \( m \leq k \)

\[
\lim_{m,k \to \infty} \langle B(u(t(k)) - u(t(m))), u(t(k)) - u(t(m)) \rangle = 0 \quad (21.5.36)
\]
Hence \( \{ Bu(t(k)) \}_{k=1}^{\infty} \) is a convergent sequence in \( W' \) because
\[
\| B[u(t(k)) - u(t(m))] \| \\
\leq \langle Bu(t(k)) - u(t(m)), u(t(k)) - u(t(m)) \rangle^{1/2} \langle By, y \rangle^{1/2} \\
\leq \langle Bu(t(k)) - u(t(k)), u(t(k)) - u(t(m)) \rangle^{1/2} \| B \|^{1/2} \| y \|_W
\]

Does it converge to \( Bu(t) \)? Let \( \xi(t) \in W' \) be what it does converge to. Let \( v \in V \). Then
\[
\langle \xi(t), v \rangle = \lim_{k \to \infty} \langle Bu(t(k)), v \rangle = \lim_{k \to \infty} \langle Bu(t(k)), v \rangle = \langle Bu(t), v \rangle
\]
because it is known that \( t \to Bu(t) \) is continuous into \( V' \). It is also known that for \( t \in N^C \), \( Bu(t) \in W' \subseteq V' \) and that the \( Bu(t) \) for \( t \in N^C \) are uniformly bounded in \( W' \). Therefore, since \( V \) is dense in \( W \), it follows that \( \xi(t) = Bu(t) \).

Now for every \( t \in D \), it was shown above that
\[
\langle Bu(t), u(t) \rangle + \int_0^t \langle Bu(t), u(t) \rangle \, dr = \langle Bu_0, u_0 \rangle + 2 \int_0^t \langle Bu(t,r), u(r) \rangle \, dr
\]
also it was just shown that \( Bu(t(k)) \to Bu(t) \) for \( t \notin N \). Then for \( t \notin N \)
\[
\langle Bu(t(k)), u(t(k)) \rangle - \langle Bu(t), u(t) \rangle \\
\leq \langle Bu(t(k)), u(t(k)) - u(t) \rangle + \langle Bu(t), u(t) \rangle
\]
Then the second term converges to 0. The first equals
\[
\langle Bu(t(k)), u(t(k)) - u(t) \rangle \\
\leq \langle Bu(t(k)), u(t(k)) - u(t) \rangle^{1/2} \langle Bu(t(k)), u(t(k)) \rangle^{1/2}
\]
From the above, this is dominated by an expression of the form
\[
\langle Bu(t(k)), u(t(k)) - u(t) \rangle^{1/2} \langle Bu(t(k)), u(t(k)) \rangle^{1/2} C
\]
Then from the choice of \( N \) and the pointwise convergence of \( u^*_n \) to \( u \) off \( N \) the above converges to 0 for each \( t \notin N \). It follows that
\[
\lim_{k \to \infty} |\langle Bu(t(k)), u(t(k)) \rangle - \langle Bu(t), u(t) \rangle| = 0
\]
Then from the formula,
\[
\langle Bu(t), u(t) \rangle = \langle Bu_0, u_0 \rangle + 2 \int_0^t \langle Bu(t, r), u(r) \rangle \, dr - \int_0^t \langle Bu(t, r), u(r) \rangle \, dr
\]
valid for \( t \in D \), it follows that the same formula holds for all \( t \notin N \). Then define \( \langle Bu(t), u(t) \rangle \) to equal \( \langle Bu(t), u(t) \rangle \) off \( N \) and the right side for \( t \in N \). Thus \( t \to \langle Bu(t), u(t) \rangle \) is continuous and for all \( t \in [0, T] \),
\[
\langle Bu(t), u(t) \rangle = \langle Bu_0, u_0 \rangle + 2 \int_0^t \langle Bu(t, r), u(r) \rangle \, dr - \int_0^t \langle Bu(t, r), u(r) \rangle \, dr
\]
Also recall that \( t \to B(t) u(t) \) was shown to be weakly continuous into \( W' \) on \( N^C \). Is it continuous on \( N^C \)? Suppose \( t \in N^C \) and let \( s_n \to t \) where \( s_n \in D \). Then \( u(s_n) = u^*_n(t) \) because \( s_n \) is one of the mesh points. Since \( s_n \to t \) one can assume that \( m_n \to \infty \). Hence \( u(s_n) = u^*_n(t) \) by the pointwise convergence implied by \( 21.5.23 \). Then obviously
\[
B(s_n) u(s_n) = B(s_n) u^*_n(t) \to B(t) u(t)
\]
Now suppose you just have \( t_n \to t \) where each of \( t_n, t \) are in \( N^C \). Does it always follow that \( B(t_n) u(t_n) \to B(t) u(t) \)? Suppose not. Then there exists such a sequence \( t_n \to t \) of points in \( N^C \) and \( \varepsilon > 0 \) such that
\[
\| B(t_n) u(t_n) - B(t) u(t) \| \geq \varepsilon
\]
However, from the density of \( D \) and what was just shown, there exists \( s_n \in D \) such that \( |s_n - t_n| < \frac{1}{2} \) and
\[
\| B(s_n) u(s_n) - B(t_n) u(t_n) \| < \frac{1}{2n}
\]
Then
\[ \varepsilon \leq \|B(t_n)u(t_n) - B(s_n)u(t_n)\| + \|B(s_n)u(t_n) - B(t)u(t)\| \]
\[ < \frac{1}{n} + \|B(s_n)u(s_n) - B(t)u(t)\| \]

Since \( s_n \to t \), what was just shown implies both terms on the right converge to 0. This is a contradiction. Thus \( t \to B(t)u(t) \) must be continuous on \( N^c \) into \( W' \).

Consider the case that \( t \to B(u(t)) \) has a weak derivative, denoted as \( (Bu)'(t) \) which is in \( L^p'([0,T];V') \). Then as shown above, there is a continuous function, denoted as \( Bu(t) \) which equals \( B(t)(u(t)) \) for a.e. \( t \) and
\[ Bu(t) = Bu(0) + \int_0^t (Bu)'(s) \, ds \]

Then the above theorem applies. Then one obtains the following corollary.

**Corollary 21.5.5** Let \( V \subseteq W, W' \subseteq V' \) be separable Banach spaces, and \( B(t) \in \mathcal{L}(W, W') \) is nonnegative and self adjoint, \( B \in C^1([0,T];W') \). Also suppose \( t \to B(u(t)) \) has a weak derivative \( (Bu)'(t) \in L^p'([0,T];V') \) for \( u \in L^p([0,T];V) \cap L^2([0,T];W) \). Then there is a continuous function denoted as \( Bu(t) \) which equals \( B(t)(u(t)) \) a.e. \( t \). Say for \( t \notin N \). Suppose \( Bu(0) = Bu_0, u_0 \in W \). Then
\[ Bu(t) = Bu_0 + \int_0^t (Bu)'(s) \, ds \quad \text{in } V' \]
(21.5.37)

Then \( t \to Bu(t) \) is in \( C(N^c, W') \) and also for such \( t \),
\[ \frac{1}{2} \langle Bu(t), u(t) \rangle + \frac{1}{2} \int_0^t \langle B'(s)u(s), u(s) \rangle \, ds = \frac{1}{2} \langle Bu_0, u_0 \rangle + \int_0^t \langle (Bu)'(s), u(s) \rangle \, ds \]

There exists a continuous function \( t \to \langle Bu, u \rangle(t) \) which equals the right side of the above for all \( t \) and equals \( \langle B(t)u(t), u(t) \rangle \) off \( N \). This satisfies
\[ \sup_{t \in [0,T]} \langle Bu, u \rangle(t) \leq C(\|u\|_{K'}, \|u\|_K) \]

Note how if everything is nice and smooth, this integration by parts formula is what you would be expected to get. To see this, assume \( u \) is smooth and formally work on the right side.
\[ \frac{d}{dt} \langle Bu, u \rangle = \langle (Bu)', u \rangle + \langle Bu, u' \rangle \]
\[ = \langle (Bu)', u \rangle + \langle Bu', u \rangle \]
\[ = 2 \langle (Bu)', u \rangle - \langle Bu', u \rangle \]

Thus
\[ \frac{1}{2} \langle Bu_0, u_0 \rangle + \int_0^t \langle (Bu)'(s), u(s) \rangle \, ds \]
\[ = \frac{1}{2} \langle Bu_0, u_0 \rangle + \frac{1}{2} \left[ \int_0^t \frac{d}{ds} \langle Bu, u \rangle \, ds + \int_0^t \langle B'u, u \rangle \, ds \right] \]
\[ = \frac{1}{2} \langle Bu(t), u(t) \rangle + \frac{1}{2} \int_0^t \langle B'u, u \rangle \, ds \]
which equals the left side.

A related topic is the continuity of pointwise evaluation of \( Bu \). Let \( Lu = (Bu)' \).
\[ u = D(L) \equiv \left\{ u \in L^p(0,T;V) : Lu \in L^p'(0,T,V') \right\} \]
\[ \|u\|_X = \|u\|_{L^p(0,T;V)} + \|Lu\|_{L^p'(0,T,V')} \]

Since \( L \) is closed, this \( X \) is a Banach space. Then the following theorem is obtained.
Theorem 21.5.6 In the above corollary, the map \( u \to Bu(t) \) is continuous as a map from \( X \) to \( V' \). Also if \( Y \) denotes those \( f \in L^p([0,T];V) \) for which \( f' \in L^p([0,T];V) \), so that \( f \) has a representative such that \( f(t) = f(0) + \int_0^t f'(s) \, ds \), then if \( \|f\|_Y \equiv \|f\|_{L^p([0,T];V)} + \|f'\|_{L^p([0,T];V)} \) the map \( f \to f(t) \) is continuous.

Proof: First, why is \( u \to Bu(0) \) continuous? Say \( u, v \in X \) and say \( p \geq 2 \) first.

\[
Bu(t) - Bv(t) = Bu(0) - Bv(0) + \int_0^t (Bu)'(s) - (Bv)'(s) \, ds
\]

and so,

\[
\left( \int_0^T \|Bu(0) - Bv(0)\|_{V'}^p \, dt \right)^{1/p'} \leq \left( \int_0^T \|Bu(t) - Bv(t)\|_{V'}^p \, dt \right)^{1/p'}
\]

so

\[
\|Bu(0) - Bv(0)\|_{V', T^{1/p'}} \leq \left( \|B\| \|u - v\|_{L^p([0,T];V')} + T^{1/p'} \left\| (Bu)' - (Bv)' \right\|_{L^p([0,T];V')} \right)
\]

\[
\leq C \left( \|B\|, T \right) \|u - v\|_X
\]

Thus \( u \to Bu(0) \) is continuous into \( V' \). If \( p < 2 \), then you do something similar.

\[
\left( \int_0^T \|Bu(0) - Bv(0)\|_{V'}^p \, dt \right)^{1/p} \leq \left( \int_0^T \|Bu(t) - Bv(t)\|_{V'}^p \, dt \right)^{1/p}
\]

\[
\|Bu(0) - Bv(0)\|_{V', T^{1/p}} \leq \|B\| \|u - v\|_{L^p} + C(T) \left\| (Bu)' - (Bv)' \right\|_{L^p([0,T];V')}
\]

\[
\leq C \left( \|B\|, T \right) \|u - v\|_X.
\]

However, one could just as easily have done this for an arbitrary \( s < T \) by repeating the argument for

\[
Bu(t) = Bu(s) + \int_s^t (Bu)'(r) \, dr
\]

Thus this mapping is certainly continuous into \( V' \). The last assertion is similar. ■

21.6 Some Imbedding Theorems

The next theorem is very useful in getting estimates in partial differential equations. It is called Erling’s lemma.

Definition 21.6.1 Let \( E, W \) be Banach spaces such that \( E \subseteq W \) and the injection map from \( E \) into \( W \) is continuous. The injection map is said to be compact if every bounded set in \( E \) has compact closure in \( W \). In other words, if a sequence is bounded in \( E \) it has a convergent subsequences converging in \( W \). This is also referred to by saying that bounded sets in \( E \) are precompact in \( W \).

Theorem 21.6.2 Let \( E \subseteq W \subseteq X \) where the injection map is continuous from \( W \) to \( X \) and compact from \( E \) to \( W \). Then for every \( \varepsilon > 0 \) there exists a constant, \( C_\varepsilon \) such that for all \( u \in E \),

\[
\|u\|_W \leq \varepsilon \|u\|_E + C_\varepsilon \|u\|_X
\]
Proof: Suppose not. Then there exists $\varepsilon > 0$ and for each $n \in \mathbb{N}$, $u_n$ such that

$$||u_n||_W > \varepsilon ||u_n||_E + n ||u_n||_X$$

Now let $v_n = u_n/||u_n||_E$. Therefore, $||v_n||_E = 1$ and

$$||v_n||_W > \varepsilon + n ||v_n||_X$$

It follows there exists a subsequence, still denoted by $v_n$, such that $v_n$ converges to $v$ in $W$. However, the above inequality shows that $||v_n||_X \to 0$. Therefore, $v = 0$. But then the above inequality would imply that $||v_n|| > \varepsilon$ and passing to the limit yields $0 > \varepsilon$, a contradiction. ■

**Definition 21.6.3** Define $C([a, b]; X)$ the space of functions continuous at every point of $[a, b]$ having values in $X$.

You should verify that this is a Banach space with norm

$$||u||_{\infty, X} = \max \{||u_n(t) - u(t)||_X : t \in [a, b]\}.$$ 

The following theorem is an infinite dimensional version of the Ascoli Arzelà theorem. It is like a well known result due to Simon. It is the appropriate generalization to stochastic problems in which you do not have weak derivatives. See Theorem 17.12 on the Holder continuity of the stochastic integral.

**Theorem 21.6.4** Let $q > 1$ and let $E \subseteq W \subseteq X$ where the injection map is continuous from $W$ to $X$ and compact from $E$ to $W$. Let $S$ be defined by

$$\bigg\{ u \text{ such that } ||u(t)||_E \leq R \text{ for all } t \in [a, b], \text{ and } ||u(s) - u(t)||_X \leq R|t - s|^{1/q}\bigg\}.$$ 

Thus $S$ is bounded in $L^\infty(0, T, E)$ and in addition, the functions are uniformly Holder continuous into $X$. Then $S \subseteq C([a, b]; W)$ and if $\{u_n\} \subseteq S$, there exists a subsequence, $\{u_{n_k}\}$ which converges to a function $u \in C([a, b]; W)$ in the following way.

$$\lim_{k \to \infty} ||u_{n_k} - u||_{\infty, W} = 0.$$ 

Proof: First consider the issue of $S$ being a subset of $C([a, b]; W)$. Let $\varepsilon > 0$ be given. Then by Theorem 21.6.2 there exists a constant, $C_\varepsilon$ such that for all $u \in W$

$$||u||_W \leq \varepsilon \frac{6R}{R} ||u||_E + C_\varepsilon ||u||_X.$$ 

Therefore, for all $u \in S$, 

$$||u(t) - u(s)||_W \leq \varepsilon \frac{6R}{R} ||u(t) - u(s)||_E + C_\varepsilon ||u(t) - u(s)||_X \leq \frac{\varepsilon}{3} + C_\varepsilon R|t - s|^{1/q}. \quad (21.6.38)$$

Since $\varepsilon$ is arbitrary, it follows $u \in C([a, b]; W)$.

Let $D = \mathbb{Q} \cap [a, b]$ so $D$ is a countable dense subset of $[a, b]$. Let $D = \{t_n\}_{n=1}^{\infty}$. By compactness of the embedding of $E$ into $W$, there exists a subsequence $u_{(n,1)}$ such that $n \to \infty$, $u_{(n,1)}(t_1)$ converges to a point in $W$. Now take a subsequence of this, called $(n, 2)$ such that $n \to \infty$, $u_{(n,2)}(t_2)$ converges to a point in $W$. It follows that $u_{(n,2)}(t_2)$ also converges to a point of $W$. Continue this way. Now consider the diagonal sequence, $u_k \equiv u(k, k)$. This sequence is a subsequence of $u_{(n,l)}$ whenever $k > l$. Therefore, $u_k(t_j)$ converges for all $t_j \in D$.

Claim: Let $\{u_k\}$ be as just defined, converging at every point of $D \equiv [a, b] \cap \mathbb{Q}$. Then $\{u_k\}$ converges at every point of $[a, b]$.

Proof of claim: Let $\varepsilon > 0$ be given. Let $t \in [a, b]$. Pick $t_m \in D \cap [a, b]$ such that in 21.6.38 $C_\varepsilon R|t - t_m| < \varepsilon/3$. Then there exists $N$ such that if $l, n > N$, then $||u_l(t_m) - u_n(t_m)||_X < \varepsilon/3$. It follows that for $l, n > N$,

$$||u_l(t) - u_n(t)||_W \leq ||u_l(t) - u_l(t_m)||_W + ||u_l(t_m) - u_n(t_m)||_W + ||u_n(t_m) - u_n(t)||_W \leq \frac{2\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{2\varepsilon}{3} = 2\varepsilon$$
Since \( \varepsilon \) was arbitrary, this shows \( \{ u_k(t) \}_{k=1}^{\infty} \) is a Cauchy sequence. Since \( W \) is complete, this shows this sequence converges.

Now for \( t \in [a, b] \), it was just shown that if \( \varepsilon > 0 \) there exists \( N_t \) such that if \( n, m > N_t \), then
\[
\| u_n(t) - u_m(t) \|_W < \frac{\varepsilon}{3}.
\]

Now let \( s \neq t \). Then
\[
\| u_n(s) - u_m(s) \|_W \leq \| u_n(s) - u_n(t) \|_W + \| u_n(t) - u_m(t) \|_W + \| u_m(t) - u_m(s) \|_W
\]

From (21.6.39)
\[
\| u_n(s) - u_m(s) \|_W \leq 2 \left( \frac{\varepsilon}{3} + C \varepsilon R |t-s|^{1/q} \right) + \| u_n(t) - u_m(t) \|_W
\]

and so it follows that if \( \delta \) is sufficiently small and \( s \in B(t, \delta) \), then when \( n, m > N_t \)
\[
\| u_n(s) - u_m(s) \| < \varepsilon.
\]

Since \([a, b]\) is compact, there are finitely many of these balls, \( \{ B(t_i, \delta) \}_{i=1}^{p} \), such that for \( s \in B(t_i, \delta) \) and \( n, m > N_t \), the above inequality holds. Let \( N > \max \{ N_1, \ldots, N_p \} \). Then if \( m, n > N \) and \( s \in [a, b] \) is arbitrary, it follows the above inequality must hold. Therefore, this has shown the following claim.

**Claim:** Let \( \varepsilon > 0 \) be given. Then there exists \( N \) such that if \( m, n > N \), then \( \| u_n - u_m \|_{\infty, W} < \varepsilon \).

Now let \( u(t) = \lim_{k \to \infty} u_k(t) \).
\[
\| u(t) - u(s) \|_W \leq \| u(t) - u_n(t) \|_W + \| u_n(t) - u_n(s) \|_W + \| u_n(s) - u(s) \|_W
\]

(21.6.39)

Let \( N \) be in the above claim and fix \( n > N \). Then
\[
\| u(t) - u_n(t) \|_W = \lim_{m \to \infty} \| u_m(t) - u_n(t) \|_W \leq \varepsilon
\]

and similarly, \( \| u_n(s) - u(s) \|_W \leq \varepsilon \). Then if \( |t-s| \) is small enough, (21.6.39) shows the middle term in 21.6.39 is also smaller than \( \varepsilon \). Therefore, if \( |t-s| \) is small enough,
\[
\| u(t) - u(s) \|_W < 3\varepsilon.
\]

Thus \( u \) is continuous. Finally, let \( N \) be as in the above claim. Then letting \( m, n > N \), it follows that for all \( t \in [a, b] \),
\[
\| u_m(t) - u_n(t) \|_W < \varepsilon.
\]

Therefore, letting \( m \to \infty \), it follows that for all \( t \in [a, b] \),
\[
\| u(t) - u_n(t) \|_W \leq \varepsilon.
\]

and so \( \| u - u_n \|_{\infty, W} \leq \varepsilon \). ■

Here is an interesting corollary. Recall that for \( E \) a Banach space \( C^{0,\alpha}([0, T], E) \) is the space of continuous functions \( u \) from \([0, T]\) to \( E \) such that
\[
\| u \|_{\alpha, E} \equiv \| u \|_{\infty, E} + \rho_{\alpha, E}(u) < \infty
\]

where here
\[
\rho_{\alpha, E}(u) \equiv \sup_{t \neq s} \frac{\| u(t) - u(s) \|_E}{|t-s|^\alpha}
\]

**Corollary 21.6.5** Let \( E \subseteq W \subseteq X \) where the injection map is continuous from \( W \) to \( X \) and compact from \( E \) to \( W \). Then if \( \gamma > \alpha \), the embedding of \( C^{0,\gamma}([0, T], E) \) into \( C^{0,\alpha}([0, T], X) \) is compact.

**Proof:** Let \( \phi \in C^{0,\gamma}([0, T], E) \)
\[
\frac{\| \phi(t) - \phi(s) \|_X}{|t-s|^\gamma} \leq \left( \frac{\| \phi(t) - \phi(s) \|_W}{|t-s|^\gamma} \right)^{\alpha/\gamma} \| \phi(t) - \phi(s) \|_W^{1-\alpha/\gamma}
\]
\[
\leq \left( \frac{\| \phi(t) - \phi(s) \|_E}{|t-s|^\gamma} \right)^{\alpha/\gamma} \| \phi(t) - \phi(s) \|_W^{1-\alpha/\gamma} \leq \rho_{\gamma, E}(\phi) \| \phi(t) - \phi(s) \|_W^{1-\alpha/\gamma}
\]
Now suppose \( \{u_n\} \) is a bounded sequence in \( C^{0,\gamma}([0, T], E) \). By Theorem 4.1.13 above, there is a subsequence still called \( \{u_n\} \) which converges in \( C^0([0, T], W) \). Thus from the above inequality

\[
\|u_n(t) - u_m(t) - (u_n(s) - u_m(s))\|_X |t - s|^\alpha \\
\leq \rho_{\gamma,E}(u_n - u_m) \|u_n(t) - u_m(t) - (u_n(s) - u_m(s))\|_W^{1 - (\alpha/\gamma)} \\
\leq C \{\{u_n\}\} \left(2\|u_n - u_m\|_{X,W}^{1 - (\alpha/\gamma)}\right)
\]

which converges to 0 as \( n, m \to \infty \). Thus

\[
\rho_{\alpha,X}(u_n - u_m) \to 0 \text{ as } n, m \to \infty
\]

Also \( \|u_n - u_m\|_{X,W} \to 0 \) as \( n, m \to \infty \) so this is a Cauchy sequence in \( C^{0,\alpha}([0, T], X) \). ■

The next theorem is a well known result probably due to Lions.

**Theorem 21.6.6** Let \( E \subseteq W \subseteq X \) where the injection map is continuous from \( W \) to \( X \) and compact from \( E \) to \( W \). Let \( p \geq 1 \), let \( q > 1 \), and define

\[
S = \{u \in L^p([a,b]; E) : \text{for some } C, \|u(t) - u(s)\|_X \leq C|t - s|^{1/q} \}
\]

and \( \|u\|_{L^p([a,b]; E)} \leq R \).

Thus \( S \) is bounded in \( L^p([a,b]; E) \) and Holder continuous into \( X \). Then \( S \) is precompact in \( L^p([a,b]; W) \). This means that if \( \{u_n\}_{n=1}^\infty \subseteq S \), it has a subsequence \( \{u_{n_k}\} \) which converges in \( L^p([a,b]; W) \).

**Proof:** By Proposition 4.3.3 on Page 12 it suffices to show \( S \) has an \( \eta \) net in \( L^p([a,b]; W) \) for each \( \eta > 0 \).

If not, there exists \( \eta > 0 \) and a sequence \( \{u_n\} \subseteq S \), such that

\[
\|u_n - u_m\| \geq \eta
\]

for all \( n \neq m \) and the norm refers to \( L^p([a,b]; W) \). Let

\[
a = t_0 < t_1 < \cdots < t_k = b, \quad t_i - t_{i-1} = (b - a)/k.
\]

Now define

\[
\pi_n(t) = \sum_{i=1}^k \pi_{n_i} X_{[t_{i-1}, t_i)}(t), \quad \pi_{n_i} = \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} u_n(s) \, ds.
\]

The idea is to show \( \pi_n \) approximates \( u_n \) well and then to argue that a subsequence of the \( \{\pi_n\} \) is a Cauchy sequence yielding a contradiction to 21.6.40.

Therefore,

\[
u_n(t) - \pi_n(t) = \sum_{i=1}^k u_n(t) X_{[t_{i-1}, t_i)}(t) - \sum_{i=1}^k \pi_{n_i} X_{[t_{i-1}, t_i)}(t)
\]

\[
= \sum_{i=1}^k \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} u_n(t) \, ds X_{[t_{i-1}, t_i)}(t) - \sum_{i=1}^k \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} u_n(s) \, ds X_{[t_{i-1}, t_i)}(t)
\]

\[
= \sum_{i=1}^k \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} (u_n(t) - u_n(s)) \, ds X_{[t_{i-1}, t_i)}(t).
\]

It follows from Jensen’s inequality that

\[
\|u_n(t) - \pi_n(t)\|^p \leq \sum_{i=1}^k \left| \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} (u_n(t) - u_n(s)) \, ds \right|^p X_{[t_{i-1}, t_i)}(t)
\]

\[
\leq \sum_{i=1}^k \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} \|u_n(t) - u_n(s)\|_W^p \, ds X_{[t_{i-1}, t_i)}(t)
\]
21.6. SOME IMBEDDING THEOREMS

and so

\[ \int_a^b \| (u_n (t) - \overline{u}_n(s)) \|^p_W \, ds \]
\[ \leq \int_a^b \sum_{i=1}^k \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} \| u_n (t) - u_n (s) \|^p_W \, ds \, \chi_{[t_{i-1}, t_i)} (t) \, dt \]
\[ = \sum_{i=1}^k \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t_i} \| u_n (t) - u_n (s) \|^p_W \, ds \, dt. \] (21.6.41)

From Theorem 21.6.2 if \( \varepsilon > 0 \), there exists \( C_\varepsilon \) such that

\[ \| u_n (t) - u_n (s) \|^p_W \leq \varepsilon \| u_n (t) - u_n (s) \|^p_E + C_\varepsilon \| u_n (t) - u_n (s) \|^p_X \]
\[ \leq 2^{p-1} \varepsilon (\| u_n (t) \|^p + \| u_n (s) \|^p) + C_\varepsilon |t - s|^{p/q} \]

This is substituted in to (21.6.40) to obtain

\[ \int_a^b \| (u_n (t) - \overline{u}_n(s)) \|^p_W \, ds \leq \]
\[ \sum_{i=1}^k \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t_i} \left( 2^{p-1} \varepsilon (\| u_n (t) \|^p + \| u_n (s) \|^p) + C_\varepsilon |t - s|^{p/q} \right) \, ds \, dt \]
\[ = \sum_{i=1}^k 2^{p} \varepsilon \int_{t_{i-1}}^{t_i} \| u_n (t) \|^p_W \, dt + C_\varepsilon \sum_{i=1}^k \frac{1}{(t_i - t_{i-1})} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t_i} |t - s|^{p/q} \, ds \, dt \]
\[ \leq 2^{p} \varepsilon \int_a^b \| u_n (t) \|^p_W \, dt + C_\varepsilon \sum_{i=1}^k \frac{1}{(t_i - t_{i-1})} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t_i} (t_i - t_{i-1})^{p/q} \, ds \, dt \]
\[ \leq 2^{p} \varepsilon R^p + C_\varepsilon \sum_{i=1}^k (t_i - t_{i-1})^{1+p/q} = 2^{p} \varepsilon R^p + C_\varepsilon k \left( \frac{b-a}{k} \right)^{1+p/q}. \]

Taking \( \varepsilon \) so small that \( 2^{p} \varepsilon R^p < \eta^p / 8p \) and then choosing \( k \) sufficiently large, it follows

\[ \| u_n - \overline{u}_n \|_{L^p([a,b];W)} < \frac{\eta}{4}. \]

Thus \( k \) is fixed and \( \overline{u}_n \) at a step function with \( k \) steps having values in \( E \). Now use compactness of the embedding of \( E \) into \( W \) to obtain a subsequence such that \( \{ \overline{u}_n \} \) is Cauchy in \( L^p (a,b; W) \) and use this to contradict (21.6.41). The details follow.

Suppose \( \overline{u}_n (t) = \sum_{i=1}^k u_i^n \chi_{[t_{i-1}, t_i)} (t) \). Thus

\[ \| \overline{u}_n (t) \|_E = \sum_{i=1}^k \| u_i^n \|_E \chi_{[t_{i-1}, t_i)} (t) \]

and so

\[ R \geq \int_a^b \| \overline{u}_n (t) \|^p_W \, dt = \frac{T}{k} \sum_{i=1}^k \| u_i^n \|^p_E. \]

Therefore, the \( \{ u_i^n \} \) are all bounded. It follows that if taking subsequences \( k \) times there exists a subsequence \( \{ u_{n_k} \} \) such that \( u_{n_k} \) is a Cauchy sequence in \( L^p (a,b; W) \). You simply get a subsequence such that \( u_i^{n_k} \) is a Cauchy sequence in \( W \) for each \( i \). Then denoting this subsequence by \( n \),

\[ \| u_n - u_m \|_{L^p(a,b; W)} \leq \| u_n - \overline{u}_n \|_{L^p(a,b; W)} + \| \overline{u}_n - \overline{u}_m \|_{L^p(a,b; W)} + \| \overline{u}_m - u_m \|_{L^p(a,b; W)} \leq \frac{\eta}{4} \]

provided \( m, n \) are large enough, contradicting (21.6.40).
Chapter 22

Measurability Without Uniqueness

With the Ito formula which holds for a single space, it is time to consider stochastic ordinary differential equations. First is a general theory which allows one to consider measurable solutions to stochastic equations in which there is no uniqueness available. Unfortunately, it does not include obtaining adapted solutions. Instead, it includes measurability of functions with respect to a single $\sigma$ algebra. Then when path uniqueness is available, one can include the concept of adapted solutions rather easily and this will be done for ordinary differential equations. First is a general result about multifunctions.

22.1 Multifunctions And Their Measurability

Let $X$ be a separable complete metric space and let $(\Omega, C, \mu)$ be a set, a $\sigma$ algebra of subsets of $\Omega$, and a measure $\mu$ such that this is a complete $\sigma$ finite measure space. Also let $\Gamma : \Omega \rightarrow \mathcal{P}_c(X)$, the closed subsets of $X$.

Definition 22.1.1 We define $\Gamma^{-}(S) \equiv \{\omega \in \Omega : \Gamma(\omega) \cap S \neq \emptyset\}$

We will consider a theory of measurability of set valued functions. The following theorem is the main result in the subject. In this theorem the third condition is what we will refer to as measurable. The second condition is called strongly measurable. More can be said than what we will prove here.

Theorem 22.1.2 In the following, 1. $\Rightarrow$ 2. $\Rightarrow$ 3. $\Rightarrow$ 4.

1. For all $B$ a Borel set in $X$, $\Gamma^{-}(B) \in C$.
2. For all $F$ closed in $X$, $\Gamma^{-}(F) \in C$.
3. For all $U$ open in $X$, $\Gamma^{-}(U) \in C$.
4. There exists a sequence, $\{\sigma_n\}$ of measurable functions satisfying $\sigma_n(\omega) \in \Gamma(\omega)$ such that for all $\omega \in \Omega$,

$$\Gamma(\omega) = \{\sigma_n(\omega) : n \in \mathbb{N}\}$$

These functions are called measurable selections.

Also 4. $\Rightarrow$ 3. If $\Gamma(\omega)$ is compact for each $\omega$, then also 3. $\Rightarrow$ 2.

Proof: It is obvious that 1.) $\Rightarrow$ 2.). To see that 2.) $\Rightarrow$ 3.) note that $\Gamma^{-}(\bigcup_{i=1}^{\infty} F_i) = \bigcup_{i=1}^{\infty} \Gamma^{-}(F_i)$. Since any open set in $X$ can be obtained as a countable union of closed sets, this implies 2.) $\Rightarrow$ 3.).

Now we verify that 3.) $\Rightarrow$ 4.). Let $\{x_n\}_{n=1}^{\infty}$ be a countable dense subset of $X$. For $\omega \in \Omega$, let $\psi_1(\omega) = x_n$ where $n$ is the smallest integer such that $\Gamma(\omega) \cap B(x_n, 1) \neq \emptyset$. Therefore, $\psi_1(\omega)$ has countably many values, $x_{n_1}, x_{n_2}, \cdots$ where $n_1 < n_2 < \cdots$. Now

$$\{\omega : \psi_1 = x_n\} = \{\omega : \Gamma(\omega) \cap B(x_n, 1) \neq \emptyset\} \cap \Omega \setminus \bigcup_{k<n} \{\omega : \Gamma(\omega) \cap B(x_k, 1) \neq \emptyset\} \in C.$$ 

Thus we see that $\psi_1$ is measurable and $\text{dist}(\psi_1(\omega), \Gamma(\omega)) < 1$. Let

$$\Omega_n \equiv \{\omega \in \Omega : \psi_1(\omega) = x_n\}.$$
Then \( \Omega_n \in \mathcal{C} \) and \( \Omega_n \cap \Omega_m = \emptyset \) for \( n \neq m \) and \( \bigcup_{n=1}^{\infty} \Omega_n = \Omega \). Let \( D_n = \{ x_k : x_k \in B(x_n, 1) \} \). Now for each \( n \), and \( \omega \in \Omega_n \), let \( \psi_2(\omega) = x_k \) where \( k \) is the smallest index such that \( x_k \in D_n \) and \( B(x_k, \frac{1}{2}) \cap \Gamma(\omega) \neq \emptyset \). Thus \( \text{dist} (\psi_2(\omega), \Gamma(\omega)) < \frac{1}{2} \) and \( d(\psi(\omega), \psi_1(\omega)) < 1 \). Continue this way obtaining \( \psi_k \) a measurable function such that

\[
\text{dist} (\psi_k(\omega), \Gamma(\omega)) < \frac{1}{2^{k-1}}, \quad d(\psi_k(\omega), \psi_{k+1}(\omega)) < \frac{1}{2^{k-2}}.
\]

Then for each \( \omega \), \( \{ \psi_k(\omega) \} \) is a Cauchy sequence converging to a point, \( \sigma(\omega) \in \Gamma(\omega) \). This has shown that if \( \Gamma \) is measurable there exists a measurable selection, \( \sigma(\omega) \in \Gamma(\omega) \). It remains to show there exists a sequence of these measurable selections, \( \sigma_n \) such that the conclusion of 4.) holds. To do this we define

\[
\Gamma_{n_i}(\omega) = \begin{cases} 
\Gamma(\omega) \cap B(x_n, 2^{-i}) & \text{if } \Gamma(\omega) \cap B(x_n, 2^{-i}) \neq \emptyset \\
\Gamma(\omega) & \text{otherwise.}
\end{cases}
\]

First we show that \( \Gamma_{n_i} \) is measurable. Let \( U \) be open. Then

\[
\{ \omega : \Gamma_{n_i}(\omega) \cap U \neq \emptyset \} = \{ \omega : \Gamma(\omega) \cap B(x_n, 2^{-i}) \cap U \neq \emptyset \} \cup \\
\left( \{ \omega : \Gamma(\omega) \cap B(x_n, 2^{-i}) = \emptyset \} \cap \{ \omega : \Gamma(\omega) \cap U \neq \emptyset \} \right) \cup \\
\left( \{ \omega : \Gamma(\omega) \cap B(x_n, 2^{-i}) \neq \emptyset \} \cap \{ \omega : \Gamma(\omega) \cap U \neq \emptyset \} \right),
\]

a measurable set. By what was just shown there exists \( \sigma_{n_i} \), a measurable function such that \( \sigma_{n_i}(\omega) \in \Gamma_{n_i}(\omega) \subseteq \Gamma(\omega) \) for all \( \omega \in \Omega \). If \( x \in \Gamma(\omega) \), then \( x \in B(x_n, 2^{-i}) \) whenever \( x_n \) is close enough to \( x \). Therefore, \( |\sigma_{n_i}(\omega) - x| < 2^{-i} \).

And it follows that 4.) holds.

Now we verify that 4.) \( \Rightarrow \) 3.). Suppose there exist measurable selections \( \sigma_n(\omega) \in \Gamma(\omega) \) satisfying condition 4.). Let \( U \) be open. Then

\[
\{ \omega : \Gamma(\omega) \cap U \neq \emptyset \} = \bigcup_{n=1}^{\infty} \sigma_n^{-1}(U) \in \mathcal{C}.
\]

Now suppose \( \Gamma(\omega) \) is compact for every \( \omega \) and that \( \Gamma^-(U) \in \mathcal{C} \) for every \( U \) open. Then let \( F \) be a closed set and let \( \{ U_n \} \) be a decreasing sequence of open sets whose intersection equals \( F \) such that also, for all \( n, U_n \supseteq U_{n+1} \).

Then

\[
\Gamma(\omega) \cap F = \bigcap_n \Gamma(\omega) \cap U_n = \bigcap_n \Gamma(\omega) \cap \overline{U_n}
\]

Now because of compactness, the set on the left is nonempty if and only if each set on the right is also nonempty. Thus \( \Gamma^-(F) = \bigcap_n \Gamma^-(U_n) \in \mathcal{C} \).

Actually these are all equivalent in the case of complete measure spaces but we do not need this and it is much harder to show.

### 22.2 A Measurable Selection

This section deals with the problem of getting product measurable functions in a context of no uniqueness. The following is the main result. It is stated in great generality because it has fairly wide application although it will be used first in finite dimensions.

**Theorem 22.2.1** Let \( V \) be a reflexive separable Banach space and \( V' \) its dual and \( \frac{1}{p} + \frac{1}{p'} = 1 \) where \( p > 1 \) as usual. For \( n \in \mathbb{N} \) let the functions \( t \rightarrow u_n(t, \omega) \) be in \( L^p([0, T]; V') \) and \( t, \omega \rightarrow u_n(t, \omega) \) be \( B([0, T]) \times \mathcal{F} \equiv \mathcal{P} \) measurable into \( V' \). Suppose there is a set of measure zero \( N \) such that if \( \omega \notin N \), then for all \( n \),

\[
\sup_{t \in [0, T]} \| u_n(t, \omega) \|_{V'} \leq C(\omega).
\]

Also suppose for each \( \omega \notin N \), each subsequence of \( \{ u_n \} \) has a further subsequence which converges weakly in \( L^p([0, T]; V') \) to \( u(\cdot, \omega) \in L^p([0, T]; V') \) such that \( t \rightarrow u(t, \omega) \) is weakly continuous into \( V' \). Then there exists a product measurable, with \( t \rightarrow u(t, \omega) \) being weakly continuous into \( V' \). Moreover, there exists, for each \( \omega \notin N \), a subsequence \( u_n(\omega) \) such that \( u_n(\cdot, \omega) \rightarrow u(\cdot, \omega) \) weakly in \( L^p([0, T]; V') \).
22.2. A MEASURABLE SELECTION

Note that the exceptional set is given. It could be the empty set with no change in the conclusion of the theorem. Let \( X = \prod_{k=1}^{\infty} C ([0, T]) \) with the product topology. One can consider this as a metric space using the metric

\[
d(f, g) = \sum_{k=1}^{\infty} 2^{-k} \frac{\| f_k - g_k \|}{1 + \| f_k - g_k \|},
\]

where the norm is the maximum norm in \( C ([0, T]) \). With this metric, \( X \) is complete and separable.

Lemma 22.2.2 Let \( \{ f_n \} \) be a sequence in \( X \) and suppose that the \( k \)-th components \( f_{nk} \) are bounded in \( C^{0,1} ([0, T]) \). (This refers to the Hölder space with \( \gamma = 1 \).) Then there exists a subsequence converging to some \( f \in X \). Thus if \( \{ f_n \} \) has each component bounded in \( C^{0,1} ([0, T]) \), then \( \{ f_n \} \) is pre-compact in \( X \).

Proof: By the Ascoli–Arzelà theorem, there exists a subsequence \( n_1 \) such that the first component \( f_{n_1 k} \) converges in \( C ([0, T]) \). Taking a subsequence, one can obtain \( n_2 \) a subsequence of \( n_1 \) such that both the first and second components of \( f_{n_2} \) converge. Continuing this way one obtains a sequence of subsequences, each a subsequence of the previous one such that \( f_{n_j} \) has the first \( j \) components converging to functions in \( C ([0, T]) \). Therefore, the diagonal subsequence has the property that it has every component converging to a function in \( C ([0, T]) \). The resulting function in \( \prod_{k=1}^{\infty} C ([0, T]) \) is \( f \). \( \blacksquare \)

Now for \( m \in \mathbb{N} \) and \( \phi \in V' \), define \( l_m (t) \equiv \max (0, t - (1/m)) \) and \( \psi_{m, \phi} : L^p \to C ([0, T]) \) as follows

\[
\psi_{m, \phi} (u) (t) \equiv \int_{l_m (t)}^{t} \langle m \phi X_{[l_m (t), t]} (s), u (s) \rangle_{V, V'} ds = m \int_{l_m (t)}^{t} \langle \phi, u (s) \rangle_{V, V'} ds.
\]

Let \( D = \{ \phi \}_{j=1}^{\infty} \) denote a countable dense subset of \( V \). Then the pairs \( (\phi, m) \) for \( \phi \in D \) and \( m \in \mathbb{N} \) yield a countable set. Let \( \{ m_k, \phi_k \} \) denote an enumeration of these pairs \((m, \phi) \in \mathbb{N} \times D \). To save notation, we denote

\[
f_k (u) (t) \equiv \psi_{m_k, \phi_k} (u) (t) = m_k \int_{l_m (t)}^{t} \langle \phi_k, u (s) \rangle_{V, V'} ds.
\]

For fixed \( \omega \notin N \) and \( k \), the functions \( \{ t \mapsto f_k (u_j (\cdot, \omega))(t) \} \) are uniformly bounded and equicontinuous because they are in \( C^{0,1} ([0, T]) \). Indeed,

\[
|f_k (u_j (\cdot, \omega))(t)| = m_k \int_{l_m (t)}^{t} \langle \phi_k, u_j (s, \omega) \rangle_{V, V'} ds \leq C (\omega) \| \phi_k \|_V,
\]

and for \( t \leq t' \)

\[
|f_k (u_j (\cdot, \omega))(t) - f_k (u_j (\cdot, \omega))(t')| \\
\leq m_k \int_{l_m (t)}^{t} \langle \phi_k, u_j (s, \omega) \rangle_{V, V'} ds - m_k \int_{l_m (t')}^{t'} \langle \phi_k, u_j (s, \omega) \rangle_{V, V'} ds \\
\leq 2m_k |t' - t| \| \phi_k \|_V C (\omega).
\]

By Lemma \textit{compact}, the set of functions \( \{ f (u_j (\cdot, \omega))) \}_{j=n}^{\infty} \) is pre-compact in \( X = \prod_{k} C ([0, T]) \). Then define a set valued map \( \Gamma^n : \Omega \to X \) as follows.

\[
\Gamma^n (\omega) \equiv \bigcup_{j \geq n} \{ f (u_j (\cdot, \omega)) \},
\]

where the closure is taken in \( X \). Then \( \Gamma^n (\omega) \) is the closure of a pre-compact set in \( \prod_{k} C ([0, T]) \) and so \( \Gamma^n (\omega) \) is compact in \( \prod_{k} C ([0, T]) \). From the definition, a function \( f \) is in \( \Gamma^n (\omega) \) if and only if \( d (f, f (w_l)) \to 0 \) as \( l \to \infty \), where each \( w_l \) is one of the \( u_j (\cdot, \omega) \) for \( j \geq n \).

From the topology on \( X \) this happens if and only if for every \( k \),

\[
f_k (t) = \lim_{l \to \infty} m_k \int_{l_m (t)}^{t} \langle \phi_k, w_l (s, \omega) \rangle_{V, V'} ds,
\]

where the limit is the uniform limit in \( t \).

Note that in the case of a filtration, instead of a single σ-algebra \( F \) where each \( u_j \) is progressively measurable, if the sequence \( u_l \) does not have the index \( l \) dependent on \( \omega \), then if such a limit holds for each \( \omega \), it follows that \( (t, \omega) \to f_k (t, \omega) \) will inherit progressive measurability from the \( u_l \). This situation will be typical when dealing with stochastic equations with path uniqueness known. Thus this is a reasonable way to attempt to consider measurability and the more difficult question of whether a process is adapted.
Lemma 22.2.3 $\omega \rightarrow \Gamma^n(\omega)$ is a $\mathcal{F}$ measurable set valued map with values in $X$. If $\sigma$ is a measurable selection, $(\sigma(\omega) \in \Gamma^n(\omega))$ so $\sigma = \sigma(\cdot,\omega)$ a continuous function. To have this measurable would mean that $\sigma^{-1}(\emptyset)$ is open in $\mathcal{F}$ such that for each $t$, $\omega \rightarrow \sigma(t,\omega)$ is $\mathcal{F}$ measurable and $(t,\omega) \rightarrow \sigma(t,\omega)$ is $\mathcal{B}\{[0,T]\} \times \mathcal{F} \equiv \mathcal{P}$ measurable.

Proof: Let $O$ be a basic open set in $X$. Thus $O = \prod_{k=1}^{\infty} O_k$ where $O_k$ is a proper open set of $C\{[0,T]\}$ only for $k \in \{k_{1,\cdots,k_r}\}$. We need to consider whether

$$\Gamma^n(O) = \{ \omega : \Gamma^n(\omega) \cap O \neq \emptyset \} \in \mathcal{F}.$$ 

Now $\Gamma^n(O)$ equals

$$\bigcap_{i=1}^{n} \{ \omega : \Gamma^n(\omega)_{k_i} \cap O_{k_i} \neq \emptyset \}$$

Thus we consider whether

$$\{ \omega : \Gamma^n(\omega)_{k_i} \cap O_{k_i} \neq \emptyset \} \in \mathcal{F}$$

From the definition of $\Gamma^n(\omega)$, this is equivalent to the condition that for some $j \geq n$,

$$f_{k_i}(u_j(\cdot,\omega)) = (f(u_j(\cdot,\omega)))_{k_i} \in O_{k_i}$$

and so the above set in $\mathcal{P}$ is of the form

$$\bigcup_{j=n}^{\infty} \{ \omega : (f(u_j(\cdot,\omega)))_{k_i} \in O_{k_i} \}$$

Now $\omega \rightarrow (f(u_j(\cdot,\omega)))_{k_i}$ is $\mathcal{F}$ measurable into $C\{[0,T]\}$ and so the above set is in $\mathcal{F}$. To see this, let $g \in C\{[0,T]\}$ and consider the inverse image of the ball $B(g,r)$,

$$\{ \omega : \| (f(u_j(\cdot,\omega)))_{k_i} - g \|_{C([0,T])} < r \}.$$ 

By continuity considerations,

$$\| (f(u_j(\cdot,\omega)))_{k_i} - g \|_{C([0,T])} = \sup_{t \in [0,T]} \| (f(u_j(t,\omega)))_{k_i} - g(t) \|$$

which is the sup of countably many $\mathcal{F}$ measurable functions. Thus it is $\mathcal{F}$ measurable. Since every open set is the countable union of such balls, it follows that the claim about $\mathcal{F}$ measurability is valid. Thus $\Gamma^n(O)$ is $\mathcal{F}$ measurable whenever $O$ is a basic open set.

Now $X$ is a separable metric space and so every open set is a countable union of these basic sets. Let $U$ be an open set in $X$ and let $U = \bigcup_{i=1}^{\infty} O^i$ where $O^i$ is a basic open set as above. Then

$$\Gamma^n(U) = \bigcup_{i=1}^{\infty} \Gamma^n(O^i) \in \mathcal{F}.$$ 

That there exists a measurable selection follows from the standard theory of measurable multi-functions $\mathcal{E}$, $\mathcal{E}$. This is proved in Theorem 22.2.2 above. For $\sigma$ one of these measurable selections, the evaluation at $t$ is $\mathcal{F}$ measurable. Thus $\omega \rightarrow \sigma(t,\omega)$ is $\mathcal{F}$ measurable with values in $\mathbb{R}^n$. Also $t \rightarrow \sigma(t,\omega)$ is continuous, and so it follows that in fact $\sigma$ is product measurable as claimed. 

Definition 22.2.4 Let $\Gamma(\omega) = \cap_{i=1}^{\infty} \Gamma^n(\omega)$.

Lemma 22.2.5 $\Gamma$ is a nonempty $\mathcal{F}$ measurable set valued function having values in the compact sub-sets of $X$. There exists a measurable selection $\gamma$. For $\gamma$ a $\mathcal{F}$ measurable selection, $(t,\omega) \rightarrow \gamma(t,\omega)$ is $\mathcal{P}$ measurable. Also, for each $\omega$, there exists a subsequence, $u_{n(\omega)}(\cdot,\omega)$ such that for each $k$,

$$\gamma_k(t,\omega) = \lim_{n(\omega) \rightarrow \infty} f(u_{n(\omega)}(t,\omega))_k = \lim_{n(\omega) \rightarrow \infty} m_k \int_{l_{m_k}(t)}^{t} \langle \phi_{rk}, u_{n(\omega)}(s,\omega) \rangle_{V',V} \, ds$$

Proof: Consider $\Gamma(\omega) = \cap_{i=1}^{\infty} \Gamma^n(\omega)$. Then $\omega \rightarrow \Gamma(\omega)$ is a compact set valued map in $X$. It is nonempty because each $\Gamma^n(\omega)$ is nonempty and compact, and these sets are nested. Is it $\mathcal{F}$ measurable? Each $\Gamma^n$ is compact valued and $\mathcal{F}$ measurable. Hence if $F$ is closed,

$$\Gamma(\omega) \cap F = \cap_{i=1}^{\infty} \Gamma^n(\omega) \cap F.$$
and the left is non empty if and only if each $\Gamma^n(\omega) \cap F \neq \emptyset$. Hence for $F$ closed,
\[
\{ \omega : \Gamma(\omega) \cap F \neq \emptyset \} = \cap_n \{ \omega : \Gamma^n(\omega) \cap F \neq \emptyset \}
\]

and so
\[
\Gamma^-(F) = \cap_n \Gamma^n-(F) \in \mathcal{F}
\]
The last claim follows from the theory of multi-functions Theorem 22.2.6. Since $\Gamma^n(\omega)$ is compact, the measurability of $\Gamma^n$, that $\Gamma^n-(U) \in \mathcal{F}$ for $U$ open implies the strong measurability of $\Gamma^n$, that $\Gamma^n-(F) \in \mathcal{F}$. Thus $\omega \to \Gamma(\omega)$ is non empty compact valued in $X$ and $\mathcal{F}$ measurable.

From standard theory of measurable multi-functions, Theorem 22.1.2, there exists a $\mathcal{F}$ measurable selection $\omega \to \gamma(\omega)$ with $\gamma(\omega) \in \Gamma(\omega)$ for each $\omega$. Now it follows that $t \to \gamma_k(t,\omega)$ is continuous. This is what it means for $\gamma(\omega) \in X$. What of the product measurability of $\gamma_k$? We know that $\omega \to \gamma_k(\omega)$ is $\mathcal{F}$ measurable into $C([0,T])$ and so since pointwise evaluation is continuous, $\omega \to \gamma_k(t,\omega)$ is $\mathcal{F}$ measurable. Then since $t \to \gamma_k(t,\omega)$ is continuous, it follows that $\gamma_k$ is a $\mathcal{P}$ measurable real valued function and that $\gamma$ is a $\mathcal{P}$ measurable $\mathbb{R}^\infty$ valued function.

Since $\gamma(\omega) \in \Gamma(\omega)$, it follows that for each $n$, $\gamma(\omega) \in \Gamma^n(\omega)$. Therefore, there exists $j_n \geq n$ such that for each $\omega$,
\[
d(\{f(u_{j_n}(\cdot,\omega))\},\gamma(\omega)) < 2^{-n}
\]
It follows that, taking a suitable subsequence, denoted as $\{u_{n_k}(\cdot,\omega)\}$,
\[
\gamma(\omega) = \lim_{n(\omega) \to \infty} f(u_{n_k}(\cdot,\omega))
\]
for each $\omega$. In particular, for each $k$
\[
\gamma_k(t,\omega) = \lim_{n_k(\omega) \to \infty} f(u_{n_k}(t,\omega)) = \lim_{n(\omega) \to \infty} m_k \int_{l_{m_k}(t)}^t \langle \phi_{r_k},u_{n_k}(s,\omega) \rangle_{V,V'} ds
\]
for each $t$.

Note that it is not clear that $(t,\omega) \to f(u_{n_k}(t,\omega))$ is $\mathcal{P}$ measurable although $(t,\omega) \to \gamma(t,\omega)$ is $\mathcal{P}$ measurable.

**Proof of the theorem:** By assumption, there exists a further subsequence still denoted by $n(\omega)$ such that, in addition to 22.2.2 above, the weak limit $\lim_{n(\omega) \to \infty} u_{n_k}(\cdot,\omega) = u(\cdot,\omega)$ exists in $L^{p'}([0,T];V')$ such that $t \to u(t,\omega)$ is weakly continuous into $V'$. Then the above equation continues to hold for this further subsequence and in addition to this,
\[
m_k \int_{l_{m_k}(t)}^t \langle \phi_{r_k},u(s,\omega) \rangle_{V,V'} ds = \lim_{n(\omega) \to \infty} m_k \int_{l_{m_k}(t)}^t \langle \phi_{r_k},u_{n_k}(s,\omega) \rangle_{V,V'} ds = \gamma_k(t,\omega)
\]
Letting $\phi \in \mathcal{D}$ given, there exists a subsequence denoted by $k$ such that $m_k \to \infty$ and $\phi_{r_k} = \phi$ for all $k$. Then passing to a limit and using the assumed continuity of $s \to u(s,\omega)$, the left side of this equation converges to $\langle \phi,u(t,\omega) \rangle_{V,V'}$, and so the right side, $\gamma_k(t,\omega)$ must also converge, this for each $\omega$. Since the right side is a product measurable function of $(t,\omega)$, it follows that the pointwise limit is also product measurable. Hence $(t,\omega) \to \langle \phi,u(t,\omega) \rangle_{V,V'}$ is product measurable, this for each $\phi \in \mathcal{D}$. Since $\mathcal{D}$ is a dense set, it follows that $(t,\omega) \to \langle \phi,u(t,\omega) \rangle_{V,V'}$ is $\mathcal{P}$ measurable for all $\phi \in V$ and so by the Pettis theorem, 22.2, $(t,\omega) \to u(t,\omega)$ is $\mathcal{P}$ measurable into $V'$.

One can say more about the measurability of the approximating sequence. In fact, we can obtain one for which $\omega \to u_{n_k}(t,\omega)$ is also $\mathcal{F}$ measurable.

**Lemma 22.2.6** Suppose, $u_{n(\omega)}(t,\omega)$ weakly in $L^{p'}([0,T];V')$ where $u$ is product measurable measurable and $\{u_{n(\omega)}\}$ is a subsequence of $\{u_n\}$ where
\[
\sup_{t \in [0,T]} \|u_n(t,\omega)\|_{V'} < C(\omega), \text{ for } \omega \notin N \text{ a set of measure zero},
\]
Then for each $\omega \notin N$, there exists a subsequence of $\{u_n\}$ denoted as $\{u_k(\omega)\}$ such that $u_k(\omega) \to u$ weakly in $L^{p'}([0,T];V')$, $\omega \to k(\omega)$ is $\mathcal{F}$ measurable, and $\omega \to u_k(\omega)(t,\omega)$ is also $\mathcal{F}$ measurable, the last assertions holding for all $\omega \notin N$. 

CHAPTER 22. MEASURABILITY WITHOUT UNIQUENESS

**Proof:** For \( f, g \in L^p ([0, T]; V') \equiv V' \), \( L^p ([0, T]; V) \equiv V \), let \( \{ \phi_k \} \) be a countable dense subset of \( L^p ([0, T]; V) \). Then a bounded set in \( L^p ([0, T]; V') \) with the weak topology can be considered a complete metric space using the following metric.

\[
 d (f, g) = \sum_{j=1}^{\infty} 2^{-j} \frac{|\langle \phi_k, f - g \rangle_{V, V'}|}{1 + |\langle \phi_k, f - g \rangle_{V, V'}|}
\]

Now let \( k (\omega) \) be the first index from the indices of \( \{ u_n \} \) at least as large as \( k \) such that

\[
 d (u_k (\omega), u) \leq 2^{-k}
\]

Such an index exists because there exists a convergent sequence \( u_{n(\omega)} \) which does converge weakly to \( u \). This is just picking another one which happens to also retain measurability. In fact,

\[
 \{ \omega : k (\omega) = l \} = \{ \omega : d (u_l, u) \leq 2^{-k} \} \cap \cap_{j=k}^{\infty} \{ \omega : d (u_j, u) > 2^{-k} \}
\]

Since \( u \) is product measurable and each \( u_l \) is also product measurable, these are all measurable sets with respect to \( \mathcal{F} \) and \( \omega \to k (\omega) \) is \( \mathcal{F} \) measurable. Now we have \( u_{k(\omega)} \to u \) weakly in \( L^p ([0, T]; V') \) for each \( \omega \) with each function being \( \mathcal{F} \) measurable because

\[
 u_{k(\omega)} (t, \omega) = \sum_{j=1}^{\infty} \mathcal{Y}_{[k(\omega)=j]} u_j (t, \omega)
\]

and every term in the sum is \( \mathcal{F} \) measurable. \( \blacksquare \)

The following obvious corollary shows the significance of this lemma.

**Corollary 22.2.7** Let \( V \) be a reflexive separable Banach space and \( V' \) its dual and \( \frac{1}{p} + \frac{1}{p'} = 1 \) where \( p > 1 \) as usual. Let the functions \( t \to u_{n(\omega)} (t, \omega) \) be in \( L^p ([0, T]; V') \) and \( t \to u_{n(\omega)} (t, \omega) \) be \( \mathcal{B} ([0, T]) \times \mathcal{F} \equiv \mathcal{P} \) measurable into \( V' \). Here \( \{ u_{n(\omega)} \}_{n=1}^{\infty} \) is a sequence, one for each \( \omega \). Suppose there is a set of measure zero \( N \) such that if \( \omega \notin N \), then for all \( n \),

\[
 \sup_{t \in [0, T]} \| u_{n(\omega)} (t, \omega) \|_{V'} \leq C (\omega).
\]

Also suppose for each \( \omega \notin N \), each subsequence of \( \{ u_{n(\omega)} \} \) has a further subsequence which converges weakly in \( L^p ([0, T]; V') \) to \( u (\cdot, \omega) \in L^p ([0, T]; V') \) such that \( t \to u (t, \omega) \) is weakly continuous into \( V' \). Then there exists \( u \) product measurable, with \( t \to u (t, \omega) \) being weakly continuous into \( V' \). Moreover, there exists, for each \( \omega \notin N \), a subsequence \( u_{n(\omega)} \) such that \( u_{n(\omega)} (\cdot, \omega) \to u (\cdot, \omega) \) weakly in \( L^p ([0, T]; V') \).

**Proof:** It suffices to consider the functions \( v_n (t, \omega) = u_{n(\omega)} (t, \omega) \) and use the result of Theorem 22.2.6. \( \blacksquare \)

Of course when you have all functions having values in \( H \) a separable Hilbert space, there is no change in the argument to obtain the following theorem.

**Theorem 22.2.8** Let \( H \) be a real separable Hilbert space. For \( n \in \mathbb{N} \) let the functions \( t \to u_n (t, \omega) \) be in \( L^2 ([0, T]; H) \) and \( t \to u_n (t, \omega) \) be \( \mathcal{B} ([0, T]) \times \mathcal{F} \equiv \mathcal{P} \) measurable into \( H \). Suppose there is a set of measure zero \( N \) such that if \( \omega \notin N \), then for all \( n \),

\[
 \sup_{t \in [0, T]} \| u_n (t, \omega) \|_H \leq C (\omega).
\]

Also suppose for each \( \omega \notin N \), each subsequence of \( \{ u_n \} \) has a further subsequence which converges weakly in \( L^2 ([0, T]; H) \) to \( u (\cdot, \omega) \in L^2 ([0, T]; H) \) such that \( t \to u (t, \omega) \) is weakly continuous into \( H \). Then there exists \( u \) product measurable, with \( t \to u (t, \omega) \) being weakly continuous into \( H \). Moreover, there exists, for each \( \omega \notin N \), a subsequence \( u_{n(\omega)} \) such that \( u_{n(\omega)} (\cdot, \omega) \to u (\cdot, \omega) \) weakly in \( L^2 ([0, T]; H) \).

### 22.3 Measurability In Finite Dimensional Problems

What follows is like the Peano existence theorem from ordinary differential equations except that it provides a solution which retains product measurability. It is a nice example of the above theory. It will be used in the next section in the Galerkin method.
Lemma 22.3.1 Suppose \( N(t, u, v, w, \omega) \in \mathbb{R}^d \) for \( u, v, w \in \mathbb{R}^d, t \in [0, T] \) and \( (t, u, v, w, \omega) \rightarrow N(t, u, v, w, \omega) \) is progressively measurable relative to the filtration consisting of the single \( \sigma \) algebra \( F \). Also suppose \((t, u, v, w) \rightarrow N(t, u, v, w, \omega)\) is continuous and that \( N(t, u, v, w, \omega) \) is uniformly bounded in \((t, u, v, w)\) by \( M(\omega) \). Let \( f \) be \( P \)-measurable and \( f(\cdot, \omega) \in L^2([0,T]; \mathbb{R}^d) \). Then for \( h > 0 \), there exists a \( P \)-measurable solution \( u \) to the integral equation

\[
 u(t, \omega) - u_0(\omega) + \int_0^t N(s, u(s, \omega), u(s-h, \omega), w(s, \omega), \omega) \, ds = \int_0^t f(s, \omega) \, ds.
\]

Here \( u_0 \) has values in \( \mathbb{R}^d \) and is \( F \)-measurable, \( u(s-h, \omega) \equiv u_0(\omega) \) if \( s-h < 0 \) and for \( w_0 \) a given \( F \)-measurable function,

\[
 w(t, \omega) \equiv w_0(\omega) + \int_0^t u(s, \omega) \, ds.
\]

Proof: Let \( u_n \) be the solution to the following equation:

\[
 u_n(t, \omega) - u_0(\omega) + \int_0^t N(s, \tau_{1/n} u_n(s, \omega), u_n(s-h, \omega), \tau_{1/n} w_n(s, \omega), \omega) \, ds = \int_0^t f(s, \omega) \, ds.
\]

where here \( \tau_{1/n} \) is defined as follows. For \( \delta > 0 \),

\[
 \tau_{\delta} u(s) \equiv \begin{cases} u(s-\delta) & \text{if } s > \delta \\ 0 & \text{if } s - \delta \leq 0 \end{cases}
\]

It follows that \((t, \omega) \rightarrow u_n(t, \omega) \) is \( P \)-measurable. From the assumptions on \( N \), it follows that for fixed \( \omega \), \( \{u_n(\cdot, \omega)\} \) is uniformly bounded:

\[
 \sup_{t \in [0,T]} |u_n(t, \omega)| \leq |u_0(\omega)| + \int_0^T M(\omega) \, ds + \int_0^T |f(s, \omega)| \, ds = C(\omega),
\]

and is also equicontinuous because for \( s < t \),

\[
 |u_n(t, \omega) - u_n(s, \omega)| \leq \int_s^t \left| N(r, \tau_{1/n} u_n(r, \omega), u_n(r-h, \omega), \tau_{1/n} w_n(r, \omega), \omega) \right| \, dr 
\]

\[
 + \int_s^t |f(r, \omega)| \, dr \leq C(\omega, f) |t-s|^{1/2}.
\]

Therefore, by the Ascoli–Arzelà theorem, for each \( \omega \), there exists a subsequence \( \tilde{u}(\omega) \) depending on \( \omega \) and a function \( \tilde{u}(t, \omega) \) such that

\[
 u_{n(\omega)}(t, \omega) \rightarrow \tilde{u}(t, \omega) \text{ uniformly in } C([0,T]; \mathbb{R}^d).
\]

This verifies the assumptions of Theorem 22.2.8.

It follows that there exists \( \tilde{u} \) product measurable and a subsequence \( \{u_{n(\omega)}\} \) for each \( \omega \) such that

\[
 \lim_{n(\omega) \rightarrow \infty} u_{n(\omega)}(\cdot, \omega) = \tilde{u}(\cdot, \omega) \text{ weakly in } L^2([0,T]; \mathbb{R}^d)
\]

and that \( t \rightarrow \tilde{u}(t, \omega) \) is continuous. (Note that weak continuity is the same as continuity in \( \mathbb{R}^d \).) The same argument given above applied to the \( u_{n(\omega)} \) for a fixed \( \omega \) yields a further subsequence, denoted as \( \{u_{n_{\delta}(\omega)}(\cdot, \omega)\} \) which converges uniformly to a function \( u(\cdot, \omega) \) on \([0,T]\). So \( \tilde{u}(t, \omega) = u(t, \omega) \) in \( L^2([0,T]; \mathbb{R}^d) \). Since both of these functions are continuous in \( t \), they must be equal for all \( t \). Hence, \((t, \omega) \rightarrow u(t, \omega) \) is product measurable. Passing to the limit in the equation solved by \( \{u_{n(\omega)}(\cdot, \omega)\} \) using the dominated convergence theorem, we obtain

\[
 u(t, \omega) - u_0(\omega) + \int_0^t N(s, u(s, \omega), u(s-h, \omega), w(s, \omega), \omega) \, ds = \int_0^t f(s, \omega) \, ds.
\]

Thus \( t \rightarrow u(t, \omega) \) is a product measurable solution to the integral equation. ■
Remark 22.3.2 When \( w(t) = w_0(\omega) + \int_0^t u(s, \omega) \, ds \),

\[
  v(t) = \begin{cases} 
    u(t - h) & \text{if } t \geq h \\
    u_0 & \text{if } t < h 
  \end{cases}
\]

and when the estimate

\[
  (N(t, u, v, w, \omega), u) \geq -C(t, \omega) - \mu \left( |u|^2 + |v|^2 + |w|^2 \right)
\]

holds, it follows that

\[
  \int_0^t (N(t, u, v, w, \omega), u) \, ds \geq -C \left( C(\omega) + \int_0^t |u|^2 \, ds \right)
\]

for some constant \( C \) depending on the initial data but not on \( u \).

To see this,

\[
  \int_0^t |u(s - h)|^2 \, ds = \int_0^h |u_0|^2 \, ds + \int_h^t |u(s - h)|^2 \, ds
\]

\[
  = |u_0|^2 h + \int_0^{t-h} |u(s)|^2 \, ds \leq |u_0|^2 h + \int_0^t |u(s)|^2 \, ds
\]

if \( t \geq h \) and if \( s < h \), this is dominated by

\[
  |u_0|^2 t \leq |u_0|^2 h \leq |u_0|^2 h + \int_0^t |u(s)|^2 \, ds
\]

As to the terms from \( w \),

\[
  \int_0^t |w(s)|^2 \, ds \leq \int_0^t \left| w_0 + \int_0^s u(r) \, dr \right|^2 \, ds \leq \int_0^t \left( |w_0| + \left| \int_0^s u(r) \, dr \right| \right)^2 \, ds
\]

\[
  \leq \int_0^t \left( |w_0|^2 + 2 |w_0| \left| \int_0^s u(r) \, dr \right| + \left| \int_0^s u(r) \, dr \right|^2 \right) \, ds
\]

\[
  \leq T |w_0|^2 + T |w_0|^2 + \int_0^t \left| \int_0^s u(r) \, dr \right|^2 \, ds + \int_0^t \left| \int_0^s u(r) \, dr \right|^2 \, ds
\]

\[
  \leq 2T |w_0|^2 + 2 \int_0^t \left( \int_0^s |u(r)| \, dr \right)^2 \, ds \leq 2T |w_0|^2 + 2 \int_0^t \int_0^s |u(r)|^2 \, drds
\]

\[
  \leq 2T |w_0|^2 + 2T \int_0^t \int_0^s |u(r)|^2 \, drds \leq 2T |w_0|^2 + 2T^2 \int_0^t |u(r)|^2 \, dr
\]

From this, the claimed result follows.

Theorem 22.3.3 Suppose \( N(t, u, v, w, \omega) \in \mathbb{R}^d \) for \( u, v, w \in \mathbb{R}^d, t \in [0, T] \) and \( (t, u, v, w, \omega) \to N(t, u, v, w, \omega) \) is progressively measurable with respect to a constant filtration \( F_t = F \). Also suppose \( (t, u, v, w) \to N(t, u, v, w, \omega) \) is continuous and satisfies the following conditions for \( C(\cdot, \omega) \geq 0 \) in \( L^1([0, T]) \) and some \( \mu > 0 \):

\[
  (N(t, u, v, w, \omega), u) \geq -C(t, \omega) - \mu \left( |u|^2 + |v|^2 + |w|^2 \right).
\]

Also let \( f \) be product measurable and \( f(\cdot, \omega) \in L^2([0, T]; \mathbb{R}^d) \). Then for \( h > 0 \), there exists a product measurable solution \( u \) to the integral equation

\[
  u(t, \omega) = u_0(\omega) + \int_0^t N(s, u(s, \omega), u(s - h, \omega), w(s, \omega), \omega) \, ds = \int_0^t f(s, \omega) \, ds, \quad (22.3.3)
\]

where \( u_0 \) has values in \( \mathbb{R}^d \) and is \( F \) measurable. Here \( u(s - h, \omega) \equiv u_0(\omega) \) for all \( s - h \leq 0 \) and for \( w_0 \) a given \( F \) measurable function,

\[
  w(t, \omega) \equiv w_0(\omega) + \int_0^t u(s, \omega) \, ds
\]
These are product measurable and then define the functions \( \omega, \tau \) and note that

\[
\inf_{\emptyset} \equiv \inf \left\{ t \in [0, T] : |u_m(t, \omega)|^2 + |w_m(t, \omega)|^2 > 2^m \right\},
\]

where \( \inf \emptyset \equiv T \). Localizing with the stopping time,

\[
u_m(t, \omega) = \int_0^t \mathcal{X}_{[0, \tau_m]}(s, u_m^\tau_m(s, \omega), u_m^\tau_m(s - h, \omega), w_m^\tau_m(s, \omega), w_m^\tau_m(s, \omega), \omega) ds = \int_0^t \mathcal{X}_{[0, \tau_m]} f(s, \omega) ds.
\]

Note how the stopping time allowed the elimination of the projection map in the equation. Then we get

\[
\frac{1}{2} |u_m^\tau_m(t, \omega)|^2 - \frac{1}{2} |u_0(\omega)|^2 + \int_0^t \left( \mathcal{X}_{[0, \tau_m]}(s, u_m^\tau_m(s, \omega), u_m^\tau_m(s - h, \omega), w_m^\tau_m(s, \omega), w_m^\tau_m(s, \omega), u_m^\tau_m(s, \omega)) ds \right.
\]

From the estimate,

\[
\frac{1}{2} |u_m^\tau_m(t, \omega)|^2 - \frac{1}{2} |u_0(\omega)|^2 \leq \int_0^t \left( \mu \left( |u_m^\tau_m(s, \omega)|^2 + |u_m^\tau_m(s - h, \omega)|^2 + |w_m^\tau_m(s, \omega)|^2 \right) + C(s, \omega) + \frac{1}{2} |f(s, \omega)|^2 \right) ds + \frac{1}{2} \int_0^t |u_m^\tau_m(s, \omega)|^2 ds.
\]

Note that

\[
|u_0|^2 + \int_0^t |u_m^\tau_n(s)|^2 ds \geq \int_0^t |u_m^\tau_n(s - h, \omega)|^2 ds
\]

and

\[
\int_0^t |w_n^\tau_m(s, \omega)|^2 ds = \int_0^t \left| w_0 + \int_0^s \mathcal{X}_{[0, \tau_m]} u_n(r) dr \right|^2 ds
\]

By Gronwall’s inequality,

\[
|u_m^\tau_m(t, \omega)|^2 \leq C \left( |u_0(\omega), w_0(\omega), \mu, \|C(\cdot, \omega)\|_{L^1([0, T]; \mathbb{R}^d)}, T, \|f(\cdot, \omega)\|_{L^2([0, T]; \mathbb{R}^d)} \right)
\]

\[
= : C(\omega).
\]

Thus, for a.e. \( \omega, \tau_m = T \) for all \( m \) large enough, say for \( m \geq M(\omega) \) where

\[
C(\omega) \leq 2^{M(\omega)}.
\]

Then define the functions

\[
y_n(t, \omega) \equiv u_m^\tau_n(t, \omega).
\]

These are product measurable and

\[
y_n(t, \omega) - u_0(\omega) + \int_0^t \mathcal{X}_{[0, \tau_n]}(s, y_n(s, \omega), y_n(s - h, \omega), w_0(\omega) + \int_0^s y_n(r, \omega) dr, \omega) ds
\]

\[
= \int_0^t \mathcal{X}_{[0, \tau_n]} f(s, \omega) ds.
\]
So each is continuous in \( t \). For large enough \( n \), \( \tau_n = T \) and hence

\[
y_n(t, \omega) - u_0(\omega) + \int_0^t N \left( s, y_n(s, \omega), y_n(s - h, \omega), w_0(\omega) + \int_0^s y_n(r, \omega) \, dr, \omega \right) \, ds = \int_0^t f(s, \omega) \, ds.
\]

Also these satisfy the inequality

\[
\sup_{t \in [0, T]} |y_n(t, \omega)|^2 \leq C(\omega) \leq 2^{M(\omega)} < 9^{M(\omega)},
\]

the constant on the right not depending on \( n \). Thus for fixed \( \omega \), we can regard \( N \) as bounded and the same reasoning used in the above lemma involving the Ascoli–Arzelà theorem implies that every subsequence has a further subsequence which converges to a solution to the integral equation for that \( \omega \). Thus it is continuous into \( \mathbb{R}^d \).

It follows from the measurable selection theorem above that there exists \( u \) product measurable and continuous in \( t \) such that \( u(\cdot, \omega) = \lim_{n \to \infty} y_n(\cdot, \omega) \) in \( L^2([0, T]; \mathbb{R}^d) \). By the reasoning of the above lemma, there is a further subsequence, denoted the same way, for which \( \lim_{n \to \infty} y_n(\omega) \) in \( C([0, T]; \mathbb{R}^d) \) solves the integral equation for a fixed \( \omega \). Thus \( u \) is a product measurable solution to the integral equation as claimed.

We made use of an estimate in order to get the conclusion of this theorem. However, all that is really needed is the following.

**Corollary 22.3.4** Suppose \( N(t, u, v, w, \omega) \in \mathbb{R}^d \) for \( u, v, w \in \mathbb{R}^d \) and \( t \in [0, T] \) and \( (t, u, v, w, \omega) \to N(t, u, v, w, \omega) \) is progressively measurable with respect to a constant filtration \( \mathcal{F}_t = \mathcal{F} \). Also suppose \( (t, u, v, w, \omega) \to N(t, u, v, w, \omega) \) is continuous. Suppose for each \( \omega \), there exists an estimate for any solution \( u(\cdot, \omega) \) to the integral equation

\[
u(t, \omega) - u_0(\omega) + \int_0^t N(s, u(s, \omega), u(s - h, \omega), w(s, \omega), \omega) \, ds = \int_0^t f(s, \omega) \, ds,
\]

which is of the form

\[
\sup_{t \in [0, T]} |u(t, \omega)| \leq C(\omega) < \infty
\]

Also let \( f \) be product measurable and \( f(\cdot, \omega) \in L^1([0, T]; \mathbb{R}^d) \). Here \( u_0 \) has values in \( \mathbb{R}^d \) and is \( \mathcal{F} \) measurable and \( u(s - h, \omega) \equiv u_0(\omega) \) whenever \( s - h \leq 0 \) and

\[
w(t, \omega) \equiv w_0(\omega) + \int_0^t u(s, \omega) \, ds
\]

where \( w_0 \) is a given \( \mathcal{F} \) measurable function. Then for \( h > 0 \), there exists a product measurable solution \( u \) to the integral equation.

Of course the same conclusions apply when there is no dependence in the integral equation on \( u(s - h, \omega) \) or the integral \( w(t, \omega) \). Note that these theorems hold for all \( \omega \).

### 22.4 The Navier–Stokes Equations

In this section, we study the stochastic Navier–Stokes equations of arbitrary dimension. We prove there exists a global solution which is product measurable. The main result is Theorem [22.4.1.1]. We use the Galerkin method and Theorem [22.4.6.2] to get product measurable approximate solutions. Then we take weak limits and get path solutions. After this, we apply Theorem [22.4.6.2.2] to get product measurable global solutions.

As in [21.6.6], an important part of our argument is the theorem in Lions [41] which follows. See Theorem [22.4.6.4.4].

**Theorem 22.4.1** Let \( W, H, \) and \( V' \) be separable Banach spaces. Suppose \( W \subseteq H \subseteq V' \) where the injection map is continuous from \( H \) to \( V' \) and compact from \( W \) to \( H \). Let \( q_1 \geq 1, q_2 > 1 \), and define

\[
S \equiv \{ u \in L^{q_1}([a, b]; W) : u' \in L^{q_2}([a, b]; V') \text{ and } ||u||_{L^{q_1}([a, b]; W)} + ||u'||_{L^{q_2}([a, b]; V')} \leq R \}.
\]

Then \( S \) is pre-compact in \( L^{q_1}([a, b]; H) \). This means that if \( \{ u_n \}_{n=1}^\infty \subseteq S \), it has a subsequence \( \{ u_{n_k} \} \) which converges in \( L^{q_1}([a, b]; H) \).
22.4. THE NAVIER–STOKES EQUATIONS

A proof of a generalization of this theorem is found on Page 512. Let $U$ be a bounded open set in $\mathbb{R}^d$ and let $S$ denote the functions which are infinitely differentiable having zero divergence and also having compact support in $U$. We have in mind $d = 3$, but the approach is not limited by dimension. We use the same Galerkin method found in [13], but the approach is not limited by dimension. We use the same Galerkin method found in [13]. The details being included in slightly abbreviated form for convenience of the reader. The difference is that we switch the roles of $V$ and $W$ along with a few other minor modifications. This is the part of the argument which gives a solution for each $\omega$ and it is standard material. Define

$$V \equiv S \text{ in } (H^d(U))^d, \quad W \equiv S \text{ in } (H^1(U))^d, \quad \text{and } H \equiv S \text{ in } (L^2(U))^d,$$

where $d^*$ is such that for $w \in (H^{d^*}(U))^d$ then $\|Dw\|_{L^\infty(U)} < \infty$. For example, you could take $d^* = 3$ for $d = 3$. In [13], they take $d^* = 8$ which is large enough to work for all dimensions of interest.

Let $A : W \rightarrow W'$ and $N : W \rightarrow V'$ be defined by

$$\langle Au, v \rangle \equiv \int_U \nabla u_i \cdot \nabla v_i \, dx, \quad \langle Nu, v \rangle \equiv -\int_U u_i u_j v_j \, dx.$$

Then $N$ is a continuous function. Indeed, pick $v \in V$ and suppose $u_n \rightarrow u$ in $W$, then

$$|\langle Nu - Nu_n, v \rangle| \leq \int_U \left| \sum_{i,j} (u_{ni} u_{nj} - u_{ii} u_{jj}) v_{ij} \right| \, dx \leq C \|v\|_V \int_U (|u_n| + |u|)(|u_n - u|) \, dx \leq C \|v\|_V \left( \int_U |u_n|^2 + |u|^2 \, dx \right)^{1/2} \left( \int_U |u_n - u|^2 \, dx \right)^{1/2},$$

where what multiplies $\|v\|_V$ clearly converges to 0.

An abstract form for the incompressible Navier–Stokes equations is

$$u' + \nu A u + Nu = f, \quad u(0) = u_0,$$

where $f \in L^2([0,T];W')$, for some fixed $T > 0$. As in [13], we will let $\nu = 1$ to simplify the presentation. A stochastic version of this would be the integral equation in $V'$

$$u(t, \omega) - u_0(\omega) + \int_0^t A(u(s, \omega)) \, ds + \int_0^t N(u(s, \omega)) \, ds = \int_0^t f(s, \omega) \, ds + q(t, \omega),$$

where $q(\cdot, \omega)$ will be continuous into $V$; $(t, \omega) \rightarrow q(t, \omega)$ will be product measurable having values in $V$, and $q(0, \omega) = 0$. So $q$ here is a fixed stochastic process, which serves as the random source. Also $(t, \omega) \rightarrow f(t, \omega)$ will be product measurable into $W'$ as well as having $t \rightarrow f(t, \omega)$ in $L^2([0,T];W')$. Our problem is to show the existence of a product measurable solution.

Let $T$ be any fixed positive number and let $q$ be any fixed process satisfying the above.

**Definition 22.4.2** A **global solution** to the above integral equation is a process $u(t, \omega)$, for which $\omega \rightarrow u(t, \omega)$ is $\mathcal{F}$ measurable and satisfies for each $\omega$ outside a set of measure zero and all $t \in [0,T]$,

$$u(t, \omega) - u_0(\omega) + \int_0^t A(u(s, \omega)) \, ds + \int_0^t N(u(s, \omega)) \, ds = \int_0^t f(s, \omega) \, ds + q(t, \omega).$$

In order to apply the earlier result, let $w(t, \omega) = u(t, \omega) - q(t, \omega)$ and write the equation in terms of $w$,

$$w(t, \omega) - w_0(\omega) + \int_0^t A(w(s, \omega) + q(s, \omega)) \, ds + \int_0^t N(w(s, \omega) + q(s, \omega)) \, ds = \int_0^t f(s, \omega) \, ds.$$

It turns out that it is convenient to define

$$\langle B(u, v), w \rangle \equiv -\int_U u_i v_j w_{ij} \, dx,$$

and write the equation in the following form:

$$w(t, \omega) - w_0(\omega) + \int_0^t A(w(s, \omega)) \, ds + \int_0^t N(w(s, \omega)) \, ds = \int_0^t f(s, \omega) \, ds,$$
where
\[
\hat{N}(w(t, \omega)) \equiv N(w(t, \omega)) + B(w(t, \omega), q(t, \omega)) + B(q(t, \omega), w(t, \omega)),
\]
\[
\hat{f}(t, \omega) \equiv f(t, \omega) - A(q(t, \omega)) - N(q(t, \omega)).
\]

This is an equation in \(V'\). Moreover, we have the following:

**Lemma 22.4.3** For fixed \(\omega \in \Omega\), \(\hat{f} \in L^2([0, T]; W')\), and
\[
(t, w) \to B(w, q(t, \omega)), \quad (t, w) \to B(q(t, \omega), w)
\]
are continuous functions having values in \(W'\). For fixed \(w \in W\),
\[
(t, \omega) \to B(w, q(t, \omega)), \quad (t, \omega) \to B(q(t, \omega), w)
\]
are product measurable. In addition to this, if \(z \in W\),
\[
\langle B(w, q(t, \omega)), z \rangle \leq C \|q(t, \omega)\|_V \|w\|_H \|z\|_H, \quad \langle B(q(t, \omega), w), z \rangle \leq C \|q(t, \omega)\|_V \|w\|_H \|z\|_H.
\]

**Proof:** The first claim is straightforward to prove from the definition of \(A\) and \(N\). Consider the next claim about continuity. Let \(z \in W\) be given. Then from the fact that all the functions are divergence free,
\[
\langle B(w, q(t)), z \rangle = \left| \int_U (w_q j(t) - \tilde{w}_q j(s)) z_{j,i} dx \right| \
\leq \left| \int_U (w_q j_i(t) - \tilde{w}_q j_i(t)) z_j dx \right| + \left| \int_U (\tilde{w}_q j_i(t) - \tilde{w}_q j_i(s)) z_j dx \right| \
\leq C \left( \|q(t)\|_V \int_U |w - \tilde{w}| |z| dx + \|q(t) - q(s)\|_V \int_U |\tilde{w}| |z| dx \right) \
\leq C \left( \|q(t)\|_V \|w - \tilde{w}\|_H + \|q(t) - q(s)\|_V \|\tilde{w}\|_H \right) \|z\|_H,
\]
where we have suppressed the dependence of \(q\) on \(\omega\) to simplify the notation. The other function is similar.

As to the claim about product measurability, this follows from the above definition and assumptions about \(q\) being product measurable. For the estimates,
\[
\|B(w, q), z\| = \left| \int_U w_q z_{j,i} \right| \leq \left| \int_U w_q z_j \right| \leq C \|q\|_V \int_U |w| |z| dx,
\]
and apply Hölder’s inequality. The other estimate is similar. \(\blacksquare\)

This has shown that it suffices to verify that there exists a global solution \(u\) to the equation
\[
\dot{u}(t, \omega) - u_0(\omega) + \int_0^t A(u(s, \omega)) ds + \int_0^t \hat{N}(u(s, \omega)) ds = \int_0^t f(s, \omega) ds,
\]
where \(f(\cdot, \omega) \in L^2([0, T]; W')\) for each \(\omega \in \Omega\).

Let \(R\) be the Riesz map from \(V\) to \(V'\), so \(\langle Rv_1, v_2 \rangle_{V', V} = \langle v_1, v_2 \rangle_V\) for any \(v_1, v_2 \in V\). The compactness of the embeddings imply that \(R^{-1}\) is a compact self-adjoint operator on \(H\) and so there is a complete orthonormal basis \(\{w_k\}\) for \(H\) such that \(Rw_k = \mu_k w_k\), where \(\{\mu_k\}\) is a decreasing sequence of positive numbers which converges to 0. Thus \(\hat{R}w_k = \lambda_k w_k\), where \(\lim_{k \to \infty} \lambda_k = 0\). \(\{w_k\}\) is a special basis. It is orthonormal in \(H\) and orthogonal in \(V\), since \(\langle w_k, w_l \rangle_V = \langle \hat{R}w_k, \hat{R}w_l \rangle_H = \langle \lambda_k w_k, \lambda_l w_l \rangle_H\).

To use the Galerkin method, let \(V_n = \operatorname{span}(w_1, \ldots, w_n)\). Clearly \(\cup_n V_n\) is dense in \(H\). This is also dense in \(V\). If not, then there exists \(\phi \in V'\) such that \(\phi \neq 0\) but \(\langle \phi, v_n \rangle \to 0\) as \(n \to \infty\). Then \(\phi = R \cdot y\). Hence for \(z \in \bigcup_{n=1}^\infty V_n\),
\[
0 = \langle R \cdot y, z \rangle = \langle \hat{R} \cdot y, z \rangle_H.
\]
But \(\hat{R}z \in \operatorname{span}(w_1, \ldots, w_M)\) and in fact, \(R\) maps \(V\) onto \(V_0\) and so this shows that \(y\) is perpendicular to \(\operatorname{span}(w_1, \ldots, w_M)\) for each \(M\) so \(y = 0\) and \(\phi = 0\) after all. Thus \(\cup_n V_n\) is also dense in \(V\) and hence it is also dense in \(W\).

Let \(u_n(t, \omega) = \sum_{k=1}^n x_k(t, \omega) w_k\), where \(x(t, \omega) = (x_1(t, \omega), \ldots, x_n(t, \omega))^T \in \mathbb{R}^n\). We consider the problem of finding \(x(t, \omega)\) such that for all \(w_k, k \leq n\),
\[
\langle u_n(t, \omega), w_k \rangle_H - \langle u_n(\omega), w_k \rangle_H + \int_0^t \langle A(u_n(s, \omega)), w_k \rangle ds + \int_0^t \langle \hat{N}(u_n(s, \omega)), w_k \rangle ds = \int_0^t \langle f(s, \omega), w_k \rangle ds,
\]
(22.4.6)
22.4. THE NAVIER–STOKES EQUATIONS

517

where \( u_{0n} \) is the orthogonal projection of \( u_0 \) onto \( V_n \).

By the continuity of the operators described above, and the orthogonality of the \( w_k \), this is nothing but an ordinary differential equation for the vector \( x(t, \omega) \). By Theorem 22.4.5, there exists a product measurable solution \( x \) and therefore, \( u_n(t, \omega) \) is also product measurable in \( H \).

Take the derivative, multiply by \( x_k(t, \omega) \), add, and integrate again in the usual way to obtain

\[
\frac{1}{2} |u_n(t, \omega)|_H^2 - \frac{1}{2} |u_{0n}(\omega)|_H^2 = -\int_0^t \langle A(u_n(s, \omega)), u_n(s, \omega) \rangle ds - \int_0^t \langle \tilde{N}(u_n(s, \omega)), u_n(s, \omega) \rangle ds + \int_0^t \langle f(s, \omega), u_n(s, \omega) \rangle ds.
\]

Recall that \( \tilde{N}(u(t, \omega)) = N(u(t, \omega)) + B(u(t, \omega), q(s, \omega)) + B(q(t, \omega), u(t, \omega)) \). From the above lemma and that all functions are divergence free, we obtain \( \int_0^t \langle B(q(s, \omega), u_n(s, \omega)), u_n(s, \omega) \rangle ds = 0 \), and

\[
\left| \int_0^t \langle B(u_n(s, \omega), q(s, \omega)), u_n(s, \omega) \rangle ds \right| \leq C \int_0^t \| q(s, \omega) \|_V |u_n(s, \omega)|^2_H ds.
\]

Then one can obtain an inequality of the following form

\[
\frac{1}{2} |u_n(t, \omega)|_H^2 + \int_0^t \| u_n(s, \omega) \|_W^2 ds \\
\leq \frac{1}{2} |u_{0n}(\omega)|_H^2 + C \int_0^t \| q(s, \omega) \|_V \| u_n(s, \omega) \|_H^2 ds + C \int_0^t \| f(s, \omega) \|_W \| u_n(s, \omega) \|_H^2 ds + \frac{1}{2} \int_0^t \| u_n(s, \omega) \|_W^2 ds.
\]

Since \( t \to \| q(t, \omega) \|_V \) is continuous, it follows from Gronwall’s inequality that there is an estimate of the form

\[
|u_n(t, \omega)|_H^2 + \int_0^t \| u_n(s, \omega) \|_W^2 ds \leq C(u_0, f, q, T, \omega).
\]

The next task is to estimate \( \| u_n(\omega) \|_{L^2([0,T];V')} \) for each fixed \( \omega \in \Omega \). We will suppress the dependence on \( \omega \) of all functions whenever it is appropriate. With \( \text{V.3.2.3.1} \), the fundamental theorem of calculus implies that for each \( \mathbf{w} \in V_n \),

\[
\langle u'_n(t), \mathbf{w} \rangle_{V', V} + \langle A(u_n(t)), \mathbf{w} \rangle_{V', V} + \langle \tilde{N}(u_n(t)), \mathbf{w} \rangle = \langle f(t), \mathbf{w} \rangle.
\]

In terms of inner products in \( V \),

\[
\left( R^{-1}u'_n(t) + R^{-1}A(u_n(t)) + R^{-1}\tilde{N}(u_n(t)) - R^{-1}f(t), \mathbf{w} \right)_V = 0
\]

for all \( \mathbf{w} \in V_n \). This is equivalent to saying that for \( P_n \) the orthogonal projection in \( V \) onto \( V_n \),

\[
\left( R^{-1}u'_n(t) + R^{-1}A(u_n(t)) + R^{-1}\tilde{N}(u_n(t)) - R^{-1}f(t), P_n \mathbf{w} \right)_V = 0
\]

for all \( \mathbf{w} \in V \). This is to say that

\[
R^{-1}u'_n(t) + P_nR^{-1}A(u_n(t)) + P_nR^{-1}\tilde{N}(u_n(t)) = P_nR^{-1}f(t).
\]

Now the projection map decreases norms and \( R^{-1} \) preserves norms. Hence

\[
\| u'_n(t) \|_V = \| R^{-1}u'_n(t) \|_V \leq \| A(u_n(t)) \|_V + \| \tilde{N}(u_n(t)) \|_V + \| f(t) \|_V,
\]

from which it follows that \( u'_n \) is bounded in \( L^2([0,T];V') \). Indeed, this is the case because \( A(u_n) \) and \( \tilde{N}(u_n) \) are both bounded in \( L^2([0,T];V') \). The term \( \| \tilde{N}(u_n(t)) \|_V \) can be split further into terms involving \( \| N(u_n(t)) \|_V \), \( \| B(u_n, q) \|_V \), and \( \| B(q, u_n) \|_V \). For example, consider \( N(u_n) \) which is the least obvious. Let \( \mathbf{w} \in L^2([0,T];V) \). From the definitions,

\[
\langle N(u_n), \mathbf{w} \rangle_{L^2([0,T];V)} = \int_0^T \int_U u_{mi}u_{nj}w_{lj} \, dx \, dt \leq C \int_0^T \| \mathbf{w}(t) \|_V |u_n|^2_H dt \leq C \| \mathbf{w} \|_{L^2([0,T];V)} C(u_0, f, q, T, \omega).
\]
We have now shown that
\[
\sup_{t \in [0,T]} u_n(t, \omega)^2 H + \int_0^T \|u_n(s, \omega)\|^2 H ds + \|u'_n(\omega)\|_{L^2([0,T];V')} \leq C(u_0, f, q, T, \omega). \tag{22.4.8}
\]
This condition holds for all \(\omega\). Now for each \(\omega\), one can take a subsequence such that a solution to the evolution equation is obtained. Then, when this is done, we will apply the measurable selection result to obtain a product measurable solution.

It follows from the above estimate that there is a subsequence, still denoted as \(n\) and a function \(u(t, \omega)\) such that
\[
\begin{align*}
\text{for each } & n, \\
\text{we have } & u_n \to u \text{ weak * in } L^\infty([0,T]; H), \\
u'_n & \to u' \text{ weakly in } L^2([0,T]; V'), \\
u_n & \to u \text{ weakly in } L^2([0,T]; V), \\
u_n & \to u \text{ strongly in } L^2([0,T]; H).
\end{align*}
\tag{22.4.10}
\]
This last convergence follows from Theorem \ref{thm:22.4.1}. The sequence is bounded in \(L^2([0,T]; W)\) and the derivative is bounded in \(L^2([0,T]; V')\) so such a strongly convergent subsequence exists. Since \(A\) is linear, we can also assume that
\[
Au_n \to Au \text{ weakly in } L^2([0,T]; W').
\tag{22.4.11}
\]
What happens with the nonlinear operator \(\hat{N}\)? Let \(w \in L^\infty([0,T]; V)\). A computation shows then that
\[
\left| \int_0^T \langle \hat{N}u_n(t) - Nu(t), w(t) \rangle dt \right| = \left| \int_0^T \int_U (u_{ni}(t)u_{nj}(t) - u_{ij}(t)) w_{ji} dx dt \right|
\leq \|w\|_{L^\infty([0,T], V)} \int_0^T \int_U \left( |u_n(t)| + |u(t)| \right) \left( |u(t) - u_n(t)| \right) dx dt
\leq \|w\|_{L^\infty([0,T], V)} \left( \int_0^T \int_U |u_n| + |u| dx dt \right)^{1/2} \left( \int_0^T \int_U |u - u_n| dx dt \right)^{1/2}.
\]
This converges to 0 thanks to the estimates and the strong convergence. Similar convergence holds for the other nonlinear terms \(B(u_n(t), q), B(q, u_n(t))\).

We have shown that for any \(n \geq m\), and \(w \in V_m\),
\[
\langle u'_n(t), w \rangle_{V', V} + \langle A(u_n(t)), w \rangle_{V', V} + \langle \hat{N}(u_n(t)), w \rangle = \langle f(t), w \rangle.
\tag{22.4.12}
\]
Let \(\zeta \in C^\infty([0,T])\) be such that \(\zeta(T) = 0\). Then
\[
\langle u'_n(t), w \zeta(t) \rangle_{V', V} + \langle A(u_n(t)), w \zeta(t) \rangle_{V', V} + \langle \hat{N}(u_n(t)), w \zeta(t) \rangle = \langle f(t), w \zeta(t) \rangle.
\]
Integrating this equation from 0 to \(T\) we obtain
\[
- \langle u_0, w \zeta(0) \rangle_H - \int_0^T \zeta'(s) \langle u_n(s, \omega), w \rangle_H ds = - \int_0^T \langle A(u_n(s, \omega)), w \zeta(s) \rangle ds - \int_0^T \langle \hat{N}(u_n(s, \omega)), w \zeta(s) \rangle ds + \int_0^T \langle f(s, \omega), w \zeta(s) \rangle ds.
\]
Now letting \(n \to \infty\), from the above list of convergent sequences,
\[
- \langle u_0, w \zeta(0) \rangle_H - \int_0^T \zeta'(s) \langle u(s, \omega), w \rangle_H ds = - \int_0^T \langle A(u(s, \omega)), w \zeta(s) \rangle ds - \int_0^T \langle \hat{N}(u(s, \omega)), w \zeta(s) \rangle ds + \int_0^T \langle f(s), w \zeta(s) \rangle ds.
\]
It follows that in the sense of \(V'\) valued distributions,
\[
u'(\omega) + A(u(\omega)) + \hat{N}(u(\omega)) = f(\omega)
\tag{22.4.13}
\]
along with the initial condition
\[
u(0) = u_0.
\tag{22.4.14}
\]
This has proved most of the following lemma:
Lemma 22.4.4 Let \( u_0 \) have values in \( H \) and be \( F \) measurable, and let \( u_n \) be a solution to \( 22.4.11 \). Then for each \( \omega \), the estimate \( 22.4.13 \) holds. Also there is a subsequence, still called \( u_n \) such that the convergence for \( 22.4.11 - 22.4.14 \) are valid. For all \( \omega \), the function \( u(\cdot, \omega) \) is a solution to \( 22.4.11 - 22.4.14 \) and satisfies
\[
\psi(t, \omega) \in L^\infty([0, T]; H) \cap L^2([0, T]; W), u' \in L^2([0, T]; V').
\]
This solution is also weakly continuous into \( H \) for each \( \omega \).

Proof: All that remains to show is the last claim about weak continuity into \( H \). The equation \( 22.4.11 \) shows that \( u(\cdot, \omega) \) is continuous into \( V' \). However, the weak convergence and the estimate \( 22.4.13 \) show that \( u(\cdot, \omega) \) is bounded in \( H \). It follows from density of \( V \) in \( H \) that \( t \to u(t, \omega) \) is weakly continuous into \( H \).

From \( 22.4.11 - 22.4.14 \) the following integral equation for a path solution holds:
\[
u(t, \omega) - u_0(\omega) + \int_0^t A(u(s, \omega)) ds + \int_0^t N(u(s, \omega)) ds = \int_0^t f(s, \omega) ds.
\]

We apply Theorem \( 22.4.6 \) to prove the above solution could be taken product measurable.

Theorem 22.4.5 Let \( f(t, \omega), q(t, \omega) \) be product measurable and \( u_0 \) be measurable, such that for each \( \omega \in \Omega \), \( f(\cdot, \omega) \in L^2([0, T]; W') \), \( q(\cdot, \omega) \in C([0, T]; V) \) with \( q(0) = 0 \), and \( u_0(\omega) \in H \). Then there exists a global solution to the integral equation
\[
u(t, \omega) - u_0(\omega) + \int_0^t A(u(s, \omega)) ds + \int_0^t N(u(s, \omega)) ds = \int_0^t f(s, \omega) ds.
\]

Proof: Letting \( u_n \) be a solution to \( 22.4.11 - 22.4.14 \), we verify the conditions of Theorem \( 22.4.6 \) for \( u_n \).

The assumption in this theorem that the \( u_n \) are bounded follows from the above estimate \( 22.4.13 \). Then it was shown in the above lemma that whenever a sequence satisfies the estimate \( 22.4.13 \), it has a subsequence which converges as in \( 22.4.11 - 22.4.14 \) to a weakly continuous \( u(\cdot, \omega) \). Therefore, by Theorem \( 22.4.6 \) there is a subsequence \( u_{n(\omega)}(\cdot, \omega) \) converging weakly to \( u(\cdot, \omega) \), such that \( (t, \omega) \to u(t, \omega) \) is a product measurable function into \( H \). Then a further subsequence converges to a path solution to the above integral equation, which must be the same function because when a sequence converges, all subsequence converge to the same thing. In addition to this, \( u \) is also product measurable into \( W \). This follows from the above estimate \( 22.4.13 \). For \( \phi \in H, (t, \omega) \to (\phi, u(t, \omega)) \) is product measurable. However, \( H \) is dense in \( W' \) and so \( \psi \in W' \), there is a sequence \( \{\phi_n\} \in H \) such that \( \phi_n \to \psi \). Then
\[
\langle \psi, u \rangle = \lim_{n \to \infty} \langle \phi_n, u \rangle,
\]
so by the Pettis theorem \( 22.1.1 \), \( u \) is product measurable into \( W \) also.

This shows much of the following theorem which is the main result.

Theorem 22.4.6 Let \( f(t, \omega), q(t, \omega) \) be product measurable and \( u_0 \) be measurable, such that for each \( \omega \in \Omega \), \( f(\cdot, \omega) \in L^2([0, T]; W') \), \( q(\cdot, \omega) \in C([0, T]; V) \) with \( q(0) = 0 \), and \( u_0(\omega) \in H \). Then there exists a global solution to the integral equation
\[
u(t, \omega) - u_0(\omega) + \int_0^t A(u(s, \omega)) ds + \int_0^t N(u(s, \omega)) ds = \int_0^t f(s, \omega) ds + q(t, \omega).
\]

In addition to this, \( t \to u(t, \omega) \) is continuous into \( H \) and satisfies
\[
u(\cdot, \omega) \in L^\infty([0, T]; H) \cap L^2([0, T]; W).
\]

If, in addition to the above, \( u_0 \in L^2(\Omega; H) \) and \( f \in L^2([0, T] \times \Omega; W') \) and \( q \in L^2([0, T] \times \Omega; V) \), then the solution \( u \) is in \( L^2([0, T] \times \Omega; H) \cap L^2([0, T] \times \Omega; W) \).

Proof: The last claim follows from the estimates used in the Galerkin method, taking expectations and passing to a limit. To verify the continuity into \( H \), one can observe that from the integral equation, \( u \) is continuous into \( V' \). One has
\[
\|u(t)\|_H^2 = \sum_{k=1}^\infty (\langle u(t), w_k \rangle)^2 = \sum_{k=1}^\infty (\langle u(t), w_k \rangle)^2,
\]
and so \( t \to \|u(t)\|_H \) is lower semi-continuous. Since it is in \( L^\infty \), this implies this function is bounded. Hence the continuity into \( V' \) and density of \( V \) in \( H \) implies that \( u(t) \) is weakly continuous into \( H \). Then one can use the formulation in Theorem \( 22.4.6 \) to verify \( t \to \|u(t)\|_H \) is continuous and apply uniform convexity of the Hilbert space \( H \).

One can replace \( q(t, \omega) \) with \( q(t, \omega, u) \) and \( f(t, \omega, u) \) in the above with no change in the argument, provided it is assumed that \( (t, \omega, u) \to (q(t, \omega, u), f(t, \omega, u)) \) are product measurable, continuous in \( (t, u) \) and bounded.
22.5 A Friction contact problem

In this section we will consider a friction contact problem which has a coefficient of friction which is dependent on the slip speed.

\[
\ddot{u}_i = \sigma_{ij,j}(u, \dot{u}) + f_i \quad \text{for} \quad (t, x) \in (0, T) \times U, \tag{22.5.15}
\]

\[
u(0, x) = u_0(x), \tag{22.5.16}
\]

\[
\dot{u}(0, x) = v_0(x), \tag{22.5.17}
\]

where \( U \) is a bounded open subset of \( \mathbb{R}^3 \) having Lipschitz boundary, along with some boundary conditions which pertain to a part of the boundary of \( U, \Gamma_C \). For \( x \in \Gamma_C \),

\[
\sigma_n = -p((u_n - g)_+) C_n, \tag{22.5.18}
\]

This is the normal compliance boundary condition.

\[
|\sigma_T| \leq F((u_n - g)_+) \mu \left( |\dot{u}_T - \dot{U}_T| \right), \tag{22.5.19}
\]

\[
|\sigma_T| < F((u_n - g)_+) \mu \left( |\dot{u}_T - \dot{U}_T| \right) \quad \text{implies} \quad \dot{u}_T - \dot{U}_T = 0, \tag{22.5.20}
\]

\[
|\sigma_T| = F((u_n - g)_+) \mu \left( |\dot{u}_T - \dot{U}_T| \right) \quad \text{implies} \quad \dot{u}_T - \dot{U}_T = -\lambda \sigma_T(u, \dot{u}). \tag{22.5.21}
\]

Here \( C_n \) is a positive function in \( L^\infty(\Gamma_C) \), which we take equal to 1 to simplify notation, \( \dot{U}_T \) is the velocity of the foundation, \( \lambda \) is non negative, and \( \mu \) is a bounded positive function having a bounded continuous derivative. We could also let \( \mu \) depend on \( x \in \Gamma_C \) to model the roughness of the contact surface but we will suppress this dependence in the interest of simpler notation. Also, \( n \) is the unit outward normal to \( \partial U \) and \( u_n, u_T, \sigma_T, \) and \( \sigma_n \) are defined by the following.

\[
\begin{align*}
u_n &= u \cdot n \\
u_T &= u - (u \cdot n)n \\
\sigma_n &= \sigma_{ij} n_j n_i \\
\sigma_T &= \sigma_{ij} n_j - \sigma_n n_i,
\end{align*}
\]

written more simply,

\[
\sigma_T = \sigma n - \sigma_n n
\]

Systems like the above model dynamic friction contact problems. The function \( g \) represents the gap between the contact surface of \( U, \Gamma_C \), and a foundation which is sliding tangent to \( \Gamma_C \) with tangential velocity \( \dot{U}_T \).

The new ingredient in this paper is that we allow

\[
g = g(t, x, \omega)
\]

where \( \omega \in (\Omega, \mathcal{F}) \) and we assume \( (t, x, \omega) \rightarrow g(x, \omega) \) is \( \mathcal{B}([0, T] \times \Gamma_C) \times \mathcal{F} \) measurable. Also, we make the reasonable assumption that

\[
0 \leq g(t, x, \omega) \leq l < \infty
\]

for all \( (t, x, \omega) \). We also assume that the given motion of the foundation \( \dot{U}_T \) is a stochastic process

\[
\dot{U}_T = \dot{U}_T(t, x, \omega)
\]

and is \( \mathcal{B}([0, T] \times \Gamma_C) \times \mathcal{F} \) measurable. Here \( \mathcal{B}([0, T] \times \Gamma_C) \) denotes the Borel sets of \( [0, T] \times \Gamma_C \). We make the reasonable assumption that \( \dot{U}_T(t, x, \omega) \) is uniformly bounded. In the interest of notation, we will often suppress the dependence on \( t, x, \) and \( \omega \).

The condition \( \mathcal{C}_{\text{in}} \) is the contact condition. It says the normal component of the traction force density is dependent on the normal penetration of the body into the foundation surface. Conditions \( \mathcal{C}_{\text{fric}} \) model friction. They say that the tangential part of the traction force density is bounded by a function determined by the normal force or penetration. No sliding takes place until \( |\sigma_T| \) reaches this bound, \( F((u_n - g)_+) \mu(0) \). When this occurs, the tangential force density has a direction opposite the relative tangential velocity. The dependence of the friction coefficient on the magnitude of the slip velocity, \( |\dot{u}_T - \dot{U}_T| \), may be experimentally verified.
and so it has been included. The new feature in this model is the assumption that the gap is a random variable for each \( x \in \Gamma_C \) and we want to consider measurability of the solutions. Thus for a fixed \( \omega \), we have a standard friction problem and it is the measurability which is of interest here.

In this paper, we assume the following on \( p \) and \( F \). The functions \( p \) and \( F \) are increasing and

\[
\delta^2 r - K \leq p(r) \leq K (1 + r), \quad r \geq 0, \tag{22.5.22}
\]

\[
p(r) = 0, \quad r < 0,
\]

\[
F(r) \leq K (1 + r), \quad r \geq 0, \tag{22.5.23}
\]

\[
F(r) = 0 \text{ if } r < 0,
\]

\[
|\mu (r_1) - \mu (r_2)| \leq \text{Lip} (\mu) |r_1 - r_2|, \quad ||\mu||_\infty \leq C, \tag{22.5.24}
\]

and for \( a = F,p,\text{and } r_1, r_2 \geq 0, \)

\[
|a(r_1) - a(r_2)| \leq K |r_1 - r_2|. \tag{22.5.25}
\]

One can consider more general growth conditions than this, but we are keeping this part simple to emphasize the new stochastic considerations.

It will be assumed that

\[
\sigma_{ij} = A_{ijkl} u_{k,l} + C_{ijkl} u_{k,l}, \tag{22.5.26}
\]

where \( A \) and \( C \) are in \( L^\infty (U) \) and for \( B = A \) or \( C \), we have the following symmetries.

\[
B_{ijkl} = B_{ijlk}, \quad B_{ijkl} = B_{ijkl}, \quad B_{ijkl} = B_{klij}, \tag{22.5.27}
\]

and we also assume for \( B = A \) or \( C \) that

\[
B_{ijkl} H_{ij} H_{kl} \geq \varepsilon H_{rs} H_{rs} \tag{22.5.28}
\]

for all symmetric \( H \).

Throughout the paper, \( V \) will be a closed subspace of \( (H^1(U))^3 \) containing the test functions \((C_0^\infty (U))^3\), \( \rightharpoonup \) will denote weak or weak * convergence while \( \rightarrow \) will mean strong convergence. \( \gamma \) will denote the trace map from \( W^{12} (U) \) into \( L^2 (\partial U) \). \( H \) will denote \( (L^2 (U))^3 \) and we will always identify \( H \) and \( H' \) to write

\[
V \subseteq H = H' \subseteq V'.
\]

We define

\[
\mathcal{V} = L^2 (0,T;V), \quad \mathcal{H} = L^2 (0,T;H), \quad \mathcal{V}' = L^2 (0,T;V')
\]

### 22.5.1 The Abstract Problem

We shall use two theorems found in Lions [21], and Simon [24] respectively. These theorems apply for fixed \( \omega \). Proofs of generalizations of these theorems begin on Page 541.

**Theorem 22.5.1** If \( p \geq 1, \quad q > 1 \), and \( W \subseteq U \subseteq Y \) where the inclusion map of \( W \) into \( U \) is compact and the inclusion map of \( U \) into \( Y \) is continuous, let

\[
S = \{ u \in L^p (0,T;W) : u' \in L^q (0,T;Y) \text{ and } \}

\[
||u||_{L^p (0,T;W)} + ||u'||_{L^q (0,T;Y)} < R \}
\]

Then \( S \) is pre compact in \( L^p (0,T;U) \).

**Theorem 22.5.2** Let \( W, U, \) and \( Y \) be as in Theorem 22.5.1 and let

\[
S = \{ u : ||u(t)||_W + ||u'||_{L^q (0,T;Y)} \leq R \text{ for } t \in [0,T] \}
\]

for \( q > 1 \). Then \( S \) is pre compact in \( C(0,T;U) \).
Now we give an abstract formulation of the problem described roughly in 22.5.15 - 22.5.21. We begin by defining several operators. Let $M, A : V \to V'$ be given by

$$\langle Mu, v \rangle = \int_U C_{ijkl} u_k l v_{i,j} dx, \quad (22.5.29)$$

$$\langle Au, v \rangle = \int_U A_{ijkl} u_k l v_{i,j} dx. \quad (22.5.30)$$

Also let the operator $v \mapsto P(u)$ map $V$ to $V'$ be given by

$$\langle P(u), w \rangle = \int_T \int_{\Gamma_C} C_{ijkl} u_k l ((u_n - g)_+) w_n d\alpha dt, \quad (22.5.31)$$

where $u(t) = u_0 + \int_0^t v(s) ds \quad (22.5.32)$

for $u_0 \in V_q$. (Technically, $P$ depends on $u_0$ but we suppress this in favor of simpler notation). Let $\gamma^*_T : L^2(0, T; L^2(\Gamma_C))^3 \to \mathcal{V}$ is defined as

$$\langle \gamma^*_T \xi, w \rangle \equiv \int_T \int_{\Gamma_C} \xi \cdot w_T d\alpha dt.$$

Now the abstract form of the problem, denoted by $P$, is the following.

$$v' + Mv + Au + Pu + \gamma^*_T \xi = f \quad \text{in} \quad V',$$

$$v(0) = v_0 \in H, \quad (22.5.33)$$

where

$$u(t) = u_0 + \int_0^t v(s) ds, \quad u_0 \in V_q, \quad (22.5.34)$$

and for all $w \in \mathcal{V}$,

$$\langle \gamma^*_T \xi, w \rangle \leq \int_T \int_{\Gamma_C} F((u_n - g)_+) \mu \left| \left| v_T - \dot{U}_T \right| \right. \left| v_T - \dot{U}_T + w_T \right. - \left| v_T - \dot{U}_T \right| \cdot$$

$$\left| \left| v_T - \dot{U}_T + w_T \right. - \left| v_T - \dot{U}_T \right| \right. d\alpha dt. \quad (22.5.36)$$

Also $f \in L^2(0, T; V')$ so $f$ can include the body force as well as traction forces on various parts of $\partial U$. If $v$ solves the above abstract problem, then $u$ can be considered a weak solution to 22.5.15 - 22.5.21 along with other variational and stable boundary conditions depending on the choice of $W$ and $f \in L^2(0, T; V')$.

In order to carry out our existence and uniqueness proofs, we assume $M$ and $A$ satisfy the following for some $\delta > 0, \lambda \geq 0$.

$$\langle Bu, u \rangle \geq \delta \| u \|^2_W - \lambda \| u \|^2_H, \quad \langle Bu, u \rangle \geq 0, \quad \langle Bu, v \rangle = \langle Bv, u \rangle, \quad (22.5.37)$$

for $B = M$ or $A$. This is the assumption that we use, and we note that 22.5.37 is a consequence of 22.5.24 and Korn’s inequality [71].

#### 22.5.2 An Approximate Problem

We will use the Galerkin method. To do this, we will first regularize that subgradient material. Let

$$\psi_\varepsilon (r) = \sqrt{|r|^2 + \varepsilon}$$

Then this is a convex, Lipschitz continuous function having bounded derivative which converges uniformly to $\psi (r) = |r|$ on $\mathbb{R}$. Also

$$|\psi_\varepsilon (x) - \psi_\varepsilon (y)| \leq |x - y|, \quad |\psi'_\varepsilon (t)| \leq 1$$
And finally, \( \psi' \) is Lipschitz continuous with a Lipschitz constant \( C/\sqrt{\epsilon} \). Here \( \psi' \) denotes the gradient or Frechet derivative of the scalar valued function.

Our approximate problem for which we will apply the Galerkin method will be \( \mathcal{P}_\epsilon \) given by

\[
\dot{v}' + M v + A u + P u + \gamma_T^* F \left( (u_n - g)_{+} \right) \mu \left( \left| v_T - \dot{U}_T \right| \right) \psi' \left( v_T - \dot{U}_T \right) = f \text{ in } V',
\]

\[
\dot{v}(0) = v_0 \in H,
\]

where

\[
v(t) = u_0 + \int_0^t v(s) ds, \quad u_0 \in V,
\]

Here the long operator on the left is defined in the following manner.

\[
\langle \gamma_T^* F \left( (u_n - g)_{+} \right) \mu \left( \left| v_T - \dot{U}_T \right| \right) \psi' \left( v_T - \dot{U}_T \right), w \rangle = \int_{\Gamma_C} F \left( (u_n - g)_{+} \right) \mu \left( \left| v_T - \dot{U}_T \right| \right) \psi' \left( v_T - \dot{U}_T \right) \cdot w_T ds.
\]

Let \( R \) denote the Riesz map from \( V \) to \( V' \) defined by \( \langle Ru, v \rangle = (u, v)_{V} \). Then \( R^{-1} : H \to V \) is a compact self-adjoint operator and so there exists a complete orthonormal basis for \( H \), \( \{e_k\} \subseteq V \) such that

\[
Re_k = \lambda_k e_k.
\]

where \( \lambda_k \to \infty \). Let \( V_n = \text{span} \{e_1, \cdots, e_n\} \). Thus \( \cup_n V_n \) is dense in \( H \). In addition \( \cup_n V_n \) is dense in \( V \) and \( \{e_k\} \) is also orthogonal in \( V \). To see first that \( \{e_k\} \) is orthogonal in \( V \),

\[
0 = (e_k, e_l)_H = \frac{1}{\lambda_k} \langle Re_k, e_l \rangle_H = \frac{1}{\lambda_k} \langle Re_k, e_l \rangle_V = \frac{1}{\lambda_k} (e_l, e_k)_V.
\]

Next consider why \( \cup_n V_n \) is dense in \( V \). If this is not so, then there exists \( f \in V' \), \( f \neq 0 \) such that \( \cup_n V_n \) is in \( \ker(f) \). But \( f = Ru \) and so

\[
0 = \langle Ru, e_k \rangle = (Re_k, u) = \lambda_k (e_k, u)_H
\]

for all \( e_k \) and so \( u = 0 \) by density of \( \cup_n V_n \) in \( H \). Hence \( Ru = 0 = f \) after all, a contradiction. Hence \( \cup_n V_n \) is dense in \( V \) as claimed.

Now we set up the Galerkin method for Problem \( \mathcal{P}_\epsilon \). Let

\[
v_k(t, \omega) = \sum_{j=1}^k x_j(t, \omega) e_j, \quad u_k(t) = u_0 + \int_0^t v_k(s) ds
\]

and let \( v_k \) be the solution to the following integral equation for each \( \omega \) and \( j \leq k \). The dependence on \( \omega \) is suppressed in most terms in order to save space.

\[
\left\langle v_k(t) - v_{0k} + \int_0^t M v_k + A u_k + P u_k + \gamma_T^* F \left( (u_{kn} - g(\omega))_{+} \right) \cdot \mu \left( \left| v_{kt} - \dot{U}_{kt} \right| \right) \psi' \left( v_{kt} - \dot{U}_{kt} \right), e_j \right\rangle = \int_0^t \langle f, e_j \rangle ds
\]

Here \( v_{0k} \to v_0 \in H \) and the equation holds for each \( e_j \) for each \( j \leq k \). Then this integral equation reduces to a system of ordinary differential equations for the vector \( x(t, \omega) \) whose \( j^{th} \) component is \( x_j(t, \omega) \) mentioned above. Differentiate, multiply by \( x_j \) and add. Then integrate. This will yield some terms which need to be estimated. Here is the one which comes from the long term.

\[
\int_0^t \int_{\Gamma_C} F \left( (u_{kn} - g(\omega))_{+} \right) \mu \left( \left| v_{kt} - \dot{U}_{kt} \right| \right) \psi' \left( v_{kt} - \dot{U}_{kt} \right) \cdot v_{kt} ds ds
\]

\[
= \int_0^t \int_{\Gamma_C} F \left( (u_{kn} - g(\omega))_{+} \right) \mu \left( \left| v_{kt} - \dot{U}_{kt} \right| \right) \psi' \left( v_{kt} - \dot{U}_{kt} \right) \cdot \left( v_{kt} - \dot{U}_{kt} \right) ds ds
\]

\[
+ \int_0^t \int_{\Gamma_C} F \left( (u_{kn} - g(\omega))_{+} \right) \mu \left( \left| v_{kt} - \dot{U}_{kt} \right| \right) \psi' \left( v_{kt} - \dot{U}_{kt} \right) \cdot \dot{U}_{kt} ds ds
\]
Thus in \( V \) we have
\[
R^{-1}v_k'(t) + P_k R^{-1} \left( \frac{M v_k + A u_k + P u_k + \gamma_2 F \left( (u_{kn} - g(\omega))_+ \right) \cdot}{\mu \left( \left| v_k T - \bar{U}_t \right| \right) \psi'_\varepsilon (v_k T - \bar{U}_T)} , \right) \right) _V = P_k R^{-1} f
\]
and \( R^{-1} \) preserves norms while \( P_k \) decreases them. Hence the estimate \( 22.5.42 \) implies that \( \| v_k' \|_V \) is also bounded independent of \( \varepsilon, \omega \) and \( k \). Then summarizing this yields
\[
|v_k (t, \omega)|_H + \|v_k (\cdot, \omega)\|_V + \|v_k' (\cdot, \omega)\|_V + \|u_k (t, \omega)\|_V \leq C (\omega)
\] (22.5.44)
where $C$ is some constant which does not depend on $\varepsilon, \omega$, and $k$. Also, integrating, it follows that
\[
\int_0^t \left( v_k(t) - v_0 + \int_0^t Mv_k ds + \int_0^t Au_k ds + \int_0^t P u_k ds + \int_0^t \gamma_+^T F \left( \left( u_{kn} - g(\omega) \right)_+ \right) \mu \left( \left| v_k - \tilde{U}_T \right| \right) \psi_\varepsilon \left( v_k - \tilde{U}_T \right) ds \right) = \int_0^t f ds
\] (22.5.45)

Where $i_k^*$ is the dual map to the inclusion map $i_k : V_k \to V$.

Let
\[
V \subseteq W, \ V \text{ dense in } W,
\]
where the embedding is compact and the trace map onto the boundary of $U$ is continuous. Using Theorem 22.5.2 and 22.5.34, it follows that for a fixed $\omega$, there exist the following convergences valid for a suitable subsequence, still denoted as $\{v_k\}$ which may depend on $\omega$.

\[
v_k \to v \text{ in } V
\] (22.5.46)

\[
v_k' \to v' \text{ in } V'
\] (22.5.47)

\[
v_k \to v \text{ strongly in } C([0,T], W')
\] (22.5.48)

\[
v_k \to v \text{ strongly in } L^2([0, T]; W)
\] (22.5.49)

\[
v_k(t) \to v(t) \text{ in } W \text{ for a.e. } t
\] (22.5.50)

\[
u_k \to u \text{ strongly in } C([0, T]; W)
\] (22.5.51)

\[
Au_k \to Au \text{ in } V'
\] (22.5.52)

\[
M v_k \to M v \text{ in } V'
\] (22.5.53)

Now from these convergences and the density of $U \cap V_0$, it follows on passing to a limit and using dominated convergence theorem and the strong convergences above in the nonlinear terms, we obtain the following equation which holds in $V'$.

\[
v(t) - v_0 + \int_0^t M v ds + \int_0^t A u ds + \int_0^t P u ds + \int_0^t \gamma_+^T F \left( \left( u_{kn} - g(\omega) \right)_+ \right) \mu \left( \left| v - \tilde{U}_T \right| \right) \psi_\varepsilon \left( v - \tilde{U}_T \right) ds = \int_0^t f ds
\] (22.5.54)

Thus $t \to v(t, \omega)$ is continuous into $V'$. This along with the estimate 22.5.34 implies that the conditions of Theorem 22.5.4 are satisfied. It follows that there is a function $\tilde{v}$ which is product measurable into $V'$ and weakly continuous in $t$ and for each $\omega$, a subsequence $v_{k(\omega)}$ such that $v_{k(\omega)} \to \tilde{v} (\cdot, \omega)$ in $V'$. Then by a repeat of the above argument, for each $\omega$, there exists a further subsequence still denoted as $v_{k(\omega)}$ which converges in $V'$ to $v (\cdot, \omega)$ which is a solution to the above integral equation which is continuous into $V'$. Hence, $v (\cdot, \omega) = \tilde{v} (\cdot, \omega)$ and since these are both weakly continuous into $V'$ they must be the same function. Hence, there is a product measurable solution $v$.

Next we pass to a limit as $\varepsilon \to 0$. Denoting the product measurable solution to the above integral equation as $v_k$, where $\varepsilon = 1/k$. The estimate 22.5.34 is obtained as before. Then we get a subsequence, still denoted as $v_k$ which has the same convergences as in 22.5.5. Thus we obtain these convergences along with the fact that $v_k$ is product measurable and for each $\omega$, it is a solution of

\[
v_k(t) - v_0 + \int_0^t M v_k ds + \int_0^t A u_k ds + \int_0^t P u_k ds + \int_0^t \gamma_+^T F \left( \left( u_{kn} - g(\omega) \right)_+ \right) \mu \left( \left| v_k - \tilde{U}_T \right| \right) \psi_\varepsilon \left( v_k - \tilde{U}_T \right) ds = \int_0^t f ds
\] (22.5.55)

Now in addition to these convergences, we can also obtain

\[
\psi_\varepsilon \left( v_k - \tilde{U}_T \right) \to \xi \text{ in } L^\infty ([0, T]; L^\infty (\Gamma_C)^3)
\]

We have also

\[
\psi_\varepsilon \left( v_k - \tilde{U}_T \right) \cdot w_T \leq \psi_\varepsilon \left( v_k - \tilde{U}_T + w_T \right) - \psi_\varepsilon \left( v_k - \tilde{U}_T \right)
\]
CHAPTER 22. MEASURABILITY WITHOUT UNIQUENESS

and so, passing to a limit, using the strong convergence of \( v_{kT} \) to \( v_T \) in \( L^2([0,T];W) \), uniform convergence of \( \psi_{1/k} \) to \( \| \cdot \| \), and pointwise convergence in \( W \), we obtain using the dominated convergence theorem that for \( w \in V \),

\[
\int_0^t \int_{\Gamma_C} F \left( (u_n - g(\omega))_+ \right) \mu \left( |v_{kT} - \dot{U}_T| \right) \psi_{1/k} \left( v_{kT} - \dot{U}_T \right) \cdot w_T dxds
\]

\[
\rightarrow \int_0^t \int_{\Gamma_C} F \left( (u_n - g(\omega))_+ \right) \mu \left( |v_T - \dot{U}_T| \right) \xi \cdot w_T dxds
\]

where

\[
\int_0^t \int_{\Gamma_C} \xi \cdot w_T dxds \leq \int_0^t \int_{\Gamma_C} |v_{kT} - \dot{U}_T + w_T| - |v_{kT} - \dot{U}_T| d\alpha ds
\] (22.5.56)

Then passing to the limit in the integral equation \textit{22.5.55}, we obtain that \( v \) is a solution for each \( \omega \) to the integral equation

\[
v(t) - v_0 + \int_0^t Mvds + \int_0^t Au ds + \int_0^t Pu ds + \int_0^t \gamma^s_F \left( (u_n - g(\omega))_+ \right) \mu \left( |v_T - \dot{U}_T| \right) \xi ds = \int_0^t f ds
\] (22.5.57)

where \( \xi \) satisfies the inequality \textit{22.5.56}. In particular, \( v \) is continuous into \( V' \) and now, the conclusion of the measurable selection theorem applies and yields the existence of a measurable solution to the integral equation just displayed for each \( \omega \). Taking a weak derivative, it follows that we have obtained a measurable solution to the system \textit{22.5.33} - \textit{22.5.36}.

In this case of Lipschitz \( \mu \) one can show that the solution for each \( \omega \) to the above integral equation is unique although this it is not an obvious theorem. This follows standard procedures involving Gronwall’s inequality and estimates. Therefore, it is possible to obtain the measurability using more elementary methods. In addition, it becomes possible to include a stochastic integral of the form \( \int_0^t \Phi dW \). In this case one must consider a filtration and obtain solutions which are adapted to the filtration. In the next section we consider the case of discontinuous friction coefficient and in this case it is not clear whether there is uniqueness but we have still obtained a measurable solution.

\subsection{22.5.3 Discontinuous coefficient of friction}

In this section we consider the case where the coefficient of friction is a discontinuous function of the slip speed. This is the case described in elementary physics courses which state that the coefficient of sliding friction is less than the coefficient of static friction. Specifically, we assume the function \( \mu \), has a jump discontinuity at 0, becoming smaller when the speed is positive.

![Fig. 2. The graph of \( \mu \) vs. the slip rate \(|v_*|\), and \( \nu \).](image_url)

We assume the function \( \mu_* \) of the picture is Lipschitz continuous and decreasing just as shown. The new function \( \nu \) is extended for \( r < 0 \) as shown and is just \( \mu_* (r) + \eta \) for \( r > 0 \).

Let

\[
h_{\varepsilon} (r) \equiv (\eta^2 r^2 + \varepsilon)^{1/2}
\]

\[
\mu_{\varepsilon} (r) = \nu (r) - h'_{\varepsilon} (r)
\]
Thus \( \mu_e \) is bounded, Lipschitz continuous and as \( \varepsilon \to 0, \mu_e (r) \to \mu (r) \) for \( r > 0 \). Thus, for each \( \varepsilon = 1/k \), there exists a measurable solution to the integral equation

\[
v_k (t) - v_0 + \int_0^t Mv_k ds + \int_0^t Au_k ds + \int_0^t Pu_k ds + \int_0^t \gamma^*_T \left((u_{kn} - g(\omega))_+ \right) \mu_{1/k} (\|v_{kT} - \tilde{U}_T\|) \xi_k ds = \int_0^t f ds
\]

where

\[
\int_0^t \int_{\Gamma_C} \xi_k \cdot w_T d\alpha dt \leq \int_0^t \int_{\Gamma_C} \left|v_{kT} - \tilde{U}_T + w_T\right| - \left|v_{kT} - \tilde{U}_T\right| d\alpha ds
\]

Now for a given \( \omega \), the same estimate obtained earlier, \( 22.5.53 \) is available. Thus

\[
|v_k (t)|^2_H + \int_0^T \|v_k\|^2_V ds + \|u_k (t)\|^2_V \leq C
\]

where \( C \) is not dependent on \( k \). Recall also that \( \xi_k \) is bounded. Hence from \( 22.5.53 \), and this estimate, it also follows that \( v_k' \) is bounded in \( V' \). Thus

\[
|v_k (t)|^2_H + \int_0^T \|v_k\|^2_V ds + \|u_k (t)\|^2_V + \|v_k'\|_{V'} \leq C
\]

As earlier, we can take \( C \) independent of \( k \) and \( \omega \) although we do not need this constant to be independent of \( \omega \). Now for fixed \( \omega \), there exists a subsequence, still denoted as \( \{v_k\} \) such that the convergences obtained earlier all hold, that is \( \eta_{22.5.42, 58} \). Taking a further subsequence, we may assume also that

\[
\psi - h'_{1/k} \left(\left|v_{kT} - \tilde{U}_T\right|\right) \to 0 \text{ in } L^\infty ([0, T], L^\infty (\Gamma_C)),
\]

\[
\xi_k \rightharpoonup \xi \text{ weak* in } L^\infty ([0, T], L^\infty (\Gamma_C))^3.
\]

That is, \( h'_{1/k} \left(\left|v_{(1/k)T} - \tilde{U}_T\right|\right) \) converges weak* in \( L^\infty ([0, T], L^\infty (\Gamma_C)) \) to some \( \psi \). This is because

\[
h'_{e} (r) = \frac{\eta^2 \sqrt{r^2 + \eta^2}}{r^2 + \eta^2 + \varepsilon}
\]

and this is bounded. Letting \( w \in L^1 ([0, T]; L^1 (\Gamma_C)) \),

\[
\int_0^T \int_{\Gamma_C} h'_{1/k} \left(\left|v_{kT} - \tilde{U}_T\right|\right) w d\alpha dt \leq \int_0^T \int_{\Gamma_C} h_{1/k} \left(\left|v_{kT} - \tilde{U}_T\right| + w\right) - h_{1/k} \left(\left|v_{kT} - \tilde{U}_T\right|\right) d\alpha dt
\]

Thanks to the strong convergences and the uniform convergence of \( h_{1/k} (r) \) to \(|\eta r|\),

\[
\int_0^T \int_{\Gamma_C} \psi w d\alpha dt \leq \int_0^T \int_{\Gamma_C} |\eta \left(\left|v_{T} - \tilde{U}_T\right| + w\right) - \eta \left|v_{T} - \tilde{U}_T\right|| d\alpha dt
\]

Therefore, for \( a.e.t. \), \( \psi (t, x, \omega) \) is in the subgradient of the function \( \phi_\eta (r) = |\eta r| \) for \( a.e.x \in \Gamma_C \) at the point \( r = \left|v_{kT} - \tilde{U}_T\right| \). In particular, \( \psi \in [-\eta, \eta] \) so that \( \nu \left(\left|v_{kT} - \tilde{U}_T\right|\right) - \psi \) is between \( \mu_\varepsilon (0) \) and \( \mu_0 \) if \( \left|v_{kT} - \tilde{U}_T\right| = 0 \). If this quantity is positive, then \( \psi = \eta \) and \( \nu \left(\left|v_{kT} - \tilde{U}_T\right|\right) - \psi \) reduces to \( \mu_\varepsilon \left(\left|v_{kT} - \tilde{U}_T\right|\right) \). Thus

\[
\left(\left|v_{kT} - \tilde{U}_T\right|, \nu \left(\left|v_{kT} - \tilde{U}_T\right|\right) - \psi \right)
\]

is in the graph of \( \mu \ a.e. \). Similar reasoning based on strong convergence and \( \eta_{22.5.42} \) implies that for \( a.e.t. \), \( \xi \in \partial \gamma \) where \( \gamma (y) = |y| \) at the point \( v_{kT} - \tilde{U}_T \) for \( a.e.x \in \Gamma_C \).
Consider the friction terms in (22.5.58). Letting \( w \in \mathcal{V} \) and recalling that \( \mu_{(1/k)}(r) = \nu(r) - h'_{(1/k)}(r) \),

\[
\int_0^T \int_{\Gamma_C} F \left( (u_{kn} - g)^+ \right) \mu_{(1/k)} \left( |v_{kT} - \hat{U}_T| \right) \xi_k \cdot w_t \, dt \, ds
\]

\[
= \int_0^T \int_{\Gamma_C} F \left( (u_{kn} - g)^+ \right) \left( \nu \left( |v_{kT} - \hat{U}_T| \right) - h'_{(1/k)} \left( |v_{kT} - \hat{U}_T| \right) \right) \xi_k \cdot w_t \, dt \, ds
\]

\[
= \int_0^T \int_{\Gamma_C} F \left( (u_{kn} - g)^+ \right) \left( \nu \left( |v_{kT} - \hat{U}_T| \right) - \psi \right) \xi_k \cdot w_t \, dt \, ds
\]

\[
+ \int_0^T \int_{\Gamma_C} F \left( (u_{kn} - g)^+ \right) \left( \psi - h'_{(1/k)} \left( |v_{kT} - \hat{U}_T| \right) \right) \xi_k \cdot w_t \, dt \, ds
\]

(22.5.60)

Now consider the first integral. The strong convergence yields that this integral in (22.5.60) converges to

\[
\int_0^T \int_{\Gamma_C} F \left( (u_{n} - g)^+ \right) \left( \nu \left( |v_{T} - \hat{U}_T| \right) - \psi \right) \xi \cdot w_t \, dt \, ds
\]

where \( \nu \left( |v_{T} - \hat{U}_T| \right) - \psi \) is in the graph of \( \mu \) a.e.

Consider the second integral in (22.5.60).

\[
\int_0^T \int_{\Gamma_C} F \left( (u_{kn} - g)^+ \right) \left( \psi - h'_{(1/k)} \left( |v_{kT} - \hat{U}_T| \right) \right) \xi_k \cdot w_t \, dt \, ds
\]

\[
\leq \int_0^T \int_{\Gamma_C} F \left( (u_{kn} - g)^+ \right) \left( \psi - h'_{(1/k)} \left( |v_{kT} - \hat{U}_T| \right) \right) \left( |v_{kT} - \hat{U}_T + w_t| - |v_{kT} - \hat{U}_T| \right) \, dt \, ds
\]

Similarly,

\[
- \int_0^T \int_{\Gamma_C} F \left( (u_{kn} - g)^+ \right) \left( \psi - h'_{(1/k)} \left( |v_{kT} - \hat{U}_T| \right) \right) \xi_k \cdot w_t \, dt \, ds
\]

\[
\leq \int_0^T \int_{\Gamma_C} F \left( (u_{kn} - g)^+ \right) \left( \psi - h'_{(1/k)} \left( |v_{kT} - \hat{U}_T| \right) \right) \left( |v_{kT} - \hat{U}_T - w_t| - |v_{kT} - \hat{U}_T| \right) \, dt \, ds
\]

Each of these integrals on the right side converge to 0 because, from the strong convergence results,

\[
F \left( (u_{kn} - g)^+ \right) \left( |v_{kT} - \hat{U}_T| \pm w_t| - |v_{kT} - \hat{U}_T| \right)
\]

converges in \( L^1([0,T], L^1(\Gamma_C)) \) and so the weak * convergence to 0 of

\[
\psi - h'_{(1/k)} \left( |v_{kT} - \hat{U}_T| \right)
\]

implies that these integrals converge to 0. Thus the integral in (22.5.60)

\[
\int_0^T \int_{\Gamma_C} F \left( (u_{kn} - g)^+ \right) \left( \psi - h'_{(1/k)} \left( |v_{kT} - \hat{U}_T| \right) \right) \xi_k \cdot w_t \, dt \, ds
\]

is between two sequences each of which converges to 0 so it also converges to 0.

To save space, denote by

\[
\hat{\mu} = \nu \left( |v_{T} - \hat{U}_T| \right) - \psi
\]

Then passing to the limit in this subsequence, we obtain for fixed \( \omega \) the existence of a solution to the following integral equation.

\[
v(t) - v_0 + \int_0^t Mvds + \int_0^t Auds + \int_0^t Puds + \int_0^t \gamma^*_TF \left( (u_{n} - g)^+ \right) \hat{\mu} \xi ds = \int_0^t fds
\]

(22.5.62)

where

\[
u(t) = u(t) + \int_0^t v(s) \, ds
\]

(22.5.63)
and \( \left| v_k T - \hat{U}_T \right| \), \( \mu \) is contained in the graph of \( \mu \) a.e. Also for each \( w \in V \),
\[
\int_0^T \int_{\Gamma_C} \xi \cdot w_T d\alpha ds \leq \int_0^T \int_{\Gamma_C} \left| v_k T - \hat{U}_T + w_T \right| - \left| v_k T - \hat{U}_T \right| d\alpha ds \tag{22.5.64}
\]

The remaining issue concerns the existence of a measurable solution. However, this follows in the same way as before from the measurable selection theorem, Theorem 22.2.1. From the above reasoning, for fixed \( \omega \) any sequence has a subsequence which leads to a solution to the integral equation 22.5.62 - 22.5.64 which is continuous into \( V' \). There is also an estimate of the right sort for all of the \( v_k \). Therefore, from this theorem, there is a function \( v (\cdot, \omega) \) in \( V' \) which is weakly continuous into \( V' \) and a sequence \( v_{k(\omega)} (\cdot, \omega) \) converging to \( v (\cdot, \omega) \). Then from the above argument, a subsequence converges to a solution to the integral equation and since both are weakly continuous into \( V' \), it follows that the solution to the integral equation equals this measurable function for all \( t \), this for each \( \omega \). Thus there is a measurable solution to the stochastic friction problem. The result is stated in the following theorem.

**Theorem 22.5.3** For each \( \omega \) let \( u_0 (\omega) \in V, v_0 (\omega) \in H \). Let \( f \in V' \). Also assume the gap \( g \) and sliding velocity \( \hat{U}_T \) are \( \mathcal{F} \) measurable. Then there exists a solution \( v \), to the problem summarized in 22.5.62 - 22.5.64 for each \( \omega \). This solution \( (t, \omega) \rightarrow v (t, \omega) \) is measurable into \( V', H' \) and \( V' \).

It only remains to check the last claim about measurability into the other spaces. By density of \( V \) into \( H \), it follows that \( H' \) is dense in \( V' \) and so a simple Pettis theorem argument implies right away that \( \omega \rightarrow v (t, \omega) \) is \( \mathcal{F} \) measurable into both \( V \) and \( H \).
Chapter 23

Stochastic O.D.E. One Space

23.1 Adapted Solutions With Uniqueness

Instead of a single \( \sigma \) algebra \( \mathcal{F} \), one can generalize to the case of a normal filtration \( \mathcal{F}_t \) and obtain adapted solutions to finite dimensional theorems, provided one also knows path uniqueness of the solutions. Recall that a filtration is normal includes the following condition which is what we will use.

\[
\mathcal{F}_t = \cap_{s > t} \mathcal{F}_s \tag{23.1.1}
\]

**Theorem 23.1.1** Suppose \( N(t, u, v, w, \omega) \in \mathbb{R}^d \) for \( u, v, w \in \mathbb{R}^d, t \in [0, T] \) and \( (t, u, v, w, \omega) \rightarrow N(t, u, v, w, \omega) \) is progressively measurable with respect to a normal filtration or more generally one which satisfies Theorem 23.1.1. Also suppose \( (t, u, v, w) \rightarrow N(t, u, v, w, \omega) \) is continuous. Suppose for each \( \omega \), there exists an estimate for any solution \( u(\cdot, \omega) \) to the integral equation

\[
u(t, \omega) - u_0(\omega) + \int_0^t N(s, u(s, \omega), u(s - h, \omega), w(s, \omega), \omega) ds = \int_0^t f(s, \omega) ds, \tag{23.1.2}
\]

which is of the form

\[
\sup_{t \in [0, T]} |u(t, \omega)| \leq C(\omega) < \infty
\]

Also let \( f \) be progressively measurable and \( f(\cdot, \omega) \in L^1([0, T]; \mathbb{R}^d) \). Here \( u_0 \) has values in \( \mathbb{R}^d \) and is \( \mathcal{F}_0 \) measurable and \( u(s - h, \omega) \equiv u_0(\omega) \) whenever \( s - h \leq 0 \) and

\[
w(t, \omega) \equiv w_0(\omega) + \int_0^t u(s, \omega) ds
\]

where \( w_0 \) is a given \( F_0 \) measurable function. Also assume that for each \( \omega \) there is at most one solution to the integral equation [23.1.2]. Then for \( h > 0 \), there exists a progressively measurable solution \( u \) to the integral equation [23.1.2].

**Proof:** Let \( 0 = t_0 < t_1 < \cdots < t_n = T \). From Theorem 22.3.3, there exists a solution to the integral equation \( u \) which has the property that \( u(t \wedge t_j) \) is \( \mathcal{F}_{t_j} \) measurable. One simply applies this theorem to the succession of intervals determined by the given partition. Now suppose \( \mathcal{P} \) consists of the points \( k2^{-n}T \equiv t^n_j \) so that these satisfy \( \mathcal{P} \subseteq \mathcal{P}^{n+1} \) and the lengths of the sub-intervals decreases to 0 with increasing \( n \). Let \( u_n \) denote the solution just described corresponding to \( \mathcal{P} \) such that \( u_n(t \wedge t^n_j) \) is \( \mathcal{F}_{t^n_j} \) measurable. As before, using the estimate, these \( u_n(\cdot, \omega) \) for a fixed \( \omega \) are uniformly bounded and equicontinuous. This is because it is a solution to the integral equation for each \( \omega \) and so by assumption, there is an estimate. Therefore, for fixed \( \omega \), there exists \( u(\cdot, \omega) \) and a subsequence, denoted as \( u_n(\cdot, \omega) \) which converges uniformly to \( u(\cdot, \omega) \) on \([0, T] \). Therefore, \( u(\cdot, \omega) \) will be a solution to the integral equation for that \( \omega \). It follows from the uniqueness assumption, that it is not necessary to take a subsequence. Thus

\[
u(t, \omega) = \lim_{n \to \infty} u_n(t, \omega)
\]

For \( t \in (t^n_{j-1}, t^n_j) \), it follows that \( \omega \to u(t, \omega) \) is \( \mathcal{F}_{t^n_j} \) measurable. Since this is true for each \( n \) and the filtration is assumed to be a normal filtration, we conclude that \( \omega \to u(t, \omega) \) is \( \mathcal{F}_t \) measurable.\( \blacksquare \)

Why can’t this be generalized to the situation where no uniqueness is known? We have been unable to do this. It appears that the difficulty is related to the need to use theorems about measurable selections and these theorems pertain to a single \( \sigma \) algebra. Attempts to use the \( \sigma \)-algebra of progressively measurable sets have not been successful either.
23.2 Including Stochastic Integrals

It is not surprising that Theorem 23.2.4 is sufficient to allow the inclusion of a stochastic integral. Thus, with the same descriptions of the symbols used in that theorem, one could consider the following integral equation.

\[ u(t, \omega) - u_0(\omega) + \int_0^t N(s, u(s, \omega), u(s-h, \omega), w(s, \omega), \omega) \, ds \]

\[ = \int_0^t f(s, \omega) \, ds + \int_0^t \Phi dW \]

where, as usual \( \Phi \in L^2 ([0,T] \times \Omega; L^2 (Q^{1/2}U, \mathbb{R}^d)) \) where \( U \) is a Hilbert space. It could be \( \mathbb{R}^d \) of course. To include a stochastic integral, you define a new variable.

\[ \hat{u}(t) = u(t) - \int_0^t \Phi dW \]

Then in terms of this new variable, the integral equation is

\[ \hat{u}(t) - u_0(\omega) + \int_0^t N(s, \hat{u}(s, \omega), \int_0^s \Phi dW, \hat{u}(s-h, \omega) + \int_0^{s-h} \Phi dW, \]

\[ \int_0^s \left( \hat{u}(r) + \int_0^r \Phi dW \right) \, dr, \omega) \, ds = \int_0^t f(s, \omega) \, ds \]

This is in the situation of Theorem 23.2.1 provided \( N \) is progressively measurable with respect to the normal filtration \( \mathcal{F}_t \) determined by the Wiener process and there exists an estimate of the sort in this theorem and for a given \( \omega \) there is at most one solution \( t \to \hat{u}(t, \omega) \) to the above integral equation.

**Theorem 23.2.1** Suppose \( N(t, u, v, w, \omega) \in \mathbb{R}^d \) for \( u, v, w \in \mathbb{R}^d, t \in [0,T] \) and \( (t, u, v, w, \omega) \to N(t, u, v, w, \omega) \) is progressively measurable with respect to the normal filtration \( \mathcal{F}_t \) determined by a given Wiener process \( W(t) \). Also suppose \( (t, u, v, w) \to N(t, u, v, w, \omega) \) is continuous and satisfies the following conditions for \( C(\cdot, \omega) \geq 0 \) in \( L^1 ([0,T]) \) and some \( \mu > 0 \):

\[ (N(t, u, v, w, \omega), u) \geq -C(t, \omega) - \mu \left( |u|^2 + |v|^2 + |w|^2 \right). \tag{23.2.3} \]

Also let \( f \) be progressively measurable and \( f(\cdot, \omega) \in L^2 ([0,T]; \mathbb{R}^d) \). Let \( \Phi \in L^2 ([0,T] \times \Omega; L^2 (Q^{1/2}U, \mathbb{R}^d)) \) where \( U \) is some Hilbert space, \( \mathbb{R}^d \), for example. Also suppose path uniqueness. That is, for each \( \omega \), there is at most one solution to the integral equation

\[ u(t, \omega) - u_0(\omega) + \int_0^t N(s, u(s, \omega), u(s-h, \omega), w(s, \omega), \omega) \, ds \]

\[ = \int_0^t f(s, \omega) \, ds + \int_0^t \Phi dW, \tag{23.2.4} \]

Then for \( h > 0 \), there exists a unique progressively measurable solution \( u \) to the integral equation [23.2.4] where \( u_0 \) has values in \( \mathbb{R}^d \) and is \( \mathcal{F}_0 \) measurable. Here \( u(s-h, \omega) \equiv u_0(\omega) \) for all \( s-h \leq 0 \) and for \( w_0 \) a given \( \mathcal{F}_0 \) measurable function,

\[ w(t, \omega) \equiv w_0(\omega) + \int_0^t u(s, \omega) \, ds \]

**Proof:** The only thing left is to observe that the given estimate is sufficient to obtain an estimate for the solutions to the integral equation for \( \hat{u} \) defined above. Then from Theorem 23.2.1 there exists a unique progressively measurable solution for \( \hat{u} \) and hence for \( u \).

Note that the integral equation holds for all \( t \) for each \( \omega \). There is no exceptional set of measure zero which might depend on the initial condition needed.

What is a sufficient condition for path uniqueness? Suppose the following weak monotonicity condition for \( \mu = \mu(\omega) \).

\[ (N(t, u_1, v_1, w_1, \omega) - N(t, u_2, v_2, w_2, \omega), u_1 - u_2) \]

\[ \geq -\mu \left( |u_1 - u_2|^2 + |v_1 - v_2|^2 + |w_1 - w_2|^2 \right) \tag{23.2.5} \]
Then path uniqueness will hold. This follows from subtracting the two integral equations, one for $u_1$ and one for $u_2$, using the estimate and then applying Gronwall’s inequality.

Recall the Ito formula
\[ u(t) - u_0 + \int_0^t N ds = \int_0^t f ds + \int_0^t \Phi dW \]
where $u(t) \in H$ a Hilbert space. Consider $F(u) = \frac{1}{2} |u|^2$. Also let $R$ denote the Riesz map from $H \to H'$ such that $(Rx, y) \equiv (x, y)_H$. Then proceeding formally, to see what the Ito formula says,
\[ dF = DF(u) du + \frac{1}{2} D^2 F(u) (du, du) + O(du^3) \]
Recall then that $du = -N dt + f dt + \Phi dW$ and so recalling $(dW, dW) = dt$,
\[ R(u) (-N dt + f dt + \Phi dW) + \frac{1}{2} \| \Phi \|^2 dt \]
Hence
\[ \frac{1}{2} |u(t)|^2_H - \frac{1}{2} |u_0|^2_H + \int_0^t (N, u)_H ds - \frac{1}{2} \int_0^t \| \Phi \|^2_{L^2} ds = \int_0^t (f, u) ds + \int_0^t R u (\Phi) dW \]
The last term is a martingale or local martingale $M$ whose quadratic variation is given by
\[ [M](t) = \int_0^t \| \Phi \|^2_{L^2} |u|^2 ds \]
This is all that is of importance in what follows. Therefore, this martingale may be simply denoted as $M(t)$ in what follows.

Under the assumption you can include instead of the term $\int_0^t \Phi dW$, the more general term $\int_0^t \sigma (s, u, \omega) dW$. This will be shown by doing the argument and indicating what extra assumptions are needed as this is done. Let $z$ be progressively measurable and in $L^2 (\Omega; C ([0, T]; \mathbb{R}^n))$. Also assume that $\sigma$ has linear growth. That is
\[ \| \sigma (s, u, \omega) \|_{L^2} \leq a + b |u|_{\mathbb{R}^n} \]
Then from the above theorem, there exists a unique progressively measurable solution $u$ to
\[ u(t, \omega) - u_0(\omega) + \int_0^t \mathcal{N} (s, u(s, \omega), u(s - h, \omega), w(s, \omega), \omega) ds = \int_0^t f (s, \omega) ds + \int_0^t \sigma (s, z) dW, \]
This holds for all $\omega$. There is no exceptional set needed. Now assume
\[ u_0 \in L^2 (\Omega) \]
and also a Lipschitz condition
\[ \| \sigma (s, u, \omega) - \sigma (s, \hat{u}, \omega) \|_{L^2} \leq K |u - \hat{u}| \]
Then let $u$ coincide with $z$ and $\hat{u}$ come from $\hat{z}$. Then applying the Ito formula, one can obtain the following for a constant $C$ which does not depend on $u, \hat{u}$.
\[ \frac{1}{2} |u(t) - \hat{u}(t)|^2 - C \int_0^t |u(s) - \hat{u}(s)|^2 ds - K \int_0^t |u(s) - \hat{u}(s)|^2 ds = M(t) \]
where $M(t)$ is a local martingale whose quadratic variation satisfies
\[ [M](t) = \int_0^t \| \sigma (s, z, \omega) - \sigma (s, \hat{z}, \omega) \|_{L^2}^2 |u - \hat{u}|^2 ds \]
Thus, simplifying the constants,
\[ \sup_{s \in [0, t]} |u(s) - \hat{u}(s)|^2 \leq C \int_0^t |u(s) - \hat{u}(s)|^2 ds + M^*(t) \]
where \( M^*(t) = \sup_{s \in [0,t]} |M(s)| \). Then by Gronwall’s inequality,
\[
\sup_{s \in [0,t]} |u(s) - \hat{u}(s)|^2 \leq C M^*(t)
\]
Then take the expectation of both sides. Using the Burkholder Davis Gundy inequality,
\[
E \left( \sup_{s \in [0,t]} |u(s) - \hat{u}(s)|^2 \right) \leq CE \left( \int_0^t K |z - \hat{z}|^2 |u - \hat{u}|^2 \, ds \right)^{1/2}
\]
Then adjusting the constant again,
\[
\leq \frac{1}{2} E \left( \sup_{s \in [0,t]} |u(s) - \hat{u}(s)|^2 \right) + CE \left( \int_0^t K |z - \hat{z}|^2 \, ds \right)
\]
and so,
\[
E \left( \sup_{s \in [0,t]} |u(s) - \hat{u}(s)|^2 \right) \leq C \int_0^t E \left( \sup_{r \in [0,s]} |z(r) - \hat{z}(r)|^2 \right) \, ds
\]
Letting \( \mathcal{T}z = u \) where \( u \) is defined from \( z \) in the integral equation \( \ref{eq:23.2.1} \), the above inequality implies that
\[
E \left( \sup_{s \in [0,t]} |\mathcal{T}^n z_1(s) - \mathcal{T}^n z_2(s)|^2_H \right) \leq C \int_0^t E \left( \sup_{r \in [0,s]} |\mathcal{T}^{n-1} z_1(r) - \mathcal{T}^{n-1} z_2(r)|^2 \right) \, ds
\]
\[
\leq C^2 \int_0^t \int_0^s E \left( \sup_{r_1 \in [0,r]} |\mathcal{T}^{n-2} z_1(r_1) - \mathcal{T}^{n-2} z_2(r_1)|^2 \right) \, dr \, ds
\]
One can iterate this, eventually finding that
\[
E \left( \sup_{s \in [0,t]} |\mathcal{T}^n z_1(s) - \mathcal{T}^n z_2(s)|^2_H \right) \leq C^n \int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} dt_{n-1} \cdots dt_1 E \left( \sup_{s \in [0,t]} z_1(s) - z_2(s)|^2_H \right)
\]
\[
= \frac{C^n T^n}{(n!)} E \left( \sup_{s \in [0,t]} z_1(s) - z_2(s)|^2_H \right)
\]
In particular, this holds for \( t = T \) and so, letting \( z \in L^2(\Omega, C([0,T], \mathbb{R}^n)) \), \( \{\mathcal{T}^n z\} \) is a Cauchy sequence in this space because a high enough power is a contraction map, so it converges to a unique fixed point \( u \). Each \( \mathcal{T}^n z \) is progressively measurable and so the fixed point is also. In \( L^2(\Omega, C([0,T], \mathbb{R}^n)) \), you get the integral equation
\[
u(t, \omega) - u_0(\omega) + \int_0^t N(s, u(s, \omega), u(s-h, \omega), w(s, \omega), \omega) \, ds
\]
\[
= \int_0^t f(s, \omega) \, ds + \int_0^t \sigma(s, u) \, dW_s \tag{23.2.10}
\]
Thus off a set of measure zero, the equation holds for all \( t \) and \( u \) is progressively measurable. The place where \( u_0 \in L^2(\Omega) \) is needed is in having \( \mathcal{T}z \in L^2(\Omega, C([0,T]; \mathbb{R}^n)) \). One uses a similar procedure involving the Ito formula, the growth condition
\[
(N(t, u, v, w, \omega, u) \geq -C(t, \omega) - \mu(\|u\|^2 + |v|^2 + |w|^2).
\]
and the Burkholder Davis Gundy inequality to verify this. Since \( \mathcal{T} \) depends on \( u_0 \), it appears that the set of measure zero, off which the integral equation holds, will also depend on \( u_0 \). It appears that this ultimately results from the need to take an expectation in order to deal with the stochastic integral. If this integral could be generalized in such a way that it made sense for each \( \omega \) as in the usual Riemann Stieltjes integral, then likely this restriction could be removed. It is a problem because the Wiener process is not of bounded variation.
Theorem 23.2.2 Suppose the weak monotonicity condition \( N(t,u,v,w,\omega) \) and the growth estimate \( \hat{N}(t,u,v,w,\omega) \). Also assume \( N(t,u,v,w,\omega) \in \mathbb{R}^d \) for \( u,v,w \in \mathbb{R}^d, t \in [0,T] \) and \( (t,u,v,w,\omega) \rightarrow N(t,u,v,w,\omega) \) is progressively measurable with respect to the normal filtration \( \mathcal{F}_t \) determined by a given Wiener process \( W(t) \). Also suppose \( (t,u,v,w) \rightarrow N(t,u,v,w,\omega) \) is continuous. Let \( f \in L^2(\Omega, C([0,T],\mathbb{R}^n)) \) and \( (t,u,w) \rightarrow \sigma(t,u,w) \) is progressively measurable and satisfies the linear growth condition \( \hat{\sigma}(t,u,w,\omega) \) and the Lipschitz condition \( \hat{\sigma}(t,u,w,\omega) \). Also suppose \( u_0 \in \mathcal{F}_0 \) measurable and in \( L^2(\Omega, \mathbb{R}^n) \). Then there exists a progressively measurable solution \( u \) to \( \hat{u}(t) = u(t) \) for all \( t \).

\[ \text{Proof:} \] It only remains to verify the uniqueness assertion. This happens because the fixed point is unique in \( L^2(\Omega, C([0,T],\mathbb{R}^n)) \). Therefore, off a set of measure zero the two solutions are equal for all \( t \).

23.3 STOCHASTIC DIFFERENTIAL EQUATIONS IN A HILBERT SPACE

In this section, ordinary differential equations in Hilbert space which are of the form

\[ du + N(u) dt = f dt + \sigma(u) dW \]

are considered under Lipschitz assumptions on \( N \) and \( \sigma \). A very satisfactory theorem can be proved.

The assumptions made are as follows.

\[ \|\sigma(t,u,\omega) - \sigma(t,\hat{u},\omega)\|_{L_2(Q^{1/2}L_2, H)} \leq K |u-\hat{u}|_H, \quad (23.3.11) \]

\[ |N(t,u_1,v_1,w_1,\omega) - N(t,u_2,v_2,w_2,\omega)| \leq K (|u_1 - u_2| + |v_1 - v_2| + |w_1 - w_2|) \quad (23.3.12) \]

where the norms \(|\cdot|\) refer here to the Hilbert space \( H \). Assume \( N, \sigma \) are both progressively measurable. From the Lipschitz condition given above,

\[ |N(t,u,v,w,\omega) - N(t,0,0,0,\omega)| \leq K (|u| + |v| + |w|) \]

and it is assumed that

\[ t \rightarrow N(t,0,0,0,\omega) \quad (23.3.13) \]

is in \( L^2(\Omega, C([0,T]; H)) \). Also consider the growth condition which is implied by the above condition and the Lipschitz assumption.

\[ (N(t,u,v,w,\omega),u) \geq -C(t,\omega) - \mu (|u|^2 + |v|^2 + |w|^2) \quad (23.13.14) \]

where \( C \in L^1([0,T] \times \Omega) \) and the linear growth condition for \( \sigma \),

\[ \|\sigma(t,u,\omega)\| \leq a + b |u|_H \quad (23.3.15) \]

23.3.1 The Lipschitz Case

Theorem 23.3.1 Suppose \( \hat{u}_0(\cdot,\omega) \) and let

\[ w(t) = w_0 + \int_0^t u(s) ds, \quad w_0 \in L^2(\Omega), \quad w_0 \text{ is } \mathcal{F}_0 \text{ measurable}. \]

Then there exists a unique progressively measurable solution \( u(t,\omega) \) to the integral equation

\[ u(t,\omega) - u_0(\omega) + \int_0^t N(s,u(s,\omega),u(s-h,\omega),w(s,\omega),\omega) ds \]

\[ = \int_0^t f(s,\omega) ds + \int_0^t \sigma(s,u,\omega) dW. \quad (23.13.16) \]

where \( u \in L^2(\Omega, C([0,T];H)), u_0 \in L^2(\Omega), u_0 \text{ is } \mathcal{F}_0 \text{ measurable, } f \text{ is progressively measurable and in } L^2([0,T] \times \Omega; H). \) Here there is a set of measure zero such that if \( \omega \) is not in this set, then \( u(\cdot,\omega) \) solves the above integral equation \( (23.13.16) \) and furthermore, if \( \hat{u}(\cdot,\omega) \) is another solution to it, then \( u(t,\omega) = \hat{u}(t,\omega) \) for all \( t \) if \( \omega \) is off some set of measure zero.
Proof: Let \( v \in L^2(\Omega; C([0,T];H)) \) where \( v \) is also progressively measurable. Then let \( u \) be given by

\[
    u(t,\omega) - u_0(\omega) + \int_0^t N(s,v(s,\omega),v(s-h,\omega),w(s,\omega),\omega) \, ds
    = \int_0^t f(s,\omega) \, ds + \int_0^t \sigma(s,v,\omega) \, dW.
\]

The Lipschitz condition, the assumption, and the linear growth assertion implies that \( u \) is also in \( L^2(\Omega; C([0,T];H)) \). The proof of this involves the same arguments about to be given in order to show that this determines a mapping which has a sufficiently high power a contraction map. They are also the same arguments to be used in the following theorem to establish estimates which imply a stopping time is eventually infinity.

Let \( v_1, v_2 \) be two given functions of this sort and let the corresponding \( u \) be denoted by \( u_1, u_2 \) respectively. Then

\[
    u_1(t) - u_2(t) + \int_0^t N(s,v_1(s),v_1(s-h),u_1(s)) - N(s,v_2(s),v_2(s-h),w_2(s)) \, ds
    = \int_0^t \sigma(s,v_1,\omega) - \sigma(s,v_2,\omega) \, dW.
\]

Use the Ito formula and the Lipschitz condition on \( N \) to obtain an expression of the form

\[
    \frac{1}{2} |u_1(t) - u_2(t)|^2 - C \int_0^t |v_1 - v_2|^2 \, ds - C \int_0^t |u_1 - u_2|^2 \, ds
    - \frac{1}{2} \int_0^t \| \sigma(s,v_1,\omega) - \sigma(s,v_2,\omega) \|^2 \, ds
    \leq |M(t)|
\]

where \( M(t) \) is a martingale whose quadratic variation is dominated by

\[
    C \int_0^t \| \sigma(s,v_1,\omega) - \sigma(s,v_2,\omega) \|^2 |u_1 - u_2|^2 \, ds
\]

Therefore, using the Lipschitz condition on \( \sigma \) and the Burkholder-Davis-Gundy inequality, the above implies

\[
    E \left( \sup_{s \in [0,t]} |u_1(s) - u_2(s)|^2 \right) \leq CE \int_0^t \sup_{r \in [0,s]} |v_1(r) - v_2(r)|^2 \, ds + CE \int_0^t \sup_{r \in [0,s]} |v_1(r) - v_2(r)|^2 \, ds
    + CE \left( \left( \int_0^t \| \sigma(s,v_1,\omega) - \sigma(s,v_2,\omega) \|^2 |u_1 - u_2|^2 \, ds \right)^{1/2} \right)
\]

Then a use of Gronwall’s inequality allows this to be simplified to an expression of the form

\[
    E \left( \sup_{s \in [0,t]} |u_1(s) - u_2(s)|^2 \right) \leq CE \int_0^t \sup_{r \in [0,s]} |v_1(r) - v_2(r)|^2 \, ds
    + CE \left( \left( \int_0^t \| \sigma(s,v_1,\omega) - \sigma(s,v_2,\omega) \|^2 |u_1 - u_2|^2 \, ds \right)^{1/2} \right)
    \leq C \int_0^t E \left( \sup_{r \in [0,s]} |v_1(r) - v_2(r)|^2 \right) \, ds + \frac{1}{2} E \left( \sup_{s \in [0,t]} |u_1(s) - u_2(s)|^2 \right)
    + CE \left( \int_0^t \| \sigma(s,v_1,\omega) - \sigma(s,v_2,\omega) \|^2 \, ds \right)
\]

Now using the Lipschitz condition on \( \sigma \), this simplifies further to give an inequality of the form

\[
    E \left( \sup_{s \in [0,t]} |u_1(s) - u_2(s)|^2 \right) \leq C \int_0^t E \left( \sup_{r \in [0,s]} |v_1(r) - v_2(r)|^2 \right) \, ds
\]
23.3. STOCHASTIC DIFFERENTIAL EQUATIONS IN A HILBERT SPACE

Letting $Tv = u$ where $u$ is defined from $v$ in the integral equation, the above inequality implies that

$$E \left( \sup_{\sigma \in [0,t]} \left| \mathcal{T}^n v_1 (s) - \mathcal{T}^2 v_2 (s) \right|^2_H \right) \leq C \int_0^t E \left( \sup_{\tau \in [0,s]} \left| \mathcal{T}^{n-1} v_1 (r) - \mathcal{T}^{n-1} v_2 (r) \right|^2 \right) ds$$

$$\leq C^2 \int_0^t \int_0^s E \left( \sup_{r \in [0,t]} \left| \mathcal{T}^{n-2} v_1 (r_1) - \mathcal{T}^{n-2} v_2 (r_1) \right|^2 \right) dr ds$$

One can iterate this, eventually finding that

$$E \left( \sup_{\sigma \in [0,t]} \left| \mathcal{T}^n v_1 (s) - \mathcal{T}^2 v_2 (s) \right|^2_H \right) \leq C^n \prod_{i=0}^{n-1} E \left( \sup_{\sigma \in [0,t]} \left| v_1 (s) - v_2 (s) \right|^2_H \right)$$

$$= \frac{C^n}{(n!)} E \left( \sup_{\sigma \in [0,t]} \left| v_1 (s) - v_2 (s) \right|^2_H \right)$$

In particular, one could take $t = T$. This shows that for all $n$ large enough, $\mathcal{T}^n$ is a contraction map on $L^2 (\Omega, C ([0,T] ; H))$. Therefore, picking $v \in L^2 (\Omega, C ([0,T] ; H))$, such that $v$ is also progressively measurable, $\{ \mathcal{T}^k v \}_{k=1}^\infty$ converges in $L^2 (\Omega, C ([0,T] ; H))$ to the unique fixed point of $\mathcal{T}$ denoted as $u$. Thus $T u = u$ in $L^2 (\Omega; C ([0,T] ; H))$. That is,

$$\int_\Omega \sup_t \left| Tu - u \right|^2 dP = 0$$

It follows that there is a set of measure zero such that for $\omega$ not in this set,

$$u (t, \omega) - u_0 (\omega) + \int_0^t N (s, u(s, \omega), u (s - h, \omega), w (s, \omega), \omega) ds$$

$$= \int_0^t f (s, \omega) ds + \int_0^t \sigma (s, u, \omega) dW. \quad (23.3.18)$$

The function $u$ is progressively measurable because each $\mathcal{T}^n v$ is progressively measurable and there exists a subsequence still indexed with $n$ such that for $\omega$ off a set of measure zero, $\mathcal{T}^n v (\cdot, \omega) \rightarrow u (\cdot, \omega)$ in $C ([0,T] ; H)$.

Note that the fixed point of $\mathcal{T}$ is unique in the space $L^2 (\Omega; C ([0,T] ; H))$ and so any solution to the integral equation in this space must equal this one. Hence, there exists a set of measure zero such that for $\omega$ off this set, the two solutions are equal for all $t$. 

### 23.3.2 The Locally Lipschitz Case

Now replace the Lipschitz assumption with the locally Lipschitz assumption which says that if $\max (|u|, |v|, |w|) < R$, then there is a constant $K_R$ such that

$$|N (t, u_1, v_1, w_1, \omega) - N (t, u_2, v_2, w_2, \omega)| \leq K (R) (|u_1 - u_2| + |v_1 - v_2| + |w_1 - w_2|) \quad (23.3.19)$$

Also assume the growth condition

$$(N (t, u, v, w, \omega), u) \geq - C (t, \omega) - \mu \left( |u|^2 + |v|^2 + |w|^2 \right) \quad (23.3.20)$$

and the linear growth condition on $\sigma$

$$\| \sigma (t, u, \omega) \| \leq a + b |u|_H$$

and the Lipschitz condition on $\sigma$. This can likely be relaxed as in the case of the Lipschitz condition for $N$ but for simplicity, we keep it.

**Theorem 23.3.2** Suppose $\mu \geq 0$, $a \geq 0$, $b \geq 0$ and let

$$w (t) = w_0 + \int_0^t u (s) ds, \quad w_0 \in L^2 (\Omega), \quad w_0 \text{ is } \mathcal{F}_0 \text{ measurable.}$$
Then there exists a unique progressively measurable solution $u$ to the integral equation

$$u(t, \omega) - u_0(\omega) + \int_0^t N(s, u(s, \omega), u(s-h, \omega), w(s, \omega), \omega) \, ds$$

$$= \int_0^t f(s, \omega) \, ds + \int_0^t \sigma(s, u, \omega) \, dW.$$  \hspace{1cm} (23.3.21)

where $u \in L^2(\Omega, C([0,T]; H))$, $u_0 \in L^2(\Omega, H)$, $u_0$ is $\mathcal{F}_0$ measurable, $f$ is progressively measurable and in $L^2([0,T] \times \Omega; H)$. Here there is a set of measure zero such that if $\omega$ is not in this set, then $u(\cdot, \omega)$ solves the above integral equation and furthermore, if $\hat{u}(\cdot, \omega)$ is another solution to it, then $u(t, \omega) = \hat{u}(t, \omega)$ for all $t$ if $\omega$ is off some set of measure zero.

**Proof:** Let $u_n$ be the unique solution to the integral equation

$$u_n(t, \omega) - u_0(\omega) + \int_0^t N(s, P_n u_n(s, \omega), P_n u_n(s-h, \omega), P_n w_n(s, \omega), \omega) \, ds$$

$$= \int_0^t f(s, \omega) \, ds + \int_0^t \sigma(s, u_n, \omega) \, dW.$$  \hspace{1cm} (23.3.22)

where $P_n$ is the projection onto $B(0, 9^n)$. Thus the modified problem is in the situation of Theorem \ref{thm:43.3.1} so there exists such a solution. Then let

$$\tau_n = \inf \{ t : |u_n(t)| + |w_n(t)| > 2^n \}$$

Then stopping the equation with this stopping time, we can write

$$u_n^{\tau_n}(t, \omega) - u_0(\omega) + \int_0^t X_{[0, \tau_n]} N(s, u_n^{\tau_n}(s, \omega), u_n^{\tau_n}(s-h, \omega), w_n^{\tau_n}(s, \omega), \omega) \, ds$$

$$= \int_0^t X_{[0, \tau_n]} f(s, \omega) \, ds + \int_0^t X_{[0, \tau_n]} \sigma(s, u_n^{\tau_n}, \omega) \, dW.$$  \hspace{1cm} (23.3.23)

Then using the growth condition and the Itô formula,

$$\frac{1}{2} |u_n^{\tau_n}(t)|_H^2 \leq C(u_n, w_n, f) + C \int_0^t |u_n^{\tau_n}(r)|_H^2 \, dr + \sup_{s \in [0,t]} |M(t)|$$

where $M(t)$ is a martingale whose quadratic variation is dominated by

$$\int_0^t ||\sigma(s, u_n^{\tau_n})||^2 |u_n^{\tau_n}|^2 \, ds$$

Then it follows by the Burkholder-Davis-Gundy inequality

$$E \left( \sup_{s \in [0,t]} |u_n^{\tau_n}(s)|_H^2 \right) \leq E(C(u_0, w_0, f)) + C \int_0^t E \left( \sup_{r \in [0,s]} |u_n^{\tau_n}(r)|^2 \, dr \right) \, ds$$

$$+ CE \left( \left( \int_0^t ||\sigma(s, u_n^{\tau_n})||^2 |u_n^{\tau_n}|^2 \, ds \right)^{1/2} \right)$$

Now apply Gronwall’s inequality and modify the constants so that

$$E \left( \sup_{s \in [0,t]} |u_n^{\tau_n}(s)|_H^2 \right) \leq E(C(u_0, w_0, f)) + CE \left( \left( \int_0^t ||\sigma(s, u_n^{\tau_n})||^2 |u_n^{\tau_n}|^2 \, ds \right)^{1/2} \right)$$

$$\leq E(C(u_0, w_0, f)) + \frac{1}{2} E \left( \sup_{s \in [0,t]} |u_n^{\tau_n}(s)|_H^2 \right) + CE \left( \left( \int_0^t ||\sigma(s, u_n^{\tau_n})||^2 \, ds \right)^{1/2} \right)$$
Then, using the linear growth condition on $\sigma$, it follows on modification of the constants again that
\[
E \left( \sup_{s \in [0,t]} |u^*_n(s)|^2_H \right) \leq E \left( C(u_0, w_0, f) + CE \left( \int_0^t |u^*_n|^2 ds \right) \right)
\]
\[
\leq E \left( C(u_0, w_0, f) + CE \left( \int_0^t \sup_{r \in [0,s]} |u^*_n|^2 dr \right) \right)
\]
and so, another application of Gronwall’s inequality implies that
\[
E \left( \sup_{s \in [0,T]} |u^*_n(s)|^2_H \right) \leq E \left( C(u_0, w_0, f) \right) < \infty
\]
Then
\[
P \left( \sup_{s \in [0,T]} |u^*_n(s)|^2_H > \left( \frac{3}{2} \right)^n \right) \leq E \left( C(u_0, w_0, f) \right) \left( \frac{2}{3} \right)^n
\]
Now an application of the Borel Cantelli lemma shows that there exists a set of measure zero $\tilde{N}$ such that for $\omega \notin \tilde{N}$, it follows that for all $n$ large enough,
\[
\sup_{s \in [0,T]} |u^*_n(s)|^2_H < (3/2)^n
\]
and so $\tau_n = \infty$ for all $n$ large enough.

**Claim:** For $m < n$, there is a set of measure zero $N_{mn}$ such that if $\omega \notin N_{mn}$, then $u^*_m(s) = u^*_m(s)$ on $[0, T \wedge \tau_m]$.

**Proof of the claim:** Note that $\tau_m \leq \tau_n$. Therefore, these are both progressively measurable solutions to the integral equation
\[
u(t \wedge \tau_m, \omega) - u_0(\omega) + \int_0^t \mathcal{X}_{[0,\tau_m]} N(s, u(s, \omega), u(s - h, \omega), w_u(s, \omega), \omega) ds
\]
\[
= \int_0^t \mathcal{X}_{[0,\tau_m]} f(s, \omega) ds + \int_0^t \mathcal{X}_{[0,\tau_m]} \sigma(s, u, \omega) dW.
\] (23.3.24)
where
\[
w_u(t) = w_0 + \int_0^t u(s) ds.
\]
To save notation, refer to these functions as $u, v$ and let $\tau_m = \tau$. Subtract and use the Ito formula to obtain
\[
\frac{1}{2} |u(t \wedge \tau) - v(t \wedge \tau)|^2_H \leq \int_0^t \mathcal{X}_{[0,\tau_m]} \left( N(s, u(s), u(s - h), w_u(s)) - N(s, v(s), v(s - h), w_u - v(s)) \right) ds + \sup_{s \in [0,t]} |M(t)|
\]
where the quadratic variation of the martingale $M(t)$ is dominated by
\[
\int_0^t \mathcal{X}_{[0,\tau]} \|\sigma(s, u, \omega) - \sigma(s, v, \omega)\|^2 |u - v|^2 ds
\]
Then from the assumption that $N$ is locally Lipschitz and routine manipulations,
\[
\frac{1}{2} |u(t \wedge \tau) - v(t \wedge \tau)|^2_H \leq C_m \int_0^t \mathcal{X}_{[0,\tau]} |u - v|^2 ds + \sup_{s \in [0,t]} |M(s)|
\]
and so, adjusting the constants yields
\[
\sup_{s \in [0,t]} |u(s \wedge \tau) - v(s \wedge \tau)|^2_H \leq C_m \int_0^t \mathcal{X}_{[0,\tau]} \sup_{r \in [0,s]} |u(r \wedge \tau) - v(r \wedge \tau)|^2 ds + \sup_{s \in [0,t]} |M(s)|
\]
and so, by Gronwall’s inequality followed by the Burkholder-Davis-Gundy inequality,
\[
E \left( \sup_{s \in [0,t]} |u(s \wedge \tau) - v(s \wedge \tau)|^2_H \right) \leq
\]
\[
CE \left( \left( \int_0^t X_{[0,\tau]} \| \sigma(s, u, \omega) - \sigma(s, v, \omega) \|^2 | u(s \wedge \tau) - v(s \wedge \tau)|^2 \, ds \right)^{1/2} \right) \\
\leq \frac{1}{2} E \left( \sup_{s \in [0,t]} | u(s \wedge \tau) - v(s \wedge \tau)|^2 \right) + CE \left( \int_0^t X_{[0,\tau]} \| \sigma(s, u) - \sigma(s, v) \|^2 \, ds \right)
\]

and so, adjusting the constant again,

\[
E \left( \sup_{s \in [0,t]} | u(s \wedge \tau) - v(s \wedge \tau)|^2 \right) \leq CE \left( \int_0^t X_{[0,\tau]} \| \sigma(s, u(s \wedge \tau)) - \sigma(s, v(s \wedge \tau)) \|^2 \, ds \right)
\]

\leq CE \left( \int_0^t X_{[0,\tau]} K | u(s \wedge \tau) - v(s \wedge \tau)|^2 \, ds \right)

\leq C \int_0^t E \left( \sup_{r \in [0,s]} | u(r \wedge \tau) - v(r \wedge \tau)|^2 \right) \, ds
\]

and so, Gronwall’s inequality shows that for every \( t \),

\[
E \left( \sup_{s \in [0,t]} | u(s \wedge \tau) - v(s \wedge \tau)|^2 \right) = 0
\]

In particular, for \( t = T \) this holds. Hence

\[
E \left( \sup_{s \in [0,T]} | u(s \wedge \tau) - v(s \wedge \tau)|^2 \right) = 0
\]

It follows that

\[
E \left( \sup_{s \in [0,T \wedge T]} | u(s) - v(s)|^2 \right) = 0
\]

so that off a set of measure zero, \( u(s) = v(s) \) for all \( s \in [0, \tau] \). This proves the claim.

Now let the set of measure zero \( N \) be given by \( N = \bigcup_{m < n} N_{mn} \cup \hat{N} \) where \( \hat{N} \) is the set of measure zero off which \( \tau_m = \infty \) for all \( m \) large enough. Then for \( \omega \notin N \), it follows that \( u^{m_n}(s) = u^{m_n}(s) \) on \( [0, \tau_m \wedge T] \) and, for all \( m \) large enough, \( \tau_m = \infty \). Hence for all \( m \) large enough, and such \( \omega \), \( u_n(s, \omega) = u_m(s, \omega) \) for all \( s \in [0, T] \). Thus, for \( \omega \) off \( N \), it follows that \( \lim_{m \to \infty} u^{m_m}(s, \omega) = u(s, \omega) \) exists, this for each \( s \in [0, T] \) and \( \omega \) off a fixed set of measure zero. In fact, this convergence is uniform on \( [0, T] \) because for all \( n \) sufficiently large and for such a fixed \( \omega \notin N \), there is no change in increasing \( m \). Hence, \( u \) is progressively measurable and satisfies the integral equation [435\text{a}].

It remains to verify uniqueness. Suppose there are two solutions \( u, v \) each progressively measurable solutions of the given integral equation. Then let \( \tau_n \) be a stopping time

\[
\tau_n = \inf \{ t : |u(t)| + |v(t)| > 2^n \}
\]

Then a repeat of the arguments given in the above claim shows that on \( [0, \tau_n \wedge T] \) the two functions \( u^{\tau_n}, v^{\tau_n} \) are equal on \( [0, \tau_n \wedge T] \) off a set of measure zero \( N_n \). Let \( N \) be the union of the exceptional sets. Then for \( \omega \notin N \), \( u(t, \omega) = v(t, \omega) \) for all \( t \in [0, \tau_n \wedge T] \). However, \( \tau_n(\omega) = \infty \) for all \( n \) large enough because each of these functions is continuous. Hence, the two functions are equal on \( [0, T] \) for such \( \omega \). This shows uniqueness. \( \blacksquare \)
Chapter 24

The Hard Ito Formula

Recall the following definition of stochastically continuous.

$X$ is stochastically continuous at $t_0 \in I$ means: for all $\varepsilon > 0$ and $\delta > 0$ there exists $\rho > 0$ such that

$$P (||X(t) - X(t_0)|| \geq \varepsilon) \leq \delta \text{ whenever } |t - t_0| < \rho, t \in I.$$ 

Note the above condition says that for each $\varepsilon > 0$,

$$\lim_{t \to t_0} P (||X(t) - X(t_0)|| \geq \varepsilon) = 0.$$ 

24.1 Predictable And Stochastic Continuity

Definition 24.1.1 Let $\mathcal{F}_t$ be a filtration. The predictable sets consists of those sets which are in the smallest $\sigma$-algebra which contains the sets $E \times \{0\}$ for $E \in \mathcal{F}_0$ and $E \times (a,b]$ where $E \in \mathcal{F}_a$. Thus every predictable set is a progressively measurable set.

First of all, here is an important observation.

Proposition 24.1.2 Let $X(t)$ be a stochastic process having values in $E$ a complete metric space and let it be $\mathcal{F}_t$ adapted and left continuous where $\mathcal{F}_t$ is a normal filtration. Then it is predictable. If $t \to X(t,\omega)$ is continuous for all $\omega /\in N, P(N) = 0$, then $(t,\omega) \to X(t,\omega)\mathcal{X}_{\text{NC}}(\omega)$ is predictable. Also, if $X(t)$ is stochastically continuous and adapted on $[0,T]$, then it has a predictable version. If $X \in C([0,T];L^p(\Omega;F))$, $p \geq 1$ for $F$ a Banach space, then $X$ is stochastically continuous.

Proof: First suppose $X$ is continuous for all $\omega \in \Omega$. Define

$$I_{m,k} \equiv ((k-1)2^{-m}T, k2^{-m}T]$$

if $k \geq 1$ and $I_{m,0} = \{0\}$ if $k = 1$. Then define

$$X_m(t) = \sum_{k=1}^{2^m} X(T(k-1)2^{-m})\mathcal{X}_{((k-1)2^{-m}T, k2^{-m}T]}(t) + X(0)\mathcal{X}_{[0,0]}(t)$$

Here the sum means that $X_m(t)$ has value $X(T(k-1)2^{-m})$ on the interval $((k-1)2^{-m}T, k2^{-m}T]$. Thus $X_m$ is predictable because each term in the formal sum is. Thus

$$X_m^{-1}(U) = \bigcup_{k=1}^{2^m} (X(T(k-1)2^{-m})\mathcal{X}_{((k-1)2^{-m}T, k2^{-m}T]}^{-1}(U)$$

$$= \bigcup_{k=1}^{2^m} ((k-1)2^{-m}T, k2^{-m}T] \times (X(T(k-1)2^{-m}))^{-1}(U),$$
a finite union of predictable sets. Since \( X \) is left continuous,

\[
X (t, \omega) = \lim_{m \to \infty} X_m (t, \omega)
\]

and so \( X \) is predictable.

Now suppose that for \( \omega \not\in N \), \( P (N) = 0 \), \( t \to X (t, \omega) \) is continuous. Then applying the above argument to \( X (t) \mathcal{X}_{N, C} \) it follows \( X (t) \mathcal{X}_{N, C} \) is predictable by completeness of \( \mathcal{F}_t \), \( X (t) \mathcal{X}_{N, C} \) is \( \mathcal{F}_t \) measurable.

Next consider the other claim. Since \( \mathcal{X} \) is stochastically continuous on \([0, T] \) it is uniformly stochastically continuous on this interval by Lemma \((\ref{lem:stoch_cont})\). Therefore, there exists a sequence of partitions of \([0, T] \), the \( m \)-th being

\[
0 = t_m, 0 < t_m, 1 < \cdots < t_m, n (m) = T
\]

such that for \( X_m \) defined as above, then for each \( t \)

\[
P \left( \left[ d (X_m (t), X (t)) \geq 2^{-m} \right] \right) \leq 2^{-m} \quad (24.1.1)
\]

Then as above, \( X_m \) is predictable. Let \( A \) denote those points of \( \mathcal{P}_T \) at which \( X_m (t, \omega) \) converges. Thus \( A \) is a predictable set because it is just the set where \( X_m (t, \omega) \) is a Cauchy sequence. Now define the predictable function \( Y \)

\[
Y (t, \omega) = \left\{ \begin{array}{ll}
\lim_{m \to \infty} X_m (t, \omega) & \text{if } (t, \omega) \in A \\
0 & \text{if } (t, \omega) \not\in A
\end{array} \right.
\]

From \((\ref{lem:stoch_cont})\) it follows from the Borel Cantelli lemma that for fixed \( t \), the set of \( \omega \) which are in infinitely many of the sets,

\[
[d (X_m (t), X (t)) \geq 2^{-m}]
\]

has measure zero. Therefore, for each \( t \), there exists a set of measure zero, \( N (t) \) such that for \( \omega \not\in N (t) \) and all \( m \) large enough

\[
[d (X_m (t, \omega), X (t, \omega)) < 2^{-m}]
\]

Hence for \( \omega \not\in N (t) \), \( (t, \omega) \in A \) and so \( X_m (t, \omega) \to Y (t, \omega) \) which shows

\[
d (Y (t, \omega), X (t, \omega)) = 0 \text{ if } \omega \not\in N (t) .
\]

The predictable version of \( X (t) \) is \( Y (t) \).

Finally consider the claim about the specific example where

\[
X \in C ([0, T] ; L^p (\Omega; F)).
\]

\[
P (\| X (t) - X (s) \|_F \geq \varepsilon) \varepsilon^p \leq \int_{\Omega} \| X (t) - X (s) \|_F^p dP \leq \varepsilon^p \delta
\]

provided \(|s - t|\) sufficiently small. Thus

\[
P (\| X (t) - X (s) \|_F \geq \varepsilon) < \delta
\]

when \(|s - t|\) is small enough. \( \blacksquare \)

### 24.2 Approximating With Step Functions

This Ito formula seems to be the fundamental idea which allows one to obtain solutions to stochastic partial differential equations using a variational point of view. I am following the treatment found in \((\ref{book:ito})\). The following lemma is fundamental to the presentation. It approximates a function with a sequence of two step functions \( X_k^r, X_k^l \) where \( X_k^r \) has the value of \( X \) at the right end of each interval and \( X_k^l \) gives the value \( X \) at the left end of the interval. The lemma is very interesting for its own sake. You can obviously do this sort of thing for a continuous function but here the function is not continuous and in addition, it is a stochastic process depending on \( \omega \) also. This lemma was proved earlier Lemma \((\ref{lem:step_functions})\).

**Lemma 24.2.1** Let \( \Phi : [0, T] \times \Omega \to V \), be \( \mathcal{B} ([0, T]) \times \mathcal{F} \) measurable and suppose

\[
\Phi \in K \equiv L^p ([0, T] \times \Omega; E), \ p \geq 1
\]
24.2. APPROXIMATING WITH STEP FUNCTIONS

Then there exists a sequence of nested partitions, $P_k \subseteq P_{k+1}$,

$$P_k = \{t_{k,0}, \ldots, t_{k,m_k}\}$$

such that the step functions given by

$$\Phi_k^L(t) = \sum_{j=1}^{m_k} \Phi(t_{j-1}^k) \chi_{(t_{j-1}^k, t_j^k]}(t)$$

$$\Phi_k^R(t) = \sum_{j=1}^{m_k} \Phi(t_{j}^k) \chi_{(t_{j-1}^k, t_j^k]}(t)$$

both converge to $\Phi$ in $K$ as $k \to \infty$ and

$$\lim_{k \to \infty} \max \{|t_j^k - t_{j+1}^k| : j \in \{0, \ldots, m_k\}\} = 0.$$

Also, each $\Phi(t_j^k)$, $\Phi(t_{j-1}^k)$ is in $L_p(\Omega; E)$. One can also assume that $\Phi(0) = 0$. The mesh points $\{t_j^k\}_{j=0}^{m_k}$ can be chosen to miss a given set of measure zero. In addition to this, we can assume that

$$|t_j^k - t_{j-1}^k| = 2^{-nk}$$

except for the case where $j = 1$ or $j = m_{nk}$ when this might not be so. In the case of the last subinterval defined by the partition, we can assume

$$|t_{m_k}^k - t_{m-1}^k| = |T - t_{m-1}^k| \geq 2^{-(n+1)}$$

The following lemma is convenient.

**Lemma 24.2.2** Let $f_n \to f$ in $L^p([0, T] \times \Omega, E)$. Then there exists a subsequence $n_k$ and a set of measure zero $N$ such that if $\omega \notin N$, then

$$f_{n_k} (\cdot, \omega) \to f (\cdot, \omega)$$

in $L^p([0, T] \times \Omega, E)$ and for a.e. $t$.

**Proof:** We have

$$P \left( \|f_n - f\|_{L^p([0, T] \times \Omega, E)} > \lambda \right) \leq \frac{1}{\lambda} \int_{\Omega} \|f_n - f\|_{L^p([0, T], E)} \, dP \leq \frac{1}{\lambda} \|f_n - f\|_{L^p([0, T] \times \Omega, E)}$$

Hence there exists a subsequence $n_k$ such that

$$P \left( \|f_{n_k} - f\|_{L^p([0, T], E)} > 2^{-k} \right) \leq 2^{-k}$$

Then by the Borel Cantelli lemma, it follows that there exists a set of measure zero $N$ such that for all $k$ large enough and $\omega \notin N$,

$$\|f_{n_k} - f\|_{L^p([0, T], E)} \leq 2^{-k} \quad \blacksquare$$

Because of this lemma, it can also be assumed that for a.e. $\omega$ pointwise convergence is obtained on $[0, T]$ as well as convergence in $L^p([0, T])$. This kind of assumption will be tacitly made whenever convenient in the context of the above lemma.

Also recall the diagram for the definition of the integral.

$$U \uparrow Q^{1/2}$$

$$U_1 \supseteq JQ^{1/2}U \uparrow J^{1/2}Q^{1/2}U$$

$$\Phi_n \downarrow \Phi$$

$$H$$

The idea was to get $\int_0^t \Phi \, dW$ where $\Phi \in L^2([0, T] \times \Omega; \mathcal{L}_2(Q^{1/2}U, H))$. Here $W(t)$ was a cylindrical Wiener process. This meant that it was a $Q_1$ Wiener process on $U_1$ for $Q_1 = JJ^*$ and $J$ was a Hilbert Schmidt operator mapping $Q^{1/2}U$ to $U_1$. 
24.3 The Situation

Now consider the following situation.

**Situation 24.3.1** Let \( X \) satisfy the following.

\[
X(t) = X_0 + \int_0^t Y(s) \, ds + \int_0^t Z(s) \, dW(s),
\]

(24.3.2)

\( X_0 \in L^2(\Omega; H) \) and is \( \mathcal{F}_0 \) measurable, where \( Z \) is \( L_2(Q^{1/2}U, H) \) progressively measurable and

\[
\int_0^T \int_0^T \|Z(s)\|^2_{L_2(Q^{1/2}U, H)} \, dP \, dt < \infty
\]

so that the stochastic integral makes sense. Also \( X \) has a measurable representative \( \bar{X} \) which has values in \( V \). (For a.e. \( X(t) = X(t) \) for \( P \) a.e. \( \omega \).) This representative satisfies

\[
\bar{X} \in L^2([0,T] \times \Omega, \mathcal{B}([0,T] \times \mathcal{F}, H)) \cap L^p([0,T] \times \Omega, \mathcal{B}([0,T]) < \infty
\]

Assume \( Y(s) \) satisfies

\[
Y \in K' = L^p'([0,T] \times \Omega; V')
\]

where \( 1/p' + 1/p = 1 \) and \( Y \) is \( V' \) progressively measurable. The situation in which the equation holds is as follows. For a.e. \( \omega \), the equation holds for all \( t \in [0,T] \) in \( V' \). Thus it follows that \( X(t) \) is automatically progressively measurable into \( V' \) from Proposition 24.1.2. Also \( W(t) \) is a Wiener process on \( U_1 \) in the above diagram. Thus \( X \) is continuous into \( V' \) off a set of measure zero, and it is also \( V' \) predictable.

The goal is to prove the following Itô formula.

\[
|X(t)|^2 = |X_0|^2 + \int_0^t \left( 2 \langle Y(s), \bar{X}(s) \rangle + \|Z(s)\|^2_{L_2(Q^{1/2}U, H)} \right) \, ds
\]

\[
+ 2 \int_0^t \mathcal{R} \left( (Z(s) \circ J^{-1})^* \bar{X}(s) \right) \circ JdW(s)
\]

(24.3.3)

where \( \mathcal{R} \) is the Riesz map which takes \( U_1 \) to \( U_1' \). The main thing is that the last term above be a local martingale.

In all that follows, the mesh points \( t_j \) will be points where \( \bar{X}(t_j) = X(t_j) \) a.e. \( \omega \).

**Lemma 24.3.2** Let \( X \) be as in Situation 24.3.1 and let \( X_k^j \) be as in Lemma 24.1.2 corresponding to \( \bar{X} \) above. Say

\[
X_k^j(t) = \sum_{j=0}^{m_k} \bar{X}(t_j) \chi_{(t_j,t_{j+1})} (t), \quad X_k^j(0) \equiv 0.
\]

Then each term in the above sum for which \( t_j > 0 \) is predictable into \( H \). As mentioned earlier, we can take \( X(0) \equiv 0 \) in the definition of the “left step function”. Since, at the mesh points, \( \bar{X} = X \) a.e., it makes no difference off a set of measure zero whether we use \( \bar{X}(t_j) \) or \( X(t_j) \) at the left end point.

**Proof:** This is a step function and a typical term is of the form \( X(a) \chi_{(a,b)} (t) \). I will try and show this is predictable. Let \( a_n \) be an increasing sequence converging to \( a \) and let \( b_n \) be an increasing sequence converging to \( b \). Then for a.e. \( \omega \),

\[
X(a_n) \chi_{(a_n,b_n)} (t) \to X(a) \chi_{(a,b)} (t)
\]

in \( V \) due to the fact that \( t \to X(t) \) is continuous into \( V' \) for a.e. \( \omega \). Therefore, letting \( v \in V \) be given, it follows that for a.e. \( \omega \)

\[
\langle X(a_n) \chi_{(a_n,b_n)} (t), v \rangle \to \langle X(a) \chi_{(a,b)} (t), v \rangle,
\]

and since the filtration is a normal filtration in which all sets of measure zero from \( \mathcal{F}_T \) are in \( \mathcal{F}_0 \), this shows

\[
(t, \omega) \mapsto \langle X(a) \chi_{(a,b)} (t), v \rangle
\]

is real predictable because it is the pointwise limit of real predictable functions, those in the sequence being real predictable because of the continuity of \( X(t) \) into \( V' \) and Proposition 24.1.2. Now since \( H \subseteq V' \) it follows that for all \( v \in V \),

\[
(t, \omega) \mapsto \langle X(a) \chi_{(a,b)} (t), v \rangle
\]

is real predictable. This holds for \( h \in H \) replacing \( v \) in the above because \( V \) is dense in \( H \). By the Pettis theorem, this proves the lemma. ■
Lemma 24.3.3 In Situation 24.3.1 the following formula holds for a.e. $\omega$ for $0 < s < t$ where $M(t) = \int_0^t Z(u) \, dW(u)$. Here and elsewhere, $\cdot$ denotes the norm in $H$ and $\langle \cdot, \cdot \rangle$ denotes the duality pairing between $V, V'$. Also $X = \bar{X}$ for a.e. $\omega$ at $t, s$ so that it makes no difference off a set of measure zero whether we write $\langle Y(u), X(t) \rangle$ or $\langle Y(u), \bar{X}(t) \rangle$.

\[ |X(t)|^2 = |X(s)|^2 + 2 \int_s^t \langle Y(u), X(t) \rangle \, du + 2 \langle X(s), M(t) - M(s) \rangle \]

Also for $t > 0$

\[ |X(t)|^2 = |X_0|^2 + 2 \int_0^t \langle Y(u), X(t) \rangle \, du + 2 \langle X_0, M(t) \rangle \]

\[ + |M(t)|^2 - |X(t) - X_0 - M(t)|^2 \] (24.3.4)

Proof: The formula is a straight forward computation which holds a.e. $\omega$.

\[
|M(t) - M(s)|^2 - |X(t) - X(s) - (M(t) - M(s))|^2 + 2 \langle X(s), M(t) - M(s) \rangle
\]

\[
= |M(t) - M(s)|^2 - |X(t) - X(s)|^2 - |M(t) - M(s)|^2 + 2 \langle X(t) - X(s), M(t) - M(s) \rangle + 2 \langle X(s), M(t) - M(s) \rangle
\]

\[
= - |X(t) - X(s)|^2 + 2 \langle X(t), M(t) - M(s) \rangle - 2 \langle \int_s^t Y(u) \, du, X(t) \rangle
\]

\[
= - |X(t)|^2 - |X(s)|^2 + 2 \langle X(t), X(s) \rangle + 2 |X(t)|^2 - 2 \langle X(t), X(s) \rangle - 2 \int_s^t \langle Y(u), X(t) \rangle \, du
\]

\[
= |X(t)|^2 - |X(s)|^2 - 2 \int_s^t \langle Y(u), X(t) \rangle \, du
\]

Comparing the ends of this string of equations,

\[
|X(t)|^2 = |X(s)|^2 + 2 \int_s^t \langle Y(u), X(t) \rangle \, du + 2 \langle X(s), M(t) - M(s) \rangle + |M(t) - M(s)|^2 - |X(t) - X(s) - (M(t) - M(s))|^2
\]

which is what was to be shown.

Now it is time to prove the other assertion.

\[
|M(t)|^2 - |X(t) - X_0 - M(t)|^2 + 2 \langle X_0, M(t) \rangle
\]

\[
= - |X(t) - X_0|^2 + 2 \langle X(t) - X_0, M(t) \rangle + 2 \langle X_0, M(t) \rangle
\]

\[
= - |X(t) - X_0|^2 + 2 \langle X(t), M(t) \rangle
\]

\[
= - |X(t) - X_0|^2 + 2 \langle X(t), X(t) - X_0 \rangle - 2 \left\langle \int_0^t Y(s) \, ds, X(t) \right\rangle
\]

\[
= |X(t)|^2 - |X_0|^2 - 2 \left\langle \int_0^t Y(s), X(t) \right\rangle \, ds \]

Noting that $X(0) = X_0 \in L^2(\Omega, H)$ and is $\mathcal{F}_0$ measurable, the first formula works in both cases.
24.4 The Main Estimate

The following phenomenal estimate holds and it is this estimate which is the main idea in proving the Ito formula. The last assertion about continuity is like the well known result that if \( y \in L^p (0,T; V) \) and \( y' \in L^p (0,T; V') \), then \( y \) is actually continuous with values in \( H \). Later, this continuity result is strengthened further to give strong continuity.

**Lemma 24.4.1** In the Situation 24.3.4, let \( t \) denote a point of \( \mathcal{P}_k \) from Lemma 24.3.1. Then for \( t_j > 0 \), \( X (t_j) \) is just the value of \( X \) at \( t_k \) but when \( t = 0 \), the definition of \( X (0) \) in this step function is \( X (0) \equiv 0 \). Thus

\[
|X (t_m)|^2 - |X_0|^2 = \sum_{j=1}^{m-1} |X (t_j)|^2 - |X_0|^2 + |X (t_1)|^2 - |X_0|^2
\]

Using the formula in Lemma 24.3.3, for \( t = t_m \) this yields

\[
|X (t_m)|^2 - |X_0|^2 = 2 \sum_{j=1}^{m-1} \int_{t_j}^{t_j+1} \langle Y (u), X_k (u) \rangle \, du + 2 \sum_{j=1}^{m-1} \int_{t_j}^{t_j+1} \langle Z (u), X (t) \rangle \, du
\]

\[
+ 2 \sum_{j=1}^{m-1} \left( \int_{t_j}^{t_j+1} Z (u) \, dW, X (t_j) \right)_H + 2 \sum_{j=1}^{m-1} \left( X (t_j), X (t_j) \right)
\]

\[
- \sum_{j=1}^{m-1} \left( X (t_j+1), X (t) - (M (t_j+1) - M (t_j)) \right)^2
\]

\[
+ 2 \int_0^{t_1} \langle Y (u), X (t_1) \rangle \, du + 2 \left( X_0, \int_0^{t_1} Z (u) \, dW \right) + |M (t_1)|^2
\]

Of course

\[
2 \int_0^{t_1} \langle Y (u), X (t_1) \rangle \, du + 2 \left( X_0, \int_0^{t_1} Z (u) \, dW \right) + |M (t_1)|^2
\]
converges to 0 for a.e. $\omega$ as $k \to \infty$ because the norms of the partitions converge to 0 and the stochastic integral is continuous off a set of measure zero. Actually this is not completely clear for the first of the above terms. This term is dominated by

$$
\left( \int_0^{t_1} \|Y(u)\|^p' \, du \right)^{1/p} \left( \int_0^T \|X_k(u)\|^p \, du \right)^{1/p}
$$

By Theorem 24.4.7, this equals

$$
C(\omega) \left( \int_0^{t_1} \|Y(u)\|^p' \, du \right)^{1/p}
$$

Hence this converges to 0 for a.e. $\omega$. At this time, not much is known about the last term in 24.4.7, but it is negative and is about to be neglected anyway. The Ito isometry implies the other two terms converge to 0 in $L^1(\Omega)$ also, in addition to converging for a.e. $\omega$. At this time, not much is known about the last term in 24.4.7, but it is negative and is about to be neglected anyway.

The term involving the stochastic integral equals

$$
2 \sum_{j=1}^{m-1} \left( \int_{t_j}^{t_{j+1}} Z(u) \, dW, X(t_j) \right)_H
$$

By Theorem 24.4.4, this equals

$$
2 \int_0^{t_1} \mathcal{R} \left( (Z(u) \circ J^{-1})^* X_k^*(u) \right) \circ JdW,
$$

to $\int_0^{t_1} \mathcal{R} \left( (Z(u) \circ J^{-1})^* X_k^*(u) \right) \circ JdW$ being a local martingale. Therefore, 24.4.4 equals

$$
|X(t_m)|^2 - |X_0|^2 = 2 \int_0^{t_1} \langle Y(u), X_k^*(u) \rangle \, du + e(k)
$$

$$
2 \int_0^{t_1} \mathcal{R} \left( (Z(u) \circ J^{-1})^* X_k^*(u) \right) \circ JdW + \sum_{j=1}^{m-1} |M(t_{j+1}) - M(t_j)|^2
$$

$$
- \sum_{j=1}^{m-1} |X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j))|^2 - |X(t_1) - X_0 - M(t_1)|^2
$$

where $e(k)$ converges to 0 in $L^1(\Omega)$ and for a.e. $\omega$. Note that $X_k(u) = 0$ on $[0, t_1)$ and so that stochastic integral equals

$$
\int_0^{t_1} \mathcal{R} \left( (Z(u) \circ J^{-1})^* X_k^*(u) \right) \circ JdW.
$$

Therefore, from the above,

$$
|X(t_m)|^2 - |X_0|^2 = 2 \int_0^{t_1} \langle Y(u), X_k^*(u) \rangle \, du + e(k)
$$

$$
2 \int_0^{t_1} \mathcal{R} \left( (Z(u) \circ J^{-1})^* X_k^*(u) \right) \circ JdW + \sum_{j=0}^{m-1} |M(t_{j+1}) - M(t_j)|^2 - |M(t_1)|^2
$$

$$
- \sum_{j=1}^{m-1} |X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j))|^2 - |X(t_1) - X_0 - M(t_1)|^2
$$

Then since $|M(t_1)|$ converges to 0 in $L^1(\Omega)$ and for a.e. $\omega$, as discussed above,

$$
|X(t_m)|^2 - |X_0|^2 = 2 \int_0^{t_1} \langle Y(u), X_k^*(u) \rangle \, du + e(k)
$$

$$
+ 2 \int_0^{t_1} \mathcal{R} \left( (Z(u) \circ J^{-1})^* X_k^*(u) \right) \circ JdW + \sum_{j=0}^{m-1} |M(t_{j+1}) - M(t_j)|^2
$$
- \left| X(t_1) - X_0 - M(t_1) \right|^2 \\
- \sum_{j=1}^{m-1} \left| X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j)) \right|^2 \quad (24.4.8)

where \( e(k) \to 0 \) for a.e. \( \omega \) and also in \( L^1(\Omega) \).

Now it follows on discarding the negative terms,

\[
\sup_{t_j \in P_k} \left| X(t_j) \right|^2 \leq |X_0|^2 + 2 \int_0^T \left| \langle Y(u) , X^k(u) \rangle \right| du \\
+ 2 \sup_{t \in [0,T]} \left| \int_0^t \mathcal{R} \left( (Z(u) \circ J^{-1})^* X^k(u) \right) \circ J dW \right| + \sum_{j=0}^{m-1} \int_{t_j}^{t_{j+1}} \left| Z(u) dW \right|^2
\]

where there are \( m_k + 1 \) points in \( P_k \).

Do \( \int_\Omega \) to both sides. Using the Ito isometry, this yields

\[
\int_\Omega \left( \sup_{t_j \in P_k} \left| X(t_j) \right|^2 \right) dP \leq E\left( |X_0|^2 \right) + 2 ||Y||_{K'} ||X^k||_K \\
+ \sum_{j=0}^{m-1} \int_{t_j}^{t_{j+1}} \int_\Omega \left| Z(u) \right|^2 dP du
\]

\[
+ 2 \int_\Omega \left( \sup_{t \in [0,T]} \left| \int_0^T \mathcal{R} \left( (Z(u) \circ J^{-1})^* X^k(u) \right) \circ J dW \right| \right) dP + E\left( |e(k)| \right)
\]

\[
\leq C + \int_0^T \int_\Omega \left| Z(u) \right|^2 dP du + \\
+ 2 \int_\Omega \left( \sup_{t \in [0,T]} \left| \int_0^T \mathcal{R} \left( (Z(u) \circ J^{-1})^* X^k(u) \right) \circ J dW \right| \right) dP
\]

\[
\leq C + 2 \int_\Omega \left( \sup_{t \in [0,T]} \left| \int_0^T \mathcal{R} \left( (Z(u) \circ J^{-1})^* X^k(u) \right) \circ J dW \right| \right) dP
\]

where the result of Lemma 24.2.1 that \( X^k \) converges to \( \tilde{X} \) in \( K \) shows the term \( 2 ||Y||_{K'} ||X^k||_K \) is bounded. Note that the constant \( C \) is a continuous function of

\[
||Y||_{K'}, ||\tilde{X}||_{K'}, ||Z||_J, ||X_0||_{L^2(\Omega,H)}
\]

which equals zero when all are equal to zero. The term involving the stochastic integral is next.

Applying the Burkholder-Davis-Gundy inequality, Theorem 24.2 for \( F(r) = r \) along with the description of the quadratic variation of the Ito integral found in Corollary 17.1.1

\[
\int_\Omega \sup_{t_j \in P_k} |X(t_k)|^2 dP \leq C + C \int_\Omega \left( \int_0^T \left| \mathcal{R} \left( (Z(u) \circ J^{-1})^* X^k(u) \right) \circ J \right|^2 du \right)^{1/2} dP
\]

\[
\leq C + C \int_\Omega \left( \int_0^T \left| Z(u) \right|^2 |X^k(u)|^2 du \right)^{1/2} dP
\]

Now for each \( \omega \), there are only finitely many values of \( X^k(u) \) and they equal \( X(t_j) \) for \( t_j \in P_k \) with the convention that \( X(0) = 0 \). Therefore, the above is dominated by

\[
C + C \int_\Omega \left( \sup_{t_j \in P_k} |X(t_j)|^2 \right)^{1/2} \left( \int_0^T \left| Z(u) \right|^2 du \right)^{1/2} dP
\]
24.4. THE MAIN ESTIMATE

\[ \leq C + \frac{1}{2} \int_{\Omega} \sup_{t \in \mathcal{P}_k} |X(t_j)|^2 + C \int_{\Omega} \int_0^T \|Z(u)\|^2_{L^2(Q^{1/2}_U,H)} \, du \, dP \]

and so

\[ \frac{1}{2} \int_{\Omega} \sup_{t \in \mathcal{P}_k} |X(t_k)|^2 \, dP \leq C \]

for some constant \( C \) independent of \( \mathcal{P}_k \) dependent on \( \int_{\Omega} \int_0^T \|Z(u)\|_{L^2(Q^{1/2}_U,H)} \, du \, dP \). This constant is dependent on \( \|Y\|_{K'}, \|\tilde{X}\|_{K'}, \|Z\|_{J'}, \|X_0\|_{L^2(\Omega,H)} \) and equals zero when all of these quantities equal 0.

Let \( D \) denote the union of all the \( \mathcal{P}_k \). Thus \( D \) is a dense subset of \([0,T]\) and it has just been shown that for a constant \( C \) independent of \( \mathcal{P}_k \),

\[ E \left( \sup_{t \in D} |X(t)|^2 \right) \leq C. \]

Let \( \{e_j\} \) be an orthonormal basis for \( H \) which is also contained in \( V \) and has the property that \( \text{span} \{e_k\}_{k=1}^\infty \) is dense in \( V \). I claim that for \( y \in V' \)

\[ |y|^2_H = \sup_n \sum_{j=1}^n \langle y, e_j \rangle^2 \]

This is certainly true if \( y \in H \) because in this case

\[ \langle y, e_j \rangle = (y, e_j) \]

If \( y \notin H \), then the series must diverge. If not, you could consider the infinite sum

\[ z \equiv \sum_{j=1}^\infty \langle y, e_j \rangle e_j \in H \]

and argue that \( \langle z - y, v \rangle = 0 \) for all \( v \in \text{span} \{e_k\}_{k=1}^\infty \) which would also imply that this is true for all \( v \in V \). Then since \( z = y \) in \( V' \), it follows that \( y \in H \) contrary to the assumption that \( y \notin H \).

It follows

\[ |X(t)|^2 = \sup_n \sum_{j=1}^n |\langle X(t), e_j \rangle|^2 \]

and for a.e. \( \omega \), this is just the sup of continuous functions of \( t \). Therefore, for given \( \omega \) off a set of measure zero,

\[ t \to |X(t)|^2 \]

is lower semicontinuous. Hence letting \( t \in [0,T] \) and \( t_j \to t \) where \( t_j \in D \),

\[ |X(t)|^2 \leq \liminf_{j \to \infty} |X(t_j)|^2 \]

so it follows for a.e. \( \omega \)

\[ \sup_{t \in [0,T]} |X(t)|^2 \leq \sup_{t \in D} |X(t)|^2 \leq \sup_{t \in [0,T]} |X(t)|^2 \]

Hence

\[ E \left( \sup_{t \in [0,T]} |X(t)|^2 \right) \leq C \left( \|Y\|_{K'}, \|\tilde{X}\|_{K'}, \|Z\|_{J'}, \|X_0\|_{L^2(\Omega,H)} \right). \tag{24.4.9} \]

Note the above shows that for a.e. \( \omega \), \( \sup_{t \in [0,T]} |X(t)|^H < \infty \) so that for such \( \omega \), \( X(t) \) has values in \( H \). Note that we began by assuming it had a representative with values in \( H \) although the equation only held in \( V' \). Say

\[ |X(t,\omega)| \leq C(\omega). \]

Hence if \( v \in V \), then for a.e. \( \omega \)

\[ \lim_{t \to s} (X(t), v) = \lim_{t \to s} (X(s), v) = (X(s), v) = (X(s), v) \]

Therefore, since for such \( \omega \), \( |X(t,\omega)| \) is bounded, the above holds for all \( h \in H \) in place of \( v \) as well. Therefore, for a.e. \( \omega, t \to X(t,\omega) \) is weakly continuous with values in \( H \).
Eventually, it is shown that in fact, the function $t \to X(t, \omega)$ is continuous with values in $H$.

This lemma also provides a way to simplify one of the formulas derived earlier in the case that $X_0 \in L^p(\Omega, V)$. Refer to \ref{24.4.10}. One term there is $|X(t_1) - X_0 - M(t_1)|^2$.

$$|X(t_1) - X_0 - M(t_1)|^2 \leq 2|X(t_1) - X_0|^2 + 2|M(t_1)|^2$$

It was shown above that $2|M(t_1)|^2 \to 0$ a.e. and also in $L^1(\Omega)$ as $k \to \infty$. Apply the above lemma to $|X(t) - X_0|^2$ using $[0, t_1]$ instead of $[0, T]$. The new $X_0$ equals 0. Then from the estimate \ref{24.4.10}, it follows that

$$E\left(|X(t_1) - X_0|^2\right) \to 0$$

as $k \to \infty$. Taking a subsequence, we could also assume that $|X(t_1) - X_0|^2 \to 0$ a.e. $\omega$ as $k \to \infty$. Then, using this subsequence, it would follow from \ref{24.4.10}.

$$|X(t_m)|^2 - |X_0|^2 = 2\int_0^{t_m} \langle Y(u), X_k^+(u) \rangle du + e(k)$$

$$+ 2\int_0^{t_m} \mathcal{R}\left((Z(u) \circ J^{-1})^* X_k^+(u)\right) \circ J dW + \sum_{j=0}^{m-1} |M(t_{j+1}) - M(t_j)|^2$$

$$- \sum_{j=1}^{m-1} |X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j))|^2$$

(24.4.10)

where $e(k) \to 0$ in $L^1(\Omega)$ and a.e. $\omega$.

Can you obtain something similar even in case $X_0$ is not assumed to be in $L^p(\Omega, V)$? Let $Z_{0k} \in L^p(\Omega, V) \cap L^2(\Omega, H), Z_{0k} \to X_0$ in $L^2(\Omega, H)$. Then

$$|X(t_1) - X_0| \leq |X(t_1) - Z_{0k}| + |Z_{0k} - X_0|$$

Also, restoring the superscript to identify the parition,

$$X(t_1^k) - Z_{0k} = X_0 - Z_{0k} + \int_0^{t_1^k} Y(s) ds + \int_0^{t_1^k} Z(s) dW.$$}

Of course $\|\bar{X} - Z_{0k}\|_{K^{\prime}_{i_1 t_1^k}}$ is not bounded but for each $k$ it is at least finite. There is a sequence of partitions $\mathcal{P}_k, \|\mathcal{P}_k\| \to 0$ such that all the above holds. In the definitions of $K, K', J$ replace $[0, t]$ with $[0, t]$ and let the resulting spaces be denoted by $K_t, K'_t, J_t$. Let $n_k$ denote a subsequence of $\{k\}$ such that

$$\|\bar{X} - Z_{0k}\|_{K^{\prime}_{i_1 n_k}} < 1/k.$$}

Then from the above lemma,

$$E\left(\sup_{t \in [0, t_1^{n_k}]} |X(t_1^{n_k}) - Z_{0k}|^2_H\right)$$

$$\leq C\left(\|Y\|_{K''_{i_1 n_k}}, \|\bar{X} - Z_{0k}\|_{K^{\prime}_{i_1 n_k}}, \|Z\|_{J_{i_1 n_k}}, \|X_0 - Z_{0k}\|_{L^2(\Omega, H)}\right)$$

$$\leq C\left(\|Y\|_{K''_{i_1 n_k}}, \frac{1}{k}, \|Z\|_{J_{i_1 n_k}}, \|X_0 - Z_{0k}\|_{L^2(\Omega, H)}\right)$$

Hence

$$E\left(|X(t_1^{n_k}) - X_0|^2\right) \leq 2E\left(|X(t_1^{n_k}) - Z_{0k}|^2_H\right) + 2E\left(|Z_{0k} - X_0|^2_H\right)$$

$$\leq 2C\left(\|Y\|_{K''_{i_1 n_k}}, \frac{1}{k}, \|Z\|_{J_{i_1 n_k}}, \|X_0 - Z_{0k}\|_{L^2(\Omega, H)}\right) + 2\|Z_{0k} - X_0\|^2_H$$

which converges to 0 as $k \to \infty$. It follows that there exists a suitable subsequence such that \ref{24.4.10} holds even in the case that $X_0$ is only known to be in $L^2(\Omega, H)$. From now on, assume this subsequence for the paritions $\mathcal{P}_k$. Thus $k$ will really be $n_k$. 
24.5 Converging In Probability

I am working toward the Ito formula. In order to get this, there is a technical result which will be needed.

**Lemma 24.5.1** Let \( X (s) - X_k^t (s) \equiv \Delta_k (s) \). Then the following limit occurs.

\[
\lim_{k \to \infty} P \left( \sup_{t \in [0,T]} \left| \int_0^t R \left( (Z(s) \circ J^{-1})^* \Delta_k (s) \right) \circ JdW (s) \right| \geq \varepsilon \right) = 0. \tag{24.5.11}
\]

That is,

\[
\sup_{t \in [0,T]} \left| \int_0^t R \left( (Z(s) \circ J^{-1})^* \left( X(s) - X_k^t (s) \right) \right) \circ JdW (s) \right|
\]

converges to 0 in probability. Also the stochastic integral makes sense because \( X \) is \( H \) predictable.

**Proof:** First note that from Lemma 24.5.1, for a.e. \( \omega \), \( X(t) \) has values in \( H \) for \( t \in [0,T] \) and so it makes sense to consider it in the stochastic integral provided it is \( H \) progressively measurable. However, as noted in Situation 24.5.1, this function is automatically \( V' \) predictable. Therefore,

\[
(X(t),v) = (X(t),v)
\]

is real predictable for every \( v \in V \). Now if \( h \in H \), let \( v_n \to h \) in \( H \) and so for each \( \omega \),

\[
(X(t,\omega),v_n) \to (X(t,\omega),h)
\]

By the Pettis theorem, \( X \) is \( H \) predictable, hence progressively measurable. Also it was shown above that \( t \to X(t) \) is weakly continuous into \( H \). Therefore, the desired result follows from Lemma 24.5.1 on Page 102.

24.6 The Ito Formula

Now at long last, here is the first version of the Ito formula.

**Lemma 24.6.1** In Situation 24.3.1, let \( D \) be as above, the union of all the positive mesh points for all the \( \mathcal{P}_k \). Also assume \( X_0 \in L^2 (\Omega; H) \). Then for every \( t \in D \),

\[
|X(t)|^2 = |X_0|^2 + \int_0^t \left( 2 \langle Y(s), \bar{X}(s) \rangle + \|Z(s)\|^2_{L^2(\mathcal{Q}^{1/2}U; H)} \right) ds \\
+ 2 \int_0^t R \left( (Z(s) \circ J^{-1})^* X(s) \right) \circ JdW (s) \tag{24.6.12}
\]

Note that it was shown above that \( X(t,\omega) \) has values in \( H \) for a.e. \( \omega \).

**Proof:** Let \( t \in D \). Then \( t \in \mathcal{P}_k \) for all \( k \) large enough. Consider 24.3.11,

\[
|X(t)|^2 - |X_0|^2 = 2 \int_0^t \langle Y(u), X_k^t (u) \rangle du \\
+ 2 \int_0^t R \left( (Z(u) \circ J^{-1})^* X_k^t (u) \right) \circ JdW + \sum_{j=0}^{q_k-1} |M(t_{j+1}) - M(t_j)|^2 \\
- \sum_{j=1}^{q_k-1} \left| X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j)) \right|^2 + c(k) \tag{24.6.13}
\]

where \( t_{q_k} = t \). By Lemma 24.5.1 the second term on the right, the stochastic integral, converges to

\[
2 \int_0^t R \left( (Z(u) \circ J^{-1})^* \bar{X}(u) \right) \circ JdW
\]
in probability. The first term on the right converges to
\[ 2 \int_0^t \langle Y(u), \bar{X}(u) \rangle \, du \]
in \( L^1(\Omega) \) because \( X_k^* \to X \) in \( K \). Therefore, this also happens in probability. Consider the next term.

\[ E \left( \sum_{j=0}^{q_k-1} |M(t_{j+1}) - M(t_j)|^2 \right). \]

It is known from the theory of the quadratic variation that this term converges in probability to \( [M](t) = \int_0^t ||Z(s)||^2 \, ds \). See Theorem 4.7.1 on Page 323 and the description of the quadratic variation in Corollary 14.7.4.

Thus all the terms in (24.6.13) converge in probability except for the last term which also must converge in probability because it equals the sum of terms which do. It remains to find what this last term converges to. Thus

\[ |X(t)|^2 - |X_0|^2 = 2 \int_0^t \langle Y(u), \bar{X}(u) \rangle \, du + 2 \int_0^t R \left( (Z(u) \circ J^{-1})^\ast X(u) \right) \circ JdW + \int_0^t ||Z(s)||_2^2 \, ds - a \]

where \( a \) is the limit in probability of the term
\[ \sum_{j=1}^{q_k-1} |X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j))|^2 \]

Let \( P_n \) be the projection onto \( \text{span} \{ e_1, \cdots, e_n \} \) as before where \( \{ e_k \} \) is an orthonormal basis for \( H \) with each \( e_k \in V \). Then using
\[ X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j)) = \int_{t_j}^{t_{j+1}} Y(s) \, ds \]
the troublesome term above is of the form
\[ \sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \langle Y(s), X(t_{j+1}) - X(t_j) - P_n (M(t_{j+1}) - M(t_j)) \rangle \, ds \]

(24.6.14)

- \[ - \sum_{j=1}^{q_k-1} \langle X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j)), (I - P_n)(M(t_{j+1}) - M(t_j)) \rangle \]

(24.6.15)

The sum in (24.6.14) is dominated by
\[ \left( \sum_{j=1}^{q_k-1} |X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j))|^2 \right)^{1/2} \]

and
\[ \left( \sum_{j=1}^{q_k-1} |(I - P_n)(M(t_{j+1}) - M(t_j))|^2 \right)^{1/2} \]

(24.6.16)

Now it is known that \( \sum_{j=1}^{q_k-1} |X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j))|^2 \) converges in probability to \( a \). If you take the expectation of the other factor it is
\[ E \left( \sum_{j=1}^{q_k-1} \left( I - P_n \right) \int_{t_j}^{t_{j+1}} Z(s) \, dW(s) \right)^2 \]

\[ = \sum_{j=1}^{q_k-1} E \left( \int_{t_j}^{t_{j+1}} \left( I - P_n \right) Z(s) \, dW(s) \right)^2 \]
The integrand converges to 0 as $n \to \infty$ and is dominated by $\sum_{i=1}^{\infty} (Z(s), e_i)^2$ which is given to be in $L^1([0,T] \times \Omega)$. Therefore, it converges to 0.

Thus the expression in (24.6.10) is of the form $f_k g_{nk}$ where $f_k$ converges in probability to $a$ as $k \to \infty$ and $g_{nk}$ converges in probability to 0 as $n \to \infty$ independently of $k$. Now this implies $f_k g_{nk}$ converges in probability to 0. Here is why,

$$P (\|f_k g_{nk}\| > \varepsilon) \leq P (2\delta |f_k| > \varepsilon) + P (2C_\delta |g_{nk}| > \varepsilon) \leq P (2\delta |f_k - a| + 2\delta |a| > \varepsilon) + P (2C_\delta |g_{nk}| > \varepsilon)$$

where $\delta |f_k| + C_\delta |g_{nk}| > |f_k g_{nk}|$ and $\lim_{\delta \to 0} C_\delta = \infty$. Pick $\delta$ small enough that $\varepsilon - 2\delta |a| > \varepsilon/2$. Then this is dominated by

$$\leq P (2\delta |f_k - a| > \varepsilon/2) + P (2C_\delta |g_{nk}| > \varepsilon)$$

Fix $n$ large enough that the second term is less than $\eta$. Now taking $k$ large enough, the above is less than $\eta$. It follows the expression in (24.6.10) and consequently in (24.6.13) converges to 0 in probability.

Now consider the other term, (24.6.12) using the $n$ just determined. This term is of the form

$$\sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \langle Y(s), X_k^j(s) - X_k^j(s) - P_n (M_k^j(s) - M_k^j(s)) \rangle ds$$

$$= \int_{t_1}^{t} \langle Y(s), X_k^j(s) - X_k^j(s) - P_n (M_k^j(s) - M_k^j(s)) \rangle ds$$

where $M_k^j$ denotes the step function

$$M_k^j(t) = \sum_{i=0}^{m_k-1} M (t_{i+1}) \chi_{\{t_i, t_{i+1} \}} (t)$$

and $M_k^j$ is defined similarly. The term

$$\int_{t_1}^{t} \langle Y(s), P_n (M_k^j(s) - M_k^j(s)) \rangle ds$$

converges to 0 for a.e. $\omega$ as $k \to \infty$. This is because the integrand converges to 0 thanks to the continuity of $M(t)$ and also since this is a projection onto a finite dimensional subspace of $V$. Therefore, for each $\omega$ off a set of measure zero,

$$\int_{t_1}^{t} \langle Y(s), P_n (M_k^j(s) - M_k^j(s)) \rangle ds$$

$$\leq \int_{t_1}^{t} \|Y(s)\|_V, \|P_n (M_k^j(s) - M_k^j(s))\|_V ds$$

and this last integral converges to 0 as $k \to \infty$ because $P_n (M(s))$ is uniformly bounded in $V$ so there is no problem getting a dominating function for the dominated convergence theorem. Let

$$A_k = \left[ \int_{t_1}^{t} \|Y(s)\|_V, \|P_n (M_k^j(s) - M_k^j(s))\|_V ds > \varepsilon \right]$$
Then since the partitions are increasing, these sets are decreasing as \( k \) increases and their intersection has measure zero. Hence \( P (A_k) \to 0 \). It follows that

\[
\lim_{k \to \infty} P \left( \left| \int_{t_1}^t \langle Y (s), P_n (M_k^r (s) - M_k^l (s)) \rangle \, ds \right| > \varepsilon \right) \leq
\]

\[
\lim_{k \to \infty} P \left( \left| \int_{t_1}^t \| Y (s) \|_{L^1} \, \| P_n (M_k^r (s) - M_k^l (s)) \|_{L^1} \, ds \right| > \varepsilon \right) = 0
\]

Now consider

\[
\int_{t_1}^t \langle Y (s), X^r_k (s) - X^l_k (s) \rangle \, ds
\]

This converges to 0 in \( L^1 (\Omega) \) because it is of the form

\[
\int_{t_1}^t \langle Y (s), X^r_k (s) \rangle \, ds - \int_{t_1}^t \langle Y (s), X^l_k (s) \rangle \, ds
\]

and both \( X^r_k \) and \( X^l_k \) converge to \( X \) in \( K \). Therefore, the expression

\[
\sum_{j=1}^{q_k-1} |X (t_{j+1}) - X (t_j) - (M (t_{j+1}) - M (t_j))|^2
\]

converges to 0 in probability. ■

In fact, the formula (24.6.14) is valid for all \( t \in [0, T] \).

**Theorem 24.6.2** In Situation (24.5.1), off a set of measure zero, for every \( t \in [0, T] \),

\[
|X (t)|^2 = |X_0|^2 + \int_0^t \left( 2 \langle Y (s), X (s) \rangle + \| Z (s) \|_{L^2 (Q^{1/2} \Omega, H)}^2 \right) \, ds
\]

\[
+ 2 \int_0^t \mathcal{R} \left( (Z (s) \circ J^{-1})^* X (s) \right) \circ J dW (s)
\]

(24.6.17)

Furthermore, for \( t \in [0, T] \), \( t \to X (t) \) is continuous as a map into \( H \) for a.e. \( \omega \). In addition to this,

\[
E \left( |X (t)|^2 \right) = E \left( |X_0|^2 \right) + E \left( \int_0^t \left( 2 \langle Y (s), X (s) \rangle + \| Z (s) \|_{L^2 (Q^{1/2} \Omega, H)}^2 \right) \, ds \right)
\]

(24.6.18)

**Proof:** Let \( t \notin D \). For \( t > 0 \), let \( t (k) \) denote the largest point of \( \mathcal{P}_k \) which is less than \( t \). Suppose \( t (m) < t (k) \). Hence \( m \leq k \). Then

\[
X (t (m)) = X_0 + \int_0^{t (m)} Y (s) \, ds + \int_0^{t (m)} Z (s) \, dW (s)
\]

a similar formula holding for \( X (t (k)) \). Thus for \( t > t (m) \),

\[
X (t) - X (t (m)) = \int_{t (m)}^t Y (s) \, ds + \int_{t (m)}^t Z (s) \, dW (s)
\]

which is the same sort of thing studied so far except that it starts at \( t (m) \) rather than at 0 and \( X_0 = 0 \). Therefore, from Lemma (24.6.1) it follows

\[
|X (t (k)) - X (t (m))|^2 = \int_{t (m)}^{t (k)} \left( 2 \langle Y (s), X (s) - X (t (m)) \rangle + \| Z (s) \|^2 \right) \, ds
\]

\[
+ 2 \int_{t (m)}^{t (k)} \mathcal{R} \left( (Z (s) \circ J^{-1})^* (X (s) - X (t (m))) \right) \circ J dW (s)
\]

(24.6.19)

Consider that last term. It equals

\[
2 \int_{t (m)}^{t (k)} \mathcal{R} \left( (Z (s) \circ J^{-1})^* (X (s) - X^l_m (s)) \right) \circ J dW (s)
\]

(24.6.20)
This is dominated by

\[
2 \int_0^{t(k)} \mathcal{R} \left( (Z(s) \circ J^{-1})^* (X(s) - X_m(s)) \right) \circ J dW(s) \\
- 2 \int_0^{t(m)} \mathcal{R} \left( (Z(s) \circ J^{-1})^* (X(s) - X_m(s)) \right) \circ J dW(s) \\
\leq 2 \int_0^{t(k)} \mathcal{R} \left( (Z(s) \circ J^{-1})^* (X(s) - X_m(s)) \right) \circ J dW(s) \\
+ 2 \int_0^{t(m)} \mathcal{R} \left( (Z(s) \circ J^{-1})^* (X(s) - X_m(s)) \right) \circ J dW(s) \\
\leq 4 \sup_{t \in [0, T]} \left| \int_0^t \mathcal{R} \left( (Z(s) \circ J^{-1})^* (X(s) - X_m(s)) \right) \circ J dW(s) \right|
\]

In Lemma 24.4.1, the above expression was shown to converge to 0 in probability. Therefore, by the usual appeal to the Borel Cantelli lemma, there is a subsequence still referred to as \( \{m\} \), such that it converges to 0 pointwise in \( \omega \) for all \( \omega \) off some set of measure 0 as \( m \to \infty \). It follows there is a set of measure 0 such that for \( \omega \) not in that set, \( 24.6.2 \) converges to 0 in \( \mathbb{R} \). Note that \( t > 0 \) is arbitrary. Similar reasoning shows the first term in the non stochastic integral of \( 24.6.12 \) is dominated by an expression of the form

\[
4 \int_0^T \left| \langle Y(s), \bar{X}(s) - X_m(s) \rangle \right| ds
\]

which clearly converges to 0 for \( \omega \) not in some set of measure zero because \( X_m \) converges in \( H \) to \( \bar{X} \). Finally, it is obvious that

\[
\lim_{m \to \infty} \int_{t(m)}^{t(k)} \|Z(s)\|^2 ds = 0 \text{ for a.e. } \omega
\]

due to the assumptions on \( Z \).

This shows that for \( \omega \) off a set of measure 0

\[
\lim_{m, k \to \infty} |X(t(k)) - X(t(m))|^2 = 0
\]

and so \( \{X(t(k))\}_{k=1}^{\infty} \) is a convergent sequence in \( H \). Does it converge to \( X(t) \)? Let \( \xi(t) \in H \) be what it converges to. Let \( v \in V \) then

\[
\langle \xi(t), v \rangle = \lim_{k \to \infty} \langle X(t(k)), v \rangle = \lim_{k \to \infty} \langle X(t(k)), v \rangle = \langle X(t), v \rangle = \langle x(t), v \rangle
\]

and now, since \( V \) is dense in \( H \), this implies \( \xi(t) = X(t) \).

Now for every \( t \in D \),

\[
|X(t)|^2 = |X_0|^2 + \int_0^t \left( 2 \langle Y(s), \bar{X}(s) \rangle + \|Z(s)\|^2 \right) ds \\
+ 2 \int_0^t \mathcal{R} \left( (Z(s) \circ J^{-1})^* \bar{X}(s) \right) \circ J dW(s)
\]

and so, using what was just shown along with the obvious continuity of the functions of \( t \) on the right of the equal sign, it follows the above holds for all \( t \in [0, T] \) off a set of measure zero.

It only remains to verify \( t \to X(t) \) is continuous with values in \( H \). However, the above shows \( t \to |X(t)|^2 \) is continuous and it was shown in Lemma 24.4.1 that \( t \to X(t) \) is weakly continuous into \( H \). Therefore, from the uniform convexity of the norm in \( H \) it follows \( t \to X(t) \) is continuous. This is very easy to see in Hilbert space. Say \( a_n \to a \) and \( |a_n| \to |a| \). From the parallelogram identity,

\[
|a_n - a|^2 + |a_n + a|^2 = 2|a_n|^2 + 2|a|^2
\]
so
\[ |a_n - a|^2 = 2|a_n|^2 + 2|a|^2 - \left( |a_n|^2 + 2(a_n, a) + |a|^2 \right) \]
Then taking \( \lim \sup \) both sides,
\[ 0 \leq \lim \sup_{n \to \infty} |a_n - a|^2 \leq 2|a|^2 + 2|a|^2 - \left( |a|^2 + 2(a, a) + |a|^2 \right) = 0. \]

Of course this fact also holds in any uniformly convex Banach space.

Now consider the last claim. If the last term in 24.6.1 was a martingale, then there would be nothing to prove. This is because if \( M(t) \) is a martingale which equals 0 when \( t = 0 \), then
\[ E(M(t)) = E(E(M(t) | F_0)) = E(M(0)) = 0. \]

However, that last term is unfortunately only a local martingale. One can obtain a localizing sequence as follows.
\[ \tau_n(\omega) \equiv \inf \{ t : |X(t, \omega)| > n \} \]
where as usual \( \inf(\emptyset) = \infty \). This is all right because it was shown above that \( t \to X(t, \omega) \) is continuous into \( H \) for a.e. \( \omega \). Then stopping both processes on the two sides of 24.6.17 with \( \tau_n \),
\[ |X(t \wedge \tau_n)|^2 = |X_0|^2 + \int_0^{t \wedge \tau_n} \left( 2 \langle Y(s), X(s) \rangle + ||Z(s)||_{L_2(Q^{1/2} U, H)}^2 \right) ds \]
\[ + 2 \int_0^{t \wedge \tau_n} \mathcal{R} \left( \left( Z(s) \circ J^{-1} \right)^* X(s) \right) \circ JdW(s) \]

Now from Lemma 17.10.1,
\[ |X(t \wedge \tau_n)|^2 = |X_0|^2 + \int_0^t X_{[0, \tau_n]}(s) \left( 2 \langle Y(s), X(s) \rangle + ||Z(s)||_{L_2(Q^{1/2} U, H)}^2 \right) ds \]
\[ + 2 \int_0^t X_{[0, \tau_n]}(s) \mathcal{R} \left( \left( Z(s) \circ J^{-1} \right)^* X(s) \right) \circ JdW(s) \]
That last term is now a martingale and so you can take the expectation of both sides. This gives
\[ E \left( |X(t \wedge \tau_n)|^2 \right) = E \left( |X_0|^2 \right) \]
\[ + E \left( \int_0^t X_{[0, \tau_n]}(s) \left( 2 \langle Y(s), X(s) \rangle + ||Z(s)||_{L_2(Q^{1/2} U, H)}^2 \right) ds \right) \]
Letting \( n \to \infty \) and using the dominated convergence theorem and \( \tau_n \to \infty \) yields the desired result. \( \blacksquare \)

**Notation 24.6.3** The stochastic integrals are unpleasant to look at.
\[ \int_0^t \mathcal{R} \left( \left( Z(s) \circ J^{-1} \right)^* X(s) \right) \circ JdW(s) \]
\[ \equiv \int_0^t (X(s), Z(s) dW(s)). \]
Chapter 25

The Hard Ito Formula, Implicit Case

25.1 Approximating With Step Functions

This Ito formula seems to be the fundamental idea which allows one to obtain solutions to stochastic partial differential equations using a variational point of view. I am following the treatment found in [72]. The following lemma is fundamental to the presentation. It approximates a function with a sequence of two step functions $X_r^k, X_l^k$ where $X_r^k$ has the value of $X$ at the right end of each interval and $X_l^k$ gives the value $X$ at the left end of the interval. The lemma is very interesting for its own sake. You can obviously do this sort of thing for a continuous function but here the function is not continuous and in addition, it is a stochastic process depending on $\omega$ also. This lemma was proved earlier, Lemma 17.3.1.

Lemma 25.1.1 Let $\Phi : [0,T] \times \Omega \rightarrow V, be B ([0,T]) \times \mathcal{F}$ measurable and suppose

$$\Phi \in K \equiv L^p ([0,T] \times \Omega ; E), p \geq 1$$

Then there exists a sequence of nested partitions, $P_k \subseteq P_{k+1}$,

$$P_k \equiv \{ t_0^k, \cdots , t_{m_k}^k \}$$

such that the step functions given by

$$\Phi_r^k (t) \equiv \sum_{j=1}^{m_k} \Phi (t_{j-1}^k, t_j^k) X(t_{j-1}^k, t_j^k) (t)$$

$$\Phi_l^k (t) \equiv \sum_{j=1}^{m_k} \Phi (t_{j-1}^k, t_j^k) X(t_{j-1}^k, t_j^k) (t)$$

both converge to $\Phi$ in $K$ as $k \rightarrow \infty$ and

$$\lim_{k \rightarrow \infty} \max \{|t_j^k - t_{j+1}^k| : j \in \{0, \cdots , m_k\} \} = 0.$$  

Also, each $\Phi (t_j^k), \Phi (t_{j-1}^k)$ is in $L^p (\Omega; E)$. One can also assume that $\Phi (0) = 0$. The mesh points $\{t_j^k\}_{j=0}^{m_k}$ can be chosen to miss a given set of measure zero. In addition to this, we can assume that

$$|t_j^k - t_{j-1}^k| = 2^{-n_k}$$

except for the case where $j = 1$ or $j = m_{n_k}$ when this might not be so. In the case of the last subinterval defined by the partition, we can assume

$$|t_m^k - t_{m-1}^k| = |T - t_{m-1}^k| \geq 2^{-(n_k+1)}$$

The following lemma is convenient.

Lemma 25.1.2 Let $f_n \rightarrow f$ in $L^p ([0,T] \times \Omega, E)$. Then there exists a subsequence $n_k$ and a set of measure zero $N$ such that if $\omega \notin N$, then

$$f_{n_k} (\cdot, \omega) \rightarrow f (\cdot, \omega)$$

in $L^p ([0,T], E)$ and for a.e. $t$. 

557
Proof: We have

\[ P\left( \|f_n - f\|_{L^p([0,T],E)} > \lambda \right) \leq \frac{1}{\lambda} \int_{\Omega} \|f_n - f\|_{L^p([0,T],E)} \, dP \]

\[ \leq \frac{1}{\lambda} \|f_n - f\|_{L^p([0,T] \times \Omega, E)} \]

Hence there exists a subsequence \( n_k \) such that

\[ P\left( \|f_{n_k} - f\|_{L^p([0,T],E)} > 2^{-k} \right) \leq 2^{-k} \]

Then by the Borel Cantelli lemma, it follows that there exists a set of measure zero \( N \) such that for all \( k \) large enough and \( \omega \notin N \),

\[ \|f_{n_k} - f\|_{L^p([0,T],E)} \leq 2^{-k} \]

Now by the usual arguments used in proving completeness, \( f_{n_k}(t) \to f(t) \) for a.e. \( t \).

Because of this lemma, it can also be assumed that for a.e. \( \omega \), pointwise convergence is obtained on \([0,T]\) as well as convergence in \( L^p([0,T]) \). This kind of assumption will be tacitly made whenever convenient.

Also recall the diagram for the definition of the integral which has values in a Hilbert space \( W \).

\[
\begin{array}{c}
U \\
\downarrow \\
Q^{1/2} \\
\uparrow \\
\frac{1}{2} \\
U_1 \\
\supseteq J Q^{1/2} U \\
\downarrow \\
Q^{1/2} U \\
\uparrow \\
Z \leftarrow J \\
\uparrow \\
W \\
\downarrow \\
Z_n \searrow \\
\downarrow \\
Z \\
\downarrow \\
W
\end{array}
\]

The idea was to get \( \int_0^t Z dW \) where \( Z \in L^2([0,T] \times \Omega; L_2(Q^{1/2} U, W)) \). Here \( W(t) \) was a cylindrical Wiener process. This meant that it was a \( Q_1 \) Wiener process on \( U_1 \) for \( Q_1 = JJ^* \) and \( J \) was a Hilbert Schmidt operator mapping \( Q^{1/2} U \) to \( U_1 \). To get \( \int_0^t Z dW \), \( Z \circ J^{-1} \) was approximated by a sequence of elementary functions having values in \( \mathcal{L}(U_1, W) \). Then

\[ \int_0^t Z dW \equiv \lim_{n \to \infty} \int_0^t Z_n dW \]

and this limit existed in \( L^2(\Omega, W) \).

25.2 The Situation

Now consider the following situation. There are real separable Banach spaces \( V, W \) such that \( W \) is a Hilbert space and \( V \subseteq W \), \( W' \subseteq V' \)

where \( V \) is dense in \( W \). Also let \( B \in \mathcal{L}(W, W') \) satisfy

\[ \langle Bw, w \rangle \geq 0, \quad \langle Bu, v \rangle = \langle Bv, u \rangle \]

Note that \( B \) does not need to be one to one. Also allowed is the case where \( B \) is the Riesz map. It could also happen that \( V = W \).

Situation 25.2.1 Let \( X \) have values in \( V \) and satisfy the following

\[ BX(t) = BX_0 + \int_0^t Y(s) \, ds + B \int_0^t Z(s) \, dW(s), \quad (25.2.1) \]

\( X_0 \in L^2(\Omega; W) \) and is \( \mathcal{F}_0 \) measurable, where \( Z \) is \( \mathcal{L}_2(Q^{1/2} U, W) \) progressively measurable and

\[ \|Z\|_{L^2([0,T] \times \Omega, \mathcal{L}_2(Q^{1/2} U, W))} < \infty. \]

This is what is needed to define the stochastic integral in the above formula.
Assume $X,Y$ satisfy
\[ BX,Y \in K' \equiv L^p ([0,T] \times \Omega; V'), \]
the $\sigma$ algebra of measurable sets defining $K'$ will be the progressively measurable sets. Here $1/p' + 1/p = 1$, $p > 1$.

Also the sense in which the equation holds is as follows. For a.e. $\omega$, the equation holds in $V'$ for all $t \in [0,T]$. Thus we are considering a particular representative $X$ of $K$ for which this happens. Also it is only assumed that $BX(t) = B(X(t))$ for a.e. $t$. Thus $BX$ is the name of a function having values in $V'$ for which $BX(t) = B(X(t))$ for a.e. $t$. Assume that $X$ is progressively measurable also and
\[ X \in L^p ([0,T] \times \Omega, V) \]
Also $W(t)$ is a $JJ^*$ Wiener process on $U_1$ in the above diagram.

The goal is to prove the following Ito formula valid for a.e. $t$ for each $\omega$ off a set of measure zero.
\[
\langle BX(t), X(t) \rangle = \langle BX_0, X_0 \rangle + \int_0^t \left( 2 \langle Y(s), X(s) \rangle + \langle BZ, Z \rangle_{L_2} \right) ds
\]
\[
+ \int_0^t (Z \circ J^{-1})^* BX \circ JdW
\]
(25.2.2)

The most significant feature of the last term is that it is a local martingale. The term $\langle BZ, Z \rangle_{L_2}$ will be discussed later, as will the meaning of the stochastic integral.

The idea is that $(Z \circ J^{-1})^* BX \circ J$ has values in $L_2 \left( (Q^{1/2}U, R) \right)$ and so it makes sense to consider this stochastic integral. To see this, $BX \in W'$ and $(Z \circ J^{-1})^* \in L_2 \left( W', (JQ^{1/2}U)' \right)$ and so
\[
(Z \circ J^{-1})^* BX \in \left( JQ^{1/2}U \right)'
\]
and so $(Z \circ J^{-1})^* BX \circ J \in L_2 \left( (Q^{1/2}U, R) = (Q^{1/2}U)' \right)$. Note that in general $H' = L_2 (H, R)$ because if you have 
\[
\sum_i \langle (R^{-1}f, e_i) \rangle = \sum_i \langle (f,e_i) \rangle = \|f\|_{H'}^2.
\]

The main item of interest relative to this stochastic integral will be a statement about its quadratic variation. It appears to depend on $J$ but this is not the case because the other terms in the formula do not.

### 25.3 Preliminary Results

Here are discussed some preliminary results which will be needed. From the integral equation, if $\phi \in L^q (\Omega; V)$ and $\psi \in C_\infty^0 (0,T)$ for $q = \max (p,2)$,
\[
\int_\Omega \int_0^T \left( (BX)(t) - B \int_0^t Z(s) dW(s) - BX_0 \right) \psi' \phi dt dP
\]
\[
= \int_\Omega \int_0^T \int_0^t Y(s) \psi' (t) ds \phi dt dP
\]
Then the term on the right equals
\[
\int_\Omega \int_0^T \int_t^T Y(s) \psi' (t) ds d\phi (\omega) dP = \int_\Omega \left( \int_0^T (Y(s) \psi (s)) dP - \int_0^T (Y(s) \psi (s)) ds \right) \phi (\omega) dP
\]
It follows that, since $\phi$ is arbitrary,
\[
\int_0^T \left( (BX)(t) - B \int_0^t Z(s) dW(s) - BX_0 \right) \psi' (t) dt = - \int_0^T \int_0^T Y(s) \psi (s) ds
\]
in $L^q (\Omega; V')$ and so the weak time derivative of
\[
t \to (BX)(t) - B \int_0^t Z(s) dW(s) - BX_0
\]
equals \( Y \) in \( L^q \left( [0, T]; L^r (\Omega, V') \right) \). Thus, by Theorem 25.3.1, for a.e. \( t \), say \( t \notin \tilde{N} \subseteq [0, T] \), \( m \left( \tilde{N} \right) = 0 \),

\[
B \left( X (t) - \int_0^t Z (s) \, dW (s) \right) = B X_0 + \int_0^t Y (s) \, ds \text{ in } L^q (\Omega, V') .
\]

That is,

\[
\langle B X (t) \rangle (t) = B X_0 + \int_0^t Y (s) \, ds + B \int_0^t Z (s) \, dW (s)
\]

holds in \( L^q (\Omega, V') \) where \( \langle B X (t) \rangle (t) = B (X (t)) \) a.e. \( t \), in addition to holding for all \( t \) for each \( \omega \). Now let \( t^n_k \) be partitions for which, from Lemma 25.3.1 there are left and right step functions \( X^t_k, X^r_k \), which converge in \( L^p ([0, T] \times \Omega; V) \) to \( X \) and such that each \( \{ t^n_k \} \) has empty intersection with the set of measure zero \( N \) where, in \( L^p (\Omega; V') \), \( \langle B X (t) \rangle (t) \neq B (X (t)) \) in \( L^q (\Omega, V') \). Thus for \( t_k \) a generic partition point,

\[
BX (t_k) = B (X (t_k)) \text{ in } L^q (\Omega; V')
\]

Hence there is an exceptional set of measure zero, \( N (t_k) \subseteq \Omega \) such that for \( \omega \notin N (t_k) \), \( BX (t_k) (\omega) = B (X (t_k, \omega)) \).

We define an exceptional set \( N \subseteq \Omega \) to be the union of all these \( N (t_k) \). There are countably many and so \( N \) is also a set of measure zero. Then for \( \omega \notin N \), and \( t_k \) any mesh point at all, \( BX (t_k) (\omega) = B (X (t_k, \omega)) \). This will be important in what follows. In addition to this, from the integral equation, for each of these \( \omega \notin N \), \( BX (t) (\omega) = B (X (t, \omega)) \) for all \( t \notin N \) where \( N \) is a set of Lebesgue measure zero. Thus the \( t_k \) from the various partitions are always in \( N \). By Lemma 25.3.2 there exists a countable set \( \{ e_i \} \) of vectors in \( V \) such that

\[
\langle B e_i, e_j \rangle = \delta_{ij}
\]

and for each \( x \in W \),

\[
\langle B x, x \rangle = \sum_{i=0}^{\infty} | \langle B x, e_i \rangle |^2 , \quad B x = \sum_{i=1}^{\infty} \langle B x, e_i \rangle B e_i
\]

Thus the conclusion of the above discussion is that at the mesh points, it is valid to write

\[
\langle (BX) (t_k), X (t_k) \rangle = \langle B (X (t_k)), X (t_k) \rangle = \sum_{i} \langle (BX) (t_k), e_i \rangle^2 = \sum_{i} \langle B (X (t_k)), e_i \rangle^2
\]

just as would be the case if \( B X (t) = B (X (t)) \) for every \( t \). In all which follows, the mesh points will be like this and an appropriate set of measure zero which may be replaced with a larger set of measure zero finitely many times is being neglected. Obviously, one can take a subsequence of the sequence of partitions described above without disturbing the above observations. We will denote these partitions as \( P_k \). As a case of this, we obtain the following interesting lemma.

**Lemma 25.3.1** In the above situation, there exists a set of measure zero \( N \subseteq \Omega \) and a dense subset of \( [0, T] \), \( D \) such that for \( \omega \notin N \), \( BX (t, \omega) = B (X (t, \omega)) \) for all \( t \in D \).

**Theorem 25.3.2** Let \( Z \) be progressively measurable and in

\[
L^2 \left( [0, T] \times \Omega, \mathcal{L}_2 \left( Q^{1/2} U, W \right) \right) .
\]

Also suppose \( X \) is progressively measurable and in \( L^2 ([0, T] \times \Omega, W) \). Let \( \{ t^n_j \} \) be a sequence of partitions of the sort in Lemma 25.3.3 such that if

\[
X_n (t) = \sum_{j=0}^{m-1} X (t^n_j) \chi^n_j, t^n_j+1 (t) \equiv X^n (t)
\]

then \( X_n \to X \) in \( L^p ([0, T] \times \Omega, W) \). Also, it can be assumed that none of these mesh points are in the exceptional set off which \( BX (t) = B (X (t)) \) (Thus it will make no difference whether we write \( BX (t) \) or \( B (X (t)) \) in what follows for all \( t \) one of these mesh points.) Then the expression

\[
\sum_{j=0}^{m-1} \left\langle B \int_{t^n_j \wedge t}^{t^n_{j+1} \wedge t} Z dW, X (t^n_j) \right\rangle = \sum_{j=0}^{m-1} \left\langle BX (t^n_j), \int_{t^n_j \wedge t}^{t^n_{j+1} \wedge t} Z dW \right\rangle
\]

(25.3.3)
is a local martingale which can be written in the form
\[ \int_0^t (Z \circ J^{-1})^* BX_n \circ JdW \]
where
\[ X_n^t (t) = \sum_{k=0}^{m_n-1} X (t_k^n) \chi_{[t_k^n, t_{k+1}^n)} (t) \]

**Proof:** First suppose that \((BX (t_k^n), X (t_k^n)) \in L^\infty ([0, T] \times \Omega, W').\) Then from the definition of the integral, let \(U \in L^2 (\Omega)\) for each \(t \in [0, T].\)

Consider one of the terms of the sum more simply as
\[ \langle BX (t_j^n), \int_{t_j^n}^{t_{j+1}^n} ZdW \rangle \]
is in \(L^1 (\Omega)\) for each \(t\) since both entries are in \(L^2 (\Omega).\) Why is this a martingale?

\[
E \left( \langle BX (t_j^n), \int_{t_j^n}^{t_{j+1}^n} ZdW \rangle \mid \mathcal{F}_{t_j^n} \right) = E \left( \langle BX (t_j^n), \int_{t_j^n}^{t_{j+1}^n} ZdW \rangle \mid \mathcal{F}_{t_j^n} \right)
= E \left( \langle BX (t_j^n), E \left( \int_{t_j^n}^{t_{j+1}^n} ZdW \mid \mathcal{F}_{t_j^n} \right) \rangle \right)
= E \left( \langle BX (t_j^n), 0 \rangle \right) = 0
\]
because the stochastic integral is a martingale. Now let \(\sigma\) be a bounded stopping time.

\[
E \left( \langle BX (t_j^n), \int_{t_j^n}^{t_{j+1}^n} ZdW \rangle \mid \mathcal{F}_{t_j^n} \right)
= E \left( \langle BX (t_j^n), \int_{t_j^n}^{t_{j+1}^n} ZdW \rangle \mid \mathcal{F}_{t_j^n} \right)
= E \left( \langle BX (t_j^n), E \left( \int_{t_j^n}^{t_{j+1}^n} ZdW \mid \mathcal{F}_{t_j^n} \right) \rangle \right)
= E \left( \langle BX (t_j^n), 0 \rangle \right) = 0
\]
and so this is a martingale. I want to write the formula in (25.3.4) as a stochastic integral. First note that \(W\) has values in \(U_1.\)

Consider one of the terms of the sum more simply as
\[
\langle B \int_a^b ZdW, X (a) \rangle, \quad a = t_k^n \wedge t, \quad b = t_{k+1}^n \wedge t.
\]

Then from the definition of the integral, let \(Z_n\) be a sequence of elementary functions converging to \(Z \circ J^{-1}\) in \(L^2 ([a, b] \times \Omega, L_2 (JQ^{1/2} U, W))\)

\[
\left\| \int_a^t ZdW - \int_a^t Z_n dW \right\|_{L^2 (\Omega, W)} \to 0
\]

Using a maximal inequality and the fact that the two integrals are martingales along with the Borel Cantelli lemma, there exists a set of measure 0 \(N\) such that for \(\omega \notin N\), the convergence of a suitable subsequence of these integrals, still denoted by \(n\), is uniform for \(t \in [a, b].\) It follows that for such \(\omega,\)
\[
\langle B \int_a^t ZdW, X (a) \rangle = \lim_{n \to \infty} \langle B \int_a^t Z_n dW, X (a) \rangle.
\]

Say
\[
Z_n (u) = \sum_{k=0}^{m_n-1} Z_k^n \chi_{[t_k^n, t_{k+1}^n)} (u)
\]
where \( Z^n_k \) has finitely many values in \( \mathcal{L}(U_1, W)_0 \), the restrictions of maps in \( \mathcal{L}(U_1, W) \) to \( JQ^{1/2}U \), and the \( t^n_k \) refer to a partition of \([a, b]\). Then the product on the right in (25.3.3) is of the form

\[
\sum_{k=0}^{m-1} \langle BZ^n_k (W (t \wedge t^n_{k+1}) - W (t \wedge t^n_k)), X (a) \rangle_{W', W}.
\]

Note that it makes sense because \( Z^n_k \) is the restriction to \( J \left( Q^{1/2}U \right) \) of a map from \( U_1 \) to \( W \) and so \( BZ^n_k \) is a map from \( U_1 \) to \( W' \). Then the Wiener process has values in \( U_1 \) so when you apply \( BZ^n_k \) to \( W (t \wedge t^n_{k+1}) - W (t \wedge t^n_k) \), you get something in \( W' \) and so the duality pairing is between \( W' \) and \( W \) as shown. Also, \( Z^n_k (W (t \wedge t^n_{k+1}) - W (t \wedge t^n_k)) \) gives something in \( W \) because the Wiener process has values in \( U_1 \) and \( Z^n_k \) acts on these things to give something in \( W \). Thus the above equals

\[
= \sum_{k=0}^{m-1} \langle BX (a), Z^n_k (W (t \wedge t^n_{k+1}) - W (t \wedge t^n_k)) \rangle_{W', W}.
\]

Recall also that the space on the left is dense in the one on the right. Now let \( \{g_i\} \) be an orthonormal basis for \( Q^{1/2}U \), so that \( \{Jg_i\} \) is an orthonormal basis for \( JQ^{1/2}U \). Then

\[
\sum_{i=1}^{\infty} \left| \langle (Z^n)_i (a) - (Z \circ J^{-1})_i (a) \rangle (Jg_i) \right|^2 = \sum_{i=1}^{\infty} \left| \langle BX (a), (Z_n - Z \circ J^{-1})_i (Jg_i) \rangle \right|^2
\]

\[
\leq \langle BX (a), X (a) \rangle \sum_{i=1}^{\infty} \langle B (Z_n - Z \circ J^{-1}) (Jg_i), (Z_n - Z \circ J^{-1}) (Jg_i) \rangle
\]

\[
\leq \langle BX (a), X (a) \rangle \|B\| \sum_{i=1}^{\infty} \left\| (Z_n - Z \circ J^{-1}) (Jg_i) \right\|^2_{W} = \langle BX (a), X (a) \rangle \|B\| \left\| Z_n - Z \circ J^{-1} \right\|^2_{L_2(JQ^{1/2}U, W)}
\]

When integrated over \([a, b] \times \Omega\), it is given that this converges to 0, assuming that \( \langle BX (a), X (a) \rangle \in L^\infty (\Omega) \), which is assumed for now.

It follows that, with this assumption,

\[
Z^n_k BX (a) \to (Z \circ J^{-1}) (a)
\]

in \( L^2 ([a, b] \times \Omega, L_2 (JQ^{1/2}U, R)) \). Writing this differently, it says

\[
Z^n_k BX (a) \to (Z \circ J^{-1}) (a) \circ J^{-1} \text{ in } L^2 ([a, b] \times \Omega, L_2 (JQ^{1/2}U, R))
\]
It follows from the definition of the integral that the Ito integrals converge. Therefore,

\[ \left\langle B \int_a^t ZdW, X(a) \right\rangle = \int_a^t (Z \circ J^{-1})^* BX(a) \circ JdW \]

The term on the right is a martingale because the one on the left is.

Next it is necessary to drop the assumption that \( \langle BX(a), X(a) \rangle \in L^\infty(\Omega) \). Note that \( X^l \) is right continuous and \( BX^l_n \) progressively measurable. Thus,

\[ \langle BX^l_n(t), X^l_n(t) \rangle = \sum_i \langle BX^l_n(t), e_i \rangle^2 \]

where \( \{e_i\} \) is the set defined in Lemma 25.2.1 each in \( V \). Thus \( \langle BX^l_n, X^l_n \rangle \) is also progressively measurable and right continuous, and one can define the stopping time

\[ \sigma_q^n \equiv \inf \{ t : \langle BX^l_n(t), X^l_n(t) \rangle > q \}, \quad (25.3.5) \]

the first hitting time of an open set. Also, for each \( \omega \), there are only finitely many values for \( \langle BX^l_n(t), X^l_n(t) \rangle \) and so \( \sigma_q^n = \infty \) for all \( q \) large enough.

From localization,

\[ \left\langle B \int_{a \wedge \sigma_q^n}^t ZdW, X(a) \right\rangle = \left\langle B \int_a^t X_{[0,\sigma_q^n]} ZdW, X(a) \right\rangle \]

\[ = \int_a^t \left( X_{[0,\sigma_q^n]} \circ J^{-1} \right)^* BX(a) \circ JdW \]

\[ = \int_{a \wedge \sigma_q^n}^t (Z \circ J^{-1})^* BX(a) \circ JdW \]

Then it follows that, using the stopping time,

\[ \sum_{j=0}^{m_n-1} \left\langle B \int_{t^n_j \wedge \sigma_q^n}^{t^n_{j+1} \wedge \sigma_q^n} ZdW, X(t^n_j) \right\rangle = \int_0^{t \wedge \sigma_q^n} (Z \circ J^{-1})^* BX^l_n \circ JdW \]

where \( X^l_n \) is the step function

\[ X^l_n(t) = \sum_{k=0}^{m_n-1} X(t^n_k) 1_{[t^n_k, t^n_{k+1})}(t) . \]

Thus the given sum equals the local martingale

\[ \int_0^t (Z \circ J^{-1})^* BX^l_n \circ JdW. \]

Note that the sum does not depend on \( J \) or on \( U_1 \) so the same must be true of what it equals although it does not look that way. The question of convergence as \( n \to \infty \) is considered next.

What follows is the main estimate and discrete formulas.

### 25.4 The Main Estimate

The argument will be based on a formula which follows in the next lemma.

**Lemma 25.4.1** In Situation 25.2.1 the following formula holds for a.e. \( \omega \) for \( 0 < s < t \) where \( M(t) \equiv \int_0^t Z(u) dW(u) \) which has values in \( W \). In the following, \( \langle \cdot, \cdot \rangle \) denotes the duality pairing between \( V, V' \).

\[ \langle BX(t), X(t) \rangle = \langle BX(s), X(s) \rangle + \]

\[ + 2 \int_s^t \langle Y(u), X(t) \rangle du + \langle B(M(t) - M(s)), M(t) - M(s) \rangle \]
\[- (BX(t) -BX(s) - (M(t) - M(s)), X(t) - X(s) - (M(t) - M(s))) \\
+ 2\langle BX(s), M(t) - M(s)\rangle \]  \hspace{1cm} (25.4.6)

Also for \( t > 0 \)

\[
\langle BX(t), X(t) \rangle = \langle BX_0, X_0 \rangle + 2 \int_0^t \langle Y(u), X(t) \rangle \, du + 2 \langle BX_0, M(t) \rangle + \\
\langle BM(t), M(t) \rangle - (BX(t) - BX_0 - BM(t), X(t) - X_0 - M(t)) \]  \hspace{1cm} (25.4.7)

**Proof:** From the formula which is assumed to hold,

\[
BX(t) = BX_0 + \int_0^t Y(u) \, du + BM(t) \\
BX(s) = BX_0 + \int_0^s Y(u) \, du + BM(s)
\]

Then

\[
BM(t) - BM(s) + \int_s^t Y(u) \, du = BX(t) - BX(s)
\]

It follows that

\[
\langle B(M(t) - M(s)), M(t) - M(s) \rangle - \\
\langle BX(t) - BX(s) - (M(t) - M(s)), X(t) - X(s) - (M(t) - M(s)) \rangle \\
+ 2\langle BX(s), M(t) - M(s) \rangle
\]

\[
= \langle B(M(t) - M(s)), M(t) - M(s) \rangle - \langle BX(t) - BX(s), X(t) - X(s) \rangle \\
+ 2\langle BX(t) - BX(s), M(t) - M(s) \rangle
\]

\[
- \langle B(M(t) - M(s)), M(t) - M(s) \rangle + 2\langle BX(s), M(t) - M(s) \rangle
\]

Some terms cancel and this yields

\[
= - \langle BX(t) - BX(s), X(t) - X(s) \rangle + 2\langle BX(t), M(t) - M(s) \rangle
\]

\[
= - \langle BX(t) - BX(s), X(t) - X(s) \rangle + 2\langle B(M(t) - M(s)), X(t) \rangle
\]

\[
= - \langle BX(t) - BX(s), X(t) - X(s) \rangle + 2\left( BX(t) - BX(s) - \int_s^t Y(u) \, du, X(t) \right)
\]

\[
= - \langle BX(t), X(t) \rangle - \langle BX(s), X(s) \rangle + 2\langle BX(t), X(t) \rangle
\]

\[
- 2\langle BX(s), X(t) \rangle - 2\int_s^t \langle Y(u), X(t) \rangle \, du
\]

Therefore,

\[
\langle BX(t), X(t) \rangle - \langle BX(s), X(s) \rangle = 2 \int_s^t \langle Y(u), X(t) \rangle \, du + \langle B(M(t) - M(s)), M(t) - M(s) \rangle
\]

\[
- \langle BX(t) - BX(s) - (M(t) - M(s)), X(t) - X(s) - (M(t) - M(s)) \rangle + 2\langle BX(s), M(t) - M(s) \rangle
\]

The case with \( X_0 \) is similar. \( \blacksquare \)

The following phenomenal estimate holds and it is this which is the main idea in proving the Ito formula. The last assertion about continuity is like the well known result that if \( y \in L^p(0, T; V) \) and \( y' \in L^p(0, T; V') \), then \( y \) is actually continuous a.e. with values in \( H \), for \( V, H, V' \) a Gelfand triple. Later, this continuity result is strengthened further to give strong continuity. In all of this, \( X^k_\frac{1}{2} \) and \( X^k_\frac{1}{2} \) are as described above, converging in \( K \) to \( X \).
Lemma 25.4.2 In the Situation \(25.4.4\), the following holds. For a.e. \(t\)
\[
E(\langle BX(t), X(t) \rangle) < C \left( \|Y\|_{K'}, \|X\|_{K}, \|Z\|_{J}, \|BX_0, X_0\|_{L^1(\Omega)} \right) < \infty.
\]
(25.4.8)
where \(K, K'\) were defined earlier and
\[
J = L^2 \left( [0, T] \times \Omega; \mathcal{L}_2 \left( Q^{1/2}U; W \right) \right)
\]
In fact,
\[
E \left( \sup_{t \in [0, T]} \sum_i \langle BX(t), e_i \rangle^2 \right) \leq C \left( \|Y\|_{K'}, \|X\|_{K}, \|Z\|_{J}, \|BX_0, X_0\|_{L^1(\Omega)} \right)
\]
Also, \(C\) is a continuous function of its arguments, increasing in each one, and \(C(0, 0, 0, 0) = 0\). Thus for a.e. \(\omega\),
\[
\sup_{t \not\in N^C} \langle BX(t, \omega), X(t, \omega) \rangle \leq C(\omega) < \infty.
\]
Also for \(\omega\) off a set of measure zero described earlier, \(t \to BX(t)(\omega)\) is weakly continuous with values in \(W'\) on \([0, T]\). Also \(t \to \langle BX(t), X(t) \rangle\) is lower semicontinuous on \(N^C\).

**Proof:** Consider the formula in Lemma \(25.4.1\)
\[
\langle BX(t), X(t) \rangle = \langle BX(s), X(s) \rangle + 2 \int_s^t \langle Y(u), X(t) \rangle \, du + \langle B(M(t) - M(s)), M(t) - M(s) \rangle
\]
\[
- \langle B(X(t) - X(s) - (M(t) - M(s))), X(t) - X(s) - (M(t) - M(s)) \rangle + 2 \langle BX(s), M(t) - M(s) \rangle
\]
(25.4.9)
Now let \(t_j\) denote a point of \(P_k\) from Lemma \(25.4.1\). Then for \(t_j > 0\), \(X(t_j)\) is just the value of \(X\) at \(t_j\) but when \(t = 0\), the definition of \(X(0)\) in this step function is \(X(0) \equiv 0\). Thus
\[
\sum_{j=1}^{m-1} \langle BX(t_{j+1}), X(t_{j+1}) \rangle - \langle BX(t_j), X(t_j) \rangle + \langle BX(t_1), X(t_1) \rangle - \langle BX_0, X_0 \rangle = \langle BX(t_m), X(t_m) \rangle - \langle BX_0, X_0 \rangle
\]
Using the formula in Lemma \(25.4.1\) for \(t = t_m\) this yields
\[
\langle BX(t_m), X(t_m) \rangle - \langle BX_0, X_0 \rangle = 2 \sum_{j=1}^{m-1} \int_{t_j}^{t_{j+1}} \langle Y(u), X_k(u) \rangle \, du + \sum_{j=1}^{m-1} \langle B \int_{t_j}^{t_{j+1}} Z(u) \, dW, X(t_j) \rangle
\]
\[
+ 2 \sum_{j=1}^{m-1} \langle B \int_{t_j}^{t_{j+1}} Z(u) \, dW, \rangle \left( X(t_j) - M(t_j) \right)
\]
\[
+ \sum_{j=1}^{m-1} \langle B( M(t_{j+1}) - M(t_j)), M(t_{j+1}) - M(t_j) \rangle
\]
\[
- \sum_{j=1}^{m-1} \langle B(X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j))), X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j)) \rangle
\]
First consider

\[ 2 \int_0^{t_1} \langle Y(u), X(t_1) \rangle \, du + 2 \left( BX_0, \int_0^{t_1} Z(u) \, dW \right) + \langle BM(t_1), M(t_1) \rangle. \]

Each term of the above converges to 0 for a.e. \( \omega \) as \( k \to \infty \) and in \( L^1(\Omega) \). This follows right away for the second two terms from the Ito isometry and continuity properties of the stochastic integral. Consider the first term. This term is dominated by

\[
\left( \int_0^{t_1} \| Y(u) \|^{p'} \, du \right)^{1/p'} \left( \int_0^{t_1} \| X_k^{r}(u) \|^{p} \, du \right)^{1/p} \leq C(\omega) \left( \int_0^{t_1} \| Y(u) \|^{p'} \, du \right)^{1/p'}, \quad \left( \int_\Omega C(\omega)^{p} \, dP \right)^{1/p} < \infty
\]

Hence this converges to 0 for a.e. \( \omega \) and also converges to 0 in \( L^1(\Omega) \).

At this time, not much is known about the last term in (25.4.10), but it is negative and is about to be neglected anyway.

The term involving the stochastic integral equals

\[
2 \sum_{j=1}^{m-1} \left( B \int_{t_j}^{t_{j+1}} Z(u) \, dW, X(t_j) \right)
\]

By Theorem 25.3.2, this equals

\[
2 \int_{t_1}^{t_m} (Z \circ J^{-1})^* BX_k^l \circ J \, dW
\]

Also note that since \( \langle BM(t_1), M(t_1) \rangle \) converges to 0 in \( L^1(\Omega) \) and for a.e. \( \omega \), the sum involving

\[
\langle B(M(t_{j+1}) - M(t_j)), M(t_{j+1}) - M(t_j) \rangle
\]

can be started at 0 rather than 1 at the expense of adding in a term which converges to 0 a.e. and in \( L^1(\Omega) \). Thus (25.4.11) is of the form

\[
\langle BX(t_m), X(t_m) \rangle - \langle BX_0, X_0 \rangle = e(k) + 2 \int_0^{t_m} \langle Y(u), X_k^{r}(u) \rangle \, du +
\]

\[
+ 2 \int_0^{t_m} (Z \circ J^{-1})^* BX_k^l \circ J \, dW
\]

\[
+ \sum_{j=0}^{m-1} \langle B(M(t_{j+1}) - M(t_j)), M(t_{j+1}) - M(t_j) \rangle
\]

\[
- \sum_{j=1}^{m-1} \langle B(X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j))), X(t_{j+1}) - X(t_j) - (M(t_{j+1}) - M(t_j)) \rangle
\]

\[
- \langle B(X(t_1) - X_0 - M(t_1)), X(t_1) - X_0 - M(t_1) \rangle
\]

where \( e(k) \to 0 \) for a.e. \( \omega \) and also in \( L^1(\Omega) \).

By definition, \( M(t_{j+1}) - M(t_j) = \int_{t_j}^{t_{j+1}} Z \, dW \). Now it follows, on discarding the negative terms,

\[
\langle BX(t_m), X(t_m) \rangle - \langle BX_0, X_0 \rangle \leq e(k) + 2 \int_0^{t_m} \langle Y(u), X_k^{r}(u) \rangle \, du +
\]
\[ +2 \int_0^{t_m} (Z \circ J^{-1})^* B X_k^l \circ JdW + \sum_{j=0}^{m-1} \left\langle B \int_{t_j}^{t_{j+1}} ZdW, \int_{t_j}^{t_{j+1}} ZdW \right\rangle \]

Therefore, \[
\sup_{t_m \in \mathcal{P}_k} \langle BX(t_m), X(t_m) \rangle \leq \langle BX_0, X_0 \rangle + c(k) + 2 \int_0^T \|Y(u), X_k^l(u)\| \, du +
\]
\[+2 \sup_{t_m \in \mathcal{P}_k} \left| \int_0^{t_m} (Z \circ J^{-1})^* B X_k^l \circ JdW \right| + \sum_{j=0}^{m-1} \left\langle B \left( \int_{t_j}^{t_{j+1}} Z(u) \, dW \right), \int_{t_j}^{t_{j+1}} Z(u) \, dW \right\rangle \]

where there are \( m_k + 1 \) points in \( \mathcal{P}_k \).

The next task is to somehow take the expectation of both sides. However, this is problematic because the stochastic integral is only a local martingale. Let \( \tau_p = \inf \{ t : \langle BX_k^l(t), X_k^l(t) \rangle > p \} \)

By right continuity this is a well defined stopping time. Then you obtain the above inequality for \( (X_k^l)^{\tau_p} \) in place of \( X_k^l \). Take the expectation and use the Ito isometry to obtain

\[ \int_{\Omega} \left\langle \sup_{t_m \in \mathcal{P}_k} \left( B (X_k^l)^{\tau_p}(t_m), (X_k^l)^{\tau_p}(t_m) \right) \right\rangle \, dP \]
\[ \leq E (\langle BX_0, X_0 \rangle) + 2 \|Y\|_{K'} \|X_k^l\|_K + \|B\| \sum_{j=0}^{m-1} \int_{t_j}^{t_{j+1}} \|Z(u)\|^2 \, dP \, du \]
\[+2 \int_{\Omega} \left( \sup_{t \in [0,T]} \left| \int_0^t \left( Z \circ J^{-1} \right)^* B (X_k^l)^{\tau_p} \circ JdW \right| \right) \, dP + E (|e(k)|) \]
\[ \leq C + \|B\| \int_0^T \int_{\Omega} \|Z(u)\|^2 \, dP \, du + E (|e(k)|) \]
\[+2 \int_{\Omega} \left( \sup_{t \in [0,T]} \left| \int_0^t \left( Z \circ J^{-1} \right)^* B (X_k^l)^{\tau_p} \circ JdW \right| \right) \, dP \leq \]
\[C + E (|e(k)|) + 2 \int_{\Omega} \left( \sup_{t \in [0,T]} \left| \int_0^t \left( Z \circ J^{-1} \right)^* B (X_k^l)^{\tau_p} \circ JdW \right| \right) \, dP \]

(25.4.12)

where the convergence of \( X_k^l \) to \( X \) in \( K \) shows the term \( 2 \|Y\|_{K'} \|X_k^l\|_K \) is bounded. Thus the constant \( C \) can be assumed to be a continuous function of

\[ \|Y\|_{K'}, \|X\|_{K'}, \|Z\|_J, \|BX_0, X_0\|_{L^1(\Omega)} \]

which equals zero when all are equal to zero and is increasing in each. The term involving the stochastic integral is next.

Let \( M(t) = \int_0^t \left( Z \circ J^{-1} \right)^* B (X_k^l)^{\tau_p} \circ JdW \). Then thanks to Corollary [Corollary 1](#)

\[ d|M| = \left\| \left( Z \circ J^{-1} \right)^* B (X_k^l)^{\tau_p} \circ J \right\|^2 \, ds \]

Applying the Burkholder Davis Gundy inequality, Theorem [Theorem 1](#) for \( F(r) = r \) in that stochastic integral, 

\[ 2 \int_{\Omega} \left( \sup_{t \in [0,T]} \left| \int_0^t \left( Z \circ J^{-1} \right)^* B (X_k^l)^{\tau_p} \circ JdW \right| \right) \, dP \]
\[ \leq C \int_{\Omega} \left( \int_{0}^{T} \left\| (Z \circ J^{-1})^* B \left( X_k^t \right)^\tau \circ J \right\|_{\mathcal{L}_2(Q^{1/2} U, \mathbb{R})}^2 ds \right)^{1/2} dP \]

(25.4.13)

So let \( \{g_i\} \) be an orthonormal basis for \( Q^{1/2} U \) and consider the integrand in the above. It equals

\[
\sum_{i=1}^{\infty} \left( \left\| (Z \circ J^{-1})^* B \left( X_k^t \right)^\tau \right\|^2 \right) = \sum_{i=1}^{\infty} \left\langle B \left( X_k^t \right)^\tau, Z \left( g_i \right) \right\rangle^2 \]

\[
\leq \sum_{i=1}^{\infty} \left\langle B \left( X_k^t \right)^\tau, \left( X_k^t \right)^\tau \right\rangle \left\langle B Z \left( g_i \right), Z \left( g_i \right) \right\rangle \]

\[
\leq \left( \sup_{t_m \in \mathcal{P}_k} \left\langle B \left( X_k^t \right)^\tau \left( t_m \right), \left( X_k^t \right)^\tau \left( t_m \right) \right\rangle \right) \|B\| \|Z\|_{\mathcal{L}_2}^2 \]

It follows that the integral in \( 25.4.13 \) is dominated by

\[
C \int_{\Omega} \sup_{t_m \in \mathcal{P}_k} \left\langle B \left( X_k^t \right)^\tau \left( t_m \right), \left( X_k^t \right)^\tau \left( t_m \right) \right\rangle^{1/2} \|B\|^{1/2} \left( \int_{0}^{T} \|Z\|_{\mathcal{L}_2}^2 ds \right)^{1/2} dP
\]

Now return to \( 25.4.12 \). From what was just shown,

\[
E \left( \sup_{t_m \in \mathcal{P}_k} \left\langle B \left( X_k^t \right)^\tau \left( t_m \right), \left( X_k^t \right)^\tau \left( t_m \right) \right\rangle \right)
\]

\[
\leq C + E \left( |e \left( k \right)| \right) + 2 \int_{\Omega} \left( \sup_{t \in [0, T]} \left| \int_{0}^{t} \left( Z \circ J^{-1})^* B \left( X_k^t \right)^\tau \circ J \right) dW \right| \right) dP
\]

\[
\leq C + C \int_{\Omega} \sup_{t_m \in \mathcal{P}_k} \left\langle B \left( X_k^t \right)^\tau \left( t_m \right), \left( X_k^t \right)^\tau \left( t_m \right) \right\rangle^{1/2},
\]

\[
\|B\|^{1/2} \left( \int_{0}^{T} \|Z\|_{\mathcal{L}_2}^2 ds \right)^{1/2} dP + E \left( |e \left( k \right)| \right)
\]

\[
\leq C + \frac{1}{2} E \left( \sup_{t_m \in \mathcal{P}_k} \left\langle B \left( X_k^t \right)^\tau \left( t_m \right), \left( X_k^t \right)^\tau \left( t_m \right) \right\rangle \right)
\]

\[
+ C \|Z\|_{\mathcal{L}_2^2([0, T] \times \Omega, \mathcal{L}_2)} + E \left( |e \left( k \right)| \right).
\]

It follows that

\[
\frac{1}{2} E \left( \sup_{t_m \in \mathcal{P}_k} \left\langle B X_k \left( t_m \right), X_k \left( t_m \right) \right\rangle \right) \leq C + E \left( |e \left( k \right)| \right)
\]

Now let \( p \to \infty \) and use the monotone convergence theorem to obtain

\[
E \left( \sup_{t_m \in \mathcal{P}_k} \left\langle B X_k \left( t_m \right), X_k \left( t_m \right) \right\rangle \right) = E \left( \sup_{t_m \in \mathcal{P}_k} \left\langle B X \left( t_m \right), X \left( t_m \right) \right\rangle \right) \leq C + E \left( |e \left( k \right)| \right)
\]

(25.4.14)

As mentioned above, this constant \( C \) is a continuous function of

\[
\|Y\|_{K'}, \|X\|_{K}, \|Z\|_{J}, \|\langle BX_0, X_0 \rangle\|_{L^1(\Omega, H)}
\]

and equals zero when all of these quantities equal 0 and is increasing with respect to each of the above quantities.

Also, for each \( \varepsilon > 0 \),

\[
E \left( \sup_{t_m \in \mathcal{P}_k} \left\langle BX \left( t_m \right), X \left( t_m \right) \right\rangle \right) \leq C + \varepsilon
\]

whenever \( k \) is large enough.
25.4. THE MAIN ESTIMATE

Let \( D \) denote the union of all the \( P_k \). Thus \( D \) is a dense subset of \([0, T]\) and it has just been shown, since the \( P_k \) are nested, that for a constant \( C \) dependent only on the above quantities which is independent of \( P_k \),

\[
E \left( \sup_{t \in D} \langle BX(t), X(t) \rangle \right) \leq C + \varepsilon.
\]

Since \( \varepsilon > 0 \) is arbitrary,

\[
E \left( \sup_{t \in D} \langle BX(t), X(t) \rangle \right) \leq C
\]  \hspace{1cm} (25.4.15)

Thus, enlarging \( N \), for \( \omega \notin N \),

\[
\sup_{t \in D} \langle BX(t), X(t) \rangle = C(\omega) < \infty
\]  \hspace{1cm} (25.4.16)

where \( \int_\Omega C(\omega) \, dP < \infty \). By Lemma 24.4, there exists a countable set \( \{e_i\} \) of vectors in \( V \) such that

\[
\langle Be_i, e_j \rangle = \delta_{ij}
\]

and for each \( x \in W \),

\[
\langle Bx, x \rangle = \sum_{i=0}^{\infty} \langle Bx, e_i \rangle^2, \quad Bx = \sum_{i=1}^{\infty} \langle Bx, e_i \rangle Be_i
\]

Thus for \( t \) not in a set of measure zero off which \( BX(t) = B(X(t)) \),

\[
\langle BX(t), X(t) \rangle = \sum_{i=0}^{\infty} \langle BX(t), e_i \rangle^2 = \sum_{m=1}^{\infty} \sum_{k=1}^{m} \langle BX(t), e_i \rangle^2
\]

Now from the formula for \( BX(t) \), it follows that \( BX \) is continuous into \( V' \). For any \( t \notin \hat{N} \) so that \( (BX)(t) = B(X(t)) \) in \( L^q( \Omega; V') \) and letting \( t_k \to t \) where \( t_k \in D \), Fatou’s lemma implies

\[
E(\langle BX(t), X(t) \rangle) = \sum_i E(\langle BX(t), e_i \rangle^2) = \sum_i \liminf_{k \to \infty} E(\langle BX(t_k), e_i \rangle^2)
\]

\[
\leq \liminf_{k \to \infty} \sum_i E(\langle BX(t_k), e_i \rangle^2) = \liminf_{k \to \infty} E(\langle BX(t_k), X(t_k) \rangle)
\]

\[
\leq C \left( \|Y\|_{K'}, \|X\|_{K}, \|Z\|_{J'}, \|\langle BX_0, X_0 \rangle\|_{L^1(\Omega)} \right)
\]

In addition to this, for arbitrary \( t \in [0, T] \), and \( t_k \to t \) from \( D \),

\[
\sum_i \langle BX(t), e_i \rangle^2 \leq \liminf_{k \to \infty} \sum_i \langle BX(t_k), e_i \rangle^2 \leq \sup_{s \in D} \langle BX(s), X(s) \rangle
\]

Hence

\[
\sup_{t \in [0, T]} \sum_i \langle BX(t), e_i \rangle^2 \leq \sup_{s \in D} \langle BX(s), X(s) \rangle
\]

\[
= \sup_{s \in D} \sum_i \langle BX(s), e_i \rangle^2 \leq \sup_{t \in [0, T]} \sum_i \langle BX(t), e_i \rangle^2
\]

It follows that \( \sup_{t \in [0, T]} \sum_i \langle BX(t), e_i \rangle^2 \) is measurable and

\[
E \left( \sup_{t \in [0, T]} \sum_i \langle BX(t), e_i \rangle^2 \right) \leq E \left( \sup_{s \in D} \langle BX(s), X(s) \rangle \right)
\]

\[
\leq C \left( \|Y\|_{K'}, \|X\|_{K}, \|Z\|_{J'}, \|\langle BX_0, X_0 \rangle\|_{L^1(\Omega)} \right)
\]

And so, for \( \omega \) off a set of measure zero, \( \sup_{t \in [0, T]} \sum_i \langle BX(t), e_i \rangle^2 \) is bounded above.
Also for \( t \notin N_\omega \) and a given \( \omega \notin N \), letting \( t_k \to t \) for \( t_k \in D \),
\[
\langle BX(t), X(t) \rangle = \sum_i \langle BX(t), e_i \rangle^2 \leq \lim \inf_{k \to \infty} \sum_i \langle BX(t_k), e_i \rangle^2
\]
\[
= \lim \inf_{k \to \infty} \langle BX(t_k), X(t_k) \rangle \leq \sup_{t \in D} \langle BX(t), X(t) \rangle
\]
and so
\[
\sup_{t \notin N_\omega} \langle BX(t), X(t) \rangle \leq \sup_{t \in D} \langle BX(t), X(t) \rangle \leq \sup_{t \notin N_\omega} \langle BX(t), X(t) \rangle
\]
From (25.4.10),
\[
\sup_{t \notin N_\omega} \langle BX(t), X(t) \rangle = C(\omega) \text{ a.e.}\omega
\]
where \( \int C(\omega) dP < \infty \). In particular, \( \sup_{t \notin N_\omega} \langle BX(t), X(t) \rangle \) is bounded for a.e. \( \omega \) say for \( \omega \notin N \) where \( N \) includes the earlier sets of measure zero. This shows that \( BX(t) \) is bounded in \( W' \) for \( t \in N_\omega \).

If \( v \in V \), then for \( \omega \notin N \),
\[
\lim_{t \to s} \langle BX(t), v \rangle = \langle BX(s), v \rangle, \quad t, s
\]
Therefore, since for such \( \omega \), \( \|BX(t)\|_{W'} \) is bounded for \( t \notin N_\omega \), the above holds for all \( v \in W \) also. Therefore, for a.e. \( \omega \), \( t \to BX(t, \omega) \) is weakly continuous with values in \( W' \) for \( t \notin N_\omega \).

Note also that
\[
\int_0^T \int_\Omega \|BX(t)\|^2 dP dt \leq \int_0^T \int_\Omega \|B\|^{1/2} \langle BX(t), X(t) \rangle dt dP
\]
\[
\leq C \left( \|Y\|_{K'}, \|X\|_{K}, \|Z\|_{J'}, \|\langle BX_1, X_0 \rangle\|_{L^1(\Omega)} \right) \|B\|^{1/2} T
\]
(25.4.17)

Eventually, it is shown that in fact, the function \( t \to BX(t, \omega) \) is continuous with values in \( W' \). The above shows that \( BX \in L^2([0, T] \times \Omega, W') \).

Finally consider the claim of weak continuity of \( BX \) into \( W' \). From the integral equation, \( BX \) is continuous into \( W' \). Also \( BX \) is bounded on \( N_\omega' \). Let \( s \in [0, T] \) be arbitrary. I claim that if \( t_n \to s, t_n \in D \), it follows that \( BX(t_n) \to BX(s) \) weakly in \( W' \). If not, then there is a subsequence, still denoted as \( t_n \) such that \( BX(t_n) \to Y \) weakly in \( W' \) but \( Y \neq BX(s) \). However, the continuity into \( W' \) means that for all \( v \in V \),
\[
\langle Y, v \rangle = \lim_{n \to \infty} \langle BX(t_n), v \rangle = \langle BX(s), v \rangle
\]
which is a contradiction since \( V \) is dense in \( W \). This establishes the claim. Also this shows that \( BX(s) \) is bounded in \( W' \).
\[
\|BX(s, w)\| = \lim_{n \to \infty} \|BX(t_n, w)\| \leq \lim \inf_{n \to \infty} \|BX(t_n)\|_{W'} \|w\|_{W} \leq C(\omega) \|w\|_{W}
\]
Now a repeat of the above argument shows that \( s \to BX(s) \) is weakly continuous into \( W' \).

## 25.5 A Simplification Of The Formula

This lemma also provides a way to simplify one of the formulas derived earlier in the case that \( X_0 \in L^p(\Omega, V) \) so that \( X - X_0 \in L^p([0, T] \times \Omega, V) \). Refer to (25.4.11). One term there is
\[
\langle B(X(t_1) - X_0 - M(t_1)), X(t_1) - X_0 - M(t_1) \rangle
\]
Also,
\[
\langle B(X(t_1) - X_0 - M(t_1)), X(t_1) - X_0 - M(t_1) \rangle
\]
\[
\leq 2 \langle B(X(t_1) - X_0), X(t_1) - X_0 \rangle + 2 \langle BM(t_1), M(t_1) \rangle
\]
It was observed above that \( 2 \langle BM(t_1), M(t_1) \rangle \to 0 \) a.e. and also in \( L^1(\Omega) \) as \( k \to \infty \). Apply the above lemma to \( \langle B(X(t_1) - X_0), X(t_1) - X_0 \rangle \) using \([0, t_1] \) instead of \([0, T] \). The new \( X_0 \) equals 0. Then from the estimate (25.4.13), it follows that
\[
E(\langle B(X(t_1) - X_0), X(t_1) - X_0 \rangle) \to 0
\]
as \( k \to \infty \). Taking a subsequence, we could also assume that
\[
\langle B(X(t_1) - X_0), X(t_1) - X_0 \rangle \to 0
\]
a.e. \( \omega \) as \( k \to \infty \). Then, using this subsequence, it would follow from (20.9),

\[
\langle BX (t_m), X (t_m) \rangle - \langle BX_0, X_0 \rangle = e (k) + 2 \int_0^{t_m} \langle Y (u), X_k^k (u) \rangle \, du + 2 \int_0^{t_m} (Z \circ J^{-1})^* BX_k^k \circ J \, dW
\]

\[
+ \sum_{j=0}^{m-1} \langle B (M (t_{j+1}) - M (t_j)), M (t_{j+1}) - M (t_j) \rangle
\]

\[
- \sum_{j=1}^{m-1} \langle B (\Delta X (t_j) - \Delta M (t_j)), \Delta X (t_j) - \Delta M (t_j) \rangle
\]

(25.5.18)

where \( e (k) \to 0 \) in \( L^1 (\Omega) \) and a.e. \( \omega \) and

\[\Delta X (t_j) \equiv X (t_{j+1}) - X (t_j)\]

\[\Delta M (t_j) \text{ being defined similarly. Note how this eliminated the need to consider the term}\]

\[\langle B (X (t_1) - X_0 - M (t_1)), X (t_1) - X_0 - M (t_1) \rangle\]

in passing to a limit. This is a very desirable thing to be able to conclude.

Can you obtain something similar even in case \( X_0 \) is not assumed to be in \( L^p (\Omega, V) \)? Let \( Z_{0k} \in L^p (\Omega, V) \cap L^2 (\Omega, W) \), \( Z_{0k} \to X_0 \) in \( L^2 (\Omega, W) \). Then from the usual arguments involving the Cauchy Schwarz inequality,

\[\langle B (X (t_1) - X_0), X (t_1) - X_0 \rangle^{1/2} \leq \langle B (X (t_1) - Z_{0k}), X (t_1) - Z_{0k} \rangle^{1/2} + \langle B (Z_{0k} - X_0), Z_{0k} - X_0 \rangle^{1/2}\]

Also, restoring the superscript to identify the partition,

\[B (X (t_k^t) - Z_{0k}) = B (X_0 - Z_{0k}) + \int_{t^0}^{t_k^t} Y (s) \, ds + B \int_0^{t_k^t} Z (s) \, dW.\]

Of course \( \|X - Z_{0k}\|_{K^p_t} \) is not bounded, but for each \( k \) it is finite. There is a sequence of partitions \( P_k, \|P_k\| \to 0 \)

such that all the above holds. In the definitions of \( K, K', J \) replace \([0, T]\) with \([0, t]\) and let the resulting spaces be denoted by \( K_t, K'_t, J_t \). Let \( n_k \) denote a subsequence of \( \{k\} \) such that

\[\|X - Z_{0k}\|_{K^p_{t_k^t}} < 1/k.\]

Then from the above lemma,

\[E (\langle B (X (t_{1_{1}^k}) - Z_{0k}), X (t_{1_{1}^k}) - Z_{0k} \rangle)\]

\[\leq C \left( \|Y\|_{K'_{t_k^t}^{p_k}}, \|X - Z_{0k}\|_{K_{t_k^t}^{p_k}}, \|Z\|_{J_{t_k^t}^{p_k}}, \langle B (X_0 - Z_{0k}), X_0 - Z_{0k} \rangle_{L^1 (\Omega)} \right)\]

\[\leq C \left( \|Y\|_{K'_{t_k^t}^{p_k}}, \frac{1}{k}, \|Z\|_{J_{t_k^t}^{p_k}}, \langle B (X_0 - Z_{0k}), X_0 - Z_{0k} \rangle_{L^1 (\Omega)} \right)\]

Hence

\[E (\langle B (X (t_{1_{1}^k}) - X_0), X (t_{1_{1}^k}) - X_0 \rangle)\]

\[\leq 2E (\langle B (X (t_{1_{1}^k}) - Z_{0k}), X (t_{1_{1}^k}) - Z_{0k} \rangle) + 2E (\langle B (Z_{0k} - X_0), Z_{0k} - X_0 \rangle)\]

\[\leq 2C \left( \|Y\|_{K'_{t_k^t}^{p_k}}, \frac{1}{k}, \|Z\|_{J_{t_k^t}^{p_k}}, \langle B (X_0 - Z_{0k}), X_0 - Z_{0k} \rangle_{L^1 (\Omega)} \right)\]

\[+ 2 \|B\| \|Z_{0k} - X_0\|_{L^2 (\Omega, W)}^2\]

which converges to 0 as \( k \to \infty \). It follows that there exists a suitable subsequence such that (20.9) holds even in the case that \( X_0 \) is only known to be in \( L^2 (\Omega, W) \). From now on, assume this subsequence for the paritions \( P_k \).

Thus \( k \) will really be \( n_k \) and it suffices to consider the limit as \( k \to \infty \) of the equation of (20.9). To emphasize this point again, the reason for the above observations is to argue that, even when \( X_0 \) is only in \( L^2 (\Omega, W) \), one can neglect

\[\langle B (X (t_1) - X_0 - M (t_1)), X (t_1) - X_0 - M (t_1) \rangle\]

in passing to the limit as \( k \to \infty \) provided a suitable subsequence is used.
25.6 Convergence

The question is whether the above stochastic integral \( \int_0^t (Z \circ J^{-1})^* BX \circ JdW \) converges as \( n \to \infty \) in some sense to
\[
\int_0^t (Z \circ J^{-1})^* BX \circ JdW
\]
and whether the above is also a local martingale. Maybe it is well to pause and consider the integral and and what it means. \( Z \circ J^{-1} \) maps \( JQ^{1/2}U \) to \( W \) and so \( (Z \circ J^{-1})^* \) maps \( W' \) to \( (JQ^{1/2}U)' \). Thus
\[
(Z \circ J^{-1})^* BX \in \left(JQ^{1/2}U\right)', \quad \text{so} \quad (Z \circ J^{-1})^* BX \circ J \in Q^{1/2}(U)' = \mathcal{L}_2(Q^{1/2}U, \mathbb{R})
\]
Thus it has the right values.

Does the stochastic integral just written even make sense? The integrand is Hilbert Schmidt and has values in \( \mathbb{R} \) so it seems like we ought to be able to define an integral. The problem is that the integrand is not in \( L^2([0, T] \times \Omega; \mathcal{L}_2(Q^{1/2}U, \mathbb{R})) \).

By assumption, \( t \to BX(t) \) is continuous into \( V' \) thanks to the integral equation solved, and also \( BX(t) = B(X(t)) \) for \( t \notin N_\omega \) a set of measure zero. For such \( t \), it follows from Lemma 25.4.2, 25.5.2,
\[
\langle BX(t), X(t) \rangle = \sum_i \langle BX(t), e_i \rangle^2_{V', V} \ a.e. \omega
\]
and so \( t \to \sum_i \langle BX(t), e_i \rangle^2 \) is lower semicontinuous and as just explained, it equals \( \langle BX(t), X(t) \rangle \) for a.e. \( t \), this for each \( \omega \notin N \), a single set of measure zero. Also, \( t \to \sum_i \langle BX(t), e_i \rangle^2_{V', V} \) is progressively measurable and lower semicontinuous in \( t \) so by Proposition 25.6.3, one can define a stopping time
\[
\tau_p = \inf \left\{ t : \sum_i \langle BX(t), e_i \rangle^2_{V', V} > p \right\}, \tau_0 = 0
\]
Instead of referring to this Proposition, you could consider
\[
\tau^m_p = \inf \left\{ t : \sum_{i=1}^m \langle BX(t), e_i \rangle^2_{V', V} > p \right\}
\]
which is clearly a stopping time because \( t \to \sum_{i=1}^m \langle BX(t), e_i \rangle^2_{V', V} \) is a continuous process. Then observe that \( \tau_p = \sup_m \tau^m_p \). Then
\[
[t \leq \tau_p] = \bigcup_m [t \leq \tau^m_p] \in \mathcal{F}_t.
\]
Is it the case that \( \tau_p = \infty \) for all \( p \) large enough? Yes, this follows from Lemma 25.6.4.

Lemma 25.6.1 Suppose \( \tau_p = \infty \) for all \( p \) large enough off a set of measure zero, then
\[
P \left( \int_0^T \left| (Z \circ J^{-1})^* BX \circ J \right|^2 dt < \infty \right) = 1
\]
Also \( \int_0^T (Z \circ J^{-1})^* BX \circ JdW \) can be defined as a local martingale.

Proof: Let
\[
A \equiv \left\{ \omega : \int_0^T \left| (Z \circ J^{-1})^* BX \circ J \right|^2 dt = \infty \right\}
\]
Then from the assumption that \( \tau_p = \infty \) for all \( p \) large enough, it follows that
\[
A = \bigcup_{m=1}^\infty A \cap \{[\tau_m = \infty] \setminus [\tau_{m-1} < \infty]\}
\]
Now
\[
P(A \cap [\tau_m = \infty]) \leq P \left( \omega : \int_0^T X_{[0, \tau_m]} \left| (Z \circ J^{-1})^* BX \circ J \right|^2 dt = \infty \right)
\]
25.6. CONVERGENCE

Look at the integrand. What is the meaning of \( (Z \circ J^{-1})^* BX \circ J \)? You have \( (Z \circ J^{-1})^* \in \mathcal{L}_2 \left( W', J (Q^{1/2}U)^t \right) \) while \( BX \in W' \) and so \( (Z \circ J^{-1})^* BX \in \mathcal{L}_2 \left( J (Q^{1/2}U)^t, \mathbb{R} \right) \) which is just \( J (Q^{1/2}U)^t \). Thus \( (Z \circ J^{-1})^* BX \circ J \) would be in \( (Q^{1/2}U)^t \) and to get the \( \mathcal{L}_2 \) norm, you would take an orthonormal basis in \( Q^{1/2}U \) denoted as \( \{g_i\} \) and the square of this norm is just

\[
\sum_i \left[ \left( (Z \circ J^{-1})^* BX \circ J \right) (g_i) \right]^2 = \sum_i \left[ (Z \circ J^{-1})^* BX (Jg_i) \right]^2 \\
= \sum_i \left[ BX (Z \circ J^{-1} (Jg_i)) \right]^2 \\
= \sum_i \|BX\|^2 \|Zg_i\|^2 \|w\| \|
\]

Now incorporating the stopping time, you know that for a.e. \( t, \langle BX, X \rangle (t) = \langle BX (t), X (t) \rangle \leq m \) and so \( \|BX (t)\| \) can be estimated in terms of \( m \) as follows.

\[
|\langle B(X(t)), w \rangle| \leq \langle B(X(t)), X(t) \rangle^{1/2} \|B\|^{1/2} \|w\| \|
\]

Thus the integrand satisfies for a.e. \( t \)

\[
X_{[0, \tau_m]} (Z \circ J^{-1})^* BX \circ J \leq m \|B\| \|Z\|^2 \mathcal{L}_2
\]

Hence, from 25.6.21: \( P (A \cap [\tau_m \leq \infty]) \)

\[
\leq P \left( \omega : \int_0^T \|Z\|^2 \mathcal{L}_2 \|B\| \|w\| dt = \infty \right)
\]

However,

\[
\int_0^T \int \|Z\|^2 \mathcal{L}_2 \|B\| \|w\| dt dP < \infty
\]

by the assumptions on \( Z \). Therefore, \( P (A \cap [\tau_m \leq \infty]) \leq 0 \). It follows that

\[
P (A) = \sum_m P (A \cap ([\tau_m \leq \infty] \setminus [\tau_{m-1} < \infty])) = \sum_m 0 = 0
\]

It follows that \( P \left( \int_0^T \left( (Z \circ J^{-1})^* BX \circ J \right)^2 dt < \infty \right) = 1 \) and so from Definition 23.4.13, one can define \( \int_0^t (Z \circ J^{-1})^* BX \circ J dW \) as a local martingale.

Convergence will be shown for a subsequence and from now on every sequence will be a subsequence of this one. As part of Lemma 25.6.1 see 25.4.11 it was shown that \( BX \in L^2 ([0, T] \times \Omega, W') \). Therefore, there exist partitions of \([0, T]\) like the above such that

\[
BX^t_k, BX^t \rightarrow BX \text{ in } L^2 ([0, T] \times \Omega, W')
\]

in addition to the convergence of \( X^t_k, X^t \) to \( X \) in \( K \). From now on, the argument will involve a subsequence of these.

**Lemma 25.6.2** There exists a subsequence still denoted with the subscript \( k \) and an enlarged set of measure zero \( N \) including the earlier one such that \( BX^t_k (t), BX^t_k (t) \) also converges pointwise a.e. \( t \) to \( BX (t) \) in \( W' \) and \( X^t_k (t), X^t_k (t) \) converge pointwise a.e. in \( V \) to \( X (t) \) for \( \omega \notin N \) as well as having convergence of \( X^t_k (\cdot, \omega) \) to \( X (\cdot, \omega) \) in \( L^p ([0, T] ; V) \) and \( BX^t_k (\cdot, \omega) \) to \( BX (\cdot, \omega) \) in \( L^2 ([0, T] ; W) \).
**CHAPTER 25. THE HARD ITO FORMULA, IMPLICIT CASE**

**Proof:** To see that such a sequence exists, let \( n_k \) be such that

\[
\int_\Omega \int_0^T \|B X_{n_k}^t (t) - B X (t)\|_W^2 \, dt \, dP + \int_\Omega \int_0^T \|X_{n_k}^t (t) - X (t)\|_V^p \, dt \, dP < 4^{-k}.
\]

Then

\[
P \left( \int_0^T \|B X_{n_k}^t (t) - B X (t)\|_W^2 \, dt + \int_0^T \|X_{n_k}^t (t) - X (t)\|_V^p \, dt \right) + \int_0^T \|B X_{n_k}^t (t) - B X (t)\|_W^2 \, dt + \int_0^T \|X_{n_k}^t (t) - X (t)\|_V^p \, dt > 2^{-k} \right) \leq 2^k (4^{-k}) = 2^{-k}
\]

and so by Borel Cantelli lemma, there is a set of measure zero \( N \) such that if \( \omega \notin N \),

\[
\int_0^T \|B X_{n_k}^t (t) - B X (t)\|_W^2 \, dt + \int_0^T \|X_{n_k}^t (t) - X (t)\|_V^p \, dt \leq 2^{-k}
\]

for all \( k \) large enough. By the usual proof of completeness of \( L^p \), it follows that \( X_{n_k}^t (t) \to X (t) \) for a.e. \( t \), this for each \( \omega \notin N \), a similar assertion holding for \( X_{n_k}^t \). We denote these subsequences as \( \{X_k^n\}_{k=1}^\infty \cdot \{X_k^l\}_{k=1}^\infty \). ■

Now with this preparation, it is possible to show the desired convergence.

**Lemma 25.6.3** In the above context, let \( X (s) - X_k^l (s) \equiv \Delta_k (s) \). Then the integral

\[
\int_0^t (Z \circ J^{-1})^* B X \circ J \, dW
\]

exists as a local martingale and the following limit is valid for the subsequence of Lemma 25.6.2:

\[
\lim_{k \to \infty} P \left( \sup_{t \in [0,T]} \left| \int_0^t (Z \circ J^{-1})^* B \Delta_k \circ J \, dW \right| \geq \varepsilon \right) = 0.
\]

That is,

\[
\sup_{t \in [0,T]} \left| \int_0^t (Z \circ J^{-1})^* B \Delta_k \circ J \, dW \right|
\]

converges to 0 in probability.

**Proof:** In the argument \( \tau_m \) will be defined in 25.10.21. Let

\[
A_k \equiv \left\{ \omega : \sup_{t \in [0,T]} \left| \int_0^t (Z \circ J^{-1})^* B \Delta_k \circ J \, dW \right| \geq \varepsilon \right\}
\]

then

\[
A_k \cap \{ \omega : \tau_m = \infty \} \subseteq \left\{ \omega : \sup_{t \in [0,T]} \left| \int_0^t (Z \circ J^{-1})^* B \Delta_k^{\tau_m} \circ J \, dW \right| \geq \varepsilon \right\}
\]

By Burkholder Davis Gundy inequality,

\[
P \left( A_k \cap \{ \omega : \tau_m = \infty \} \right) \leq \frac{C}{\varepsilon} \int_\Omega \sup_{t \in [0,T]} \left| \int_0^t (Z \circ J^{-1})^* B \Delta_k^{\tau_m} \circ J \, dW \right| \, dP \leq \frac{C}{\varepsilon} \int_\Omega \left( \int_0^T \|Z\|_{L_2}^2 \|B \Delta_k^{\tau_m}\|^2 \, dt \right)^{1/2} \, dP \leq \frac{C}{\varepsilon} \left( \int_\Omega \int_0^T \|Z\|_{L_2}^2 \|B \Delta_k^{\tau_m}\|^2 \, dt \, dP \right)^{1/2}
\]
Recall that if \( \langle Bx, x \rangle \leq m \), then \( \| Bx \|_{W'} \leq m^{1/2} \| B \|^{1/2} \). Then the integrand is bounded for a.e. \( t \) by \( \| Z \|_{L_2}^{2} 4m \| B \| \). Next use the result of Lemma 25.6.4 and the dominated convergence theorem to conclude that the above converges to 0 as \( k \to \infty \). Then from the assumption that \( \tau_m = \infty \) for all \( m \) large enough,

\[
P( A_k ) = \sum_{m=1}^{\infty} P( A_k \cap ( [\tau_m = \infty] \setminus [\tau_{m-1} < \infty] ) )
\]

Now \( \sum_m P( [\tau_m = \infty] \setminus [\tau_{m-1} < \infty] ) = 1 \) and so, one can apply the dominated convergence theorem to conclude that

\[
\lim_{k \to \infty} P( A_k ) = \lim_{m \to \infty} \lim_{k \to \infty} P( A_k \cap ( [\tau_m = \infty] \setminus [\tau_{m-1} < \infty] ) ) = 0 \]

**Lemma 25.6.4** Let \( X \) be as in Situation 25.6.5 and let \( X_k \) be as in Lemma 25.6.4 corresponding to \( X \) above. Let \( X_k \) and \( X_k' \) both converge to \( X \) in \( K \) and also

\[
BX_k, BX_k' \to BX \text{ in } L^2([0,T] \times \Omega, W')
\]

Say

\[
X_k(t) = \sum_{j=0}^{m_k} X(t_j) \mathbb{1}_{[t_j, t_{j+1})} (t), \quad (25.6.22)
\]

\[
BX_k(t) = \sum_{j=0}^{m_k} BX(t_j) \mathbb{1}_{[t_j, t_{j+1})} (t) \quad (25.6.23)
\]

Then the sum in 25.6.24 is progressively measurable into \( W' \). As mentioned earlier, we can take \( X(0) = 0 \) in the definition of the “left step function”.

**Proof:** This follows right away from the definition of progressively measurable. ■

One can take a further subsequence such that uniform convergence is obtained.

**Lemma 25.6.5** Let \( X(s) - X_k'(s) \equiv \Delta_k(s) \). Then the following limit occurs.

\[
\lim_{k \to \infty} P \left( \sup_{t \in [0,T]} \left| \int_0^t (Z \circ J^{-1})^* BX_k \circ JdW \right| \geq \varepsilon \right) = 0
\]

The stochastic integral

\[
\int_0^t (Z \circ J^{-1})^* BX \circ JdW
\]

makes sense because \( BX \) is \( W' \) progressively measurable and is in \( L^2([0,T] \times \Omega; W') \). Also, there exists a further subsequence, still denoted as \( k \) such that

\[
\int_0^t (Z \circ J^{-1})^* BX_k \circ JdW \to \int_0^t (Z \circ J^{-1})^* BX \circ JdW
\]

uniformly on \([0,T]\) for a.e. \( \omega \).

**Proof:** This follows from Lemma 25.6.4. The last conclusion follows from the usual use of the Borel Cantelli lemma. There exists a further subsequence, still denoted with subscript \( k \) such that

\[
P \left( \sup_{t \in [0,T]} \left| \int_0^t (Z \circ J^{-1})^* BX_k \circ JdW \right| \geq \frac{1}{k} \right) < 2^{-k}
\]

Then by the Borel Cantelli lemma, one can enlarge the set of measure zero such that for \( \omega \notin N \),

\[
\sup_{t \in [0,T]} \left| \int_0^t (Z \circ J^{-1})^* BX_k \circ JdW \right| < \frac{1}{k}
\]

for all \( k \) large enough. That is, the claimed uniform convergence holds. ■

From now on, the sequence will either be this subsequence or a further subsequence.
25.7 The Ito Formula

Now at long last, here is the first version of the Ito formula valid on the partition points.

**Lemma 25.7.1** In Situation 25.7, let $D$ be as above, the union of all the positive mesh points for all the $\mathcal{P}_k$. Also assume $X_0 \in L^2(\Omega; W)$. Then for $\omega \notin N$ the exceptional set of measure zero in $\Omega$ and every $t \in D$,

$$
(BX (t), X(t)) = (BX_0, X_0) + \int_0^t (2 \langle Y(s), X(s) \rangle + \langle BZ, Z \rangle_{\mathcal{L}_2}) \, ds
$$

$$
+ 2 \int_0^t (Z \circ J^{-1})^* BX \circ JdW
$$

where, in the above formula,

$$
(BZ, Z)_{\mathcal{L}_2} \equiv (R^{-1}BZ, Z)_{\mathcal{L}_2(\Omega; W)}
$$

for $R$ the Riesz map from $W$ to $W'$.

Note first that for $\{g_i\}$ an orthonormal basis for $Q^{1/2}(U)$,

$$
(R^{-1}BZ, Z)_{\mathcal{L}_2} \equiv \sum_i (R^{-1}BZ(g_i), Z(g_i))_W = \sum_i (BZ(g_i), Z(g_i))_{W'W} \geq 0
$$

**Proof:** Let $t \in D$. Then $t \in \mathcal{P}_k$ for all $k$ large enough. Consider $25.7.15$.

$$
(BX(t), X(t)) - (BX_0, X_0) = \epsilon(k) + 2 \int_0^t \langle Y(u), X_k'(u) \rangle \, du
$$

$$
+ 2 \int_0^t (Z \circ J^{-1})^* BX_k \circ JdW + \sum_{j=0}^{q_k-1} \langle B(M(t_{j+1}) - M(t_j)), M(t_{j+1}) - M(t_j) \rangle
$$

$$
- \sum_{j=1}^{q_k-1} \langle B(\Delta X(t_j) - \Delta M(t_j)), \Delta X(t_j) - \Delta M(t_j) \rangle
$$

where $t_{q_k} = t$, $\Delta X(t_j) = X(t_{j+1}) - X(t_j)$ and $\epsilon(k) \to 0$ in probability. By Lemma 25.7.2 the stochastic integral on the right converges uniformly for $t \in [0, T]$ to

$$
2 \int_0^t (Z \circ J^{-1})^* BX \circ JdW
$$

for $\omega$ off a set of measure zero. The deterministic integral on the right converges uniformly for $t \in [0, T]$ to

$$
2 \int_0^t \langle Y(u), X(u) \rangle \, du
$$

thanks to Lemma 25.1.4.

$$
\left| \int_0^t \langle Y(u), X(u) \rangle \, du - \int_0^t \langle Y(u), X_k'(u) \rangle \, du \right| \leq \int_0^T \|Y(u)\|_{W} \|X(u) - X_k'(u)\|_W
$$

$$
\leq \|Y\|_{L^p([0,T])} 2^{-k}
$$

for all $k$ large enough. Consider the fourth term. It equals

$$
\sum_{j=0}^{q_k-1} \langle R^{-1}B(M(t_{j+1}) - M(t_j)), M(t_{j+1}) - M(t_j) \rangle_W
$$

(25.7.26)

where $R^{-1}$ is the Riesz map from $W$ to $W'$. This equals

$$
\frac{1}{4} \left( \sum_{j=0}^{q_k-1} \|R^{-1}BM(t_{j+1}) + M(t_{j+1}) - (R^{-1}BM(t_j) + M(t_j))\|^2
$$

$$
- \sum_{j=0}^{q_k-1} \|R^{-1}BM(t_{j+1}) - M(t_{j+1}) - (R^{-1}BM(t_j) - M(t_j))\|^2 \right)
$$
From Theorem [25.7.22] as \( k \to \infty \), the above converges in probability to \( (t_{q_k} = t) \)

\[
\frac{1}{4} \left( [R^{-1}BM + M](t) - [R^{-1}BM - M](t) \right)
\]

However, from the description of the quadratic variation of \( M \), the above equals

\[
\frac{1}{4} \left( \int_0^t \|R^{-1}BZ + Z\|_{L^2}^2 \, ds - \int_0^t \|R^{-1}BZ - Z\|_{L^2}^2 \, ds \right)
\]

which equals

\[
\int_0^t (R^{-1}BZ, Z)_{L^2} \, ds = \int_0^t \langle BZ, Z \rangle_{L^2} \, ds
\]

This is what was desired.

Note that in the case of a Gelfand triple, when \( W = H = H' \), the term \( \langle BZ, Z \rangle_{L^2} \) will end up reducing to nothing more than \( \|Z\|_{L^2}^2 \).

Thus all the terms in \( 25.7.22 \) converge in probability except for the last term which also must converge in probability because it equals the sum of terms which do. It remains to find what this last term converges to. Thus

\[
\{BX(t), X(t)\} - \{BX_0, X_0\} = 2 \int_0^t \langle Y(u), X(u) \rangle \, du
\]

\[
+ 2 \int_0^t \langle Z \circ J^{-1}(s)BX \circ JdW + \int_0^t \langle BZ, Z \rangle_{L^2} \, ds - a
\]

where \( a \) is the limit in probability of the term

\[
\sum_{j=1}^{q_k-1} \langle B(\Delta X(t_j) - \Delta M(t_j)), \Delta X(t_j) - \Delta M(t_j) \rangle
\]

(25.7.27)

Let \( P_n \) be the projection onto \( \text{span}(e_1, \cdots, e_n) \) where \( \{e_k\} \) is an orthonormal basis for \( W \) with each \( e_k \in V \). Then using

\[
BX(t_{j+1}) - BX(t_j) = BM(t_{j+1}) - BM(t_j) = \int_{t_j}^{t_{j+1}} Y(s) \, ds
\]

the troublesome term of \( 25.7.22 \) above is of the form

\[
\sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \langle Y(s), \Delta X(t_j) - \Delta M(t_j) \rangle \, ds
\]

\[
= \sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \langle Y(s), \Delta X(t_j) - P_n \Delta M(t_j) \rangle \, ds
\]

\[
+ \sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \langle Y(s), -(I - P_n) \Delta M(t_j) \rangle \, ds
\]

which equals

\[
\sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \langle Y(s), X(t_{j+1}) - X(t_j) - P_n \Delta M(t_{j+1}) - M(t_j) \rangle \, ds
\]

(25.7.28)

\[
+ \sum_{j=1}^{q_k-1} \langle B(\Delta X(t_j) - \Delta M(t_j)), (I - P_n)(M(t_{j+1}) - M(t_j)) \rangle
\]

(25.7.29)

The reason for the \( P_n \) is to get \( P_n (M(t_{j+1}) - M(t_j)) \) in \( V \). The sum in \( 25.7.22 \) is dominated by

\[
\left( \sum_{j=1}^{q_k-1} \langle B(\Delta X(t_j) - \Delta M(t_j)), (\Delta X(t_j) - \Delta M(t_j)) \rangle \right)^{1/2}
\]
\[
\left(\sum_{j=1}^{q_k-1} |\langle B ( I - P_n) \Delta M (t_j), ( I - P_n) \Delta M (t_j) \rangle|^2\right)^{1/2}
\]  
(25.7.30)

Now it is known from the above that \(\sum_{j=1}^{q_k-1} |\langle B ( \Delta X (t_j) - \Delta M (t_j)), (\Delta X (t_j) - \Delta M (t_j)) \rangle|\) converges in probability to \(a \geq 0\). If you take the expectation of the square of the other factor, it is no larger than

\[
\|B\| E \left(\sum_{j=1}^{q_k-1} \| ( I - P_n) \Delta M (t_j) \|^2_W\right)
\]

\[
= \|B\| E \left(\sum_{j=1}^{q_k-1} \left\| ( I - P_n) \int_{t_j}^{t_{j+1}} Z (s) dW (s) \right\|^2_W\right)
\]

\[
= \|B\| \left(\sum_{j=1}^{q_k-1} E \left(\left\| \int_{t_j}^{t_{j+1}} Z (s) dW (s) \right\|^2\right)\right)
\]

\[
\leq \|B\| \left(\sum_{j=1}^{q_k-1} \left\| \int_{t_j}^{t_{j+1}} ( I - P_n) Z (s) dW (s) \right\|^2\right)
\]

\[
\leq \|B\| \left(\int_0^T \left\| ( I - P_n) Z (s) \right\|^2_{L_2(Q^{1/2}U,H)} ds\right)
\]

Now letting \(\{g_i\}\) be an orthonormal basis for \(Q^{1/2}U\),

\[
= \|B\| \int_0^T \sum_{i=1}^\infty \left\| ( I - P_n) Z (s) (g_i) \right\|^2_W ds dP
\]  
(25.7.31)

The integrand \(\sum_{i=1}^\infty \left\| ( I - P_n) Z (s) (g_i) \right\|^2_W\) converges to 0. Also, it is dominated by

\[
\sum_{i=1}^\infty \|Z (s) (g_i)\|^2_W \equiv \|Z\|_{L_2(Q^{1/2}U,W)}^2
\]

which is given to be in \(L^1([0, T] \times \Omega)\). Therefore, from the dominated convergence theorem, the expression in (25.7.31) converges to 0 as \(n \to \infty\).

Thus the expression in (25.7.31) is of the form \(f_k g_{nk}\) where \(f_k\) converges in probability to \(a^{1/2}\) as \(k \to \infty\) and \(g_{nk}\) converges in probability to 0 as \(n \to \infty\) independently of \(k\). Now this implies \(f_k g_{nk}\) converges in probability to 0. Here is why,

\[
P\left(\|f_k g_{nk}\| > \varepsilon\right) \leq P\left(2\delta |f_k| > \varepsilon\right) + P\left(2C_\delta |g_{nk}| > \varepsilon\right)
\]

\[
\leq P\left(2\delta |f_k| - a^{1/2}\right) + 2\delta \left|a^{1/2}\right| > \varepsilon\right) + P\left(2C_\delta |g_{nk}| > \varepsilon\right)
\]

where \(\delta |f_k| + C_\delta |g_{nk}| > |f_k g_{nk}|\) and \(\lim_{\delta \to 0} C_\delta = \infty\). Pick \(\delta\) small enough that \(\varepsilon - 2\delta a^{1/2} > \varepsilon/2\). Then this is dominated by

\[
\leq P\left(2\delta |f_k| - a^{1/2}\right) > \varepsilon/2\right) + P\left(2C_\delta |g_{nk}| > \varepsilon\right)
\]

Fix \(n\) large enough that the second term is less than \(\eta\) for all \(k\). Now taking \(k\) large enough, the above is less than \(\eta\). It follows the expression in (25.7.30) and consequently in (25.7.29) converges to 0 in probability.

Now consider the other term (25.7.28) using the \(n\) just determined. This term is of the form

\[
\sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \langle Y (s), X (t_{j+1}) - X (t_j) - P_n (M (t_{j+1}) - M (t_j)) \rangle ds =
\]
\[ \sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \langle Y(s), X_k^\tau(s) - X_k^l(s) - P_n(M_k^* (s) - M_k^l (s)) \rangle ds \]

\[ = \int_{t_1}^t \langle Y(s), X_k^\tau(s) - X_k^l(s) - P_n(M_k^* (s) - M_k^l (s)) \rangle ds \]

where \( M_k^* \) denotes the step function

\[ M_k^* (t) = \sum_{i=0}^{m_k-1} M(t_{i+1}) \chi_{[t_i,t_{i+1}]} (t) \]

and \( M_k^l \) is defined similarly. The term

\[ \int_{t_1}^t \langle Y(s), P_n(M_k^* (s) - M_k^l (s)) \rangle ds \]

converges to 0 for a.e. \( \omega \) as \( k \to \infty \) thanks to continuity of \( t \to M(t) \). However, more is needed than this. Define the stopping time

\[ \tau_p = \inf \{ t > 0 : \| M(t) \|_W > p \} . \]

Then \( \tau_p = \infty \) for all \( p \) large enough, this for a.e. \( \omega \). Let

\[ A_k = \left[ \left| \int_{t_1}^t \langle Y(s), P_n(M_k^* (s) - M_k^l (s)) \rangle ds \right| > \varepsilon \right] \]

\[ P(A_k) = \sum_{p=0}^{\infty} P(A_k \cap (\{ \tau_p = \infty \} \setminus \{ \tau_{p-1} < \infty \})) \] (25.7.32)

Now

\[ P(A_k \cap (\{ \tau_p = \infty \} \setminus \{ \tau_{p-1} < \infty \})) \leq P(A_k \cap (\{ \tau_p = \infty \}))) \]

\[ \leq P \left( \left| \int_{t_1}^t \langle Y(s), P_n((M^{\tau_p})_k^* (s) - (M^{\tau_p})_k^l (s)) \rangle ds \right| > \varepsilon \right) \]

This is so because if \( \tau_p = \infty \), then it has no effect but also it could happen that the defining inequality may hold even if \( \tau_p < \infty \) hence the inequality. This is no larger than an expression of the form

\[ \frac{C_n}{\varepsilon} \int_\Omega \int_0^T \| Y(s) \|_{L^p} \left\| (M^{\tau_p})_k^* (s) - (M^{\tau_p})_k^l (s) \right\|_W d\sigma dP \] (25.7.33)

The inside integral converges to 0 by continuity of \( M \). Also, thanks to the stopping time, the inside integral is dominated by an expression of the form

\[ \int_0^T \| Y(s) \|_{L^p} 2pd\sigma \]

and this is a function in \( L^1 (\Omega) \) by assumption on \( Y \). It follows that the integral in 25.7.33 converges to 0 as \( k \to \infty \) by the dominated convergence theorem. Hence

\[ \lim_{k \to \infty} P(A_k \cap (\{ \tau_p = \infty \})) = 0. \]

Since the sets \( \{ \tau_p = \infty \} \setminus \{ \tau_{p-1} < \infty \} \) are disjoint, the sum of their probabilities is finite. Hence there is a dominating function in 25.7.33 and so, by the dominated convergence theorem applied to the sum,

\[ \lim_{k \to \infty} P(A_k) = \sum_{p=0}^{\infty} \lim_{k \to \infty} P(A_k \cap (\{ \tau_p = \infty \} \setminus \{ \tau_{p-1} < \infty \})) = 0 \]

Thus \( \int_{t_1}^t \langle Y(s), P_n(M_k^* (s) - M_k^l (s)) \rangle ds \) converges to 0 in probability as \( k \to \infty \).
Now consider
\[
\left| \int_t^T \langle Y(s), X_k^t(s) - X_k^t(s) \rangle \, ds \right| \leq \int_0^T |\langle Y(s), X_k^t(s) - X(s) \rangle| \, ds \\
+ \int_0^T |\langle Y(s), X_k^t(s) - X(s) \rangle| \, ds \\
\leq 2 \|Y(\cdot,\omega)\|_{L^p(0,T)} 2^{-k}
\]
for all \(k\) large enough, this by Lemma 25.7.1. Therefore,
\[
\sum_{j=1}^{q_{n-1}} \langle B(\Delta X(t_j) - \Delta M(t_j)), \Delta X(t_j) - \Delta M(t_j) \rangle
\]
converges to 0 in probability. This establishes the desired formula for \(t \in D\). \(\blacksquare\)

Theorem 25.7.2 In Situation 25.2.1, for \(\omega\) off a set of measure zero, for every \(t \notin N_\omega\),
\[
\langle BX(t), X(t) \rangle = \langle BX_0, X_0 \rangle + \int_0^t \left( 2 \langle Y(s), X(s) \rangle + \langle BZ, Z \rangle_{L^2} \right) \, ds \\
+ 2 \int_0^t (Z \circ J^{-1})^* BX \circ JdW
\]
(25.7.34)
Also, there exists a unique continuous, progressively measurable function denoted as \(\langle BX, X \rangle\) such that it equals \(\langle BX(t), X(t) \rangle\) for a.e. \(t\) and \(\langle BX, X \rangle(t)\) equals the right side of the above for all \(t\). In addition to this,
\[
E(\langle BX, X \rangle(t)) = E(\langle BX_0, X_0 \rangle) + E \left( \int_0^t \left( 2 \langle Y(s), X(s) \rangle + \langle BZ, Z \rangle_{L^2} \right) \, ds \right)
\]
(25.7.35)
Also the quadratic variation of the stochastic integral in 25.7.34 is dominated by
\[
C \int_0^t \|Z\|_{L^2}^2 \|BX\|_{W'}^2 \, ds
\]
(25.7.36)
for a suitable constant \(C\). Also \(t \to BX(t)\) is continuous with values in \(W'\) for \(t \in N_\omega^C\).

**Proof:** Let \(t \in N_\omega^C \setminus D\). For \(t > 0\), let \(t(k)\) denote the largest point of \(\mathcal{P}_k\) which is less than \(t\). Suppose \(t(m) < t(k)\). Hence \(m \leq k\). Then
\[
BX(t(m)) = BX_0 + \int_0^{t(m)} Y(s) \, ds + B \int_0^{t(m)} Z(s) \, dW(s)
\]
a similar formula holding for \(X(t(k))\). Thus for \(t > t(m), t \notin N_\omega\),
\[
B(X(t) - X(t(m))) = \int_{t(m)}^t Y(s) \, ds + B \int_{t(m)}^t Z(s) \, dW(s)
\]
which is the same sort of thing studied so far except that it starts at \(t(m)\) rather than at 0 and \(BX_0 = 0\). Therefore, from Lemma 25.7.1 it follows
\[
\langle B(X(t(k)) - X(t(m))), X(t(k)) - X(t(m)) \rangle = \int_{t(m)}^{t(k)} \left( 2 \langle Y(s), X(s) \rangle + \langle BZ, Z \rangle_{L^2} \right) \, ds \\
+ 2 \int_{t(m)}^{t(k)} (Z \circ J^{-1})^* BX(t(m)) \circ JdW
\]
(25.7.37)
Consider that last term. It equals
\[
2 \int_{t(m)}^{t(k)} (Z \circ J^{-1})^* B\left(X(s) - X_m^l(s)\right) \circ J dW
\] (25.7.38)
This is dominated by
\[
2 \left| \int_{t(m)}^{t(k)} (Z \circ J^{-1})^* B\left(X(s) - X_m^l(s)\right) \circ J dW \right| - \int_{t(m)}^{t(k)} (Z \circ J^{-1})^* B\left(X(s) - X_m^l(s)\right) \circ J dW \leq 4 \sup_{t \in [0,T]} \int_{t(m)}^{t} (Z \circ J^{-1})^* B\left(X(s) - X_m^l(s)\right) \circ J dW \right|
\]
In Lemma 25.6.3 the above expression was shown to converge to 0 in probability. Therefore, by the usual appeal to the Borel Cantelli lemma, there is a subsequence still referred to as \( \{m\} \), such that it converges to 0 pointwise in \( \omega \) for all \( \omega \) off some set of measure 0 as \( m \to \infty \). It follows there is a set of measure 0 including the earlier one such that for \( \omega \) not in that set, \( \{X_m^l\} \) converges to 0 in \( \mathbb{R} \). Similar reasoning shows the first term on the right in the non stochastic integral of (25.7.41) is dominated by an expression of the form
\[
4 \int_{0}^{T} |\langle Y(s), X(s) - X_m^l(s) \rangle| \, ds
\]
which clearly converges to 0 thanks to Lemma 25.6.2. Finally, it is obvious that
\[
\lim_{m \to \infty} \int_{t(m)}^{t(k)} \langle BZ, Z \rangle_{L^2} \, ds = 0 \text{ for a.e. } \omega
\]
due to the assumptions on \( Z \). For \( \{g_i\} \) an orthonormal basis of \( Q^{1/2}(U) \),
\[
\langle BZ, Z \rangle_{L^2} = \sum_i \langle R^{-1} B Z(g_i), Z(g_i) \rangle = \sum_i \langle BZ(g_i), Z(g_i) \rangle \\
\leq \|B\| \sum_i \|Z(g_i)\|_{W}^2 \in L^1(0,T) \text{ a.e.}
\]
This shows that for \( \omega \) off a set of measure 0
\[
\lim_{m,k \to \infty} \langle B\left(X(t(k)) - X(t(m))\right), X(t(k)) - X(t(m)) \rangle = 0
\]
Then for \( x \in W \),
\[
|\langle B\left(X(t(k)) - X(t(m))\right), x \rangle| \\
\leq \langle B\left(X(t(k)) - X(t(m))\right), X(t(k)) - X(t(m)) \rangle^{1/2} \langle Bx, x \rangle^{1/2} \\
\leq \langle B\left(X(t(k)) - X(t(m))\right), X(t(k)) - X(t(m)) \rangle^{1/2} \|B\|^{1/2} \|x\|_{W}
\]
and so
\[
\lim_{m,k \to \infty} \|BX(t(k)) - BX(t(m))\|_{W'} = 0
\]
Recall \( t \) was arbitrary in \( N_0^c \) and \( \{t(k)\} \) is a sequence converging to \( t \). Then the above has shown that \( \{BX(t(k))\}_{k=1}^\infty \) is a convergent sequence in \( W' \). Does it converge to \( BX(t) \)? Let \( \xi(t) \in W' \) be what it converges to. Letting \( v \in V \) then, since the integral equation shows that \( t \to BX(t) \) is continuous into \( V' \),
\[
\langle \xi(t), v \rangle = \lim_{k \to \infty} \langle BX(t(k)), v \rangle = \langle BX(t), v \rangle,
\]
and now, since \( V \) is dense in \( W \), this implies \( \xi(t) = BX(t) = B(X(t)) \). Recall also that it was shown earlier that \( BX \) is weakly continuous into \( W' \) hence the strong convergence of \( \{BX(t(k))\}_{k=1}^\infty \) in \( W' \) implies that it converges to \( BX(t) \), this for any \( t \in N_0^c \).
For every \( t \in D \) and for \( \omega \) off the exceptional set of measure zero described earlier,

\[
\langle B (X (t)) , X (t) \rangle = \langle BX_0, X_0 \rangle + \int_0^t \left( 2 \langle Y (s) , X (s) \rangle + \langle BZ, Z \rangle \right) ds \, ds \\
+ 2 \int_0^t (Z \circ J^{-1})^* BX \circ J dW
\]

(25.7.39)

Does this formula hold for all \( t \in [0, T] \)? Maybe not. However, it will hold for \( t \notin N_\omega \). So let \( t \notin N_\omega \).

\[
|\langle BX (t (k)) , X (t (k)) \rangle - \langle BX (t) , X (t) \rangle| \\
\leq |\langle BX (t (k)) , X (t (k)) \rangle - \langle BX (t) , X (t (k)) \rangle| \\
+ |\langle BX (t) , X (t (k)) \rangle - \langle BX (t) , X (t) \rangle| \\
= |\langle BX (t (k)) - X (t) , X (t (k)) \rangle + |\langle BX (t (k)) - X (t) , X (t) \rangle|
\]

Then using the Cauchy Schwarz inequality on each term,

\[
\leq \langle B (X (t (k)) - X (t)) , X (t (k)) - X (t) \rangle^{1/2} \\
\cdot \left( \langle BX (t (k)) , X (t (k)) \rangle^{1/2} + \langle BX (t) , X (t) \rangle^{1/2} \right)
\]

As before, one can use the lower semicontinuity of

\[
t \to \langle B (X (t (k)) - X (t)) , X (t (k)) - X (t) \rangle
\]

on \( N_\omega^C \) along with the boundedness of \( \langle BX (t) , X (t) \rangle \) also shown earlier off \( N_\omega \) to conclude

\[
|\langle BX (t (k)) , X (t (k)) \rangle - \langle BX (t) , X (t) \rangle| \\
\leq C \langle B (X (t (k)) - X (t)) , X (t (k)) - X (t) \rangle^{1/2} \\
\leq C \lim_{m \to \infty} \inf \langle B (X (t (m)) - X (t (m))) , X (t (m)) - X (t (m)) \rangle^{1/2} < \varepsilon
\]

provided \( k \) is sufficiently large. Since \( \varepsilon \) is arbitrary,

\[
\lim_{k \to \infty} \langle BX (t (k)) , X (t (k)) \rangle = \langle BX (t) , X (t) \rangle.
\]

It follows that the formula (25.7.39) is valid for all \( t \notin N_\omega \). Now define the function \( \langle BX, X \rangle (t) \) as

\[
\langle BX, X \rangle (t) \equiv \begin{cases} \\
\langle B (X (t)) , X (t) \rangle, t \notin N_\omega \\
The right side of (25.7.39) if \( t \in N_\omega \)
\end{cases}
\]

Then in short, \( \langle BX, X \rangle (t) \) equals the right side of (25.7.39) for all \( t \in [0, T] \) and is consequently progressively measurable and continuous. Furthermore, for a.e. \( t \), this function equals \( \langle B (X (t)) , X (t) \rangle \). Since it is known on a dense subset, it must be unique.

This implies that \( t \to BX (t) \) is continuous with values in \( W' \) for \( t \notin N_\omega \). Here is why. The fact that the formula (25.7.39) holds for all \( t \notin N_\omega \) implies that \( t \to \langle BX (t) , X (t) \rangle \) is continuous on \( N_\omega^C \). Then for \( x \in W \),

\[
|\langle BX (t) - BX (s) , x \rangle| \leq \langle B (X (t) - X (s)) , X (t) - X (s) \rangle^{1/2} \|B\|^{1/2} \|x\|_{W}.
\]

(25.7.40)

Also

\[
\langle B (X (t) - X (s)) , X (t) - X (s) \rangle \\
= \langle BX (t) , X (t) \rangle + \langle BX (s) , X (s) \rangle - 2 \langle BX (t) , X (s) \rangle
\]

By weak continuity of \( t \to BX (t) \) shown earlier,

\[
\lim_{t \to s} \langle BX (t) , X (s) \rangle = \langle BX (s) , X (s) \rangle.
\]
Therefore,
\[
\lim_{t \to s} \langle B (X (t) - X (s)), X (t) - X (s) \rangle = 0
\]
and so the inequality \ref{25.7.40} implies the continuity of \( t \to BX (t) \) into \( W' \) for \( t \notin N_\omega \). Note that by assumption this function is continuous into \( V' \) for all \( t \). It was also shown that it is weakly continuous into \( W' \) on \([0, T]\) and hence it is bounded in \( W' \).

Now consider the claim about the expectation. Since the stochastic integral
\[
2 \int_0^t (Z \circ J^{-1})^* BX \circ JdW
\]
is only a local martingale, it is necessary to employ a stopping time. We use the function \( \langle BX, X \rangle \) to define this stopping time as
\[
\tau_p \equiv \inf \{ t > 0 : \langle BX, X \rangle (t) > p \}
\]
This is the first hitting time of a continuous process and so it is a valid stopping time. Using this, leads to
\[
\langle BX, X \rangle^{\tau_p} (t) = \langle BX_0, X_0 \rangle + \int_0^t \mathcal{X}_{[0, \tau_p]} (s) (2 \langle Y (s), X (s) \rangle + \langle BZ, Z \rangle \mathcal{L}_2 ds) ds
\]
\[
+ 2 \int_0^t \mathcal{X}_{[0, \tau_p]} (s) (Z \circ J^{-1})^* BX^{\tau_p} \circ JdW
\]
By continuity of \( \langle BX, X \rangle, \tau_p = \infty \) for all \( p \) large enough. Take expectation of both sides of the above. In the integrand of the last term, \( BX \) refers to the function \( BX (t, \omega) \equiv B (X (t, \omega)) \) and so it is progressively measurable because \( X \) is assumed to be so. Hence \( BX^{\tau_p} \) is also progressively measurable and for a.e. Also, for a.e. \( s, \|BX (s \wedge \tau_p)\|_{W'}, \leq \sqrt{p} \sqrt{\|B\|} \). Therefore, one can take expectations and get
\[
E (\langle BX, X \rangle^{\tau_p} (t)) = E (\langle BX_0, X_0 \rangle)
\]
\[
+ E \left( \int_0^t \mathcal{X}_{[0, \tau_p]} (s) (2 \langle Y (s), X (s) \rangle + \langle BZ, Z \rangle \mathcal{L}_2 ds) ds \right)
\]
Now let \( p \to \infty \) and use the monotone convergence theorem on the left and the dominated convergence theorem on the right to obtain the desired result \ref{25.7.35}. The claim about the quadratic variation follows from Corollary \ref{17.11.1}.
Chapter 26

A More Attractive Version

The following lemma is convenient.

**Lemma 26.0.3** Let \( f_n \to f \) in \( L^p ([0, T] \times \Omega, E) \). Then there exists a subsequence \( n_k \) and a set of measure zero \( N \) such that if \( \omega \not\in N \), then

\[
f_{n_k} (\cdot, \omega) \to f (\cdot, \omega)
\]

in \( L^p ([0, T], E) \) and for a.e. \( t \).

**Proof:** We have

\[
P \left( \left\| f_n - f \right\|_{L^p([0,T],E)} > \lambda \right) \leq \frac{1}{\lambda} \int_{\Omega} \left\| f_n - f \right\|_{L^p([0,T],E)} dP \leq \frac{1}{\lambda} \left\| f_n - f \right\|_{L^p([0,T] \times \Omega, E)}
\]

Hence there exists a subsequence \( n_k \) such that

\[
P \left( \left\| f_{n_k} - f \right\|_{L^p([0,T],E)} > 2^{-k} \right) \leq 2^{-k}
\]

Then by the Borel Cantelli lemma, it follows that there exists a set of measure zero \( N \) such that for all \( k \) large enough and \( \omega \notin N \),

\[
\left\| f_{n_k} - f \right\|_{L^p([0,T],E)} \leq 2^{-k}
\]

Now by the usual arguments used in proving completeness, \( f_{n_k} (t) \to f (t) \) for a.e. \( t \).

Also, we have the approximation lemma proved earlier, Lemma 17.3.3.

**Lemma 26.0.4** Let \( \Phi : [0, T] \times \Omega \to V \), be \( B ([0, T]) \times \mathcal{F} \) measurable and suppose

\[
\Phi \in K \equiv L^p ([0, T] \times \Omega; E), \ p \geq 1
\]

Then there exists a sequence of nested partitions, \( \mathcal{P}_k \subseteq \mathcal{P}_{k+1}, \)

\[
\mathcal{P}_k \equiv \{ t^k_0, \cdots, t^k_{m_k} \}
\]

such that the step functions given by

\[
\Phi^+_k (t) = \sum_{j=1}^{m_k} \Phi (t_j^k) \chi_{[t_{j-1}^k, t_j^k]} (t)
\]

\[
\Phi^-_k (t) = \sum_{j=1}^{m_k} \Phi (t_{j-1}^k) \chi_{[t_{j-1}^k, t_j^k)} (t)
\]

both converge to \( \Phi \) in \( K \) as \( k \to \infty \) and

\[
\lim_{k \to \infty} \max \left\{ \left| t_j^k - t_{j+1}^k \right| : j \in \{ 0, \cdots, m_k \} \right\} = 0.
\]
Also, each $\Phi (t_j^k), \Phi (t_{j-1}^k)$ is in $L^p (\Omega; E)$. One can also assume that $\Phi (0) = 0$. The mesh points $\{t_j^k\}_{j=0}^{m_k}$ can be chosen to miss a given set of measure zero. In addition to this, we can assume that $|t_j^k - t_{j-1}^k| = 2^{-nk}$ except for the case where $j = 1$ or $j = m_{nk}$ when this might not be so. In the case of the last subinterval defined by the partition, we can assume $|t_m^k - t_{m-1}^k| = |T - t_{m-1}^k| \geq 2^{-(n_k+1)}$

26.1 The Situation

Now consider the following situation. There are real separable Banach spaces $V, W$ such that $W$ is a Hilbert space and $V \subseteq W, \ W' \subseteq V'$ where $V$ is dense in $W$. Also let $B \in \mathcal{L}(W, W')$ satisfy

$$\langle Bw, w \rangle \geq 0, \langle Bu, v \rangle = \langle Bv, u \rangle$$

Note that $B$ does not need to be one to one. Also allowed is the case where $B$ is the Riesz map. It could also happen that $V = W$. Assume that $B = B(\omega)$ where $B$ is $\mathcal{F}_0$ measurable into $\mathcal{L}(W, W')$. This dependence on $\omega$ will be suppressed in the interest of simpler notation. For convenience, assume $\|B(\omega)\|$ is bounded. This is assumed mainly so that an estimate can be made on $\langle BX_0, X_0 \rangle$ for $X_0$ given in $L^2 (\Omega)$. It probably suffices to simply give an estimate on $\|\langle BX_0, X_0 \rangle\|_{L^1(\Omega)}$ along with something else on the Ito integral. However, it seems at this time like this is more trouble than it is worth.

Situation 26.1.1 Let $X$ have values in $V$ and satisfy the following

$$BX (t) = BX_0 + \int_0^t Y (s) \, ds + BM (t), \tag{26.1.1}$$

$X_0 \in L^2 (\Omega; W)$ and is $\mathcal{F}_0$ measurable. Here $M (t)$ is a continuous $L^2$ martingale having values in $W$. By this is meant that $\lim_{t \to 0^+} \|M (t)\|_{L^2 (\Omega)} = 0$ and for each $\omega, \lim_{t \to 0^+} M (t) = 0, \|M\|_W \in L^2 ([0, T] \times \Omega)$. Assume that $d\|M\| = kdm$ for $k \in L^1 ([0, T] \times \Omega)$, that is, the measure determined by the quadratic variation for the martingale is absolutely continuous with respect to Lebesgue measure as just described.

Assume $Y$ satisfies

$$Y \in K' \equiv L^p' ([0, T] \times \Omega; V'),$$

the $\sigma$ algebra of measurable sets defining $K'$ will be the progressively measurable sets. Here $1/p' + 1/p = 1, p > 1$.

Also the sense in which the equation holds is as follows. For a.e. $\omega$, the equation holds in $V'$ for all $t \in [0, T]$. Thus we are considering a particular representative $X$ for which this happens. Also it is only assumed that $BX (t) = B (X (t))$ for a.e. $t$. Thus $BX$ is the name of a function having values in $V'$ for which $BX (t) = B (X (t))$ for a.e. $t$. Assume that $X$ is progressively measurable also and $X \in L^p ([0, T] \times \Omega, V')$.

The goal is to prove the following Ito formula valid for a.e. $t$ for each $\omega$ off a set of measure zero.

$$\langle BX (t), X (t) \rangle = \langle BX_0, X_0 \rangle + \int_0^t (2 \langle Y (s), X (s) \rangle) \, ds + \left[R^{-1}BM, M\right] (t) + 2 \int_0^t \langle BX, dM \rangle \tag{26.1.2}$$

where $R$ is the Riesz map from $W$ to $W'$. The most significant feature of the last term is that it is a local martingale. The third term on the right is the covariation of the two martingales $R^{-1}BM$ and $M$. It will follow from the argument that this will be nonnegative.

Note that the assumptions on $M$ imply that $\|M\| \in L^1 ([0, T] \times \Omega)$. 
26.2 Preliminary Results

Here are discussed some preliminary results which will be needed. From the integral equation, if $\phi \in L^q (\Omega; V)$ and $\psi \in C^\infty_c (0, T)$ for $q = \max (p, 2)$,

$$
\int_\Omega \int_0^T ((BX) (t) - BM (t) - BX_0) \psi' \phi dt dP = \int_\Omega \int_0^T \int_0^t Y (s) \psi' (t) ds \phi dt dP
$$

Then the term on the right equals

$$
\int_\Omega \int_0^T \int_0^t Y (s) \psi' (t) dt ds \phi (\omega) dP = \int_\Omega \left( - \int_0^T Y (s) \psi (s) ds \right) \phi (\omega) dP
$$

It follows that, since $\phi$ is arbitrary,

$$
\int_0^T ((BX) (t) - BM (t) - BX_0) \psi' (t) dt = - \int_0^T Y (s) \psi (s) ds
$$

in $L^{q'} (\Omega; V')$ and so the weak time derivative of

$$
t \to (BX) (t) - BM (t) - BX_0
$$

equals $Y$ in $L^{q'} \left([0, T]; L^{q'} (\Omega, V')\right)$. Thus, by Theorem 21.4.1, for a.e. $t$,

$$
B (X (t) - M (t)) = BX_0 + \int_0^t Y (s) ds \text{ in } L^{q'} (\Omega, V').
$$

That is,

$$
(BX) (t) = BX_0 + \int_0^t Y (s) ds + BM (t), \quad t \notin \tilde{N}, \quad m (\tilde{N}) = 0
$$

holds in $L^{q'} (\Omega; V')$ where $(BX) (t) = B (X (t))$ a.e. $t$ in this space, for all $t \notin \tilde{N}$, a set of Lebesgue measure zero, in addition to holding for all $t$ for each $\omega$. Now let $\{t_k^n\}_{k=1}^{n=1}$ be partitions for which, from Lemma 26.2.4 there are left and right step functions $X_k^l, X_k^r$, which converge in $L^q ([0, T] \times \Omega; V)$ to $X$ and such that each $\{t_k^n\}_{k=1}^{n=\infty}$ has empty intersection with the set of measure zero $\tilde{N}$ where, in $L^{q'} (\Omega, V')$, $(BX) (t) \neq B (X (t))$ in $L^{q'} (\Omega, V')$. Thus for $t_k$ a generic partition point,

$$
BX (t_k) = B (X (t_k)) \text{ in } L^{q'} (\Omega; V')
$$

Hence there is an exceptional set of measure zero, $N (t_k) \subseteq \Omega$ such that for $\omega \notin N (t_k)$, $BX (t_k) (\omega) = B (X (t_k, \omega))$. Define an exceptional set $N \subseteq \Omega$ to be the union of all these $N (t_k)$. There are countably many and so $N$ is also a set of measure zero. Then for $\omega \notin N$, and $t_k$ any mesh point at all, $BX (t_k) (\omega) = B (X (t_k, \omega))$. This will be important in what follows. In addition to this, from the integral equation, for each of these $\omega \notin N$, $BX (t) (\omega) = B (X (t, \omega))$ for all $t \notin N_\omega \subseteq [0, T]$ where $N_\omega$ is a set of Lebesgue measure zero. Thus the $t_k$ from the various partitions are always in $N_\omega$. By Lemma 21.3.2 there exists a countable set $\{e_i\}$ of vectors in $V$ such that

$$
\langle Be_i, e_j \rangle = \delta_{ij}
$$

and for each $x \in W$,

$$
\langle Bx, x \rangle = \sum_{i=0}^\infty |\langle Bx, e_i \rangle|^2, \quad Bx = \sum_{i=1}^\infty \langle Bx, e_i \rangle B e_i
$$

By this lemma, if $B = B (\omega)$ where $B$ is $F_0$ measurable into $L (W, W')$, then the $e_i$ are also $F_0$ measurable into $V$. Thus the conclusion of the above discussion is that at the mesh points, it is valid to write

$$
\langle (BX) (t_k), X (t_k) \rangle = \langle B (X (t_k)), X (t_k) \rangle
$$

$$
= \sum_i \langle (BX) (t_k), e_i \rangle^2 = \sum_i \langle B (X (t_k)), e_i \rangle^2
$$
just as would be the case if \((BX)(t) = B(X(t))\) for every \(t\). In all which follows, the mesh points will be like this and an appropriate set of measure zero which may be replaced with a larger set of measure zero finitely many times is being neglected. Obviously, one can take a subsequence of the sequence of partitions described above without disturbing the above observations. We will denote these partitions as \(P_k\). Thus we obtain the following interesting lemma.

**Lemma 26.2.1** In the above situation, there exists a set of measure zero \(N \subseteq \Omega\) and a dense subset of \([0, T]\), \(D\) such that for \(\omega \notin N\), \(BX(t, \omega) = B(X(t, \omega))\) for all \(t \in D\). This set \(D\) is the union of nested partitions \(\{P_k\} = \{\{t_j^k\}_{j=1}^{m_k} \}_{k=1}^{\infty}\) such that the left and right step functions \(\{X_k^l\}, \{X_k^r\}\) converge to \(X\) in \(L^p([0, T] \times \Omega; V)\). There is also a set of Lebesgue measure zero \(\hat{N} \subseteq [0, T]\) such that \(BX(t) = B(X(t))\) in \(L^p'([\Omega; V'])\) for all \(t \notin \hat{N}\). Thus for such \(t\), \(BX(t)(\omega) = B(X(t, \omega))\) for a.e.\(\omega\). In particular, for such \(t \notin \hat{N}\),
\[
(\langle BX(t)(\omega), X(t, \omega) \rangle) = \sum_i \langle B(X(t)), e_i \rangle^2 \text{ a.e.}\omega.
\]

\(D\) has empty intersection with \(\hat{N}\). There is also a set of Lebesgue measure zero \(N_\omega\) for each \(\omega \notin N\) defined by \(BX(t, \omega) = B(X(t, \omega))\) for all \(t \notin N_\omega\).

Now define a stopping time.
\[
\sigma^m_\omega = \inf \left\{ t : \langle BX_n^l(t), X_n^l(t) \rangle > q \right\} \quad (26.2.3)
\]
Thus this pertains to the \(n^{th}\) partition. Since \(X_n^l\) is right continuous, this will be a well defined stopping time. Thus for \(t\) one of the partition points,
\[
\langle BX^\sigma^m(t, \omega), X^\sigma^m(t, \omega) \rangle \leq q \quad (26.2.4)
\]
From the definition of \(X_n^l\) and the observation that these partitions are nested,
\[
\lim_{n \to \infty} \sigma^m_\omega = \sigma_\omega
\]
exists because this is a decreasing sequence. There are more available times to consider as \(n\) gets larger and so when the inf is taken, it can only get smaller. Thus
\[
[\sigma_\omega \leq t] = \bigcap_{m=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{n \geq k} \left\{ \sigma^m_\omega \leq t + \frac{1}{m} \right\} \in \bigcap_{m=1}^{\infty} F_{t+(1/m)} = F_t
\]
since it is assumed that the filtration is normal. Thus this appears to be a stopping time. However, I don’t see how to use this.

**Theorem 26.2.2** Let \(\{t_j^n\}_{j=0}^{m_n}\) be the above sequence of partitions of the sort in Lemma [26.0.4] such that if
\[
X_n(t) = \sum_{j=0}^{m_n-1} X(t_j^n)^n \mathcal{X}(t_j^n, t_{j+1}^n)(t)
\]
then \(X_n \to X\) in \(L^p([0, T] \times \Omega, V)\) with the other conditions holding which were discussed above. In particular, \(BX(t) = B(X(t))\) for one of these mesh points. Then the expression
\[
\sum_{j=0}^{m_n-1} \langle B(M(t_{j+1}^n \wedge t) - M(t_j^n \wedge t)), X(t_j^n) \rangle = \sum_{j=0}^{m_n-1} \langle BX(t_j^n), (M(t_{j+1}^n \wedge t) - M(t_j^n \wedge t)) \rangle \quad (26.2.5)
\]
is a local martingale
\[
\int_0^t \langle BX^k, dM \rangle
\]
with \(\{\sigma^n_\omega\}_{q=1}^\infty\) being a localizing sequence.

**Proof:** This follows from Lemma [26.0.4]. This can be seen because, thanks to the fact that \(BX^k = \text{id}\) is bounded, the function \(BX^k\) is in the set \(G\) described there. This is a place where we use that \(d[M] = kdt\).
26.3 The Main Estimate

The argument will be based on a formula which follows in the next lemma.

**Lemma 26.3.1** In Situation [26.1.3] the following formula holds for a.e. \( \omega \) for \( 0 < s < t \). In the following, \( \langle \cdot, \cdot \rangle \) denotes the duality pairing between \( V, V' \).

\[
\langle BX(t), X(t) \rangle = \langle BX(s), X(s) \rangle + 2 \int_s^t \langle Y(u), X(t) \rangle \, du + \langle B(M(t) - M(s)), M(t) - M(s) \rangle
\]

\[
- \langle BX(t) - BX(s) - (M(t) - M(s)), X(t) - X(s) - (M(t) - M(s)) \rangle + 2 \langle BX(s), M(t) - M(s) \rangle
\]

(26.3.6)

Also for \( t > 0 \)

\[
\langle BX(t), X(t) \rangle = \langle BX_0, X_0 \rangle + 2 \int_0^t \langle Y(u), X(t) \rangle \, du + 2 \langle BX_0, M(t) \rangle + \langle BM(t), M(t) \rangle - \langle BX(t) - BX_0 - BM(t), X(t) - X_0 - M(t) \rangle
\]

(26.3.7)

**Proof:** From the formula which is assumed to hold,

\[
BX(t) = BX_0 + \int_0^t Y(u) \, du + BM(t)
\]

\[
BX(s) = BX_0 + \int_0^s Y(u) \, du + BM(s)
\]

Then

\[
BM(t) - BM(s) + \int_s^t Y(u) \, du = BX(t) - BX(s)
\]

It follows that

\[
\langle B(M(t) - M(s)), M(t) - M(s) \rangle - \langle BX(t) - BX(s) - (M(t) - M(s)), X(t) - X(s) - (M(t) - M(s)) \rangle + 2 \langle BX(s), M(t) - M(s) \rangle
\]

\[
= \langle B(M(t) - M(s)), M(t) - M(s) \rangle - \langle BX(t) - BX(s), X(t) - X(s) \rangle + 2 \langle BX(t) - BX(s), M(t) - M(s) \rangle
\]

\[
- \langle B(M(t) - M(s)), M(t) - M(s) \rangle + 2 \langle BX(s), M(t) - M(s) \rangle
\]

Some terms cancel and this yields

\[
= -2 \langle BX(t) - BX(s), X(t) - X(s) \rangle + 2 \langle BX(s), M(t) - M(s) \rangle
\]

\[
= -\langle B(X(t) - X(s)), X(t) - X(s) \rangle + 2 \left( BX(t) - BX(s) - \int_s^t Y(u) \, du, X(t) \right)
\]

\[
= -\langle BX(t), X(t) \rangle - \langle BX(s), X(s) \rangle + 2 \langle BX(t), X(t) \rangle + 2 \langle BX(t), X(t) \rangle
\]

\[
-2 \langle BX(s), X(t) \rangle - 2 \int_s^t \langle Y(u), X(t) \rangle \, du
\]

\[
= \langle BX(t), X(t) \rangle - \langle BX(s), X(s) \rangle - 2 \int_s^t \langle Y(u), X(t) \rangle \, du
\]
Therefore,

\[ \langle BX (t) , X (t) \rangle - \langle BX (s) , X (s) \rangle = 2 \int_s^t \langle Y (u) , X (t) \rangle du + \langle B (M (t) - M (s)) , M (t) - M (s) \rangle \]

\[ - \langle BX (t) - BX (s) , M (t) - M (s) \rangle , X (t) - X (s) \rangle - \langle M (t) - M (s) \rangle \]

\[ + 2 \langle BX (s) , M (t) - M (s) \rangle \]

The following phenomenal estimate holds and it is this estimate which is the main idea in proving the Ito formula. The last assertion about continuity is like the well known result that if \( y \in L^p (0, T; V) \) and \( y' \in L^p (0, T; V') \), then \( y \) is actually continuous a.e. with values in \( H \), for \( V, H, V' \) a Gelfand triple. Later, this continuity result is strengthened further to give strong continuity.

**Lemma 26.3.2** In the Situation \[ \text{[26.3.1]} \], the following holds for all \( t \not\in \hat{N} \),

\[ E (\langle BX (t) , X (t) \rangle) < C \left( ||Y||_{K'}, ||X||_K , E (\| [M] (T) \|) , \| [BX_0, X_0] \|_{L^1(\Omega)} \right) < \infty. \tag{26.3.8} \]

where \( K, K' \) were defined earlier. In fact,

\[ E \left( \sup_{t \in [0,T]} \sum_i \langle BX (t) , e_i \rangle^2 \right) \leq C \left( ||Y||_{K'}, ||X||_K , E (\| [M] (T) \|) , \| [BX_0, X_0] \|_{L^1(\Omega)} \right) \]

Also, \( C \) is a continuous function of its arguments, increasing in each one, and \( C (0, 0, 0, 0) = 0 \). Thus for a.e. \( \omega \),

\[ \sup_{t \in \hat{N}^c} \langle BX (t, \omega) , X (t, \omega) \rangle \leq C (\omega) < \infty. \]

Also for \( \omega \) off a set of measure zero described earlier, \( t \to BX (t) (\omega) \) is weakly continuous with values in \( W' \) on \( [0, T] \). Also \( t \to \langle BX (t) , X (t) \rangle \) is lower semicontinuous on \( \hat{N}^c \).

**Proof:** Consider the formula in Lemma \[ \text{[26.3.1]} \]

\[ \langle BX (t) , X (t) \rangle = \langle BX (s) , X (s) \rangle \]

\[ + 2 \int_s^t \langle Y (u) , X (t) \rangle du + \langle B (M (t) - M (s)) , M (t) - M (s) \rangle \]

\[ - \langle B (X (t) - X (s) - (M (t) - M (s))) , X (t) - X (s) - (M (t) - M (s)) \rangle \]

\[ + 2 \langle BX (s) , M (t) - M (s) \rangle \]

(26.3.9)

Now let \( t_j \) denote a point of \( \mathcal{P}_k \) from Lemma \[ \text{[26.0.4]} \]. Then for \( t_j > 0 \), \( X (t_j) \) is just the value of \( X \) at \( t_j \) but when \( t = 0 \), the definition of \( X (0) \) in this step function is \( X (0) \equiv 0 \). Thus

\[ \sum_{j=1}^{m-1} \langle BX (t_{j+1}) , X (t_{j+1}) \rangle - \langle BX (t_j) , X (t_j) \rangle \]

\[ + \langle BX (t_1) , X (t_1) \rangle - \langle BX_0, X_0 \rangle \]

\[ = \langle BX (t_m) , X (t_m) \rangle - \langle BX_0, X_0 \rangle \]

Using the formula in Lemma \[ \text{[26.3.1]} \] for \( t = t_m \) this yields

\[ \langle BX (t_m) , X (t_m) \rangle - \langle BX_0, X_0 \rangle = 2 \sum_{j=1}^{t_{m+1}} \int_{t_j}^{t_{j+1}} \langle Y (u) , X_k (u) \rangle du \]
First consider
\[
2 \int_0^{t_1} \langle Y (u), X (t_1) \rangle \, du + 2 \langle BX_0, M (t_1) \rangle + \langle BM (t_1), M (t_1) \rangle
\]
(26.3.10)

Each term of the above converges to 0 for a.e. \( \omega \) as \( k \to \infty \) and in \( L^1 (\Omega) \). This follows right away for the second two terms from the assumptions on \( M \) given in the situation. Recall it was assumed that \( \| B (\omega) \| \) is bounded. This is where it is convenient to make this assumption. Consider the first term. This term is dominated by
\[
\left( \int_0^{t_1} \left\| Y (u) \right\|^{p'} \, du \right)^{1/p'} \left( \int_0^T \left\| X_k^i (u) \right\|^p \, du \right)^{1/p} \leq C (\omega) \left( \int_0^{t_1} \left\| Y (u) \right\|^{p'} \, du \right)^{1/p'} \left( \int_\Omega C (\omega)^p \, dP \right)^{1/p} < \infty
\]

Hence this converges to 0 for a.e. \( \omega \) and also converges to 0 in \( L^1 (\Omega) \).

At this time, not much is known about the last term in 26.3.10, but it is negative and is about to be neglected anyway.

The second term on the right equals
\[
2 \int_{t_1}^{t_m} \langle BX_k^i, dM \rangle = 2 \int_0^{t_m} \langle BX_k^i, dM \rangle + \epsilon (k)
\]
where \( \epsilon (k) \to 0 \) for a.e. \( \omega \) and in \( L^1 (\Omega) \). Also note that since \( \langle BM (t_1), M (t_1) \rangle \) converges to 0 in \( L^1 (\Omega) \) and for a.e. \( \omega \), the sum involving
\[
\langle B (M (t_{j+1}) - M (t_j)), M (t_{j+1}) - M (t_j) \rangle
\]
can be started at 0 rather than 1 at the expense of adding in a term which converges to 0 a.e. and in \( L^1 (\Omega) \). Thus 26.3.10 is of the form
\[
\langle BX (t_m), X (t_m) \rangle - \langle BX_0, X_0 \rangle = \epsilon (k) + 2 \int_0^{t_m} \langle Y (u), X_k^i (u) \rangle \, du +
\]
\[
+ 2 \int_0^{t_m} \langle BX_k^i, dM \rangle
\]
\[
+ \sum_{j=0}^{m-1} \langle B (M (t_{j+1}) - M (t_j)), M (t_{j+1}) - M (t_j) \rangle
\]
\[
- \sum_{j=1}^{m-1} \langle B (X (t_{j+1}) - X (t_j) - (M (t_{j+1}) - M (t_j))), X (t_{j+1}) - X (t_j) - (M (t_{j+1}) - M (t_j)) \rangle
\]
Thus the expectation of that last term in (26.3.12) is no larger than

\[ E \left( \sum_{j=0}^{m-1} [M]^{t_{j+1}} (t_m) - [M]^{t_j} (t_m) \right) = \| B \| \| E \left( \sum_{j=0}^{m-1} [M]^{t_{j+1}} (t_m) - [M]^{t_j} (t_m) \right) \]
Thus for \( x \) and for each \( \omega \), where the constant \( C \) exists

\[
\Omega \leq \sum_{i=0}^{\infty} \langle B \omega, e_i \rangle^2, \quad B \omega = \sum_{i=1}^{\infty} \langle B \omega, e_i \rangle B e_i
\]

Now let \( D \) denote the union of these nested partitions. Then from the monotone convergence theorem, \( E \left( \sup_{t \in D} \langle B X (t), X (t) \rangle \right) \) is no larger than the right side of (26.3.13). Since this is true for all \( \varepsilon > 0 \), it follows

\[
E \left( \sup_{t \in D} \langle B X (t), X (t) \rangle \right) \leq C \left( E \left( \langle B X_0, X_0 \rangle \right), \| Y \|_{K'}, \| X \|_{K} \right) + E \left( |e (k)| \right)
\]

where \( C (\cdots) \) is increasing in each argument, continuous, and \( C (0) = 0 \). Thus, enlarging \( N \), for \( \omega \notin N \), \( \sup_{t \in D} \langle B X (t), X (t) \rangle = C (\omega) < \infty \) (26.3.15) where \( \int_{\Omega} C (\omega) dP < \infty \). By Lemma (4.16.24), there exists a countable set \( \{ e_i \} \) of vectors in \( V \) such that

\[
\langle B e_i, e_j \rangle = \delta_{ij}
\]

and for each \( x \in W \),

\[
\langle B x, x \rangle = \sum_{i=0}^{\infty} \langle B x, e_i \rangle^2, \quad B x = \sum_{i=1}^{\infty} \langle B x, e_i \rangle B e_i
\]

Thus for \( t \) not in a set of measure zero off which \( B X (t) = B (X (t)) \),

\[
\langle B X (t), X (t) \rangle = \sum_{i=0}^{\infty} \langle B X (t), e_i \rangle^2 = \sup_{m} \sum_{k=1}^{m} \langle B X (t), e_i \rangle^2
\]
Now from the formula for $BX(t)$, it follows that $BX$ is continuous into $V'$. For any $t \notin \mathcal{N}$ so that $(BX)(t) = B(X(t))$ in $L^q(\Omega; V')$ and letting $t_k \to t$ where $t_k \in D$, Fatou’s lemma implies

$$E \left( \langle BX(t), X(t) \rangle \right) = \sum_i E \left( \langle BX(t), e_i \rangle^2 \right) = \sum_i \lim_{k \to \infty} E \left( \langle BX(t_k), e_i \rangle^2 \right)$$

$$\leq \lim_{k \to \infty} \inf \sum_i E \left( \langle BX(t_k), e_i \rangle^2 \right) = \lim_{k \to \infty} E \left( \langle BX(t_k), X(t_k) \rangle \right)$$

$$\leq C \left( ||Y||_{K'}, ||X||_{K'}, ||Z||_J, ||(BX_0, X_0)||_{L^1(\Omega)} \right)$$

In addition to this, for arbitrary $t \in [0,T]$, and $t_k \to t$ from $D$,

$$\sum_i \langle BX(t), e_i \rangle^2 \leq \lim_{k \to \infty} \inf \sum_i \langle BX(t_k), e_i \rangle^2 \leq \sup_{s \in D} \langle BX(s), X(s) \rangle$$

Hence

$$\sup_{t \in [0,T]} \sum_i \langle BX(t), e_i \rangle^2 \leq \sup_{s \in D} \langle BX(s), X(s) \rangle = \sup_{s \in D} \sum_i \langle BX(s), e_i \rangle^2 \leq \sup_{t \in [0,T]} \sum_i \langle BX(t), e_i \rangle^2$$

It follows that $\sup_{t \in [0,T]} \sum_i \langle BX(t), e_i \rangle^2$ is measurable and

$$E \left( \sup_{t \in [0,T]} \sum_i \langle BX(t), e_i \rangle^2 \right) \leq E \left( \sup_{s \in D} \langle BX(s), X(s) \rangle \right)$$

$$\leq C \left( ||Y||_{K'}, ||X||_{K'}, ||Z||_J, ||(BX_0, X_0)||_{L^1(\Omega)} \right)$$

And so, for $\omega$ off a set of measure zero, $\sup_{t \in [0,T]} \sum_i \langle BX(t), e_i \rangle^2$ is bounded above. Include this exceptional set in $N$.

Also for $t \notin N_{\omega}$ and a given $\omega \notin N$, letting $t_k \to t$ for $t_k \in D$,

$$\langle BX(t), X(t) \rangle = \sum_i \langle BX(t), e_i \rangle^2 \leq \lim_{k \to \infty} \inf \sum_i \langle BX(t_k), e_i \rangle^2$$

$$= \lim_{k \to \infty} \inf \langle BX(t_k), X(t_k) \rangle \leq \sup_{t \in D} \langle BX(t), X(t) \rangle$$

and so

$$\sup_{t \notin N_{\omega}} \langle BX(t), X(t) \rangle \leq \sup_{t \in D} \langle BX(t), X(t) \rangle \leq \sup_{t \notin N_{\omega}} \langle BX(t), X(t) \rangle$$

From (26.3.13),

$$\sup_{t \notin N_{\omega}} \langle BX(t), X(t) \rangle = C(\omega) \ a.e. \omega$$

where $\int \Omega C(\omega) \ dP < \infty$. In particular, $\sup_{t \notin N_{\omega}} \langle BX(t), X(t) \rangle$ is bounded for a.e. $\omega$ say for $\omega \notin N$ where $N$ includes the earlier sets of measure zero. This shows that $BX(t)$ is bounded in $W'$ for $t \in N_\omega^C$.

If $v \in V$, then for $\omega \notin N$,

$$\lim_{t \to s} \langle BX(t), v \rangle = \langle BX(s), v \rangle, \ t, s$$

Therefore, since for such $\omega$, $\|BX(t)\|_{W'}$ is bounded for $t \notin N_{\omega}$, the above holds for all $v \in W$ also. Therefore, for a.e. $\omega$, $t \to BX(t, \omega)$ is weakly continuous with values in $W'$ for $t \notin N_{\omega}$.

Note also that

$$\int_0^T \int_\Omega \|BX(t)\|^2 \ dP dt \leq \int_0^T \|B\|^{1/2} \langle BX(t), X(t) \rangle \ dtdP$$

$$\leq C \left( ||Y||_{K'}, ||X||_{K'}, ||Z||_J, ||(BX_0, X_0)||_{L^1(\Omega)} \right) \|B\|^{1/2} T$$

(26.3.13)

Eventually, it is shown that in fact, the function $t \to BX(t, \omega)$ is continuous with values in $W'$. The above shows that $BX \in L^2([0,T] \times \Omega, W')$. 


Finally consider the claim of weak continuity of $BX$ into $W'$. From the integral equation, $BX$ is continuous into $V'$. Also $t \to BX(t)$ is bounded in $W'$ on $N^C_\omega$. Let $s \in [0,T]$ be arbitrary. I claim that if $t_n \to s$, $t_n \to D$, it follows that $BX(t_n) \to BX(s)$ weakly in $W'$. If not, then there is a subsequence, still denoted as $t_n$ such that $BX(t_n) \to Y$ weakly in $W'$ but $Y \neq BX(s)$. However, the continuity into $V'$ means that for all $v \in V$, 
\[
\langle Y, v \rangle = \lim_{n \to \infty} \langle BX(t_n), v \rangle = \langle BX(s), v \rangle
\]
which is a contradiction since $V$ is dense in $W$. This establishes the claim. Also this shows that $BX(s)$ is bounded in $W'$.
\[
|\langle BX(s), w \rangle| = \lim_{n \to \infty} |\langle BX(t_n), w \rangle| \leq \liminf_{n \to \infty} \|BX(t_n)\|_{W'} \|w\|_{W} \leq C(\omega) \|w\|_{W}
\]
Now a repeat of the above argument shows that $s \to BX(s)$ is weakly continuous into $W'$. \[\blacksquare\]

### 26.4 A SIMPLIFICATION OF THE FORMULA

This estimate in Lemma 26.4 also provides a way to simplify one of the formulas derived earlier \textbf{in the case that} $X_0 \in L^p(\Omega, V)$ so that $X - X_0 \in L^p([0,T] \times \Omega, V)$. Refer to \[\text{46.3.11}\]. One term there is
\[
\langle B(X(t_1) - X_0 - M(t_1)), X(t_1) - X_0 - M(t_1) \rangle
\]
Also,
\[
\langle B(X(t_1) - X_0 - M(t_1)), X(t_1) - X_0 - M(t_1) \rangle \leq 2 \langle B(X(t_1) - X_0), X(t_1) - X_0 \rangle + 2 \langle BM(t_1), M(t_1) \rangle
\]
It was observed above that $2 \langle BM(t_1), M(t_1) \rangle \to 0$ a.e. and also in $L^1(\Omega)$ as $k \to \infty$. Apply the above lemma to $\langle B(X(t_1) - X_0), X(t_1) - X_0 \rangle$ using $[0,t_1]$ instead of $[0,T]$. The new $X_0$ equals 0. Then from the estimate \[\text{26.3.11}\], it follows that
\[
E((\langle B(X(t_1) - X_0), X(t_1) - X_0 \rangle) \to 0
\]
as $k \to \infty$. Taking a subsequence, we could also assume that
\[
\langle B(X(t_1) - X_0), X(t_1) - X_0 \rangle \to 0
\]
a.e. $\omega$ as $k \to \infty$. Then, using this subsequence, it would follow from \[\text{46.3.11}\],
\[
\langle BX(t_m), X(t_m) \rangle - \langle BX_0, X_0 \rangle = e(k) +
\]
\[
= 2 \int_0^{t_m} \langle Y(u), X_k(u) \rangle du + 2 \int_0^{t_m} \langle BX_k, dM \rangle + \sum_{j=0}^{m-1} \langle B(M(t_{j+1}) - M(t_j)), M(t_{j+1}) - M(t_j) \rangle
\]
\[
- \sum_{j=1}^{m-1} \langle B(\Delta X(t_j) - \Delta M(t_j)), \Delta X(t_j) - \Delta M(t_j) \rangle
\]
where $e(k) \to 0$ in $L^1(\Omega)$ and a.e. $\omega$ and
\[
\Delta X(t_j) \equiv X(t_{j+1}) - X(t_j)
\]
$\Delta M(t_j)$ being defined similarly. Note how this eliminated the need to consider the term
\[
\langle B(X(t_1) - X_0 - M(t_1)), X(t_1) - X_0 - M(t_1) \rangle
\]
in passing to a limit. This is a very desirable thing to be able to conclude.

Can you obtain something similar even in case $X_0$ is not assumed to be in $L^p(\Omega, V)$? Let $X_{0k} \in L^p(\Omega, V) \cap L^2(\Omega, W)$, $X_{0k} \to X_0$ in $L^2(\Omega, W)$. Then from the usual arguments involving the Cauchy Schwarz inequality,
\[
\langle B(X(t_1) - X_0), X(t_1) - X_0 \rangle^{1/2} \leq \langle B(X(t_1) - X_{0k}), X(t_1) - X_{0k} \rangle^{1/2}
\]
\[
+ \langle B(X_{0k} - X_0), X_{0k} - X_0 \rangle^{1/2}
\]
Also, restoring the superscript to identify the parition,
\[ B (X (t^n_k) - X_{0k}) = B (X_0 - X_{0k}) + \int_0^{t^n_k} Y(s) \, ds + BM (t^n_k) . \]

Of course \( \| X - X_{0k} \|_K \) is not bounded, but for each \( k \) it is finite. There is a sequence of partitions \( \mathcal{P}_k, \| \mathcal{P}_k \| \to 0 \) such that all the above holds. In the definitions of \( K, \tilde{K}, E ([M] (T)) \) replace \([0, T]\) with \([0, t]\) and let the resulting spaces be denoted by \( K_t, \tilde{K}_t \). Let \( n_k \) denote a subsequence of \( \{ k \} \) such that
\[ \| X - X_{0k} \|_{K_{t,n_k}} < 1/k. \]

Then from the above lemma,
\[ E (\langle B (X (t^n_k) - X_{0k}) , X (t^n_k) - X_{0k} \rangle) \]
\[ \leq C \left( \langle B (X_0 - X_{0k}) , X_0 - X_{0k} \rangle_{L^1(\Omega)} , \| Y \|_{\tilde{K}_{t,n_k}} , \| X - X_{0k} \|_{K_{t,n_k}} , E ([M] (t^n_k)) \right) \]
\[ \leq C \left( \langle B (X_0 - X_{0k}) , X_0 - X_{0k} \rangle_{L^1(\Omega)} , \| Y \|_{\tilde{K}_{t,n_k}} , 1/k , E ([M] (t^n_k)) \right) \]

Hence
\[ E (\langle B (X (t^n_k) - X_0) , X (t^n_k) - X_0 \rangle) \]
\[ \leq 2E (\langle B (X (t^n_k) - X_0) , X (t^n_k) - X_0 \rangle) + 2E (\langle B (X_0 - X_{0k}) , X_0 - X_{0k} \rangle) \]
\[ \leq 2C \left( \langle B (X_0 - X_{0k}) , X_0 - X_{0k} \rangle_{L^1(\Omega)} , \| Y \|_{\tilde{K}_{t,n_k}} , 1/k , E ([M] (t^n_k)) \right) \]
\[ + 2\| B \| \| X_0 - X_{0k} \|_{L^2(\Omega, W)}^2 \]

which converges to 0 as \( k \to \infty \). It follows that there exists a suitable subsequence such that \[\text{(26.4.18)}\] holds even in the case that \( X_0 \) is only known to be in \( L^2 (\Omega, W) \). From now on, assume this subsequence for the partitions \( \mathcal{P}_k \). Thus \( k \) will really be \( n_k \) and it suffices to consider the limit as \( k \to \infty \) of the equation of \[\text{(26.4.18)}\]. To emphasize this point again, the reason for the above observations is to argue that, even when \( X_0 \) is only in \( L^2 (\Omega, W) \), one can neglect
\[ \langle B (X (t_1) - X_0 - M (t_1)) , X (t_1) - X_0 - M (t_1) \rangle \]
in passing to the limit as \( k \to \infty \) provided a suitable subsequence is used.

## 26.5 Convergence

Convergence will be shown for a subsequence and from now on every sequence will be a subsequence of this one. Since \( B X \in L^2 ([0, T] \times \Omega; W') \) which was shown above, there exists a sequence of partitions of the sort described above such that also, in addition to the other claims
\[ B X^n_k \to B X, B X^n_k \to B X \]
in \( L^2 ([0, T] \times \Omega, W') \). Then the next lemma improves on this.

**Lemma 26.5.1** There exists a subsequence still denoted with the subscript \( k \) and an enlarged set of measure zero \( \tilde{N} \) including the earlier one such that \( B X^n_k (t) , B X^n_k (t) \) also converges pointwise a.e. to \( B X (t) \) in \( W' \) and \( X^n_k (t) , X^n_k (t) \) converge pointwise a.e. in \( V \) to \( X (t) \) for \( \omega \notin \tilde{N} \) as well as having convergence of \( X^n_k (\cdot, \omega) \) to \( X (\cdot, \omega) \) in \( L^p ([0, T] ; V) \) and \( B X^n_k (\cdot, \omega) \) to \( B X (\cdot, \omega) \) in \( L^2 ([0, T] ; W') \).

**Proof:** To see that such a sequence exists, let \( n_k \) be such that
\[ \int_{\Omega} \int_{0}^{T} \| B X^n_{n_k} (t) - B X (t) \|_{W'}^2 \, dt \, dP + \int_{\Omega} \int_{0}^{T} \| X^n_{n_k} (t) - X (t) \|_{V}^p \, dt \, dP + \]
\[ \int_{\Omega} \int_{0}^{T} \| B X^n_{n_k} (t) - B X (t) \|_{W'}^2 \, dt \, dP + \int_{\Omega} \int_{0}^{T} \| X^n_{n_k} (t) - X (t) \|_{V}^p \, dt \, dP < 4^{-k}. \]
Then
\[ P \left( \int_0^T \|BX^i_{n_k}(t) - BX(t)\|_{W^*}^2 \, dt + \int_0^T \|X^r_{n_k}(t) - X(t)\|_V^p \, dt \right) +
\int_0^T \|BX^i_{n_k}(t) - BX(t)\|_{W^*}^2 \, dt + \int_0^T \|X^l_{n_k}(t) - X(t)\|_V^p \, dt > 2^{-k} \right)
\leq 2^k (4^{-k}) = 2^{-k}

and so by Borel Cantelli lemma, there is a set of measure zero \( N \) such that if \( \omega \notin N \),
\[ \int_0^T \|BX^i_{n_k}(t) - BX(t)\|_{W^*}^2 \, dt + \int_0^T \|X^r_{n_k}(t) - X(t)\|_V^p \, dt \leq 2^{-k} \]

for all \( k \) large enough. By the usual proof of completeness of \( L^p \), it follows that \( X^r_{n_k}(t) \to X(t) \) for a.e. \( t \), this for each \( \omega \notin N \), a similar assertion holding for \( X^l_{n_k} \). Also \( BX^i_{n_k}(t) \to BX(t) \) for a.e. \( t \), similar for \( BX^r_{n_k}(t) \). We denote these subsequences as \( \{X^r_k\}_{k=1}^\infty, \{X^l_k\}_{k=1}^\infty \).

Define the following stopping time.
\[ \tau_p = \inf \left\{ t : \sum_i \langle BX(t), e_i \rangle^2 > p \right\} \]

By Lemma \ref{lem:borel-cantelli}, \( \tau_p = \infty \) for all \( p \) large enough off some set of measure zero. Also, \( BX(t)(\omega) = B(X(t, \omega)) \) for a.e. \( t \) and so for a.e.t, \( \langle BX(t), X(t) \rangle = \sum_i \langle BX(t), e_i \rangle^2 \) and so \( \|BX^{\tau_p}(t)\|_{W^*} \leq \|B\| \sqrt{p} \) for a.e.t. Hence \( BX^{\tau_p} \in L^\infty([0,T] \times \Omega, W^*) \).

**Lemma 26.5.2** The process \( \int_0^t \langle BX^i_k, dM \rangle \) converges in probability as \( k \to \infty \) to \( \int_0^t \langle BX, dM \rangle \) which is a local martingale. Also, there is a subsequence and an enlarged set of measure zero \( N \) such that for \( \omega \) not in this set, the convergence is uniform on \([0,T] \).

**Proof:** By assumption, \( d[M] = kdt \) for some \( k \in L^1([0,T] \times \Omega) \) and so \( BX^{\tau_p} \in G \) where \( G \) was the class of functions for which one can write \( \int_0^t \langle BX, dM \rangle \). By the Burkholder Davis Gundy inequality,
\[
P \left( \sup_t \left| \int_0^{t \wedge \tau_p} \langle B(X^i_k) - BX, dM \rangle \right| > \varepsilon \right) =
P \left( \sup_t \left| \int_0^t \langle X^i_{[0,\tau_p]}(B(X^i_k) - BX, dM) \rangle \right| > \varepsilon \right) \leq \frac{C}{\varepsilon} \int_\Omega \left( \int_0^{T \wedge \tau_p} \|B(X^i_k) - BX\|_{W^*}^2 \, kdt \right)^{1/2} dP
\]
\[
= \frac{C}{\varepsilon} \int_\Omega \left( \int_0^T \langle X^i_{[0,\tau_p]} \|B(X^i_k) - BX\|_{W^*}^2 \, kdt \right)^{1/2} dP \tag{26.5.19}
\]

Let
\[ A_k = \sup_t \left| \int_0^t \langle BX^i_k - BX, dM \rangle \right| > \varepsilon \]

Then, since \( \tau_p = \infty \) for all \( p \) large enough,
\[ A_k = \cup_{p=0}^\infty A_k \cap (\tau_p = \infty) \setminus (\tau_{p-1} \neq \infty) \]

Consider \( BX^{\tau_p}_k \). If \( t > \tau_p \), what of the values of \( BX^{\tau_p}_k \)? It equals \( BX(s) \) where \( s \) is one of the mesh points \( s \leq \tau_p \) because this is a left step function. Therefore,
\[
\langle BX^{\tau_p}_k(s), X^{\tau_p}_k(s) \rangle = \langle B(X^{\tau_p}_k(s)), X^{\tau_p}_k(s) \rangle
\]
As to $\chi_{[0, \tau_p]} BX$, it follows that for all $t \leq \tau_p$ you have $\sum_t \langle BX(t), e_i \rangle^2 \leq p$ and so, since this equals $\langle B(X(t)), X(t) \rangle$ a.e. $t$, it follows that $\|\chi_{[0, \tau_p]} BX(t)\|_W$ is bounded by a constant depending on $p$ for $a.e.$ It follows that $BX$ and $BX_k^l$ are bounded. Now by Lemma 26.6.1, $BX_k^l(t) \to BX(t)$ a.e. $t$ and the term $\|B(X_k^l) - BX\|_W^2$ is essentially bounded. Therefore, in \textnormal{\cite{26.6.1}}, the integral converges to $0$. From this formula,

$$P(A_k \cap ([\tau_p = \infty] \setminus [\tau_{p-1} \neq \infty])) \leq P(A_k \cap ([\tau_p = \infty]))$$

Thus

$$\lim_{k \to \infty} P(A_k \cap ([\tau_p = \infty] \setminus [\tau_{p-1} \neq \infty])) = 0$$

Then

$$P(A_k) = \sum_{p=1}^{\infty} P(A_k \cap ([\tau_p = \infty] \setminus [\tau_{p-1} \neq \infty]))$$

and taking limits using the dominated convergence theorem on the sum on the right,

$$\lim_{k \to \infty} P(A_k) = \sum_{p=1}^{\infty} \lim_{k \to \infty} P(A_k \cap ([\tau_p = \infty] \setminus [\tau_{p-1} \neq \infty])) = 0$$

This proves convergence in probability.

$$\lim_{k \to \infty} P\left( \sup_t \left| \int_0^t \langle B(X_k^l) - BX, dM \rangle \right| > \varepsilon \right) = 0$$

Then selecting a subsequence, still denoted with $k$, we can obtain

$$P\left( \sup_t \left| \int_0^t \langle B(X_k^l) - BX, dM \rangle \right| > \frac{1}{k} \right) < 2^{-k}$$

and so, by the Borel Cantelli lemma, there is a set of measure zero $N$ such that for this subsequence, for all $\omega \notin N$,

$$\sup_t \left| \int_0^t \langle B(X_k^l) - BX, dM \rangle \right| \leq \frac{1}{k}$$

for all $k$ large enough. Thus convergence is uniform.  

From now on, include $N$ in the exceptional set and every subsequence will be a subsequence of this one.

### 26.6  The Ito Formula

Now at long last, here is the first version of the Ito formula valid on the partition points.

**Lemma 26.6.1** In Situation 26.5.1, let $D$ be as above, the union of all the positive mesh points for all the $P_k$. Also assume $X_0 \in L^2(\Omega; W)$. Then for $\omega \notin N$ the exceptional set of measure zero in $\Omega$ and every $t \in D$,

$$\langle BX(t), X(t) \rangle = \langle BX_0, X_0 \rangle + \int_0^t 2 \langle Y(s), X(s) \rangle ds$$

$$+ [R^{-1}BM, M](t) + 2 \int_0^t \langle BX, dM \rangle$$

for $R$ the Riesz map from $W$ to $W'$. The covariation term $[R^{-1}BM, M](t)$ is nonnegative.
Proof: Let \( t \in D \). Then \( t \in \mathcal{P}_k \) for all \( k \) large enough. Consider (26.3.11),

\[
\langle BX(t), X(t) \rangle - \langle BX_0, X_0 \rangle = e(k) + 2 \int_0^t \langle Y(u), X_k(u) \rangle \, du
\]

\[
+ 2 \int_0^t \langle BX, dM \rangle + \sum_{j=0}^{q_k-1} \langle B(M(t_{j+1}) - M(t_j)), M(t_{j+1}) - M(t_j) \rangle
\]

\[- \sum_{j=1}^{q_k-1} \langle B(\Delta X(t_j) - \Delta M(t_j)), \Delta X(t_j) - \Delta M(t_j) \rangle
\]

where \( t_{q_k} = t \), \( \Delta X(t_j) = X(t_{j+1}) - X(t_j) \) and \( e(k) \to 0 \) in probability. By Lemma 26.5.1, the stochastic integral on the right converges uniformly for \( t \in [0,T] \) to

\[
2 \int_0^t \langle BX, dM \rangle
\]

for \( \omega \) off a set of measure zero. The deterministic integral on the right converges uniformly for \( t \in [0,T] \) to

\[
2 \int_0^t \langle Y(u), X(u) \rangle \, du
\]

Thanks to Lemma 26.5.4,

\[
\left| \int_0^t \langle Y(u), X(u) \rangle \, du - \int_0^t \langle Y(u), X_k(u) \rangle \, du \right|
\]

\[
\leq \int_0^T \|Y(u)\|_{L'} \|X(u) - X_k(u)\|_{V'}
\]

\[
\leq \|Y\|_{L'([0,T])} (2^{-k})^{1/p}
\]

for all \( k \) large enough. Consider the fourth term. It equals

\[
\sum_{j=0}^{q_k-1} \left( R^{-1} B(M(t_{j+1}) - M(t_j)), M(t_{j+1}) - M(t_j) \right)_W
\]

where \( R^{-1} \) is the Riesz map from \( W \) to \( W' \). This equals

\[
\frac{1}{4} \left( \sum_{j=0}^{q_k-1} \| R^{-1} BM(t_{j+1}) + M(t_{j+1}) - (R^{-1} BM(t_j) + M(t_j)) \|^2
\]

\[- \sum_{j=0}^{q_k-1} \| R^{-1} BM(t_{j+1}) - M(t_{j+1}) - (R^{-1} BM(t_j) - M(t_j)) \|^2 \right)
\]

From Theorem 26.6.24, as \( k \to \infty \), the above converges in probability to \( (t_{q_k} = t) \)

\[
\frac{1}{4} \left( [R^{-1} BM + M](t) - [R^{-1} BM - M](t) \right) \equiv [R^{-1} BM, M](t)
\]

Also note that from 26.6.22, this term must be nonnegative since it is a limit of nonnegative quantities. This is what was desired.

Thus all the terms in (26.6.24) converge in probability except for the last term which also must converge in probability because it equals the sum of terms which do. It remains to find what this last term converges to. Thus

\[
\langle BX(t), X(t) \rangle - \langle BX_0, X_0 \rangle = 2 \int_0^t \langle Y(u), X(u) \rangle \, du
\]

\[
+ 2 \int_0^t \langle BX, dM \rangle + [R^{-1} BM, M](t) - a
\]
where $a$ is the limit in probability of the term

$$\sum_{j=1}^{q_k-1} \langle B (\Delta X (t_j) - \Delta M (t_j)), \Delta X (t_j) - \Delta M (t_j) \rangle$$

(26.6.23)

Let $P_n$ be the projection onto $\text{span} (e_1, \cdots, e_n)$ where $\{e_k\}$ is an orthonormal basis for $W$ with each $e_k \in V$. Then using

$$BX (t_{j+1}) - BX (t_j) - (BM (t_{j+1}) - BM (t_j)) = \int_{t_j}^{t_{j+1}} Y (s) \, ds$$

the troublesome term of (26.6.24) above is of the form

$$\sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \langle Y (s), \Delta X (t_j) - \Delta M (t_j) \rangle \, ds$$

$$= \sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \langle Y (s), \Delta X (t_j) - P_n \Delta M (t_j) \rangle \, ds$$

$$+ \sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \langle Y (s), -(I - P_n) \Delta M (t_j) \rangle \, ds$$

which equals

$$\sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \langle Y (s), X (t_{j+1}) - X (t_j) - P_n (M (t_{j+1}) - M (t_j)) \rangle \, ds$$

(26.6.24)

$$+ \sum_{j=1}^{q_k-1} \langle B (\Delta X (t_j) - \Delta M (t_j)) - (I - P_n) (M (t_{j+1}) - M (t_j)) \rangle$$

(26.6.25)

The reason for the $P_n$ is to get $P_n (M (t_{j+1}) - M (t_j))$ in $V$. The sum in (26.6.24) is dominated by

$$\left( \sum_{j=1}^{q_k-1} \langle B (\Delta X (t_j) - \Delta M (t_j)), \Delta X (t_j) - \Delta M (t_j) \rangle \right)^{1/2}$$

$$\left( \sum_{j=1}^{q_k-1} |\langle B (I - P_n) \Delta M (t_j), (I - P_n) \Delta M (t_j) \rangle| \right)^{1/2}$$

(26.6.26)

Now it is known from the above that

$$\sum_{j=1}^{q_k-1} \langle B (\Delta X (t_j) - \Delta M (t_j)), \Delta X (t_j) - \Delta M (t_j) \rangle$$

converges in probability to $a \geq 0$. If you take the expectation of the square of the other factor, it is no larger than

$$\|B\| E \left( \sum_{j=1}^{q_k-1} \| (I - P_n) \Delta M (t_j) \|_W^2 \right)$$

$$= \|B\| \left( \sum_{j=1}^{q_k-1} \| (I - P_n) (M (t_{j+1}) - M (t_j)) \|_W^2 \right)$$

$$= \|B\| \sum_{j=1}^{q_k-1} E \left( \| (I - P_n) (M (t_{j+1}) - M (t_j)) \|_W^2 \right)$$
Then
\[ \| ( I - P_n ) ( M ( t_{j+1} \wedge t) - M ( t_j \wedge t) ) \|^2_W = \| (1 - P_n) M^{t_{j+1}} - (1 - P_n) M^t \| (t) + N(t) \]
\[ = \| (1 - P_n) M^{t_{j+1}} (t) - [(1 - P_n) M^t (t)] + N(t) \]
for \( N(t) \) a martingale. In particular, taking \( t = t_q \), the above reduces to
\[ \| B \| \sum_{j=1}^{q_k-1} E \left\{ \| ( I - P_n ) ( M ( t_{j+1} ) - M ( t_j ) ) \|^2_W \right\} \]
\[ = \| B \| \sum_{j=1}^{q_k-1} E \left\{ \| (1 - P_n) M ( t_{j+1} ) - [(1 - P_n) M ( t_j )] \right\} \]
\[ = \| B \| E \left\{ \| (1 - P_n) M ( t_{q_k} ) \right\} = \| B \| E \left\{ \| (1 - P_n) M ( t_{q_k} ) \|^2_W \right\} \]
From maximal theorems, Theorem 15.3.35,
\[ \| B \| E \left( \sup_{t_{q_k}} \| (1 - P_n) M ( t_{q_k} ) \|^2_W \right) \leq 2 \| B \| E \left( \| (1 - P_n) M ( T ) \|^2_W \right) \]
and this on the right converges to zero as \( n \to \infty \) by assumption that \( M ( t ) \) is in \( L^2 \) and the dominated convergence theorem. In particular, this shows that
\[ \left( \sum_{j=1}^{q_k-1} \left\langle B ( I - P_n ) \Delta M ( t_j ), ( I - P_n ) \Delta M ( t_j ) \right\rangle \right)^{1/2} \]
converges to 0 in \( L^2 ( \Omega ) \) independent of \( k \) as \( n \to \infty \).

Thus the expression in 26.6.26 is of the form \( f_k g_{nk} \) where \( f_k \) converges in probability to \( a^{1/2} \) as \( k \to \infty \) and \( g_{nk} \) converges in probability to 0 as \( n \to \infty \) independent of \( k \). Now this implies \( f_k g_{nk} \) converges in probability to 0. Here is why.
\[ P \left( | f_k g_{nk} | > \varepsilon \right) \leq P \left( 2 \delta | f_k | > \varepsilon \right) + P \left( 2 C_\delta | g_{nk} | > \varepsilon \right) \]
\[ \leq P \left( 2 \delta | f_k - a^{1/2} | + 2 \delta a^{1/2} > \varepsilon \right) + P \left( 2 C_\delta | g_{nk} | > \varepsilon \right) \]
where \( \delta | f_k | + C_\delta | g_{nk} | > | f_k g_{nk} | \) and \( \lim_{\delta \to 0} C_\delta = \infty \). Pick \( \delta \) small enough that \( \varepsilon - 2 \delta a^{1/2} > \varepsilon / 2 \). Then this is dominated by
\[ \leq P \left( 2 \delta | f_k - a^{1/2} | > \varepsilon / 2 \right) + P \left( 2 C_\delta | g_{nk} | > \varepsilon \right) \]
Fix \( n \) large enough that the second term is less than \( \eta \) for all \( k \). Now taking \( k \) large enough, the above is less than \( \eta \). It follows the expression in 26.6.26 and consequently in 26.6.26 converges to 0 in probability.

Now consider the other term 26.6.26 using the \( n \) just determined. This term is of the form
\[ \sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \left\langle Y ( s ), X ( t_{j+1} ) - X ( t_j ) - P_n ( M ( t_{j+1} ) - M ( t_j ) ) \right\rangle ds = \]
\[ \sum_{j=1}^{q_k-1} \int_{t_j}^{t_{j+1}} \left\langle Y ( s ), X^\delta_k ( s ) - X_k^l ( s ) - P_n ( M^\delta_k ( s ) - M_k^l ( s ) ) \right\rangle ds \]
\[ = \int_{t_1}^{t} \left\langle Y ( s ), X^\delta_k ( s ) - X_k^l ( s ) - P_n ( M^\delta_k ( s ) - M_k^l ( s ) ) \right\rangle ds \]
where \( M_k^\delta \) denotes the step function
\[ M_k^\delta ( t ) = \sum_{i=0}^{m-1} M ( t_{i+1} ) X ( t_i, t_{i+1} ) ( t ) \]
and \( M_k^l \) is defined similarly. The term
\[
\int_{t_1}^{t} \langle Y(s), P_n (M_k^r(s) - M_k^l(s)) \rangle \, ds
\]
converges to 0 for a.e. \( \omega \) as \( k \to \infty \) thanks to continuity of \( t \to M(t) \). However, more is needed than this. Define the stopping time

\[
\tau_p = \inf \{ t > 0 : \| M(t) \|_W > p \}.
\]

Then \( \tau_p = \infty \) for all \( p \) large enough, this for a.e. \( \omega \). Let
\[
A_k = \left\lfloor \int_{t_1}^{t} \langle Y(s), P_n (M_k^r(s) - M_k^l(s)) \rangle \, ds \right\rfloor > \varepsilon
\]

\[
P(A_k) = \sum_{p=0}^{\infty} P(A_k \cap (\tau_p = \infty) \setminus [\tau_{p-1} < \infty])
\]

(26.6.27)

Now
\[
P(A_k \cap (\tau_p = \infty) \setminus [\tau_{p-1} < \infty]) \leq P(A_k \cap (\tau_p = \infty)) \leq P \left( \left\lfloor \int_{t_1}^{t} \langle Y(s), P_n (M^r_k(s) - M^l_k(s)) \rangle \, ds \right\rfloor \leq \varepsilon \right) \]

This is so because if \( \tau_p = \infty \), then it has no effect but also it could happen that the defining inequality may hold even if \( \tau_p < \infty \) hence the inequality. This is no larger than an expression of the form
\[
\frac{C_n}{\varepsilon} \int_\Omega \int_0^T \| Y(s) \|_{V'} \left\| (M^r_k(s) - M^l_k(s)) \right\|_W \, ds \, dP
\]

(26.6.28)

The inside integral converges to 0 by continuity of \( M \). Also, thanks to the stopping time, the inside integral is dominated by an expression of the form
\[
\int_0^T \| Y(s) \|_{V'}, 2pds
\]

and this is a function in \( L^1(\Omega) \) by assumption on \( Y \). It follows that the integral in (26.6.25) converges to 0 as \( k \to \infty \) by the dominated convergence theorem. Hence
\[
\lim_{k \to \infty} P(A_k \cap ([\tau_p = \infty])) = 0.
\]

Since the sets \( [\tau_p = \infty] \setminus [\tau_{p-1} < \infty] \) are disjoint, the sum of their probabilities is finite. Hence there is a dominating function in (26.6.26) and so, by the dominated convergence theorem applied to the sum,
\[
\lim_{k \to \infty} P(A_k) = \sum_{p=0}^{\infty} \lim_{k \to \infty} P(A_k \cap ([\tau_p = \infty] \setminus [\tau_{p-1} < \infty])) = 0
\]

Thus \( \int_{t_1}^{t} \langle Y(s), P_n (M_k^r(s) - M_k^l(s)) \rangle \, ds \) converges to 0 in probability as \( k \to \infty \).

Now consider
\[
\left| \int_{t_1}^{t} \langle Y(s), X_k^r(s) - X_k^l(s) \rangle \, ds \right| \leq \int_0^T |\langle Y(s), X_k^r(s) - X(s) \rangle| \, ds
\]

\[
+ \int_0^T |\langle Y(s), X_k^l(s) - X(s) \rangle| \, ds
\]

\[
\leq 2 \| Y(\cdot, \omega) \|_{L^p(0,T)} (2^{-k})^{1/p}
\]

for all \( k \) large enough, this by Lemma 40.3.1. Therefore,
\[
\sum_{j=1}^{n_k} \langle B(\Delta X(t_j) - \Delta M(t_j)), \Delta X(t_j) - \Delta M(t_j) \rangle
\]

converges to 0 in probability. This establishes the desired formula for \( t \in D \). 

In fact, the formula 20.6.20 is valid for all \( t \in N_{\omega}^C \).
26.6. THE ITO FORMULA

**Theorem 26.6.2** In Situation [26.6.1], for \( \omega \) off a set of measure zero, it follows that for every \( t \in N^C_\omega \),

\[
\langle BX (t), X (t) \rangle = \langle BX_0, X_0 \rangle + \int_0^t 2 \langle Y (s), X (s) \rangle \, ds
\]

\[
\left[ R^{-1} BM, M \right] (t) + 2 \int_0^t \langle BX, dM \rangle
\]

(26.6.29)

Also, there exists a unique continuous, progressively measurable function denoted as \( \langle BX, X \rangle \) such that it equals \( \langle BX (t), X (t) \rangle \) for a.e. \( t \) and \( \langle BX, X \rangle (t) \) equals the right side of the above for all \( t \). In addition to this,

\[
E (\langle BX, X \rangle (t)) = E (\langle BX_0, X_0 \rangle) + E \left( \int_0^t 2 \langle Y (s), X (s) \rangle \, ds + \left[ R^{-1} BM, M \right] (t) \right)
\]

(26.6.30)

Also the quadratic variation of the stochastic integral in \( \langle BX, X \rangle \) is dominated by

\[
\int_0^t \|BX\|_V, d [M]
\]

(26.6.31)

Also \( t \to BX (t) \) is continuous with values in \( W \) for \( t \in N^C_\omega \).

**Proof:** Let \( t \in N^C_\omega \setminus D \). For \( t > 0 \), let \( t (k) \) denote the largest point of \( T_k \) which is less than \( t \). Suppose \( t (m) < t (k) \). Hence \( m \leq k \). Then

\[
BX (t (m)) = BX_0 + \int_0^{t (m)} Y (s) \, ds + BM (t (m)),
\]

a similar formula holding for \( X (t (k)) \). Thus for \( t > t (m), t \in N^C_\omega \),

\[
B (X (t) - X (t (m))) = \int_{t (m)}^t Y (s) \, ds + B (M (t) - M (t (m)))
\]

which is the same sort of thing studied so far except that it starts at \( t (m) \) rather than at 0 and \( BX_0 = 0 \). Therefore, from Lemma [26.6.1] it follows

\[
\langle B (X (t (k)) - X (t (m))), X (t (k)) - X (t (m)) \rangle
\]

\[= \int_{t (m)}^{t (k)} 2 \langle Y (s), X (s) - X (t (m)) \rangle \, ds
\]

\[+ \left[ R^{-1} BM, M \right] (t (k)) - \left[ R^{-1} BM, M \right] (t (m))
\]

\[+ 2 \int_{t (m)}^{t (k)} \langle BX - X (t (m)), dM \rangle
\]

(26.6.32)

Consider that last term. It equals

\[
2 \int_{t (m)}^{t (k)} \langle B (X - X_m^l), dM \rangle
\]

(26.6.33)

This is dominated by

\[
2 \int_0^{t (k)} \langle B (X - X_m^l), dM \rangle - \int_0^{t (m)} \langle B (X - X_m^l), dM \rangle
\]

\[
\leq 4 \sup_{t \in [0, T]} \left| \int_0^t \langle B (X - X_m^l), dM \rangle \right|
\]

In Lemma [26.6.3] the above expression converges to 0. It follows there is a set of measure 0 including the earlier one such that for \( \omega \) not in that set, [26.6.3] converges to 0 in \( \mathbb{R} \). Similar reasoning shows the first term on the right in the non stochastic integral of [26.6.3] is dominated by an expression of the form

\[
4 \int_0^T \left| \langle Y (s), X (s) - X_m^l (s) \rangle \right| \, ds
\]
which clearly converges to 0 thanks to Lemma 4.5.3. Finally, it is obvious that
\[
\lim_{m,k \to \infty} \left[ R^{-1} BM, M \right] (t(k)) - \left[ R^{-1} BM, M \right] (t(m)) = 0 \text{ for a.e. } \omega
\]
due to the continuity of the quadratic variation.

This shows that for \( \omega \) off a set of measure 0
\[
\lim_{m,k \to \infty} \langle B(X(t(k)) - X(t(m))), X(t(k)) - X(t(m)) \rangle = 0
\]
Then for \( x \in W \),
\[
|\langle B(X(t(k)) - X(t(m)), x)\rangle| \\
\leq \langle B(X(t(k)) - X(t(m)), X(t(k)) - X(t(m)) \rangle^{1/2} \langle Bx, x \rangle^{1/2} \\
\leq \langle B(X(t(k)) - X(t(m)), X(t(k)) - X(t(m)) \rangle^{1/2} \|B\|^{1/2} \|x\|_W
\]
and so
\[
\lim_{m,k \to \infty} \|B(X(t(k)) - B(X(t(m))\|_{W'} = 0
\]
Recall \( t \) was arbitrary and \( \{t(k)\} \) is a sequence converging to \( t \). Then the above has shown that \( \{B(X(t(k))\}_{k=1}^\infty \) is a convergent sequence in \( W' \). Does it converge to \( B(X(t)) \)? Let \( \xi(t) \in W' \) be what it converges to. Letting \( v \in V \) then, since the integral equation shows that \( t \to B(X(t) \) is continuous into \( V' \),
\[
\langle \xi(t), v \rangle = \lim_{k \to \infty} \langle B(X(t(k)) , v \rangle = \langle B(X(t), v \rangle
\]
and now, since \( V \) is dense in \( W \), this implies \( \xi(t) = B(X(t) = B(X(t)) \) since \( t \notin N_\omega \). Recall also that it was shown earlier that \( B(X \) is weakly continuous into \( W' \) on \( [0, T] \) hence the strong convergence of \( \{B(X(t(k))\}_{k=1}^\infty \) in \( W' \) implies that it converges to \( B(X(t), \) this for any \( t \in N_\omega^C \).

For every \( t \in D \) and for \( \omega \) off the exceptional set of measure zero described earlier,
\[
\langle B(X(t)), X(t) \rangle = \langle BX_0, X_0 \rangle + \int_0^t 2 \langle Y(s), X(s) \rangle \, ds + \\
\left[R^{-1} BM, M\right](t) + 2 \int_0^t \langle BX, dM \rangle
\]
(26.6.34)
Does this formula hold for all \( t \in [0, T]? \) Maybe not. However, it will hold for \( t \notin N_\omega \). So let \( t \notin N_\omega \).
\[
|\langle BX(t(k)), X(t(k)) \rangle - \langle BX(t), X(t) \rangle| \\
\leq |\langle BX(t(k)), X(t(k)) \rangle - \langle BX(t), X(t(k)) \rangle| \\
+ |\langle BX(t), X(t(k)) \rangle - \langle BX(t), X(t) \rangle| \\
= |\langle B(X(t(k)) - X(t), X(t(k)) \rangle| + |\langle B(X(t(k)) - X(t), X(t) \rangle|
\]
Then using the Cauchy Schwarz inequality on each term,
\[
\leq \langle B(X(t(k)) - X(t), X(t(k)) - X(t) \rangle^{1/2} \\
\cdot \left( \langle BX(t(k)), X(t(k)) \rangle^{1/2} + \langle BX(t), X(t) \rangle^{1/2} \right)
\]
As before, one can use the lower semicontinuity of
\[
t \to \langle B(X(t(k)) - X(t), X(t(k)) - X(t) \rangle
\]
on \( N_\omega^C \) along with the boundedness of \( \langle BX(t), X(t) \rangle \) also shown earlier off \( N_\omega \) to conclude
\[
|\langle BX(t(k)), X(t(k)) \rangle - \langle BX(t), X(t) \rangle| \\
\leq C \langle B(X(t(k)) - X(t), X(t(k)) - X(t) \rangle^{1/2}
\]
\[ \lim_{k \to \infty} \langle B (X (t (k)), X (t (k))) = \langle B (X (t), X (t)) \rangle. \]

It follows that the formula \[26.6.34\] is valid for all \( t \in \mathcal{N}_\omega \). Now define the function \( \langle B (X, X) (t) \rangle \) as
\[
\langle B (X, X) (t) \rangle = \left\{ \begin{array}{ll}
\langle B (X (t), X (t)), t \notin \mathcal{N}_\omega \\
\text{The right side of \[26.6.34\] if } t \in \mathcal{N}_\omega
\end{array} \right.
\]

Then in short, \( \langle B (X, X) (t) \rangle \) equals the right side of \[26.6.34\] for all \( t \in [0, T] \) and is consequently progressively measurable and continuous. Furthermore, for a.e. \( t \), this function equals \( \langle B (X (t)), X (t) \rangle \). Since it is known on a dense subset, it must be unique.

This implies that \( t \to B (X) (t) \) is continuous with values in \( W' \) for \( t \in \mathcal{N}_\omega \). Here is why. The fact that the formula \[26.6.34\] holds for all \( t \in \mathcal{N}_\omega \) implies that \( t \to \langle B (X (t), X (t)) \rangle \) is continuous on \( \mathcal{N}_\omega \). Then for \( x \in W, t, s \in \mathcal{N}_\omega \)
\[
|\langle B (X (t) - B (X) (s), x) \rangle| \leq \langle B (X (t) - X (s)), X (t) - X (s) \rangle^{1/2} \parallel B \parallel^{1/2} \parallel x \parallel_W. \tag{26.6.35}
\]

Also
\[
\langle B (X (t) - X (s)), X (t) - X (s) \rangle = \langle B (X (t), X (t)) + \langle B (X (s), X (s)) - 2 \langle B (X (t), X (s)) \rangle
\]

By weak continuity of \( t \to B (X) (t) \) shown earlier,
\[
\lim_{t \to s} \langle B (X (t), X (s)) = \langle B (X (s), X (s)) \rangle.
\]

Therefore,
\[
\lim_{t \to s} \langle B (X (t) - X (s)), X (t) - X (s) \rangle = 0
\]

and so the inequality \[26.6.34\] implies the continuity of \( t \to B (X) (t) \) into \( W' \) for \( t \notin \mathcal{N}_\omega \). Note that by assumption this function is continuous into \( V' \) for all \( t \).

Now consider the claim about the expectation. Use the function \( \langle B (X, X) \rangle \) to define a stopping time as
\[
\tau_p \equiv \inf \{ t > 0 : \langle B (X, X) (t) > p \}
\]

This is the first hitting time of a continuous process and so it is a valid stopping time. Using this, leads to
\[
\langle B (X, X)^{\tau_p} (t) = \langle B X_0, X_0 \rangle + \int_0^t \chi_{[0, \tau_p]} 2 \langle Y (s), X (s) \rangle \, ds +
\]
\[
[R^{-1} BM, M]^{\tau_p} (t) + 2 \int_0^t \chi_{[0, \tau_p]} \langle B X, dM \rangle \tag{26.6.36}
\]

The term at the end is now a martingale because \( \chi_{[0, \tau_p]} \) \( B X \) is bounded. Hence the expectation of the martingale at the end equals 0. Thus you obtain
\[
E (\langle B (X, X)^{\tau_p} (t) = E (\langle B X_0, X_0 \rangle)
\]
\[
+ E \left( \int_0^t \chi_{[0, \tau_p]} 2 \langle Y (s), X (s) \rangle \, ds \right) + E \left( [R^{-1} BM, M]^{\tau_p} (t) \right)
\]

Now use the monotone convergence theorem and the dominated convergence theorem to pass to a limit as \( p \to \infty \) and obtain \[26.6.36\]. The claim about the quadratic variation follows from Theorem \[26.6.29\].

What of the special case where \( W = H = H' \) and you are in the context of a Gelfand triple
\[ V \subseteq H \subseteq H' \subseteq V' \]

and \( B \) is simply the identity. Then we obtain the following theorem as a special case.
Theorem 26.6.3 In Situation 26.1.1 in which \( W = H = H' \) and \( B = I \), it follows that off a set of measure zero, for every \( t \in [0, T] \), there is a set of measure zero \( N \) such that for \( \omega \notin N \), there is a continuous function \( \langle X, X \rangle \) which equals \( |X(t)|_H^2 \) for a.e. \( t \) such that

\[
\langle X, X \rangle (t) = |X_0|_H^2 + \int_0^t 2 \langle Y(s), X(s) \rangle \, ds
+ |M|(t) + 2 \int_0^t \langle X, dM \rangle
\] (26.6.37)

Furthermore, off a set of measure zero, \( t \to X(t) \) is continuous as a map into \( H \) for a.e. \( \omega \). In addition to this,

\[
E(|X(t)|^2) = E(|X_0|^2) + E\left(\int_0^t 2 \langle Y(s), X(s) \rangle \, ds\right) + E(|M|(t))
\] (26.6.38)

The quadratic variation of the stochastic integral satisfies

\[
\left[ \int_0^t (X, dM) \right](t) \leq \int_0^t \|X\|_H^2 d[M]
\]

It is more attractive to write \( |X(t)|_H^2 \) in place of \( \langle X, X \rangle (t) \). However, I guess this is not strictly right although the discrepancy is only on a set of measure zero so it seems fairly harmless to indulge in this sloppiness. However, for \( t \notin N_\omega \),

\[
|X(t)|_H^2 = \sum_i \langle X(t), e_i \rangle^2
\]

where the orthonormal basis \( \{e_i\} \) is in \( V \). Then for \( s \in N_\omega \), you can get the following. Let \( t_n \to s \) where \( t_n \in N_\omega \). Then in the above notation,

\[
\sum_i \langle X(s), e_i \rangle^2 \leq \lim_{n \to \infty} \sum_i \langle X(t_n), e_i \rangle_H^2 = \lim_{n \to \infty} |X(t_n)|_H^2 \leq C(\omega)
\]

It follows that in fact \( X(s) \in H \) and you can take \( X(s) = \sum_i \langle X(s), e_i \rangle e_i \in H \) because \( \sum_i \langle X(s), e_i \rangle^2 < \infty \). Hence

\[
|X(s)|^2 = \sum_i \langle X(s), e_i \rangle^2 \leq \lim_{n \to \infty} |X(t_n)|_H^2
\]

so \( X \) has values in \( H \) and is lower semicontinuous on \([0, T] \).
Chapter 27

Nonlinear Operators

In this chapter is a description and properties of some standard nonlinear maps.

27.1 An Assortment Of Nonlinear Operators

**Definition 27.1.1** For $V$ a real Banach space, $A : V \to V'$ is a pseudomonotone map if whenever

$$ u_n \rightharpoonup u $$

and

$$ \limsup_{n \to \infty} \langle Au_n, u_n - u \rangle \leq 0 $$

it follows that for all $v \in V$,

$$ \liminf_{n \to \infty} \langle Au_n, u_n - v \rangle \geq \langle Au, u - v \rangle. $$

The half arrows denote weak convergence.

**Definition 27.1.2** $A : V \to V'$ is monotone if for all $v, u \in V$,

$$ \langle A(u - Av), u - v \rangle \geq 0, $$

and $A$ is Hemicontinuous if for all $v, u \in V$,

$$ \lim_{t \to 0^+} \langle A(u + t(v - u)), u - v \rangle = \langle Au, u - v \rangle. $$

**Theorem 27.1.3** Let $V$ be a Banach space and let $A : V \to V'$ be monotone and hemicontinuous. Then $A$ is pseudomonotone.

**Proof:** Let $A$ be monotone and Hemicontinuous. First here is a claim.

**Claim:** If (27.1.1) and (27.1.2) hold, then $\lim_{n \to \infty} \langle Au_n, u_n - u \rangle = 0$.

**Proof of the claim:** Since $A$ is monotone,

$$ \langle Au_n - Au, u_n - u \rangle \geq 0 $$

so

$$ \langle Au_n, u_n - u \rangle \geq \langle Au, u_n - u \rangle. $$

Therefore,

$$ 0 = \lim_{n \to \infty} \inf \langle Au, u_n - u \rangle \leq \lim_{n \to \infty} \inf \langle Au_n, u_n - u \rangle \leq \limsup_{n \to \infty} \langle Au_n, u_n - u \rangle \leq 0. $$

Now using that $A$ is monotone again, then letting $t > 0$,

$$ \langle Au_n - A(u + t(v - u)), u_n - u + t(u - v) \rangle \geq 0 $$

and so

$$ \langle Au_n, u_n - u + t(u - v) \rangle \geq \langle A(u + t(v - u)), u_n - u + t(u - v) \rangle. $$
Taking the lim inf on both sides and using the claim and \( t > 0 \),
\[
\liminf_{n \to \infty} \langle Au_n, u - v \rangle \geq t \langle A(u + t(v - u)), (u - v) \rangle.
\]

Next divide by \( t \) and use the Hemicontinuity of \( A \) to conclude that
\[
\liminf_{n \to \infty} \langle Au_n, u - v \rangle \geq \langle Au, u - v \rangle.
\]

From the claim,
\[
\liminf_{n \to \infty} \langle Au_n, u - v \rangle = \liminf_{n \to \infty} (\langle Au_n, u_n - v \rangle + \langle Au_n, u - u_n \rangle)
\]
\[
= \liminf_{n \to \infty} \langle Au_n, u - v \rangle \geq \langle Au, u - v \rangle. \quad [\text{\#}]
\]

Monotonicity is very important in the above proof. The next example shows that even if the operator is linear and bounded, it is not necessarily pseudomonotone.

**Example 27.1.4** Let \( H \) be any Hilbert space and let \( A : H \to H' \) be given by
\[
\langle Ax, y \rangle \equiv (-x, y)_H.
\]
Then \( A \) fails to be pseudomonotone.

**Proof:** Let \( \{x_n\}_{n=1}^{\infty} \) be an orthonormal set of vectors in \( H \). Then Parsevall’s inequality implies
\[
||x||^2 \geq \sum_{n=1}^{\infty} |(x_n, x)|^2
\]
and so for any \( x \in H, \lim_{n \to \infty} (x_n, x) = 0 \). Thus \( x_n \to 0 \equiv x \). Also
\[
\limsup_{n \to \infty} \langle Ax_n, x_n - x \rangle =
\]
\[
\limsup_{n \to \infty} \langle Ax_n, x_n - 0 \rangle = \limsup_{n \to \infty} (-||x_n||^2) = -1 \leq 0.
\]
If \( A \) were pseudomonotone, we would need to be able to conclude that for all \( y \in H, \)
\[
\liminf_{n \to \infty} \langle Ax_n, x_n - y \rangle \geq \langle Ax, x - y \rangle = 0.
\]
However,
\[
\liminf_{n \to \infty} \langle Ax_n, x_n - 0 \rangle = -1 < 0 = \langle A0, 0 - 0 \rangle.
\]
Now the following proposition is useful.

**Proposition 27.1.5** Suppose \( A : V \to V' \) is pseudomonotone and bounded where \( V \) is separable. Then it must be demicontinuous. This means that if \( u_n \to u \), then \( Au_n \to Au \).

**Proof:** Since \( u_n \to u \) is strong convergence and since \( Au_n \) is bounded, it follows
\[
\limsup_{n \to \infty} \langle Au_n, u_n - u \rangle = \lim_{n \to \infty} \langle Au_n, u_n - u \rangle = 0.
\]
Suppose this is not so that \( Au_n \) converges weakly to \( Au \). Since \( A \) is bounded, there exists a subsequence, still denoted by \( n \) such that \( Au_n \to \xi \) weak *. I need to verify \( \xi = Au \). From the above, it follows that for all \( v \in V \)
\[
\langle Au, u - v \rangle \leq \liminf_{n \to \infty} \langle Au_n, u - v \rangle
\]
\[
= \liminf_{n \to \infty} \langle Au_n, u - v \rangle = \langle \xi, u - v \rangle
\]
Hence \( \xi = Au \). [\text{\#}]

There is another type of operator which is more general than pseudomonotone.
Definition 27.1.6 Let $A : V \to V'$ be an operator. Then $A$ is called type M if whenever $u_n \rightharpoonup u$ and $Au_n \rightharpoonup \xi$, and

$$\limsup_{n \to \infty} \langle Au_n, u_n \rangle \leq \langle \xi, u \rangle$$

it follows that $Au = \xi$.

Proposition 27.1.7 If $A$ is pseudomonotone, then $A$ is type M.

Proof: Suppose $A$ is pseudomonotone and $u_n \rightharpoonup u$ and $Au_n \rightharpoonup \xi$, and

$$\limsup_{n \to \infty} \langle Au_n, u_n \rangle \leq \langle \xi, u \rangle$$

Then

$$\limsup_{n \to \infty} \langle Au_n, u_n - u \rangle = \limsup_{n \to \infty} \langle Au_n, u_n \rangle - \langle \xi, u \rangle \leq 0$$

Hence

$$\liminf_{n \to \infty} \langle Au_n, u_n - v \rangle \geq \langle Au, u - v \rangle$$

for all $v \in V$. Consequently, for all $v \in V$,

$$\langle Au, u - v \rangle = \liminf_{n \to \infty} \langle Au_n, u_n - v \rangle
= \liminf_{n \to \infty} (\langle Au_n, u - v \rangle + \langle Au_n, u_n - u \rangle)
= \langle \xi, u - v \rangle + \liminf_{n \to \infty} \langle Au_n, u_n - u \rangle \leq \langle \xi, u - v \rangle$$

and so $Au = \xi$. ■

An interesting result is the following which states that a monotone linear function added to a type M is also type M.

Proposition 27.1.8 Suppose $A : V \to V'$ is type M and suppose $L : V \to V'$ is monotone, bounded and linear. Then $L + A$ is type M. Let $V$ be separable or reflexive so that the weak convergences in the following argument are valid.

Proof: Suppose $u_n \rightharpoonup u$ and $Au_n + Lu_n \rightharpoonup \xi$ and also that

$$\limsup_{n \to \infty} \langle Au_n + Lu_n, u_n \rangle \leq \langle \xi, u \rangle$$

Does it follow that $\xi = Au + Lu$? Suppose not. There exists a further subsequence, still called $n$ such that $Lu_n \rightharpoonup Lu$. This follows because $L$ is linear and bounded. Then from monotonicity,

$$\langle Lu_n, u \rangle \geq \langle Lu_n, u \rangle + \langle L(u), u_n - u \rangle$$

Hence with this further subsequence, the limsup is no larger and so

$$\limsup_{n \to \infty} \langle Au_n, u_n \rangle + \limsup_{n \to \infty} (\langle Lu_n, u \rangle + \langle L(u), u_n - u \rangle) \leq \langle \xi, u \rangle$$

and so

$$\limsup_{n \to \infty} \langle Au_n, u_n \rangle \leq \langle \xi - Lu, u \rangle$$

It follows since $A$ is type M that $Au = \xi - Lu$, which contradicts the assumption that $\xi \neq Au + Lu$. ■

There is also the following useful generalization of the above proposition.

Corollary 27.1.9 Suppose $A : V \to V'$ is type M and suppose $L : V \to V'$ is monotone, bounded and linear. Then for $u_0 \in V$ define $M(u) \equiv L(u - u_0)$. Then $M + A$ is type M. Let $V$ be separable or reflexive so that the weak convergences in the following argument are valid.
CHAPTER 27. NONLINEAR OPERATORS

Proof: Suppose $u_n \to u$ and $Au_n + Mu_n \to \xi$ and also that

$$ \limsup_{n \to \infty} \langle Au_n + Mu_n, u_n \rangle \leq \langle \xi, u \rangle $$

Does it follow that $\xi = Au + Mu$? Suppose not. By assumption, $u_n - u_0 \to u - u_0$ and so, since $L$ is bounded, there is a further subsequence, still called $n$ such that

$$ Mu_n = L(u_n - u_0) \to L(u - u_0) = Mu. $$

Since $M$ is monotone,

$$ \langle Mu_n - Mu, u_n - u \rangle \geq 0 $$

Thus

$$ \langle Mu_n, u_n \rangle - \langle Mu_n, u \rangle - \langle Mu, u_n \rangle \geq 0 $$

and so

$$ \langle Mu_n, u_n \rangle \geq \langle Mu_n, u \rangle + \langle Mu, u_n - u \rangle $$

Hence with this further subsequence, the lim sup is no larger and so

$$ \limsup_{n \to \infty} \langle Au_n, u_n \rangle \leq \langle \xi - Mu, u \rangle $$

It follows since $A$ is type $M$ that $Au = \xi - Mu$, which contradicts the assumption that $\xi \neq Au + Mu$. □

The following is Browder's lemma. It is a very interesting application of the Brouwer fixed point theorem.

Lemma 27.1.10 (Browder) Let $K$ be a convex closed and bounded set in $\mathbb{R}^n$ and let $A : K \to \mathbb{R}^n$ be continuous and $f \in \mathbb{R}^n$. Then there exists $x \in K$ such that for all $y \in K$,

$$ (f - Ax, y - x) \leq 0 $$

Proof: Let $P_K$ denote the projection onto $K$. Thus $P_K$ is Lipschitz continuous.

$$ x \to P_K(f - Ax + x) $$

is a continuous map from $K$ to $K$. By the Brouwer fixed point theorem, it has a fixed point $x \in K$. Therefore, for all $y \in K$,

$$ (f - Ax + x - x, y - x) = (f - Ax, y - x) \leq 0 $$

From this lemma, there is an interesting theorem on surjectivity.

Proposition 27.1.11 Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be continuous and coercive,

$$ \lim_{|x| \to \infty} \frac{A(x + x_0), x}{|x|} = \infty $$

for some $x_0$. Then for all $f \in \mathbb{R}^n$, there exists $x \in \mathbb{R}^n$ such that $Ax = f$.

Proof: Define the closed convex sets $B_n = B(x_0, n)$. By Browder’s lemma, there exists $x_n$ such that

$$ (f - Ax_n, y - x_n) \leq 0 $$

for all $y \in B_n$. Then taking $y = x_0$, it follows from the coercivity condition that the $x_n - x_0$ are bounded. It follows that for large $n$, $x_n$ is an interior point of $B_n$. Therefore,

$$ (f - Ax_n, z) \leq 0 $$

for all $z$ in some open ball centered at $x_0$. Hence $f = Ax_n$. □

Lemma 27.1.12 Let $A : V \to V'$ be type $M$ and bounded and suppose $V$ is reflexive or $V$ is separable. Then $A$ is demicontinuous.
27.1. AN ASSORTMENT OF NONLINEAR OPERATORS

**Proof:** Suppose \( u_n \to u \) and \( Au_n \) fails to converge weakly to \( Au \). Then there is a further subsequence, still denoted as \( u_n \) such that \( Au_n \to \zeta \neq Au \). Then thanks to the strong convergence, you have

\[
\limsup_{n \to \infty} \langle Au_n, u_n \rangle = \langle \zeta, u_n \rangle
\]

which implies \( \zeta = Au \) after all. \( \blacksquare \)

With these lemmas and the above proposition, there is a very interesting surjectivity result.

**Theorem 27.1.13** Let \( A : V \to V' \) be type M, bounded, and coercive

\[
\lim_{\|u\| \to \infty} \frac{\langle A(u + u_0), u \rangle}{\|u\|} = \infty,
\]

(27.1.4)

for some \( u_0 \), where \( V \) is a separable reflexive Banach space. Then \( A \) is surjective.

**Proof:** Since \( V \) is separable, there exists an increasing sequence of finite dimensional subspaces \( \{V_n\} \) such that \( \bigcup_n V_n = V \). Say \( \text{span} (v_1, \cdots, v_n) = V_n \). Then consider the following diagram.

\[
\begin{array}{cccc}
\mathbb{R}^n & \overset{q}{\longrightarrow} & V' & \overset{\iota}{\longrightarrow} & V' \\
\mathbb{R}^n & \overset{q}{\longrightarrow} & V_n & \overset{i}{\longrightarrow} & V
\end{array}
\]

Here the map \( \theta \) is the one which does the following.

\[
\theta (x) = \sum_{i=1}^{n} x_i v_i.
\]

The map \( i \) is the inclusion map. Consider the map \( \theta^* i^* A i \theta \). By Lemma 27.1.12 this map is continuous. The map \( \theta \) is continuous, one to one, and onto. Thus its inverse is also continuous. Let \( x_0 \) correspond to \( u_0 \). Then for some constant \( C \),

\[
\frac{(\theta^* i^* A i \theta (x + x_0), x)}{|x|} \geq \frac{\langle A i \theta (x + x_0), i \theta x \rangle}{C \|i \theta x\|_V}
\]

and to say \( |x| \to \infty \) is the same as saying that \( \|i \theta x\|_V \to \infty \). Hence \( \theta^* i^* A i \theta \) is coercive. Let \( f \in V' \). Then from 27.1.11 there exists \( x_n \) such that

\[
\theta^* i^* A i \theta x_n = \theta^* i^* f
\]

Thus, \( i^* A i \theta x_n = i^* f \) and this implies that for \( v_n \equiv \theta x_n \),

\[
i^* Av_n = i^* f
\]

In other words,

\[
\langle Av_n, y \rangle = \langle f, y \rangle
\]

(27.1.5)

for all \( y \in V_n \). Then from the coercivity condition 27.1.4 the \( v_n \) are bounded independent of \( n \). Since \( V \) is reflexive, there is a subsequence, still called \( \{v_n\} \) which converges weakly to \( v \in V \). Since \( A \) is bounded, it can also be assumed that \( Av_n \to \zeta \in V' \). Then

\[
\limsup_{n \to \infty} \langle Av_n, v_n \rangle = \limsup_{n \to \infty} \langle f, v_n \rangle = \langle f, v \rangle
\]

Also, passing to the limit in 27.1.4,

\[
\langle \zeta, y \rangle = \langle f, y \rangle
\]

for any \( y \in V_n \), this for any \( n \). Since the union of these \( V_n \) is dense, it follows that the above equation holds for all \( y \in V \). Therefore, \( f = \zeta \) and so

\[
\limsup_{n \to \infty} \langle Av_n, v_n \rangle = \limsup_{n \to \infty} \langle f, v_n \rangle = \langle f, v \rangle = \langle \zeta, v \rangle
\]

Since \( A \) is type \( M \),

\[
Av = \zeta = f \quad \blacksquare
\]
27.2 Duality Maps

The duality map is an attempt to duplicate some of the features of the Riesz map in Hilbert space which is discussed in the chapter on Hilbert space.

**Definition 27.2.1** A Banach space is said to be strictly convex if whenever \( \|x\| = \|y\| \) and \( x \neq y \), then

\[
\left\| \frac{x + y}{2} \right\| < \|x\|.
\]

\( F : X \to X' \) is said to be a duality map if it satisfies the following: a.) \( \|F(x)\| = \|x\|^{p-1} \). b.) \( F(x) = \|x\|^p, \) where \( p > 1 \).

Duality maps exist. Here is why. Let

\[
F(x) \equiv \left\{ x^* : \|x^*\| \leq \|x\|^{p-1} \text{ and } x^*(x) = \|x\|^p \right\}
\]

Then \( F(x) \) is not empty because you can let \( f(\alpha x) = \alpha \|x\|^p \). Then \( f \) is linear and defined on a subspace of \( X \). Also

\[
\sup_{\|\alpha x\| \leq 1} |f(\alpha x)| = \sup_{\|\alpha x\| \leq 1} |\alpha| \|x\|^p \leq \|x\|^{p-1}
\]

Also from the definition,

\[
f(x) = \|x\|^p
\]

and so, letting \( x^* \) be a Hahn Banach extension, it follows \( x^* \in F(x) \). Also, \( F(x) \) is closed and convex. It is clearly closed because if \( x^*_n \to x^* \), the condition on the norm clearly holds and also the other one does too. It is convex because

\[
\|x^*\lambda + (1 - \lambda) y^*\| \leq \lambda \|x^*\| + (1 - \lambda) \|y^*\| \leq \lambda \|x\|^{p-1} + (1 - \lambda) \|y\|^{p-1}
\]

If the conditions hold for \( x^* \), then we can show that in fact \( \|x^*\| = \|x\|^{p-1} \). This is because

\[
\|x^*\| \geq x^* \left( \frac{x}{\|x\|} \right) = \frac{1}{\|x\|} |x^*(x)| = \|x\|^{p-1}.
\]

Now how many things are in \( F(x) \) assuming the norm on \( X' \) is strictly convex? Suppose \( x^*_1 \) and \( x^*_2 \) are two things in \( F(x) \). Then by convexity, so is \( (x^*_1 + x^*_2)/2 \). Hence by strict convexity, if the two are different, then

\[
\left\| \frac{x^*_1 + x^*_2}{2} \right\| = \|x\|^{p-1} < \frac{1}{2} \|x^*_1\| + \frac{1}{2} \|x^*_2\| = \|x\|^{p-1}
\]

which is a contradiction. Therefore, \( F \) is an actual mapping.

What are some of its properties? First is one which is similar to the Cauchy Schwarz inequality. Since \( p-1 = p/p' \),

\[
\sup_{\|y\| \leq 1} |\langle Fx, y \rangle| = \|x\|^{p/p'}
\]

and so for arbitrary \( y \neq 0 \),

\[
|\langle Fx, y \rangle| = \|y\| \left| \frac{\langle Fx, y \rangle}{\|y\|} \right| \leq \|y\| \|x\|^{p/p'}
\]

\[
= |\langle Fy, y \rangle|^{1/p} |\langle Fx, x \rangle|^{1/p'}
\]

Next we can show that \( F \) is monotone.

\[
\langle Fx - Fy, x - y \rangle = \langle Fx, x \rangle - \langle Fx, y \rangle - \langle Fy, x \rangle + \langle Fy, y \rangle \\
\geq \|x\|^p + \|y\|^p - \|y\| \|x\|^{p/p' - \|y\|^{p/p'} \|x\|}
\]

\[
\geq \|x\|^p + \|y\|^p - \left( \frac{\|y\|^p}{p} + \frac{\|x\|^p}{p'} \right) - \left( \frac{\|y\|^p}{p'} + \frac{\|x\|^p}{p} \right) = 0
\]
Next it can be shown that $F$ is hemicontinuous. By the construction, $F(x + ty)$ is bounded as $t \to 0$. Let $t \to 0$ be a subsequence such that

$$F(x + ty) \to \xi$$

Then we ask: Does $\xi$ do what it needs to do in order to be $F(x)$? The answer is yes. First of all $\|F(x + ty)\| = \|x + ty\|^{p-1} \to \|x\|^{p-1}$. The set

$$\left\{ x^* : \|x^*\| \leq \|x\|^{p-1} + \varepsilon \right\}$$

is closed and convex and so it is weak * closed as well. For all small enough $t$, it follows $F(x + ty)$ is in this set. Therefore, the weak limit is also in this set and it follows $\|\xi\| \leq \|x\|^{p-1} + \varepsilon$. Since $\varepsilon$ is arbitrary, it follows $\|\xi\| \leq \|x\|^{p-1}$. Is $\xi(x) = \|x\|^p$? We have

$$\|x\|^p = \lim_{t \to 0} \|x + ty\|^p = \lim_{t \to 0} \langle F(x + ty), x + ty \rangle$$

$$= \lim_{t \to 0} \langle F(x + ty), x \rangle = \langle \xi, x \rangle$$

and so, $\xi$ does what it needs to do to be $F(x)$. This would be clear if $\|\xi\| = \|x\|^{p-1}$. However, $\langle \xi, x \rangle = \|x\|^p$ and so $\|\xi\| \geq \langle \xi, \frac{x}{\|x\|^p} \rangle = \|x\|^{p-1}$. Thus $\|\xi\| = \|x\|^{p-1}$ which shows $\xi$ does everything it needs to do to equal $F(x)$ and so it is $F(x)$. Since this conclusion follows for any convergent sequence, it follows that $F(x + ty)$ converges to $F(x)$ weakly as $t \to 0$. This is what it means to be hemicontinuous. This proves the following theorem. One can show also that $F$ is demicontinuous which means strongly convergent sequences go to weakly convergent sequences. Here is a proof for the case where $p = 2$. You can clearly do the same thing for arbitrary $p$.

**Lemma 27.2.2** Let $F$ be a duality map for $p = 2$ where $X, X'$ are reflexive and have strictly convex norms. (If $X$ is reflexive, there is always an equivalent strictly convex norm.) Then $F$ is demicontinuous.

**Proof:** Say $x_n \to x$. Then does it follow that $Fx_n \rightarrow Fx$? Suppose not. Then there is a subsequence, still denoted as $x_n$ such that $x_n \to x$ but $Fx_n \not= Fx$ where here $\rightharpoonup$ denotes weak convergence. This follows from the Eberlein Smulian theorem. Then

$$\langle y, x \rangle = \lim_{n \to \infty} \langle Fx_n, x_n \rangle = \lim_{n \to \infty} \|x_n\|^2 = \|x\|^2$$

Also, there exists $z, \|z\| = 1$ and $\langle y, z \rangle \geq \|y\| - \varepsilon$. Then

$$\|y\| - \varepsilon = \langle y, z \rangle = \lim_{n \to \infty} \langle Fx_n, z \rangle \leq \lim \inf_{n \to \infty} \|Fx_n\| = \lim \inf_{n \to \infty} \|x_n\| = \|x\|$$

and since $\varepsilon$ is arbitrary, $\|y\| \leq \|x\|$. It follows from the above construction of $Fx$, that $y = Fx$ after all, a contradiction.

**Theorem 27.2.3** Let $X$ be a reflexive Banach space with $X'$ having strictly convex norm. Then for $p > 1$, there exists a mapping $F : X \to X'$ which is bounded, monotone, hemicontinuous, coercive in the sense that $\lim_{|x| \to \infty} \langle Fx, x \rangle / |x| = \infty$, which also satisfies the inequalities

$$|\langle Fx, y \rangle| \leq |\langle Fx, x \rangle|^{1/p'} |\langle Fy, y \rangle|^{1/p}$$

Note that these conclusions about duality maps show that they map onto the dual space.

The duality map was onto and it was monotone. This was shown above. Consider the form of a duality map for the $L^p$ spaces. Let $F : L^p \to (L^p)'$ be the one which satisfies

$$\|Ff\| = ||f||^{p-1}, \langle Ff, f \rangle = ||f||^p$$

Then in this case,

$$Ff = |f|^{p-2} f$$

This is because it does what it needs to do.

$$||Ff||_{L^{p'}} = \left( \int_{\Omega} \left( |f|^{p-1} \right)^{p'} d\mu \right)^{1/p'} = \left( \int_{\Omega} \left( |f|^{p/p'} \right)^{p'} d\mu \right)^{1/p'}$$

$$= \left( \int_{\Omega} |f|^p d\mu \right)^{1 - 1/p} = \left( \int_{\Omega} |f|^p d\mu \right)^{1/p} = ||f||_{L^p}^{p-1}$$

\(^1\text{It is known that if the space is reflexive, then there is an equivalent norm which is strictly convex. However, in most examples, this strict convexity is obvious.}\)
while it is obvious that
\[
(F f, f) = \int_\Omega |f|^p \, d\mu = |f|_{L^p(\Omega)}^p.
\]

Now here is an interesting inequality which I will only consider in the case where the quantities are real valued.

**Lemma 27.2.4** Let \( p \geq 2 \). Then for \( a, b \) real numbers,
\[
\left(|a|^{p-2} a - |b|^{p-2} b\right) (a - b) \geq C |a - b|^p
\]
for some constant \( C \) independent of \( a, b \).

**Proof:** There is nothing to show if \( a = b \). Without loss of generality, assume \( a > b \). Also assume \( p > 2 \). There is nothing to show if \( p = 2 \). I want to show that there exists a constant \( C \) such that for \( a > b \),
\[
\frac{|a|^{p-2} a - |b|^{p-2} b}{|a - b|^{p-1}} \geq C.
\]

(27.2.6)

First assume also that \( b \geq 0 \). Now it is clear that as \( a \to \infty \), the quotient above converges to \( 1 \). Take the derivative of this quotient. This yields
\[
(p - 1) |a - b|^{p-2} \frac{|a|^{p-2} a - |b|^{p-2} b}{|a - b|^{2p-2}}.
\]

Now remember \( a > b \). Then the above reduces to
\[
(p - 1) |a - b|^{p-2} \frac{|b|^{p-2} - |a|^{p-2}}{|a - b|^{2p-2}}.
\]

Since \( b \geq 0 \), this is negative and so \( 1 \) would be a lower bound. Now suppose \( b < 0 \). Then the above derivative is negative for \( b < a \leq -b \) and then it is positive for \( a > -b \). It equals \( 0 \) when \( a = -b \). Therefore the quotient in (27.2.6) achieves its minimum value when \( a = -b \). This value is
\[
\frac{|b|^{p-2} (-b) - |b|^{p-2} b}{|b - b|^{p-1}} = |b|^{p-2} \frac{-2b}{2b|b|} = |b|^{p-2} \frac{1}{2|b|^{p-2}} = \frac{1}{2|b|^{p-2}}.
\]

Therefore, the conclusion holds whenever \( p \geq 2 \). That is
\[
\left(|a|^{p-2} a - |b|^{p-2} b\right) (a - b) \geq \frac{1}{2|b|^{p-2}} |a - b|^p.
\]

This proves the lemma.

This holds for \( p > 1 \) also, but I don’t remember how to show this at this time.

However, in the context of strictly convex norms on the reflexive Banach space \( X \), the following important result holds. I will give it for the case where \( p = 2 \) since this is the case of most interest.

**Theorem 27.2.5** Let \( X \) be a reflexive Banach space and \( X' \) have strictly convex norms as discussed above. Let \( F \) be the duality map with \( p = 2 \). Then \( F \) is strictly monotone. This means
\[
\langle Fu - Fv, u - v \rangle \geq 0
\]
and it equals \( 0 \) if and only if \( u = v \).

**Proof:** First why is it monotone? By definition of \( F \), \( \langle F (u), u \rangle = \|u\|^2 \) and \( \|F (u)\| = \|u\| \). Then
\[
|\langle Fu, v \rangle| = \left| \left\langle Fu, \frac{v}{\|v\|} \right\rangle \right| \|v\| \leq \|Fu\| \|v\| = \|u\| \|v\|
\]
Hence
\[
(Fu - Fv, u - v) = \|u\|^2 + \|v\|^2 - \langle Fu, v \rangle - \langle Fu, u \rangle \geq \|u\|^2 + \|v\|^2 - 2\|u\| \|v\| \geq 0
\]
Now suppose \( \|x\| = \|y\| = 1 \) but \( x \neq y \). Then

\[
\left\langle Fx, \frac{x + y}{2} \right\rangle \leq \frac{\|x + y\|}{2} = 1
\]

It follows that

\[
\frac{1}{2} \langle Fx, x \rangle + \frac{1}{2} \langle Fx, y \rangle = \frac{1}{2} + \frac{1}{2} \langle Fx, y \rangle < 1
\]

and so

\[
\langle Fx, y \rangle < 1
\]

For arbitrary \( x, y \), \( x/\|x\| \neq y/\|y\| \)

\[
\langle Fx, y \rangle = \|x\|\|y\| \left( \frac{F(x)}{\|x\|}, \frac{y}{\|y\|} \right)
\]

It is easy to check that \( F(\alpha x) = \alpha F(x) \). Therefore,

\[
|\langle Fx, y \rangle| = \|x\|\|y\| \left( F\left( \frac{x}{\|x\|} \right), \frac{y}{\|y\|} \right) < \|x\|\|y\|
\]

Now say that \( x \neq y \) and consider \( (Fx - Fy, x - y) \)

First suppose \( x = \alpha y \). Then the above is

\[
\langle F(\alpha y) - Fy, (\alpha - 1) y \rangle = \langle (\alpha - 1) \left( \langle F(\alpha y) \rangle, y \rangle - \|y\| \right) \rangle
\]

\[
= \langle (\alpha - 1) \left( \langle \alpha F(y) \), y \rangle - \|y\| \right) \rangle
\]

\[
= \langle (\alpha - 1)^2 \|y\|^2 \rangle > 0
\]

The other case is that \( x/\|x\| \neq y/\|y\| \) and in this case,

\[
\langle Fx - Fy, x - y \rangle = \|x\|^2 + \|y\|^2 - \langle Fx, y \rangle - \langle Fy, x \rangle
\]

\[
> \|x\|^2 + \|y\|^2 - 2 \|x\|\|y\| \geq 0
\]

Thus \( F \) is strictly monotone as claimed. ■

As mentioned, this will hold for any \( p > 1 \). Here is a proof in the case that the Banach space is real which is the usual case of interest. First here is a simple observation.

**Observation 27.2.6** Let \( p > 1 \). Then \( x \to |x|^{p-2} x \) is strictly monotone. Here \( x \in \mathbb{R} \).

To verify this observation,

\[
\frac{d}{dx} \left( \left( x^2 \right)^{\frac{p-2}{2}} x \right) = \frac{1}{x^2} (p-1) \left( x^2 \right)^{\frac{p}{2}} > 0
\]

**Theorem 27.2.7** Let \( X \) be a real reflexive Banach space and \( X, X' \) have strictly convex norms as discussed above. Let \( F \) be the duality map for \( p > 1 \). Then \( F \) is strictly monotone. This means

\[
\langle Fu - Fv, u - v \rangle \geq 0
\]

and it equals 0 if and only if \( u = v \).

**Proof:** First why is it monotone? By definition of \( F \), \( \langle F(u), u \rangle = \|u\|^p \) and \( \|F(u)\| = \|u\|^{p-1} \). Then

\[
|\langle Fu, v \rangle| = \left| \left\langle Fu, \frac{v}{\|v\|} \right\rangle \right| \|v\| \leq \|Fu\| \|v\| = \|u\|^{p-1} \|v\|
\]

Hence

\[
\langle Fu - Fv, u - v \rangle = \|u\|^p + \|v\|^p - \langle Fu, v \rangle - \langle Fv, u \rangle \geq \|u\|^p + \|v\|^p - \|u\|^{p-1} \|v\| - \|u\| \|v\|^{p-1}
\]
also from similar reasoning, Thus $F$ is a multiple of the other. Thus, in particular, $x$ by the above observation that $\alpha y$ is a multiple of $x$. Then

$$\langle Fx, \frac{x + y}{2} \rangle \leq \|x\|^{p-1} \left\| \frac{x + y}{2} \right\| < \frac{\|x\| + \|y\|}{2} = 1$$

It follows that

$$\frac{1}{2} \langle Fx, x \rangle + \frac{1}{2} \langle Fx, y \rangle = \frac{1}{2} + \frac{1}{2} \langle Fx, y \rangle < 1$$

and so

$$\langle Fx, y \rangle < 1$$

It is easy to check that for nonzero $\alpha$, $F(\alpha x) = |\alpha|^{p-2} \alpha F(x)$. This is because

$$\|\alpha|^{p-2} \alpha F(x)\| = |\alpha|^{p-1} \|x\|^{p-1} = \|\alpha x\|^{p-1}$$

and so, since $|\alpha|^{p-2} \alpha F(x)$ acts like $F(\alpha x)$, it is $F(\alpha x)$. It follows that for arbitrary $x, y$, such that $x \parallel \|x\| \neq y \parallel \|y\|$

$$\langle Fx, y \rangle = \|x\|^{p-1} \|y\| \left\langle F\left(\frac{x}{\|x\|}\right), \left(\frac{y}{\|y\|}\right) \right\rangle$$

Therefore,

$$\langle Fx, y \rangle = \|x\|^{p-1} \|y\| \left\langle F\left(\frac{x}{\|x\|}\right), \left(\frac{y}{\|y\|}\right) \right\rangle < \|x\|^{p-1} \|y\| \quad (27.2.7)$$

Now say that $x \neq y$ and consider

$$\langle Fx - Fy, x - y \rangle$$

First suppose $x = \alpha y$. This is the case where $x$ is a multiple of $y$. Then the above is

$$\langle F(\alpha y) - Fy, (\alpha - 1) y \rangle = (\alpha - 1) \langle F(\alpha y), y \rangle - \|y\|^p$$

$$= (\alpha - 1) \left( |\alpha|^{p-2} \alpha \|y\|^p - \|y\|^p \right) = (\alpha - 1) \left( |\alpha|^{p-2} \alpha - 1 \right) \|y\|^p > 0$$

by the above observation that $x \mapsto |x|^{p-2} x$ is strictly monotone. Similarly, $\langle Fx - Fy, x - y \rangle > 0$ if $y = \alpha x$ for $\alpha \neq 1$.

Thus the desired result holds in the case that one vector is a multiple of the other. The other case is that neither vector is a multiple of the other. Thus, in particular, $x \parallel \|x\| \neq y \parallel \|y\|$, and in this case, it follows from (27.2.6)

$$\langle Fx - Fy, x - y \rangle = \|x\|^p + \|y\|^p - \langle Fx, y \rangle - \langle Fy, x \rangle$$

$$> \|x\|^p + \|y\|^p - \|x\|^{p-1} \|y\| - \|y\|^{p-1} \|x\|$$

$$\geq \|x\|^p + \|y\|^p - \left( \frac{\|x\|^p}{p'} + \frac{\|y\|^p}{p} \right) - \left( \frac{\|y\|^p}{p'} + \frac{\|x\|^p}{p} \right) = 0$$

Thus $F$ is strictly monotone as claimed. ■

Another useful observation about duality maps for $p = 2$ is that $\|F^{-1} y^*\|_V = \|y^*\|_{V'}$. This is because

$$\|y^*\|_{V'} = \|FF^{-1} y^*\|_{V'} = \|F^{-1} y^*\|_V$$

also from similar reasoning,

$$\langle y^*, F^{-1} y^* \rangle = \langle FF^{-1} y^*, F^{-1} y^* \rangle = \|F^{-1} y^*\|^2_V = \|y^*\|^2_{V'}$$
27.3 Penalizaton And Projection Operators

In this section, \( X \) will be a reflexive Banach space such that \( X, X' \) has a strictly convex norm. Let \( K \) be a closed convex set in \( X \). Then the following lemma is obtained.

**Lemma 27.3.1** Let \( K \) be closed and convex nonempty subset of \( X \) a reflexive Banach space which has strictly convex norm. Then there exists a projection map \( P \) such that \( Px \in K \) and for all \( y \in K \),

\[
\|y - x\| \geq \|x - Px\|
\]

**Proof:** Let \( \{y_n\} \) be a minimizing sequence for \( y \to \|y - x\| \) for \( y \in K \). Thus

\[
d = \inf \{\|y - x\| : y \in K\} = \lim_{n \to \infty} \|y_n - x\|
\]

Then obviously \( \{y_n\} \) is bounded. Hence there is a subsequence, still denoted by \( b \) such that \( y_n \to w \in K \). Then

\[
\|w - x\| \leq \lim \inf_{n \to \infty} \|y_n - x\| = d
\]

How many closest points to \( x \) are there? Suppose \( w_1 \) is another one. Then

\[
\left\| \frac{w_1 + w}{2} - x \right\| = \left\| \frac{w_1 - x + w - x}{2} \right\| < \left\| \frac{w_1 - x}{2} \right\| + \left\| \frac{w - x}{2} \right\| = d
\]

contradicting the assumption that both \( w, w_1 \) are closest points to \( x \). Therefore, \( Px \) consists of a single point. \( \blacksquare \)

Denote by \( F \) the duality map such that \( \langle Fx, x \rangle = \|x\|^2 \). This is described earlier but there is also a very nice treatment which is somewhat different in [11]. Everything can be generalized and is in [11], but here I will only consider this case. First here is a useful result.

**Proposition 27.3.2** Let \( F \) be the duality map just described. Let \( \phi(x) \equiv \frac{\|x\|^2}{2} \). Then \( F(x) = \partial \phi(x) \).

**Proof:** This follows from

\[
\langle Fx, y - x \rangle \leq \langle Fx, y \rangle - \langle Fx, x \rangle \leq \langle Fx, y \rangle^{1/2} \langle Fy, y \rangle^{1/2} - \langle Fx, x \rangle \leq \frac{\langle Fy, y \rangle}{2} - \frac{\langle Fx, x \rangle}{2} = \frac{\|y\|^2}{2} - \frac{\|x\|^2}{2}. \quad \blacksquare
\]

Next is a really nice result about the characterization of \( Px \) in terms of \( F \).

**Proposition 27.3.3** Let \( K \) be a nonempty closed convex set in \( X \) a reflexive Banach space in which both \( X, X' \) have strictly convex norms. Then \( w \in H \) is equal to \( Px \) if and only if

\[
\langle F(x - w), y - w \rangle \leq 0
\]

for every \( y \in K \).

**Proof:** First suppose the condition. Then for \( y \in K \), it follows from the above proposition about the subgradient,

\[
\frac{1}{2} \|x - y\|^2 - \frac{1}{2} \|x - w\|^2 \geq \langle F(x - w), w - y \rangle \geq 0
\]

and so since this holds for all \( y \) it follows that

\[
\|x - y\| \geq \|x - w\|
\]

for all \( y \) which says that \( w = Px \).

Next, using the subgradient idea again, for \( \theta \in [0, 1] \), suppose \( w = Px \) then for \( y \in K \) arbitrary,

\[
0 \geq \frac{1}{2} \|x - w\|^2 - \frac{1}{2} \|x - (w + \theta(y - w))\|^2 \geq \langle F(x - (w + \theta(y - w))), \theta(y - x) \rangle
\]

Now divide by \( \theta \) and let \( \theta \downarrow 0 \) and use the hemicontinuity of \( F \) given above. Then

\[
0 \geq \langle F(x - w), y - x \rangle \quad \blacksquare
\]
Definition 27.3.4 An operator of penalization is an operator \( f : X \to X' \) such that \( f = 0 \) on \( K \), \( f \) is monotone and nonzero off \( K \) as well as demicontinuous. (Strong convergence goes to weak convergence.) Actually, in applications, it is usually easy to give an ad hoc description of an appropriate penalization operator.

Proposition 27.3.5 Let \( K \) be a closed convex nonempty subset of \( X \) a reflexive Banach space such that \( X, X' \) have strictly convex norms. Then
\[
f(x) = F(x - Px)
\]
is an operator of penalization. Here \( P \) is the projection onto \( K \).

**Proof:** First, observe that \( f(x) \) is 0 on \( K \) and nonzero off \( K \). Why is it monotone?
\[
\langle F(x - Px) - F(x_1 - Px_1), x - x_1 \rangle
\]
\[
= \langle F(x - Px) - F(x_1 - Px_1), x - Px - (x_1 - Px_1) \rangle
\]
\[
+ \langle F(x - Px) - F(x_1 - Px_1), Px - Px_1 \rangle
\]
The first term is \( \geq 0 \) because \( F \) is monotone. As to the second, it equals
\[
\langle F(x - Px), Px - Px_1 \rangle + \langle F(x_1 - Px_1), Px_1 - Px \rangle
\]
and both of these are \( \geq 0 \) because of Proposition 27.3.5 which characterizes the projection map.

Now why is this hemicontinuous? Let \( x_n \to x \). Then \( Px_n \) is clearly bounded. Taking a subsequence, it can be assumed that \( Px_n \to \xi \) weakly. Is \( \xi = Px? \)
\[
\|x - Px\| \leq \|x - Px_n\| \leq \|x - x_n\| + \|x_n - Px_n\|
\]
\[
\|x_n - Px_n\| \leq \|x_n - Px\| \leq \|x - x_n\| + \|x - Px\|
\]
It follows that
\[
\|x - Px\| - \|x_n - Px_n\| \leq \|x - x_n\|
\]
\[
\|x_n - Px_n\| - \|x - Px\| \leq \|x - x_n\|
\]
Hence \( \|x_n - Px_n\| \to \|x - Px\| \). However, from convexity and strong lower semicontinuity implying weak lower semicontinuity,
\[
\|x - \xi\| \leq \liminf_{n \to \infty} \|x_n - Px_n\| = \|x - Px\|
\]
and so \( \xi = Px \) because there is only one value in \( Px \). This has shown that, thanks to uniqueness of \( Px \), \( x_n \to x \) implies \( Px_n \to Px \) weakly.

Next we show that \( f \) is demicontinuous. Suppose \( x_n \to x \). Then from what was just shown, \( Px_n \to Px \) weakly. Thus \( x_n - Px_n \to x - Px \) weakly. Then
\[
\limsup_{n \to \infty} \langle F(x_n - Px_n), x_n - Px_n - (x - Px) \rangle = \limsup_{n \to \infty} \langle F(x_n - Px_n), Px - Px_n \rangle \leq 0
\]
from Proposition 27.3.5 which characterizes the projection map. It follows that, since \( F \) is monotone hemicontinuous and bounded, it is also pseudomonotone and so for all \( v \)
\[
\liminf_{n \to \infty} \langle F(x_n - Px_n), (x_n - Px_n) - v \rangle \geq \langle F(x - Px), (x - Px) - v \rangle
\]
Now \( F(x_n - Px_n) \) is bounded. If it converges to \( \xi \), then
\[
\liminf_{n \to \infty} \langle F(x_n - Px_n), (x_n - Px_n) - v \rangle \leq \limsup_{n \to \infty} [\langle F(x_n - Px_n), (x_n - Px_n) - (x - Px) \rangle + \langle F(x_n - Px_n), (x - Px) - v \rangle]
\]
\[
\leq \langle \xi, (x - Px) - v \rangle
\]
It follows that
\[
\langle \xi, (x - Px) - v \rangle \geq \liminf_{n \to \infty} \langle F(x_n - Px_n), (x_n - Px_n) - v \rangle \geq \langle F(x - Px), (x - Px) - v \rangle
\]
Since \( v \) is arbitrary, it follows that \( \xi = F(x - Px) \). Hence \( F(x_n - Px_n) \to F(x - Px) \) weakly. Thus this is demicontinuous. 

27.4 Set-Valued Maps, Pseudomonotone Operators

In the abstract theory of partial differential equations and variational inequalities, it is important to consider set-valued maps from a Banach space to the power set of its dual. In this section we give an introduction to this theory by proving a general result on surjectivity for a class of such operators.

To begin with, if $A : X \to \mathcal{P}(Y)$ is a set-valued map, define the graph of $A$ by

$$G(A) \equiv \{(x, y) : y \in Ax\}.$$  

First consider a map $A$ which maps $\mathbb{C}^n$ to $\mathcal{P}(\mathbb{C}^n)$ which satisfies

$$Ax \text{ is compact and convex.}$$  

and also the condition that if $O$ is open and $O \supseteq Ax$, then there exists $\delta > 0$ such that if

$$y \in B(x, \delta), \text{ then } Ay \subseteq O.$$  

This last condition is sometimes referred to as upper semicontinuity. In words, $A$ is upper semicontinuous and has values which are compact and convex.

**Lemma 27.4.1** Let $A$ satisfy [27.4.3] and [27.4.6]. Then $AK$ is a subset of a compact set whenever $K$ is compact. Also the graph of $A$ is closed.

**Proof:** Let $x \in K$. Then $Ax$ is compact and contained in some open set whose closure is compact, $U_x$. By assumption there exists an open set $V_x$ containing $x$ such that if $y \in V_x$, then $Ay \subseteq U_x$. Let $V_x, \cdots, V_x$ cover $K$. Then $AK \subseteq \bigcup_{i=1}^n U_{x_i}$, a compact set. To see the graph of $A$ is closed, let $x_k \to x, y_k \to y$ where $y_k \in Ax_k$. Then letting $O = Ax+B(0,r)$ it follows from [27.4.6] that $y_k \in Ax_k \subseteq O$ for all $k$ large enough. Therefore, $y \in Ax+B(0,2r)$ and since $r > 0$ is arbitrary and $Ax$ is closed it follows $y \in Ax$. 

Also, there is a general consideration relative to upper semicontinuous functions.

**Lemma 27.4.2** If $f$ is upper semicontinuous on some set $K$ and $g$ is uniformly continuous and defined on $f(K)$, then $g \circ f$ is also upper semicontinuous.

**Proof:** Let $x_n \to x$. Let $\varepsilon > 0$ be given. Let $\delta$ be such that on $f(K)$, if $\|x - y\| < \delta$, then $\|g(x) - g(y)\| < \varepsilon$. Eventually, for all $n$ large enough,

$$f(x_n) \subseteq f(x) + B(0,\delta).$$

Thus for $z_n \in f(x_n)$, there exists $z \in f(x)$ such that $z_n \in z + B(0,\delta)$. Hence $g(z_n) \in g(z) + B(0,\varepsilon)$. Since $g(z_n)$ is a generic element of $g \circ f(x_n)$, it follows that $g \circ f$ is also upper semicontinuous. 

The next lemma is an application of the Brouwer fixed point theorem. First define an $n$ simplex, denoted by $[x_0, \cdots, x_n]$, to be the convex hull of the $n+1$ points, $\{x_0, \cdots, x_n\}$ where $\{x_i - x_0\}_{i=1}^n$ are independent. Thus

$$[x_0, \cdots, x_n] \equiv \left\{ \sum_{i=1}^n t_i x_i : \sum_{i=1}^n t_i = 1, t_i \geq 0 \right\}.$$ 

If $n \leq 2$, the simplex is a triangle, line segment, or point. If $n \leq 3$, it is a tetrahedron, triangle, line segment or point. A collection of simplices is a tiling of $\mathbb{R}^n$ if $\mathbb{R}^n$ is contained in their union and if $S_1, S_2$ are two simplices in the tiling, with

$$S_j = [x_{j0}, \cdots, x_{jn}],$$

then

$$S_1 \cap S_2 = [x_{k0}, \cdots, x_{kn}]$$

where

$$\{x_{k0}, \cdots, x_{kn}\} \subseteq \{x_{01}, \cdots, x_{n1}\} \cap \{x_{02}, \cdots, x_{n2}\}$$

or else the two simplices do not intersect. The collection of simplices is said to be locally finite if, for every point, there exists a ball containing that point which also intersects only finitely many of the simplices in the collection. It is left to the reader to verify that for each $\varepsilon > 0$, there exists a locally finite tiling of $\mathbb{R}^n$ which is composed of simplices which have diameters less than $\varepsilon$. The local finiteness ensures that for each $\varepsilon$ the vertices have no limit point. Thus one can give a function any value desired on these vertices and extend appropriately to the rest of the simplex and obtain a continuous function.

The Kakutani fixed point theorem is a generalization of the Brouwer fixed point theorem.
Theorem 27.4.3 Let $K$ be a compact convex subset of $\mathbb{R}^n$ and let $A : K \to \mathcal{P}(K)$ such that $Ax$ is a closed convex subset of $K$ and $A$ is upper semicontinuous. Then there exists $x$ such that $x \in Ax$. This is the “fixed point”.

Proof: Let there be a locally finite tiling of $\mathbb{R}^n$ consisting of simplices having diameter no more than $\varepsilon$. Let $P_{\varepsilon}x$ be the point in $K$ which is closest to $x$. For each vertex $x_k$, pick $A_{\varepsilon}x_k \in AP_{\varepsilon}x_k$ and define $A_{\varepsilon}$ on all of $\mathbb{R}^n$ by the following rule. If $x \in [x_0, \ldots, x_n]$, so $x = \sum_{i=0}^{n} t_i x_i, t_i \in [0, 1], \sum_i t_i = 1$, then

$$A_{\varepsilon}x = \sum_{k=0}^{n} t_k A_{\varepsilon}x_k,$$

Now by construction $A_{\varepsilon}x_k \in AP_{\varepsilon}x_k \in K$ and so $A_{\varepsilon}$ is a continuous map defined on $\mathbb{R}^n$ with values in $K$ thanks to the local finiteness of the collection of simplices. By the Brouwer fixed point theorem $A_{\varepsilon}$ has a fixed point $x_{\varepsilon}$ in $K$, $A_{\varepsilon}x_{\varepsilon} = x_{\varepsilon}$.

$$x_{\varepsilon} = \sum_{k=0}^{n} t_k^\varepsilon A_{\varepsilon}x_k^\varepsilon, A_{\varepsilon}x_k^\varepsilon \in AP_{\varepsilon}x_k^\varepsilon \in K$$

where a simplex containing $x_{\varepsilon}$ is

$$[x_0^\varepsilon, \ldots, x_n^\varepsilon], x_{\varepsilon} = \sum_{k=0}^{n} t_k^\varepsilon x_k^\varepsilon$$

Also, $x_{\varepsilon} \in K$ and is closer than $\varepsilon$ to each $x_k^\varepsilon$ so each $x_k^\varepsilon$ is within $\varepsilon$ of $K$. It follows that for each $k$, $|P_{\varepsilon}x_k^\varepsilon - x_k^\varepsilon| < \varepsilon$ and so

$$\lim_{\varepsilon \to 0} |P_{\varepsilon}x_k^\varepsilon - x_k^\varepsilon| = 0$$

By compactness of $K$, there exists a subsequence, still denoted with the subscript of $\varepsilon$ such that for each $k$, the following convergences hold as $\varepsilon \to 0$

$$t_k^\varepsilon \to t_k, A_{\varepsilon}x_k^\varepsilon \to y_k, P_{\varepsilon}x_k^\varepsilon \to z_k, x_k^\varepsilon \to z_k$$

Any pair of the $x_k^\varepsilon$ are within $\varepsilon$ of each other. Hence, any pair of the $P_{\varepsilon}x_k^\varepsilon$ are within $\varepsilon$ of each other because $P$ reduces distances. Therefore, in fact, $z_k$ does not depend on $k$.

$$\lim_{\varepsilon \to 0} P_{\varepsilon}x_k^\varepsilon = \lim_{\varepsilon \to 0} x_k^\varepsilon = z, \lim_{\varepsilon \to 0} x_{\varepsilon} = \lim_{\varepsilon \to 0} \sum_{k=0}^{n} t_k^\varepsilon x_k^\varepsilon = \sum_{k=0}^{n} t_k z = z$$

By upper semicontinuity of $A$, for all $\varepsilon$ small enough,

$$AP_{\varepsilon}x_k^\varepsilon \subseteq Az + B(0, r)$$

In particular, since $A_{\varepsilon}x_k^\varepsilon \in AP_{\varepsilon}x_k^\varepsilon$, $A_{\varepsilon}x_k^\varepsilon \in Az + B(0, r)$ for $\varepsilon$ small enough.

Since $r$ is arbitrary and $Az$ is closed, it follows

$$y_k \in Az.$$

It follows that since $K$ is closed,

$$x_{\varepsilon} \to z = \sum_{k=0}^{n} t_k y_k, t_k \geq 0, \sum_{k=0}^{n} t_k = 1$$

Now by convexity of $Az$ and the fact just shown that $y_k \in Az$,

$$z = \sum_{k=0}^{n} t_k y_k \in Az$$

and so $z \in Az$. This is the fixed point. ■

One can replace $\mathbb{R}^n$ with $\mathbb{C}^n$ in the above theorem because it is essentially $\mathbb{R}^{2n}$. Also the theorem holds with no change for any finite dimensional normed linear space since these are homeomorphic to $\mathbb{R}^n$ or $\mathbb{C}^n$.
Lemma 27.4.4 Suppose \( A : \mathbb{C}^n \to \mathcal{P}(\mathbb{C}^n) \) satisfies \( Ax \) is compact and convex, and \( A \) is upper semicontinuous, and \( K \) is a nonempty compact convex set in \( \mathbb{C}^n \). Then if \( y \in \mathbb{C}^n \) there exists \( [x, w] \in \mathcal{G}(A) \) such that \( x \in K \) and

\[
\Re(y - w, z - x) \leq 0
\]

for all \( z \in K \).

Proof: Tile \( \mathbb{C}^n \) with \( 2n \) simplices such that the collection is locally finite and each simplex has diameter less than \( \varepsilon < 1 \). This collection of simplices is determined by a countable collection of vertices. For each vertex \( x \), pick \( A_\varepsilon x \in Ax \) and define \( A_\varepsilon \) on all of \( \mathbb{C}^n \) by the following rule. If

\[
x \in [x_0, \cdots, x_{2n}],
\]

so

\[
x = \sum_{i=0}^{2n} t_i x_i,
\]

\[
A_\varepsilon x = \sum_{k=0}^{2n} t_k A_\varepsilon x_k.
\]

Thus \( A_\varepsilon \) is a continuous map defined on \( \mathbb{C}^n \) thanks to the local finiteness of the collection of simplices. Let \( P_K \) denote the projection on the convex set \( K \). By the Brouwer fixed point theorem, there exists a fixed point, \( x_\varepsilon \in K \) such that

\[
P_K(y - A_\varepsilon x + x_\varepsilon) = x_\varepsilon.
\]

By Corollary 27.4.11 this requires

\[
\Re(y - A_\varepsilon x, z - x_\varepsilon) \leq 0
\]

for all \( z \in K \).

Suppose \( x_\varepsilon \in [x_0, \cdots, x_{2n}] \) so \( x_\varepsilon = \sum_{k=0}^{2n} t_k^\varepsilon x_k^\varepsilon \). Then since \( x_\varepsilon \) is contained in \( K \), a compact set, and the diameter of each simplex is less than \( \varepsilon < 1 \), it follows that \( A_\varepsilon x_k^\varepsilon \) is contained in \( A(K + B(0,1)) \), which is contained in a compact set thanks to Lemma 27.4.11. The reason is that \( A \) is assumed to take bounded sets to bounded sets and \( K + B(0,1) \) is a bounded set.

From the Heine Borel theorem, there exists a sequence \( \varepsilon \to 0 \) such that

\[
t_k^\varepsilon \to t_k, x_\varepsilon \to x, A_\varepsilon x_k^\varepsilon \to y_k
\]

for \( k = 0, \cdots, 2n \). Since the diameter of the simplex containing \( x_\varepsilon \) converges to 0, it follows

\[
x_k^\varepsilon \to x, A_\varepsilon x_k^\varepsilon \to y_k.
\]

By upper semicontinuity, it follows that for all \( r > 0 \), \( Ax_k^\varepsilon \subseteq Ax + B(0, r) \) for all \( \varepsilon \) small enough. Since \( A_\varepsilon x_k^\varepsilon \in Ax_k^\varepsilon \), and \( Ax \) is closed, this implies \( y_k \in Ax \). Since \( Ax \) is convex,

\[
\sum_{k=1}^{2n} t_k y_k \in Ax.
\]

Hence for all \( z \in K \),

\[
\Re\left(y - \sum_{k=1}^{2n} t_k y_k, z - x\right) = \lim_{\varepsilon \to 0} \Re\left(y - \sum_{k=1}^{2n} t_k^\varepsilon A_\varepsilon x_k^\varepsilon, z - x_\varepsilon\right)
\]

\[
= \lim_{\varepsilon \to 0} \Re(y - A_\varepsilon x_\varepsilon, z - x_\varepsilon) \leq 0.
\]

Let \( w = \sum_{k=1}^{2n} t_k y_k \).

Lemma 27.4.5 Suppose in addition to 27.4.3 and 27.4.3. (compact convex valued and upper semicontinuous) \( A \) is coercive,

\[
\lim_{|x| \to \infty} \inf \left\{ \frac{\Re(y, x)}{|x|} : y \in Ax \right\} = \infty.
\]

Then \( A \) is onto.
**Proof:** Let \( y \in \mathbb{C}^n \) and let \( K_r \equiv B(0,r) \). By Lemma 27.4.9 there exists \( x_r \in K_r \) and \( w_r \in Ax_r \) such that

\[
\text{Re} (y - w_r, z - x_r) \leq 0
\]

(27.4.10)

for all \( z \in K_r \). Letting \( z = 0 \),

\[
\text{Re} (w_r, x_r) \leq \text{Re} (y, x_r).
\]

Therefore,

\[
\inf \left\{ \frac{\text{Re} (w, x_r)}{|x_r|} : w \in Ax_r \right\} \leq |y|.
\]

It follows from the assumption of coercivity that \(|x_r|\) is bounded independent of \( r \). Therefore, picking \( r \) strictly larger than this bound, \( 27.4.10 \) implies

\[
\text{Re} (y - w_r, v) \leq 0
\]

for all \( v \) in some open ball containing \( 0 \). Therefore, for all \( v \) in this ball

\[
\text{Re} (y - w_r, v) = 0
\]

and hence this holds for all \( v \in \mathbb{C}^n \) and so \( y = w_r \in Ax_r \). This proves the lemma.

**Lemma 27.4.6** Let \( F \) be a finite dimensional Banach space of dimension \( n \), and let \( T \) be a mapping from \( F \) to \( \mathcal{P}(F') \) such that \( 27.4.8 \) and \( 27.4.9 \) both hold for \( F' \) in place of \( \mathbb{C}^n \). Then if \( T \) is also coercive,

\[
\lim_{||u|| \to \infty} \inf \left\{ \frac{\text{Re} y^* (u)}{||u||} : y^* \in Tu \right\} = \infty,
\]

(27.4.11)

it follows \( T \) is onto.

**Proof:** Let \(|\cdot|\) be an equivalent norm for \( F \) such that there is an isometry of \( \mathbb{C}^n \) and \( F, \theta \). Now define \( A : \mathbb{C}^n \to \mathcal{P}(\mathbb{C}^n) \) by \( Ax \equiv \theta^* T\theta x \).

\[
\mathcal{P}(F') \quad \theta^* \quad \mathbb{C}^n
\]

\[
T \uparrow \quad \circ \quad \uparrow A
\]

\[
F \quad \theta \quad \mathbb{C}^n
\]

Thus \( y \in Ax \) means that there exists \( z^* \in T\theta x \) such that

\[
(w, y)_{\mathbb{C}^n} = z^* (\theta w)
\]

for all \( w \in \mathbb{C}^n \). Then \( A \) satisfies the conditions of Lemma 27.4.9 and so \( A \) is onto. Consequently \( T \) is also onto. \( \blacksquare \)

With these lemmas, it is possible to prove a very useful result about a class of mappings which map a reflexive Banach space to the power set of its dual space. For more theorems about these mappings and their applications, see \( 27.4.5 \). In the discussion below, we will use the symbol, \( \rightharpoonup \), to denote weak convergence.

**Definition 27.4.7** Let \( V \) be a Reflexive Banach space. We say \( T : V \to \mathcal{P}(V') \) is pseudomonotone if the following conditions hold.

\[
Tu \text{ is closed, nonempty, convex.}
\]

(27.4.12)

If \( F \) is a finite dimensional subspace of \( V \), then if \( u \in F \) and \( W \supseteq Tu \) for \( W \) a weakly open set in \( V' \), then there exists \( \delta > 0 \) such that

\[
v \in B(u, \delta) \cap F \text{ implies } Tv \subseteq W.
\]

(27.4.13)

If \( u_k \rightharpoonup u \) and if \( u_k^* \in Tu_k \) is such that

\[
\lim \sup_{k \to \infty} \text{Re} u_k^* (u_k - u) \leq 0,
\]

then for all \( v \in V \), there exists \( u^* (v) \in Tu \) such that

\[
\lim_{k \to \infty} \inf \text{Re} u_k^* (u_k - v) \geq \text{Re} u^* (v) (u - v).
\]

(27.4.14)

We say \( T \) is coercive if

\[
\lim_{||v|| \to \infty} \inf \left\{ \frac{\text{Re} z^* (v)}{||v||} : z^* \in Tv \right\} = \infty.
\]

(27.4.15)
In the case that \( T \) takes bounded sets to bounded sets so it is a bounded set valued operator, it turns out you don’t have to consider the second of the above conditions about the upper semicontinuity. It follows from the other conditions. It is convenient to use the notation

\[
\langle u^*, v \rangle \equiv u^* (v) , \ u^* \in V', \ v \in V.
\]

and this will be used interchangably with the earlier notation from now on.

**Lemma 27.4.8** Let \( T : X \to P (X') \) satisfy conditions \((27.4.12)\) and \((27.4.14)\) above and suppose \( T \) is bounded. Then if \( x_n \to x \) in \( X \), and if \( U \) is a weakly open set containing \( Tx \), then \( Tx_n \subseteq U \) for all \( n \) large enough. If fact the limit condition \((27.4.14)\) can be weakened to the following more general condition: If \( u_k \to u \), then there exists a subsequence still denoted as \( \{u_k\} \), such that if \( u_k^* \in Tu_k \) satisfies

\[
\limsup_{k \to \infty} \Re u_k^* (u_k - u) \leq 0,
\]

then for all \( v \in V \), there exists \( u^* (v) \in Tu \) such that

\[
\liminf_{k \to \infty} \Re u_k^* (u_k - v) \geq \Re u^* (v) (u - v).
\]

In other words, a convergent sequence has a subsequence for which the pseudomonotone limit condition holds.

**Proof:** If this is not true, there exists \( x_n \to x \), a weakly open set \( U \), containing \( Tx \) and \( z_n \in Tx_n \), but \( z_n \notin U \). Taking a subsequence if necessary, we obtain a sequence which satisfies the limit condition that if the \( \limsup \) is bounded above by \( 0 \), then the condition on the \( \liminf \) holds for that sequence, and \( z_n \to z \notin U \) in addition to this. Then

\[
\limsup_{n \to \infty} \Re (z_n, x_n - x) = 0
\]

so if \( y \in X \) there exists \( z (y) \in Tx \) such that

\[
\Re (z, x - y) = \liminf_{n \to \infty} \Re (z_n, x_n - y) \geq \Re (z (y), x - y).
\]

Letting \( w = x - y \), this shows, since \( y \in X \) is arbitrary, that the following inequality holds for every \( w \in X \).

\[
\Re (z, w) \geq \Re (z (x - w), w), \ z (x - w) \in Tx.
\]

In particular, we may replace \( w \) with \( -w \) and obtain

\[
\Re (z, -w) \geq \Re (z (x + w), -w),
\]

which implies

\[
\Re (z (x - w), w) \leq \Re (z, w) \leq \Re (z (x + w), w).
\]

Therefore, there exists

\[
z_\lambda (y) \equiv \lambda z (x - w) + (1 - \lambda) z (x + w) \in Ax
\]

such that

\[
\Re (z, w) = \Re (z_\lambda (y), w).
\]

But this is a contradiction to \( z \notin Ax \) because if \( z \notin Ax \) there exists \( w \in X \) such that for all \( z_1 \in Ax \),

\[
\Re (z, w) > \Re (z_1, w).
\]

Therefore, \( z \in Ax \) which contradicts the assumption that \( z \) and consequently \( z \) are not contained in \( U \). ■

**Definition 27.4.9** Say \( T : V \to P (V') \) is modified bounded pseudomonotone if the following conditions hold.

\( Tu \) is closed, nonempty, convex. \hspace{1cm} (27.4.17)

\( T \) is bounded meaning it takes bounded sets to bounded sets. If \( u_k \to u \) then there exists a subsequence, still denoted as \( \{u_k\} \) such that if \( u_k^* \in Tu_k \) and

\[
\limsup_{k \to \infty} \Re u_k^* (u_k - u) \leq 0,
\]

then for all \( v \in V \), there exists \( u^* (v) \in Tu \) such that

\[
\liminf_{k \to \infty} \Re u_k^* (u_k - v) \geq \Re u^* (v) (u - v).
\]

(27.4.18)
One of the nice properties of pseudomonotone maps is that when you add two of them, you get another one. I will give a proof in the case that the two pseudomonotone maps are both bounded. It is probably true in general, but as just noted, it is less trouble to verify if you don’t have to worry about as many conditions. I will also assume the spaces are all real so it will not be necessary to constantly write the real part. In addition to this, it is often convenient to use a different notation. For \( u^* \in X' \) and \( v \in X \), write

\[
\langle u^*, v \rangle_{X', X} \equiv u^*(v)
\]

This new notation will often be used.

**Theorem 27.4.10** Suppose \( A, B : X \rightarrow \mathcal{P}(X') \) are both pseudomonotone and bounded. Then so is their sum. If \( A \) is modified bounded pseudomonotone and \( B \) is bounded pseudomonotone, then \( A + B \) is modified bounded pseudomonotone.

**Proof:** It is clear that \( Ax + Bx \) is closed and convex because this is true of both of the sets in the sum. It is also bounded because both terms in the sum are bounded. It only remains to verify the limit condition. Suppose then that

\[
u_n \rightharpoonup u \text{ weakly}
\]

Will the limit condition hold for \( A + B \) when applied to this further subsequence? Suppose \( z_n \in Ax_n, w_n \in Bx_n \) and

\[
\limsup_{n \to \infty} \langle z_n + w_n, u_n - u \rangle \leq 0
\]

Will the \( \liminf \) condition hold? From the above,

\[
\limsup_{n \to \infty} \langle z_n + w_n, u_n - u \rangle \leq \limsup_{n \to \infty} \langle z_n, u_n - u \rangle + \limsup_{n \to \infty} \langle w_n, u_n - u \rangle \tag{27.4.19}
\]

and so, if the second term \( \leq 0 \), you would have, since \( B \) is modified bounded pseudomonotone,

\[
\liminf_{n \to \infty} \langle w_n, u_n - u \rangle \geq \langle w(u), u - u \rangle = 0
\]

Hence you would have \( \liminf_{n \to \infty} \langle w_n, u_n - u \rangle \geq 0 \geq \limsup_{n \to \infty} \langle w_n, u_n - u \rangle \) and so \( \lim_{n \to \infty} \langle w_n, u_n - u \rangle = 0 \). Hence

\[
\limsup_{n \to \infty} \langle z_n + w_n, u_n - u \rangle = \limsup_{n \to \infty} \langle z_n, u_n - u \rangle \leq 0
\]

Then using that \( A \) is pseudomonotone, \( \lim_{n \to \infty} \langle z_n, u_n - u \rangle = 0 \) also. Then from this it is routine to establish the pseudomonotone limit condition for the sum \( A + B \). In fact, you would have for any \( v \),

\[
\liminf_{n \to \infty} \langle z_n, u_n - v \rangle \geq \langle z(u), u - v \rangle
\]

\[
\liminf_{n \to \infty} \langle w_n, u_n - v \rangle \geq \langle w(u), u - v \rangle
\]

Then you would get

\[
\liminf_{n \to \infty} \langle z_n + w_n, u_n - v \rangle = \liminf_{n \to \infty} (\langle z_n, u_n - v \rangle + \langle w_n, u_n - v \rangle)
\]

\[
\geq \liminf_{n \to \infty} (\langle z_n, u_n - v \rangle) + \liminf_{n \to \infty} (\langle w_n, u_n - v \rangle)
\]

\[
\geq \langle z(u), u - v \rangle + \langle w(u), u - v \rangle
\]

and \( z(u) + w(u) \in (A + B)(u) \). Thus the limit condition will hold if either \( \limsup_{n \to \infty} \langle z_n, u_n - u \rangle \) or \( \limsup_{n \to \infty} \langle w_n, u_n - u \rangle \) is \( \leq 0 \). Therefore, if the limit condition fails, you must have both of these strictly positive. Take a subsequence, still denoted with subscript \( n \), such that

\[
\limsup_{n \to \infty} \langle z_n, u_n - u \rangle = \lim_{n \to \infty} \langle z_n, u_n - u \rangle = \delta > 0.
\]

Then, using this subsequence (\( \limsup \) gets smaller when you go to a subsequence.), and the assumption,

\[
0 \geq \limsup_{n \to \infty} \langle z_n + w_n, u_n - u \rangle = \delta + \limsup_{n \to \infty} \langle w_n, u_n - u \rangle \geq \delta
\]
which is a contradiction. Thus the case where both are strictly positive cannot occur. The last claim follows from first
taking a subsequence for which the pseudomonotone limit condition holds for $A$ and then repeating the argument.

It is not entirely clear whether the sum of modified bounded pseudomonotone operators is modified bounded
pseudomonotone. This is because when you go to a subsequence, the lim sup gets smaller and so it is not entirely
clear whether the subsequence for $A$ will continue to yield the limit condition if a further subsequence is taken.

The following is mostly in [10].

**Theorem 27.4.11** Let $V$ be a reflexive Banach space and let $T : V \to P(V')$ be pseudomonotone, bounded, and
coercive. Then $T$ is onto. More generally, the same holds if $T$ is modified bounded pseudomonotone.

**Proof:** The proof is for modified bounded pseudomonotone since this is more general. Let $\mathcal{F}$ be the set of finite
dimensional subspaces of $V$ and let $F \in \mathcal{F}$. Then define $T_F$ as

$$T_F \equiv i_F^* T i_F$$

where here $i_F$ is the identity map from $F$ to $V$. Then $T_F$ satisfies the conditions of Lemma 27.4.10 thanks to Lemma
27.4.15 and so $T_F$ is onto $P(F')$. Let $w^* \in V'$. Then since $T_F$ is onto, there exists $u_F \in F$ such that

$$i_F^* w^* \in i_F^* T i_F u_F.$$ 

Thus for each finite dimensional subspace $F$, there exists $u_F \in F$ such that for all $v \in F$,

$$\langle w^*, v \rangle = \langle u_F^*, v \rangle, \quad u_F^* \in Tu_F. \quad (27.4.20)$$

Replacing $v$ with $u_F$, in [4/70.4/1],

$$\frac{\langle u_F^*, u_F \rangle}{||u_F||} = \frac{\langle w^*, u_F \rangle}{||u_F||} \leq ||w^*||.$$ 

Therefore, the assumption that $T$ is coercive implies $\{u_F : F \in \mathcal{F}\}$ is bounded in $V$. Now define

$$W_F \equiv \bigcup \{u_F' : F' \supseteq F\}.$$ 

Then $W_F$ is bounded and if $\overline{W_F}$ is weak closure of $W_F$, then

$$\{\overline{W_F} : F \in \mathcal{F}\}$$

is a collection of nonempty weakly compact (since $V$ is reflexive and the $u_F$ were just shown bounded) sets having
the finite intersection property because $W_F \neq \emptyset$ for each $F$. (If $F_i, i = 1, \ldots, n$ are finite dimensional subspaces, let
$F$ be a finite dimensional subspace which contains all of these. Then $W_F \neq \emptyset$ and $W_F \subseteq \cap_{i=1}^n \overline{W_{F_i}}.$) Thus there exists

$$u \in \cap \{\overline{W_F} : F \in \mathcal{F}\}.$$ 

I will show $w^* \in Tu$. If $w^* \notin Tu$, a closed convex set, there exists $v \in V$ such that

$$\Re \langle w^*, u - v \rangle < \Re \langle u^*, u - v \rangle \quad (27.4.21)$$

for all $u^* \in Tu$. This follows from the separation theorems. (These theorems imply there exists $z \in V$ such that

$$\Re \langle w^*, z \rangle < \Re \langle u^*, z \rangle$$

for all $u^* \in Tu$. Define $u - v \equiv z$.)

Now let $F \supseteq \{u, v\}$. Since $u \in \overline{W_F}$, a weakly sequentially compact set, there exists a sequence, $\{u_k\}$, such that

$$u_k \to u, \quad u_k \in W_F.$$ 

and in addition, for this subsequence, the pseudomonotone limit condition holds.

Then since $F \supseteq \{u, v\}$, there exists $u_k^* \in Tu_k$ such that

$$\langle u_k^*, u_k - u \rangle = \langle w^*, u_k - u \rangle.$$ 

Therefore,

$$\lim_{k \to \infty} \Re \langle u_k^*, u_k - u \rangle = \lim_{k \to \infty} \Re \langle w^*, u_k - u \rangle = 0.$$
It follows by the assumption that $T$ is modified bounded pseudomonotone and the pseudomonotone limit condition holds for this sequence; the following holds for the $v$ defined above in \[27.4.21\]

$$\lim \inf_{k \to \infty} \Re \langle u_k^*, u_k - v \rangle \geq \Re \langle u^* (v), u - v \rangle, \quad u^* (v) \in Tu.$$ 

But since $v \in F, \Re \langle u_k^*, u_k - v \rangle = \Re \langle w^*, u_k - v \rangle$ and so

$$\lim \inf_{k \to \infty} \Re \langle u_k^*, u_k - v \rangle = \lim \inf_{k \to \infty} \Re \langle w^*, u_k - v \rangle = \Re \langle w^*, u - v \rangle,$$

so from \[27.4.21\] $\Re \langle w^*, u - v \rangle < \Re \langle u^* , u - v \rangle$ for all $u^* \in Tu,$

$$\Re \langle w^*, u - v \rangle = \lim \inf_{k \to \infty} \Re \langle u_k^*, u_k - v \rangle \geq \Re \langle u^* (v), u - v \rangle > \Re \langle w^*, u - v \rangle,$$

a contradiction. Thus, $w^* \in Tu.$ ■

### 27.5 Generalized Gradients

This is an interesting theorem, but one might wonder if there are easy to verify examples of such possibly set valued mappings. In what follows consider only real spaces because the essential ideas are included in this case which is also the case of most use in applications. Of course, you might with some justification, make the claim that the following is not really very easy to verify any more than the original definition.

**Definition 27.5.1** Let $V$ be a real reflexive Banach space and let $f : V \to \mathbb{R}$ be a locally Lipschitz function, meaning that $f$ is Lipschitz near every point of $V$ although $f$ need not be Lipschitz on all of $V.$ Under these conditions,

$$f^0 (x,y) = \lim \sup_{\mu \to 0+} \sup_{h \to 0} \frac{f(x+h+\mu y) - f(x+h)}{\mu} \quad (27.5.22)$$

and $\partial f (x) \subseteq X'$ is defined by

$$\partial f (x) \equiv \{ x^* \in X' : x^* (y) \leq f^0 (x,y) \text{ for all } y \in X \}. \quad (27.5.23)$$

The set just described is called the generalized gradient. In \[27.5.22\] we mean the following by the right hand side.

$$\lim_{(r,\delta) \to (0,0)} \sup \left\{ \frac{f(x+h+\mu y) - f(x+h)}{\mu} : \mu \in (0,r), h \in B (0,\delta) \right\}$$

I will show, following [26], that these generalized gradients of locally Lipschitz functions are sometimes pseudomonotone. First here is a lemma.

**Lemma 27.5.2** Let $f$ be as described in the above definition. Then $\partial f (x)$ is a closed, bounded, convex, and non empty subset of $V'$. Furthermore, for $x^* \in \partial f (x),$

$$\|x^*\| \leq \text{Lip}_x (f). \quad (27.5.24)$$

**Proof:** It is left as an exercise to verify the assertions that $\partial f (x)$ is closed, and convex. It follows directly from the definition. To verify this set is bounded, let $\text{Lip}_x (f)$ denote a Lipschitz constant valid near $x \in V$ and let $x^* \in \partial f (x).$ Then choosing $y$ with $\|y\| = 1$ and $x^* (y) \geq \frac{1}{2} \|x^*\|,$

$$\frac{1}{2} \|x^*\| = x^* (y) \leq f^0 (x,y). \quad (27.5.25)$$

Also, for small $\mu$ and $h,$

$$\left| \frac{f(x+h+\mu y) - f(x+h)}{\mu} \right| \leq \text{Lip}_x (f) \|y\| = \text{Lip}_x (f).$$

Therefore, $f^0 (x,y) \leq \text{Lip}_x (f)$ and so \[27.5.22\] shows $\|x^*\| \leq 2 \text{Lip}_x (f).$
The interesting part of this Lemma is that \( \partial f (x) \neq \emptyset \). To verify this first note that the definition of \( f^0 \) implies that \( y \to f^0 (x, y) \) is a gauge function. Now fix \( y \in V \) and define on \( \mathbb{R} y \) a linear map \( x_0^* \) by
\[
x_0^* (\alpha y) = \alpha f^0 (x, y).
\]
Then if \( \alpha \geq 0 \),
\[
x_0^* (\alpha y) = \alpha f^0 (x, y) = f^0 (x, \alpha y).
\]
If \( \alpha < 0 \),
\[
\lim_{\mu \to 0^+} \inf_{h \to 0} \frac{(-\alpha f (x + h) - (-\alpha f (x + h + \mu y))}{\mu} = \frac{(-\alpha) \inf_{\mu \to 0^+} \inf_{h \to 0} \frac{f (x + h - \mu y) - f (x + h)}{\mu}} \leq (-\alpha) f^0 (x, -y) = f^0 (x, \alpha y).
\]
Therefore, \( x_0^* (\alpha y) \leq f^0 (x, \alpha y) \) for all \( \alpha \). By the Hahn Banach theorem there is an extension of \( x_0^* \) to all of \( V \), \( x^* \) which satisfies,
\[
x^* (y) \leq f^0 (x, y)
\]
for all \( y \). It remains to verify \( x^* \) is continuous. This follows easily from
\[
|x^* (y)| = \max (x^* (-y), x^* (y)) \leq \max (f^0 (x, y), f^0 (x, -y)) \leq \text{Lip}_x (f) ||y||,
\]
which verifies 27.5.2 and proves the lemma.

This lemma has verified the first condition needed in the definition of pseudomonotone. The next lemma verifies that these generalized subgradients satisfy the second of the conditions needed in the definition. In fact somewhat more than is needed in the definition is shown.

**Lemma 27.5.3** Let \( U \) be weakly open in \( V \) and suppose \( \partial f (x) \subseteq U \). Then \( \partial f (z) \subseteq U \) whenever \( z \) is close enough to \( x \).

**Proof:** Suppose to the contrary there exists \( z_n \to x \) but \( z_n^* \in \partial f (z_n) \setminus U \). From the first lemma, we may assert that \( ||z_n^*|| \leq 2 \text{Lip} (f) \) for all \( n \) large enough. Therefore, there is a subsequence, still denoted by \( n \) such that \( z_n^* \) converges weakly to \( z^* \notin U \).

**Claim:** \( f^0 (x, y) \geq \limsup_{n \to \infty} f^0 (x_n, y) \).

**Proof of the claim:** There exists \( \delta > 0 \) such that if \( \mu, ||h|| < \delta \), then
\[
\varepsilon + f^0 (x, y) \geq \frac{f (x + h + \mu y) - f (x + h)}{\mu}.
\]
Thus, for \( ||h|| < \delta \),
\[
\varepsilon + f^0 (x, y) \geq \frac{f (x_n + (x - x_n) + h + \mu y) - f (x_n + (x - x_n) + h)}{\mu}.
\]
Now let \( ||h'|| < \frac{\delta}{2} \) and let \( n \) be so large that \( ||x - x_n|| < \frac{\delta}{2} \). Suppose \( ||h'|| < \frac{\delta}{2} \). Then choosing \( h = h' - (x - x_n) \), it follows the above inequality holds because \( ||h|| < \delta \). Therefore, if \( ||h'|| < \frac{\delta}{2} \), and \( n \) is sufficiently large,
\[
\varepsilon + f^0 (x, y) \geq \frac{f (x_n + h' + \mu y) - f (x_n + h')}{\mu}.
\]
Consequently, for all \( n \) large enough,
\[
\varepsilon + f^0 (x, y) \geq f^0 (x_n, y)
\]
which proves the claim.

Now with the claim,
\[
z^* (y) = \limsup_{n \to \infty} z_n^* (y) \leq \limsup_{n \to \infty} f^0 (x_n, y) \leq f^0 (x, y)
\]
so \( z^* \in \partial f (x) \) contradicting the assumption that \( z^* \notin U \). This proves the lemma.

It is necessary to assume more on \( f^0 \) in order to obtain the third axiom defining pseudomonotone. The following theorem describes the situation.
Theorem 27.5.4 Let \( f: V \to V' \) be locally Lipschitz and suppose it satisfies the condition that whenever
\[
x_n \text{ converges weakly to } x
\]
and
\[
\limsup_{n \to \infty} f^0(x_n, x - x_n) \geq 0
\]
it follows that
\[
\limsup_{n \to \infty} f^0(x_n, z - x_n) \leq f^0(x, z - x)
\]
for all \( z \in V \). Then \( \partial f \) is pseudomonotone.

Proof: By \[27.5.24\] and \[27.5.26\] both are satisfied thanks to Lemmas \[27.5.1\] and \[27.5.2\]. It remains to verify \[27.5.1\]. To do so, I will adopt the convention that \( x^* \in \partial f(x) \). Suppose
\[
\limsup_{n \to \infty} x^*_n(x_n - x) \leq 0.
\]
This implies \( \liminf_{n \to \infty} x^*_n(x_n - x) \geq 0 \). Thus,
\[
0 \leq \liminf_{n \to \infty} x^*_n(x_n - x) \leq \liminf_{n \to \infty} f^0(x_n, x - x_n)
\]
\[
\leq \limsup_{n \to \infty} f^0(x_n, x - x_n),
\]
which implies, by the above assumption that for all \( z \),
\[
\limsup_{n \to \infty} x^*_n(z - x_n) \leq \limsup_{n \to \infty} f^0(x_n, z - x_n) \leq f^0(x, z - x).
\]
In particular, this holds for \( z = x \) and this implies \( \limsup_{n \to \infty} x^*_n(x_n - x) \leq 0 \) which along with \[27.5.26\] yields
\[
\liminf_{n \to \infty} x^*_n(x_n - x) = 0
\]

Now let \( z \) be arbitrary. There exists a subsequence, \( n_k \), depending on \( z \) such that
\[
\lim_{k \to \infty} x^*_n(x_{n_k} - z) = \liminf_{k \to \infty} x^*_n(x_{n_k} - z)
\]
Now from Lemma \[27.5.1\] and its proof, the \( ||x^*_n|| \) are all bounded by \( \text{Lip}_x(f) \) whenever \( n \) is large enough. Therefore, there is a further subsequence, still denoted by \( n_k \) such that
\[
x^*_n \text{ converges weakly to } x^*(z).
\]
We need to verify that \( x^*(z) \in \partial f(x) \). To do so, let \( y \) be arbitrary. Then from the definition,
\[
x^*_n(y - x_n) \leq f^0(x_n, y - x_n).
\]
From \[27.5.26\], we can take the lim sup of both sides and obtain, using \[27.5.26\]
\[
x^*(z)(y - x) \leq \limsup_{n \to \infty} f^0(x_n, y - x_n) \leq f^0(x, y - x).
\]
Since \( y \) is arbitrary, this shows \( x^*(z) \in \partial f(x) \) and proves the theorem.

27.6 Maximal Monotone Operators

Here it is assumed that the spaces are all real spaces to simplify the presentation.

Definition 27.6.1 Let \( A : D(A) \subseteq X \to \mathcal{P}(X) \) be a set valued map. It is said to be monotone if whenever \( y_1, y_2 \in Ax_i \),
\[
\langle y_1 - y_2, x_1 - x_2 \rangle \geq 0
\]
Denote by \( \mathcal{G}(A) \) the graph of \( A \) consisting of all pairs \((x, y)\) where \( y \in Ax \). Such a monotone operator is said to be maximal monotone if
\[
F + A
\]
is onto where \( F \) is the duality map with \( p = 2 \).
Actually, it is more usual to say that the graph is maximal monotone if the graph is monotone and there is no monotone graph which properly contains the given graph. However, the two conditions are equivalent and I am more used to using the version in the above definition.

There is a fundamental result about these which is given next.

**Theorem 27.6.2** Let $X, X'$ be reflexive and have strictly convex norms. Let $A$ be a monotone set valued map as just described. Then if $\lambda F + A$ is onto for some $\lambda > 0$, then whenever

$$\langle y - z, x - u \rangle \geq 0$$

for all $[x, y] \in \mathcal{G}(A)$ it follows that $z \in Au$ and $u \in D(A)$. That is, the graph is maximal.

**Proof:** Suppose that for all $[x, y] \in \mathcal{G}(A)$,

$$\langle y - z, x - t \rangle \geq 0$$

Does it follow that $z \in At$? By assumption, $z + \lambda F(t) = \lambda F\hat{x} + \hat{\xi}$, $\hat{\xi} \in A\hat{x}$. Then replacing $y$ with $\hat{\xi}$ and $x$ with $\hat{x}$,

$$\langle \hat{\xi} - \left( \lambda F\hat{x} + \hat{\xi} - \lambda Ft \right), \hat{x} - t \rangle \geq 0$$

and so

$$\lambda \langle Ft - F\hat{x}, t - \hat{x} \rangle \leq 0$$

which implies from Theorem 27.2.5 that $t = \hat{x}$ and so the graph of $A$ is indeed maximal monotone.

**Note** that this would have worked with no change if the duality map had been for arbitrary $p > 1$.

### 27.6.1 Maximal Monotone, Equivalent Conditions

In fact, these two conditions are equivalent. This is shown in [11]. We give a proof of this here. First it is necessary to prove a min max theorem. The proof given follows Brezis [19] which is where I found it.

**Theorem 27.6.3** Let $E, F$ be finite dimensional normed linear spaces. (Without loss of generality, Hilbert spaces.) Also suppose that $A \subseteq E, B \subseteq F$ are compact and convex sets and that $H : A \times B \to \mathbb{R}$ is such that

$$x \to H(x, y) \text{ is convex}$$

$$y \to H(x, y) \text{ is concave}$$

Thus $H$ is continuous in each variable. Then

$$\min_{x \in A} \max_{y \in B} H(x, y) = \max_{y \in B} \min_{x \in A} H(x, y)$$

This condition is equivalent to the existence of $(x_0, y_0) \in A \times B$ such that

$$H(x_0, y) \leq H(x_0, y_0) \leq H(x, y_0) \text{ for all } x, y$$

(27.6.30)

**Proof:** Define $H_\varepsilon(x, y) \equiv H(x, y) + \varepsilon |x|^2$ where $\varepsilon > 0$. Let

$$f_\varepsilon(y) \equiv \min_{x \in A} H_\varepsilon(x, y)$$

Then by strict convexity, of $H_\varepsilon$, which results from the observation that

$$\left| \frac{x + y}{2} \right|^2 < \frac{1}{2} \left( |x|^2 + |y|^2 \right),$$

it follows that there exists a unique $x \equiv g(y)$ such that

$$f_\varepsilon(y) = H_\varepsilon(g(y), y)$$
Now this function \( y \to f_\varepsilon (y) \) is the minimum of functions which are continuous and so it is upper semi continuous. Hence there exists \( y^* \) such that
\[
f_\varepsilon (y^*) = \max_{y \in B} f_\varepsilon (y) = \max_{y \in B} H_\varepsilon (g (y), y) = H_\varepsilon (g (y^*), y^*) \quad (27.6.31)
\]
Since \( H_\varepsilon \) is concave, for \( t \in (0, 1) \), it follows for \( y \in B \),
\[
H_\varepsilon (x, (1-t) y^* + ty) \geq (1-t) H_\varepsilon (x, y^*) + t H_\varepsilon (x, y)
\]
In particular, this is true if \( x = g((1-t) y^* + ty) \). Thus
\[
f_\varepsilon (y^*) \geq f_\varepsilon ((1-t) y^* + ty) \equiv H_\varepsilon (g ((1-t) y^* + ty), (1-t) y^* + ty)
\]
It follows that
\[
f_\varepsilon (y^*) - (1-t) f_\varepsilon (y^*) = tf_\varepsilon (y^*) \geq t H_\varepsilon (g ((1-t) y^* + ty), y)
\]
and so
\[
f_\varepsilon (y^*) \geq H_\varepsilon (g ((1-t) y^* + ty), y) \quad \text{for all } y \in B \quad (27.6.32)
\]
Next consider \( g((1-t) y_1 + ty_2) \). Does it follow that \( \lim_{t \to 0} g((1-t) y_1 + ty_2) = g(y_1) \)? For all \( x \in A \), it follows from the definition of \( g \) that
\[
H_\varepsilon (g((1-t) y_1 + ty_2), (1-t) y_1 + ty_2) \leq H_\varepsilon (x, (1-t) y_1 + ty_2)
\]
By concavity of \( H \) in the second argument,
\[
(1-t) H_\varepsilon (g((1-t) y_1 + ty_2), y_1) + t H_\varepsilon (g((1-t) y_1 + ty_2), y_2)
\]
Now let \( t_n \to 0 \). By compactness, \( g((1-t_n) y_1 + t_ny_2) \) is in a compact set and so there is a further subsequence, still denoted by \( t_n \) such that
\[
g((1-t_n) y_1 + t_ny_2) \to \hat{x} \in A
\]
Then passing to a limit in \( 27.6.32 \) one obtains
\[
H_\varepsilon (\hat{x}, y_1) \leq H_\varepsilon (x, y_1)
\]
which shows that \( \hat{x} = g(y_1) \). Since \( t_n \to 0 \) was arbitrary, this shows that in fact
\[
\lim_{t \to 0} g((1-t) y_1 + ty_2) = g(y_1)
\]
Now from \( 27.6.32 \) and this observation about \( g \), let \( t \to 0 \) to conclude that
\[
f_\varepsilon (y^*) \geq H_\varepsilon (g(y^*), y)
\]
which is true for any \( y \in B \). Also by definition,
\[
f_\varepsilon (y^*) \leq H_\varepsilon (x, y^*) \quad \text{for every } x
\]
Hence for every \( x \),
\[
H_\varepsilon (x, y^*) \geq f_\varepsilon (y^*) \equiv H_\varepsilon (g (y^*), y^*) \geq H_\varepsilon (g (y^*), y)
\]
Then
\[
H_\varepsilon (x, y^*) \geq H_\varepsilon (g (y^*), y^*) \geq H_\varepsilon (g (y^*), y)
\]
By \( 27.6.31 \) the inequality on the right yields
\[
H_\varepsilon (g (y^*), y^*) = \max_{y \in B} H_\varepsilon (g (y), y) = \max_{y \in B} \min_{x \in A} H_\varepsilon (x, y) \geq H_\varepsilon (g (y^*), y)
\]
This is true for all \( y \) and so
\[
\max_{y \in B} \min_{x \in A} H_\varepsilon(x, y) \geq \max_{y \in B} H_\varepsilon(g(y^*), y) \geq \min_{x \in A} \max_{y \in B} H_\varepsilon(x, y)
\]
Thus, letting \( C \equiv \max \{|x| : x \in A\} \)
\[
\varepsilon C + \max_{y \in B} \min_{x \in A} H(x, y) \geq \min_{x \in A} \max_{y \in B} H(x, y)
\]
Since \( \varepsilon \) is arbitrary, it follows that
\[
\max_{y \in B} \min_{x \in A} H(x, y) \geq \min_{x \in A} \max_{y \in B} H(x, y)
\]
However, the other inequality is obvious because
\[
\max_{y \in B} H(x, y) \geq H(x, y) \geq \min_{x \in A} H(x, y)
\]
and so for each \( x \),
\[
\max_{y \in B} H(x, y) \geq \max_{y \in B} \min_{x \in A} H(x, y)
\]
and so
\[
\min_{x \in A} \max_{y \in B} H(x, y) \geq \max_{y \in B} \min_{x \in A} H(x, y)
\]
This proves the first part.

Now consider 27.6.30 about the existence of a “saddle point” given the equality of \( \min \max \) and \( \max \min \). Let
\[
\alpha = \min_{x \in A} \max_{y \in B} H(x, y) = \max_{y \in B} \min_{x \in A} H(x, y)
\]
Then from
\[
y \to \min_{x \in A} H(x, y)
\]
being upper semicontinuous and \( x \to \max_{y \in B} H(x, y) \) being lower semicontinuous, there exist \( y_0 \) and \( x_0 \) such that
\[
\alpha = \min_{x \in A} H(x, y_0) = \max_{y \in B} \min_{x \in A} H(x, y) = \max_{y \in B} H(x_0, y)
\]
Then
\[
\alpha = \max_{y \in B} H(x_0, y) \geq H(x_0, y_0)
\]
\[
\alpha = \min_{x \in A} H(x, y_0) \leq H(x, y_0)
\]
so in fact \( \alpha = H(x_0, y_0) \) and from the above string of equalities,
\[
H(x_0, y) \leq H(x_0, y_0) \leq H(x, y_0)
\]
Thus if the \( \min \max \) condition holds, then there exists a saddle point.

Finally suppose there is a saddle point \((x_0, y_0)\) where
\[
H(x_0, y) \leq H(x_0, y_0) \leq H(x, y_0)
\]
Then
\[
\min_{x \in A} \max_{y \in B} H(x, y) \leq \max_{y \in B} H(x_0, y) \leq \min_{x \in A} H(x, y_0) \leq \max_{y \in B} \min_{x \in A} H(x, y)
\]
However, as noted above, it is always the case that
\[
\max_{y \in B} \min_{x \in A} H(x, y) \leq \min_{x \in A} \max_{y \in B} H(x, y)
\]

You don’t need to assume \( H \) is continuous in each argument. It suffices to have \( H \) lower semicontinuous in \( x \) and upper semicontinuous in \( y \). The argument goes through just fine with these more general conditions.

Next is the theorem about the graph being maximal being equivalent to the operator being maximal monotone. It is a very convenient result to have. The proof is a modified version of one in Barbu [11]. I was unable to completely follow his argument. It is based on the following lemma also in Barbu. This is a little like the Browder lemma but is based on the \( \min \max \) theorem above. It is also a very interesting argument.
Lemma 27.6.4 Let $E$ be a finite dimensional Banach space and let $K$ be a convex and compact subset of $E$. Let $\mathcal{G}(A)$ be a monotone subset of $E \times E'$ such that $D(A) \subseteq K$ and $B$ is a single valued monotone and continuous operator from $E$ to $E'$. Then there exists $x \in K$ such that

$$\langle Bx + v, u - x \rangle_{E', E} \geq 0 \text{ for all } [u,v] \in \mathcal{G}(A).$$

If $B$ is coercive

$$\lim_{\|x\| \to \infty} \frac{\langle Bx, x \rangle}{\|x\|} = \infty,$$

and $0 \in D(A)$, then one can assume only that $K$ is convex and closed.

**Proof:** Let $T : E \to K$ be the multivalued operator defined by

$$Ty \equiv \left\{ x \in K : \langle By + v, u - x \rangle_{E', E} \geq 0 \text{ for all } [u,v] \in \mathcal{G}(A) \right\}$$

Here $y \in E$ and it is desired to show that $Ty \neq \emptyset$ for all $y \in K$. For $[u,v] \in \mathcal{G}(A)$, let

$$K_{u,v} = \left\{ x \in K : \langle By + v, u - x \rangle_{E', E} \geq 0 \right\}.$$

Then $K_{u,v}$ is a closed, hence compact subset of $K$. The thing to do is to show that $\cap_{[u,v] \in \mathcal{G}(A)} K_{u,v} \equiv Ty \neq \emptyset$ whenever $y \in K$. Then one argues that $T$ is set valued, has convex compact values and is upper semicontinuous. Then one applies the Kakutani fixed point theorem to get $x \in Tx$.

Since these sets $K_{u,v}$ are compact, it suffices to show that they satisfy the finite intersection property. Thus for $\{[u_i, v_i]\}_{i=1}^n$ a finite set of elements of $\mathcal{G}(A)$, it is necessary to show that there exists a solution $x$ to the inequalities

$$\langle u_i - x, By + v_i \rangle \geq 0, \quad i = 1, 2, \cdots, n$$

and then it follows from finite intersection property that there exists $x \in \cap_{[u,v] \in \mathcal{G}(A)} K_{u,v}$ which is what was desired. Let $P_n$ be all $\bar{x} = (\lambda_1, \cdots, \lambda_n)$ such that each $\lambda_k \geq 0$ and $\sum_{k=1}^n \lambda_k = 1$. Let $H : P_n \times P_n \to \mathbb{R}$ be given by

$$H \left( \bar{\mu}, \bar{\lambda} \right) \equiv \sum_{i=1}^n \mu_i \left( By_i + v_i, \sum_{j=1}^n \lambda_j u_j - u_i \right) \quad (27.6.34)$$

Then this is both convex and concave in both $\bar{\lambda}, \bar{\mu}$ and so by Theorem 27.6.3, there exists $\bar{\mu}_0, \bar{\lambda}_0$ both in $P_n$ such that for all $\bar{\mu}, \bar{\lambda},$

$$H \left( \bar{\mu}, \bar{\lambda}_0 \right) \leq H \left( \bar{\mu}_0, \bar{\lambda}_0 \right) \leq H \left( \bar{\mu}_0, \bar{\lambda} \right) \quad (27.6.35)$$

However, plugging $\bar{\mu} = \bar{\lambda}$ in (27.6.34),

$$H \left( \bar{\lambda}, \bar{\lambda} \right) = \sum_{i=1}^n \lambda_i \left( By_i + v_i, \sum_{j=1}^n \lambda_j u_j - u_i \right) = \sum_{i=1}^n \left( By_i + v_i, \sum_{j=1}^n \lambda_i \lambda_j u_j - \lambda_i u_i \right) = \sum_{i=1}^n \left( By_i + v_i, \sum_{j=1}^n (\lambda_i \lambda_j u_j - \lambda_i \lambda_j u_i) \right)$$

$$= \sum_{i=1}^n \left( By_i, \sum_{j=1}^n (\lambda_i \lambda_j u_j - \lambda_i \lambda_j u_i) \right) + \sum_{i=1}^n \left( v_i, \sum_{j=1}^n (\lambda_i \lambda_j u_j - \lambda_i \lambda_j u_i) \right)$$

The first term obviously equals 0. Consider the second. This term equals

$$\sum_{i} \sum_{j} \lambda_i \lambda_j \langle v_i, (u_j - u_i) \rangle$$
The terms equal 0 when \( j = i \) or they come in pairs
\[
\lambda_i \lambda_j \langle v_i, (u_j - u_i) \rangle + \lambda_i \lambda_j \langle v_j, (u_i - u_j) \rangle = \lambda_i \lambda_j (\langle v_i, (u_j - u_i) \rangle - \langle v_j, (u_j - u_i) \rangle)
\]
by monotonicity of \( A \). Hence \( H \left( \bar{x}, \bar{x} \right) \leq 0 \). Then from 27.6.33, for all \( \bar{\mu} \)
\[
H \left( \bar{\mu}, \bar{x}_0 \right) \leq H \left( \bar{\mu}_0, \bar{x}_0 \right) \leq H \left( \bar{\mu}_0, \bar{\mu}_0 \right) \leq 0
\]
It follows that
\[
\sum_{i=1}^{m} \mu_i \left( By + v_i, \sum_{j=1}^{n} \lambda_j^0 u_j - u_i \right) \leq 0
\]
\[
\sum_{i=1}^{m} \mu_i \left( By + v_i, u_i - \sum_{j=1}^{n} \lambda_j^0 u_j \right) \geq 0
\]
where \( \bar{x}_0 \equiv (\lambda_1^0, \cdots, \lambda_n^0) \). This is true for any choice of \( \bar{\mu} \). In particular, you could let \( \bar{\mu} \) equal 1 in the \( i \)th position and 0 elsewhere and conclude that for all \( i = 1, \cdots, n \),
\[
\left\langle By + v_i, u_i - \sum_{j=1}^{n} \lambda_j^0 u_j \right\rangle \geq 0
\]
so you let \( x = \sum_{j=1}^{n} \lambda_j^0 u_j \) and this shows that \( Ty \neq \emptyset \) because the sets \( K_{u,v} \) have the finite intersection property.

Thus \( T : K \to \mathcal{P} (K) \) and for each \( y \in K, Ty \neq \emptyset \). In fact this is true for any \( y \) but we are only considering \( y \in K \). Now \( Ty \) is clearly a closed subset of \( K \). It is also clearly convex. Is it upper semicontinuous? Let \( y_k \to y \) and consider \( Ty + B (0,r) \). Is \( Ty_k \in Ty + B (0,r) \) for all \( k \) large enough? If not, then there is a subsequence, denoted as \( z_k \in Ty_k \) which is outside this open set \( Ty + B (0,r) \). Then taking a further subsequence, still denoted as \( z_k \), it follows that \( z_k \to z \notin Ty + B (0,r) \). Now
\[
\langle By_k + v, u - z_k \rangle \geq 0 \text{ all } [u,v] \in \mathcal{G} (A)
\]
Therefore, from continuity of \( B \),
\[
\langle By + v, u - z \rangle \geq 0 \text{ all } [u,v] \in \mathcal{G} (A)
\]
which means \( z \in Ty \) contrary to the assumption that \( T \) is not upper semicontinuous. Since \( T \) is upper semicontinuous and maps to compact convex sets, it follows from Theorem 27.6.35 that \( T \) has a fixed point \( x \in Tx \). Hence there exists a solution \( x \) to
\[
\langle Bx + v, u - x \rangle \geq 0 \text{ all } [u,v] \in \mathcal{G} (A)
\]
Next suppose that \( K \) is only closed and convex but \( B \) is coercive and \( 0 \in D (A) \). Then let \( K_n \equiv \overline{B (0,n)} \cap K \) and let \( A_n \) be the restriction of \( A \) to \( \overline{B (0,n)} \). It follows that there exists \( x_n \in K_n \) such that for all \( [u,v] \in \mathcal{G} (A_n) \),
\[
\langle Bx_n + v, u - x_n \rangle \geq 0
\]
Then since \( 0 \in D (A) \), one can pick \( v_0 \in A_0 \) and obtain
\[
\langle Bx_n + v_0, -x_n \rangle \geq 0, \ (v_0, -x_n) \geq (Bx_n, x_n)
\]
from which it follows from coercivity of \( B \) that the \( x_n \) are bounded independent of \( n \). Say \( ||x_n|| < C \). Then there is a subsequence still denoted as \( x_n \) such that \( x_n \to x \in K \), thanks to the assumption that \( K \) is closed and convex. Let \( [u,v] \in \mathcal{G} (A) \). Then for all \( n \) large enough \( ||u|| < n \) and so
\[
\langle Bx_n + v, u - x_n \rangle \geq 0
\]
Then letting \( n \to \infty \) and using the continuity of \( B \),
\[
\langle Bx + v, u - x \rangle \geq 0
\]
Since \( [u,v] \) was arbitrary, this proves the lemma.
Observation 27.6.5 If you have a monotone set valued function, then its graph can always be considered a subset of the graph of a maximal monotone graph. If \( A \) is monotone, then let \( F \) be \( G(B) \) such that \( G(B) \supseteq G(A) \) and \( B \) is monotone. Partially order by set inclusion. Then let \( C \) be a maximal chain. Let \( G(\hat{A}) = \bigcup C \). If \( [x_i,y_i] \in G(\hat{A}) \), then both are in some \( B \in C \). Hence \( (y_i - y,x_1 - x_2) \geq 0 \) so monotone and must be maximal monotone because if \( \langle z - v, x - u \rangle \geq 0 \) for all \( [u,v] \in G(\hat{A}) \) and \( [x,z] \notin A \), then you could include this ordered pair and contradict maximality of the chain \( C \).

Next is an interesting theorem which comes from this lemma. It is an infinite dimensional version of the above lemma.

Theorem 27.6.6 Let \( X \) be a reflexive Banach space and let \( K \) be a closed convex subset of \( X \). Let \( A, B \) be monotone such that

1. \( D(A) \subseteq K, 0 \in D(A) \).
2. \( B \) is single valued, hemicontinuous, bounded and coercive mapping \( X \) to \( X' \).

Then there exists \( x \in K \) such that

\[
\langle Bx + v - u, x \rangle_{X',X} \geq 0 \text{ for all } [u,v] \in G(A)
\]

Before giving the proof, here is an easy lemma.

Lemma 27.6.7 Let \( E \) be finite dimensional and let \( B : E \to E' \) be monotone and hemicontinuous. Then \( B \) is continuous.

Proof: The space can be considered a finite dimensional Hilbert space \( (\mathbb{R}^n) \) and so weak and strong convergence are exactly the same. First it is desired to show that \( B \) is bounded. Suppose it is not. Then there exists \( \|x_k\|_E = 1 \) but \( \|Bx_k\|_{E'} \to \infty \). Since finite dimensional, there is a subsequence still denoted as \( x_k \) such that \( x_k \to x, \|x\|_E = 1 \).

\[
\langle Bx_k - Bx, x_k - x \rangle \geq 0
\]

Hence

\[
\frac{\langle Bx_k - Bx, x_k - x \rangle}{\|Bx_k\|_{E'}} \geq 0
\]

Then taking another subsequence, written with index \( k \), it can be assumed that

\[
Bx_k/\|Bx_k\| \to y^* \in E', \|y^*\|_{E'} = 1
\]

Hence,

\[
\langle y^*, x_k - x \rangle \geq 0
\]

for all \( x \in E \), but this requires that \( y^* = 0 \), a contradiction. Thus \( B \) is monotone, hemicontinuous, and bounded. It follows from Theorem 27.1.5 which says that monotone and hemicontinuous operators are pseudomonotone and Proposition 27.1.6 which says that bounded pseudomonotone operators are demicontinuous that \( B \) is demicontinuous, hence continuous because, as just noted above, weak and strong convergence are the same for finite dimensional spaces. In case \( B \) is bounded, then this follows from Proposition 27.1.6 above. It is pseudomonotone and bounded hence demicontinuous and weak and strong convergence is the same in finite dimensions.

Proof of Theorem 27.6.6. Let \( \{X_n\} \) be an increasing sequence of finite dimensional subspaces. Let \( \hat{A} \) be maximal monotone on \( \cup_n X_n \) and extending \( A \). By this is meant that the graph of \( \hat{A} \) contains the graph of \( A \) restricted to \( \cup_n X_n \). \( \hat{A} \) is monotone and there is no other larger graph with these properties. See the above observation. Let \( j_n : X_n \to X \) be the inclusion map and \( j_n^*: X' \to X_n' \) be the dual map. Then \( j_n j_n^* \hat{A} j_n \equiv A_n \) and \( j_n^* B j_n \equiv B_n \) have monotone graphs from \( X_n \) to \( \mathcal{P}(X_n) \) with \( B_n \) being continuous and single valued. This follows from the hemicontinuity and the above lemma which states that on finite dimensional spaces, hemicontinuity and monotonicity imply continuity. Then

\[
[u,v] \in G(A_n)
\]

means

\[
u \in D(A) \cap X_n \text{ and } v \in j_n^* \hat{A} j_n(u) = j_n^* \hat{A}(u) \text{ since } u \in X_n
\]
Then from Lemma [7.13.4], there exists \( x_n \in X_n \) such that
\[
(B_n x_n + v_n, u_n - x_n)_{X', X} \geq 0 \quad \text{for all } [u_n, v_n] \in \mathcal{G}(A_n)
\]

That is, there exists \( x_n \in K \cap X_n \) such that for all \( u \in D(\hat{A}) \cap X_n \), \( [u, v] \in \mathcal{G}(\hat{A}) \)
\[
(B_n x_n + v, u - x_n)_{X', X} \geq 0 \tag{27.6.36}
\]

Then
\[
\langle v, u - x_n \rangle \geq \langle B_n x_n, x_n - u \rangle \tag{27.6.37}
\]

From the assumption that \( 0 \in D(\hat{A}) \), one can let \( u = 0 \) and then pick \( v_0 \in \hat{A}0 \). Then the above reduces to
\[
\langle v_0, -x_n \rangle \geq \langle B_n x_n, x_n \rangle
\]

By coercivity of \( B \), these \( x_n \) are all bounded and so by the Eberli Smulian theorem, there is a subsequence \( \{x_n\} \)
which satisfies
\[
x_n \to x \text{ weakly in } X
\]
\[
Bx_n \to y \text{ weakly in } X'
\]

Then from [7.13.4]
\[
\langle v, u - x_n \rangle + \langle B_n x_n, u \rangle \geq \langle B_n x_n, x_n \rangle
\]

Then it follows that
\[
\langle v, u - x_n \rangle + \langle B_n x_n, u \rangle - \langle B_n x_n, x \rangle \geq \langle B_n x_n, x_n - x \rangle
\]

It follows that
\[
\lim_{n \to \infty} \sup_{x_n} \langle B_n x_n, x_n - x \rangle \leq \langle v, u - x \rangle + \langle y, u \rangle - \langle y, x \rangle
\]
\[
= \langle v + y, u - x \rangle
\]

**Claim:** \( \lim_{n \to \infty} \sup_{x_n} \langle B_n x_n, x_n - x \rangle \leq 0 \).

**Proof of claim:** This is so if \( \langle v + y, u - x \rangle \leq 0 \) for some \( [u, v] \in \mathcal{G}(\hat{A}) \). If \( \langle v + y, u - x \rangle \) is greater than 0 for all \([u, v]\), then since \( \hat{A} \) is maximal, it would follow that \( -y \in \hat{A}x \). Now consider [7.13.5.4]
\[
\langle v, u - x \rangle \geq \lim_{n \to \infty} \sup_{x_n} \langle B_n x_n, x_n \rangle - \langle y, u \rangle
\]

Since \( x \in D(\hat{A}) \), you could put in \( u = x \) in the above and obtain
\[
0 \geq \lim_{n \to \infty} \sup_{x_n} \langle B_n x_n, x_n \rangle - \langle y, x \rangle = \lim_{n \to \infty} \langle B_n x_n, x_n - x \rangle
\]

which shows the claim is true.

Since \( B \) is monotone and hemicontinuous, it satisfies the pseudomonotone condition, Theorem [7.13.5.2]. Hence for any \( z \),
\[
\langle y, x - z \rangle \geq \lim_{n \to \infty} \sup_{x_n} \langle B_n x_n, x_n - x \rangle + \lim_{n \to \infty} \sup_{x_n} \langle B_n x_n, x - z \rangle
\]
\[
\geq \lim_{n \to \infty} \sup_{x_n} \left( \langle B_n x_n, x_n - x \rangle + \langle B_n x_n, x - z \rangle \right)
\]
\[
\geq \lim_{n \to \infty} \inf_{x_n} \left( \langle B_n x_n, x_n - z \rangle \right) \geq \langle Bx, x - z \rangle
\]

Since \( z \) is arbitrary, this shows that \( y = Bx \). It follows from [7.13.36] that for any \([u, v] \in \mathcal{G}(\hat{A})\),
\[
\langle B_n x_n + v, u - x_n \rangle = \langle B_n x_n + v, u - x \rangle + \langle B_n x_n + v, x - x_n \rangle \geq 0
\]
\[
\langle B_n x_n + v, u - x \rangle \geq \langle B_n x_n, x_n - x \rangle \geq \langle Bx, x_n - x \rangle
\]
Now take a limit of both sides and use the fact that $y = Bx$ to obtain
\[ \langle Bx + v, u - x \rangle \geq 0 \]
for all $[u, v] \in \mathcal{G}(\hat{A})$. Here $\hat{A}$ extends $A$ on $\bigcup_n X_n$. Why does it follow from this that there exists an $x$ such that the inequality holds for all $[u, v] \in \mathcal{G}(A)$?

Let $V$ be a finite dimensional subspace.

$$K_V \equiv \{ x \in K : \langle Bx + v, u - x \rangle_{X', X} \geq 0 \text{ for all } [u, v] \in \mathcal{G}(A), u \in V \}.$$

Then from the above argument, $K_V \neq \emptyset$. You just choose your subspaces $X_n$ to all include $V$. Also, from coercivity of $B$ and the above argument, these $K_V$ are all bounded and weakly closed. Hence they are weakly compact. Then if you have finitely many of them, you can let your subspaces include each $V$ and conclude that these $K_V$ have finite intersection property and so there exists $x \in \bigcap V K_V$ which gives the desired $x$.

Note that there is only one place where $0 \in D(A)$ was used and it was to get the estimate. In the argument,
\[ \langle v, u - x_n \rangle \geq \langle Bx_n, x_n - u \rangle \]
and it was convenient to be able to take $u = 0$. However, you could also assume other things on $B$ such as that it satisfies an estimate of the form
\[ \|Bx\| \leq C\|x\| + C \]
and if you did this, you could also obtain the necessary estimate as follows.

\[ \langle v, u - x_n \rangle \geq \langle Bx_n, x_n - u \rangle \]
\[ \langle v, u - x_n \rangle + \langle Bx_n, u \rangle \geq \langle Bx_n, x_n \rangle \]
\[ \|v\|(\|u\| + \|x_n\|) + (\|Bx_n\| + C\|u\|) \geq \langle Bx_n, x_n \rangle \]
and then pick some $[u, v]$. Thus the following corollary comes right away. This would have worked just as well if you had an estimate of the form
\[ \|Bx\| \leq C\|x\|^{p-1} + C, \ p > 1 \]

**Corollary 27.6.8** Let $X$ be a reflexive Banach space and let $K$ be a closed convex subset of $X$. Let $A, B$ be monotone such that

1. $D(A) \subseteq K$
2. $B$ is single valued, hemicontinuous, bounded and coercive mapping $X$ to $X'$ which satisfies the estimate
   \[ \|Bx\| \leq C\|x\| + C \text{ or more generally } \|Bx\| \leq C\|x\|^{p-1} + C, \ p > 1 \]

Then there exists $x \in K$ such that
\[ \langle Bx + v, u - x \rangle_{X', X} \geq 0 \text{ for all } [u, v] \in \mathcal{G}(A) \]

Now here is the equivalence between maximal monotone graph and having $F + A$ be onto. It was already shown that if $\lambda F + A$ is onto, then the graph of $A$ is maximal monotone in the sense that there is no monotone operator whose graph properly contains the graph of $A$. This was Theorem 27.6.2 above which is stated here as a reminder of what it said.

**Theorem 27.6.9** Let $X, X'$ be reflexive and have strictly convex norms. Let $A$ be a monotone set valued map as just described. Then if
\[ \lambda F + A \text{ onto,} \]
for some $\lambda > 0$, then whenever
\[ \langle y - z, x - u \rangle \geq 0 \text{ for all } [x, y] \in \mathcal{G}(A) \]
it follows that $z \in Au$ and $u \in D(A)$. That is, the graph is maximal.
\textbf{Theorem 27.6.10} Let $X$ be a strictly convex reflexive Banach space. Suppose the graph of $A : X \to P(X)$ is maximal monotone in the sense that it is monotone and no monotone graph can properly contain the graph of $A$. Then for all $\lambda > 0, \lambda F + A$ is onto. Conversely, if for some $\lambda > 0, \lambda F + A$ is onto, then the graph of $A$ is maximal with respect to being monotone.

\textbf{Proof:} In Theorem 27.6.9, let $Bx \equiv \lambda F(x) - y_0$. Then from the properties of the duality map, Theorem 27.6.9 above, it follows that $B$ satisfies the necessary conditions to use the result of Corollary 27.6.8 with $K = X$. This $B$ is monotone hemicontinuous, and coercive. Thus there exists $x$ such that for all $|u, v| \in G(A)$,

$$
\langle \lambda F(x) - y_0 + v, u - x \rangle_{X', X} \geq 0 \\
\langle v - (y_0 - \lambda F(x)), u - x \rangle_{X', X} \geq 0
$$

By maximality of the graph, it follows that $x \in D(A)$ and

$$y_0 - \lambda F(x) \in A(x), \quad y_0 = \lambda F(x) + A(x)$$

so $\lambda F + A$ is onto as claimed. The converse was proved in Theorem 27.6.2.}

Note that this theorem holds if $F$ is a duality map for $p > 1$. That is, $\langle Fx, x \rangle = \|x\|^p, \|Fx\| = \|x\|^{p-1}$.

\textbf{27.6.2 Surjectivity Theorems}

As an interesting example of this theorem, here is another result in Barbu [11]. It is interesting because it is not assumed $B$ is bounded.

\textbf{Theorem 27.6.11} Let $B : X \to X'$ be monotone hemicontinuous. Then $B$ is maximal monotone. If $B$ is coercive, then $B$ is also onto. Here $X$ is a strictly convex reflexive Banach space.

\textbf{Proof:} Suppose $B$ is not maximal monotone. Then there exists $(x_0, x_0^*) \in X \times X'$ such that for all $x$,

$$
\langle Bx - x_0^*, x - x_0 \rangle \geq 0
$$

and yet $x_0^* \neq Bx_0$. This is going to be a contradiction. Let $u \in X$ and consider $x_t \equiv tx_0 + (1 - t)u, t \in (0, 1)$. Then consider

$$
\langle Bx_t - x_0^*, x_t - x_0 \rangle
$$

However, $x_t - x_0 = tx_0 + (1 - t)u - x_0 = (1 - t)(u - x_0)$ and so, for each $t \in (0, 1)$,

$$
0 \leq \langle Bx_t - x_0^*, x_t - x_0 \rangle = (1 - t)\langle Bx_t - x_0^*, u - x_0 \rangle
$$

Divide by $(1 - t)$ and then let $t \uparrow 1$. This yields the following by hemicontinuity.

$$
\langle Bx_0 - x_0^*, u - x_0 \rangle \geq 0
$$

which holds for all $u$. Hence $Bx_0 = x_0^*$ after all. Thus $B$ is indeed maximal monotone.

Next suppose $B$ is coercive. Let $F$ be the duality map (or the duality map for arbitrary $p > 1$). Then from Theorem 27.6.9 there exists a solution $x_\lambda$ to

$$
\lambda F x_\lambda + B x_\lambda = x_0^* \in X'
$$

(27.6.38)

Then the $x_\lambda$ are bounded because, doing both sides to $x_\lambda$,

$$
\lambda \|x_\lambda\|^2 + \langle Bx_\lambda, x_\lambda \rangle = \langle x_0^*, x_\lambda \rangle
$$

and so

$$
\frac{\langle Bx_\lambda, x_\lambda \rangle}{\|x_\lambda\|} \leq \|x_0^*\|
$$

Thus the coercivity of $B$ implies that the $x_\lambda$ are bounded. There exists a subsequence such that

$$x_\lambda \to x \text{ weakly.}
$$

Then from the equation 27.6.38 $\|\lambda F x_\lambda\| = \lambda \|x_\lambda\|$ and so,

$$Bx_\lambda \to x_0^* \text{ strongly.}$$
Since $B$ is monotone and hemicontinuous, it satisfies the pseudomonotone condition, Theorem 27.6.13. The above strong convergence implies
\[
\lim_{\lambda \to 0} (Bx_\lambda, x_\lambda - x) = 0
\]
Hence for all $y$,
\[
\liminf_{\lambda \to 0} (Bx_\lambda, x_\lambda - y) = \liminf_{\lambda \to 0} (Bx_\lambda, x - y) = (x_0^*, x - y) \geq (Bx, x - y)
\]
Since $y$ is arbitrary, this shows that $x_0^* = Bx$ and so $B$ is onto as claimed. ■

Again, note that it really didn’t matter about the particular duality map used, although the usual one was featured in the argument.

There are some more things which can be said about maximal monotone operators. To include some of these, here is a very interesting lemma found in 

**Lemma 27.6.12** Let $X$ be a Banach space and suppose that
\[
x_n \to 0, \quad \|x_n^*\| \to \infty
\]
Then denoting by $D_r$ the closed disk centered at 0 with radius $r$. It follows that for every $D_r$, there exists $y_0 \in D_r$ and a subsequence with index $n_k$ such that
\[
\langle x_{n_k}^*, x_{n_k} - y_0 \rangle \to -\infty
\]
**Proof:** Suppose this is not true. Then there exists $D_r$ which has the property that for all $u \in D_r$, 
\[
\langle x_{n_k}^*, x_n - u \rangle \geq C u
\]
for all $n$. Now let
\[
E_k \equiv \{ y \in D_r : \langle x_{n_k}^*, x_n - y \rangle \geq -k \text{ for all } n \}
\]
Then this is a closed set, being the intersection of closed sets. Also, by assumption, the union of these $E_k$ equals $D_r$ which is a complete metric space. Hence one of these $E_k$ must have nonempty interior by the Bair category theorem, say for $k_0$. Say $B(y, \varepsilon) \subseteq D_r$. Then for all $\|u - y\| < \varepsilon$,
\[
\langle x_{n_k}^*, x_n - u \rangle \geq -k_0 \text{ for all } n
\]
Of course $-y \in D_r$ also, and so there is $C$ such that
\[
\langle x_{n_k}^*, x_n + y \rangle \geq C \text{ for all } n
\]
Then
\[
\langle x_{n_k}^*, 2x_n + y - u \rangle \geq C - k_0 \text{ for all } n
\]
whenever $\|y - u\| < \varepsilon$. Now recall that $x_n \to 0$. Consider only $u$ such that $\|y - u\| < \varepsilon/2$. Therefore, for all $n$ large enough, the expression $2x_n + y - u$ for such $u$ contains a small ball centered at the origin, say $D_{\delta}$. (The set of all $y - u$ for $u$ closer to $y$ than $\varepsilon/2$ is the ball $B(0, \varepsilon/2)$ and then the $2x_n$ does not move it by much provided $n$ is large enough.) Therefore,
\[
\langle x_{n_k}^*, v \rangle \geq C - k_0
\]
for all $\|v\| \leq \delta$. This contradicts the assumption that $\|x_{n_k}^*\| \to \infty$. ■

**Corollary 27.6.13** Let $X$ be a Banach space and suppose that
\[
x_n \to x, \quad \|x_n^*\| \to \infty
\]
Then denoting by $D_r$ the closed disk centered at $x$ with radius $r$. It follows that for every $D_r$, there exists $y_0 \in D_r$ and a subsequence with index $n_k$ such that
\[
\langle x_{n_k}^*, x_{n_k} - y_0 \rangle \to -\infty
\]
**Proof:** It follows that $x_n - x \to 0$. Therefore, from Lemma 27.6.12, for every $r > 0$, there exists $\hat{y}_0 \in \overline{B(0, r)}$ and a subsequence $x_{n_k}$ such that
\[
\langle x_{n_k}^*, (x_{n_k} - x) - \hat{y}_0 \rangle \to -\infty
\]
Thus
\[
\langle x_{n_k}^*, x_{n_k} - (x + \hat{y}_0) \rangle \to -\infty
\]
Just let $y_0 = x + \hat{y}_0$. Then $y_0 \in D_r$ and satisfies the desired conditions. ■
Definition 27.6.14 A set valued mapping \( A : D (A) \to \mathcal{P} (X) \) is locally bounded at \( x \in \overline{D (A)} \) if whenever \( x_n \to x \), \( x_n \in D (A) \) it follows that

\[
\limsup_{n \to \infty} \{ \| x_n^* \| : x_n^* \in Ax_n \} < \infty.
\]

Lemma 27.6.15 A set valued operator \( A \) is locally bounded at \( x \in \overline{D (A)} \) if and only if there exists \( r > 0 \) such that \( A \) is bounded on \( B(x, r) \cap D (A) \).

Proof: Say the limit condition holds. Then if no such \( r \) exists, it follows that \( A \) is unbounded on every \( B(x, r) \cap D (A) \). Hence, you can let \( r_n \to 0 \) and pick \( x_n \in B(x, r_n) \cap D (A) \) with \( x_n^* \in Ax_n \) such that \( \| x_n^* \| > n \), violating the limit condition. Hence some \( r \) exists such that \( A \) is bounded on \( B(x, r) \cap D (A) \). Conversely, suppose \( A \) is bounded on \( B(x, r) \cap D (A) \) by \( M \). Then if \( x_n \to x \), it follows that for all \( n \) large enough, \( x_n \in B(x, r) \) and so if \( x_n^* \in Ax_n \), \( \| x_n^* \| \leq M \). Hence \( \limsup_{n \to \infty} \{ \| x_n^* \| : x_n^* \in Ax_n \} \leq M < \infty \) which verifies the limit condition. \( \blacksquare \)

With this definition, here is a very interesting result.

Theorem 27.6.16 Let \( A : D (A) \to X' \) be maximal monotone. Then if \( x \) is an interior point of \( D (A) \), it follows that \( A \) is locally bounded at \( x \).

Proof: You could use Corollary 27.6.13 If \( x \) is an interior point of \( D (A) \), and \( A \) is not locally bounded, then there exists \( x_n \to x \) and \( x_n^* \in Ax_n \) such that \( \| x_n^* \| \to \infty \). Then by Corollary 27.6.13, there exists \( y_0 \) close to \( x \), in \( D (A) \) and a subsequence \( x_{n_k} \) such that

\[
\langle x_{n_k}^*, x_{n_k} - y_0 \rangle \to -\infty
\]

Letting \( y_0^* \in Ay_0 \),

\[
\langle x_{n_k}^* - y_0^*, x_{n_k} - y_0 \rangle \geq 0
\]

and so

\[
\langle x_{n_k}^*, x_{n_k} - y_0 \rangle \geq \langle y_0^*, x_{n_k} - y_0 \rangle
\]

and the right side is bounded below because it converges to \( \langle y_0^*, x - y_0 \rangle \) and this is a contradiction. \( \blacksquare \)

Does the same proof work if \( x \) is a limit point of \( D (A) \)? No. Suppose \( x \) is a limit point of \( D (A) \). If \( A \) is not locally bounded, then there exists \( x_n \to x \), \( x_n \in D (A) \) and \( x_n^* \in Ax_n \) and \( \| x_n^* \| \to \infty \). Then there is \( y_0 \) close to \( x \) such that \( \langle x_{n_k}^*, x_{n_k} - y_0 \rangle \to -\infty \) but now everything crashes in flames because it is not known that \( y_0 \in D (A) \).

It follows from the above theorem that if \( A \) is defined on all of \( X \) and is maximal monotone, then it is locally bounded everywhere. Recall the definition of a pseudomonotone operator.

Definition 27.6.17 A set valued operator \( B \) is quasi-bounded if whenever \( x \in D (B) \) and \( x^* \in Bx \) are such that

\[
\| x^* \|, \| x \| \leq M,
\]

it follows that \( \| x^* \| \leq K_M \). Bounded would mean that if \( \| x \| \leq M \), then \( \| x^* \| \leq K_M \). Here you only know this if there is another condition.

By Proposition 27.6.15 an example of a quasi-bounded operator is a maximal monotone operator \( G \) for which \( 0 \in \text{int} (D (G)) \).

Then there is a useful result which gives examples of quasi-bounded operators [20].

Proposition 27.6.18 Let \( A : D (A) \subseteq X \to \mathcal{P} (X') \) be maximal monotone and suppose \( 0 \in \text{int} (D (A)) \). Then \( A \) is quasi-bounded.

Proof: From local boundedness, Theorem 27.6.10, there exists \( \delta, C > 0 \) such that

\[
\sup \{ \| x^* \| : x^* \in A (x) \text{ for } \| x \| \leq \delta \} < C
\]

Now suppose that \( \| x \|, \| x^* \| \leq M \). Then letting \( \| y \| \leq \delta, y^* \in Ay \),

\[
0 \leq (x^* - y^*, x - y) = \langle x^*, x \rangle - \langle x^*, y \rangle - \langle y^*, x \rangle + \langle y^*, y \rangle
\]

and so for \( \| y \| \leq \delta \),

\[
\langle x^*, y \rangle \leq \langle x^*, x \rangle - \langle y^*, x \rangle + \langle y^*, y \rangle \leq M + MC + C\delta
\]

Hence, \( \| x^* \| \leq M + MC + C\delta = K_M \). \( \blacksquare \)

This is actually quite a restrictive requirement and leaves out a lot which would be interesting.
Definition 27.6.19 Let $V$ be a reflexive Banach space. We say $T : V \to \mathcal{P}(V')$ is pseudomonotone if the following conditions hold.

$$Tu \text{ is closed, nonempty, convex.} \quad (27.6.39)$$

If $F$ is a finite dimensional subspace of $V$, then if $u \in F$ and $W \supseteq Tu$ for $W$ a weakly open set in $V'$, then there exists $\delta > 0$ such that

$$v \in B(u, \delta) \cap F \implies Tv \subseteq W. \quad (27.6.40)$$

If $u_k \to u$ and if $u_k^* \in Tu_k$ is such that

$$\limsup_{k \to \infty} u_k^*(u_k - u) \leq 0,$$

then for all $v \in V$, there exists $u^*(v) \in Tu$ such that

$$\lim_{k \to \infty} \inf u_k^*(u_k - v) \geq u^*(v)(u - v). \quad (27.6.41)$$

Then there is an interesting result [28].

Theorem 27.6.20 Suppose $A : X \to \mathcal{P}(X')$ is maximal monotone. That is, $D(A) = X$. Then $A$ is pseudomonotone.

Proof: Consider the first condition. Say $x_1^* \in Ax$. Let $u^* \in Au$. For $\lambda \in [0,1]$,

$$\langle \lambda x_1^* + (1 - \lambda) x_2^* - u^*, x - u \rangle = \lambda \langle x_1^* - u^*, x - u \rangle + (1 - \lambda) \langle x_2^* - u^*, x - u \rangle \geq 0$$

and so, since $[u, u^*]$ is arbitrary, it follows that $\lambda x_1^* + (1 - \lambda) x_2^* \in Ax$. Thus $Ax$ is convex. Is it closed? Say $x_n^* \in Ax$ and $x_n^* \to x^*$. Is it the case that $x^* \in D(A)$? Let $[u, u^*] \in G(A)$ be arbitrary. Then

$$\langle x^* - u^*, x - u \rangle = \lim_{n \to \infty} \langle x_n^* - u^*, x_n - u \rangle \geq 0$$

and so $Ax$ is also closed.

Consider the second condition. It is to show that if $x_n \to x$ in $V$ a finite dimensional subspace and if $U$ is a weakly open set containing 0, then eventually $Ax_n \subseteq Ax + U$. Suppose then that this is not the case. Then there exists $x_n^*$ outside of $Ax + U$ but in $Ax_n$. Since $A$ is locally bounded at $x$, it follows that the $\|x_n^*\|$ are bounded. Thus there is a subsequence, still denoted as $x_n$ and $x_n^*$ such that $x_n^* \to x^*$ weakly and $x^* \notin Ax + U$. Now let $[u, u^*] \in G(A)$.

$$\langle x^* - u^*, x - u \rangle = \lim_{n \to \infty} \langle x_n^* - u^*, x_n - u \rangle \geq 0$$

and since $[u, u^*]$ is arbitrary, it follows that $x^* \in Ax$ and so is inside $Ax + U$. Thus the second condition holds also.

Consider the third. Say $x_k \to x$ weakly and letting $x_k^* \in Ax_k$, suppose

$$\limsup_{k \to \infty} \langle x_k^*, x_k - x \rangle \leq 0,$$

Is it the case that there exists $x^*(y) \in Ax$ such that

$$\liminf_{k \to \infty} \langle x_k^*, x_k - y \rangle \geq \langle x^*(y), x - y \rangle?$$

The proof goes just like it did earlier in the case of single valued pseudomonotone operators. It is just a little more complicated. First, let $x^* \in Ax$.

$$\langle x_k^* - x^*, x_k - x \rangle \geq 0$$

and so

$$\lim_{k \to \infty} \langle x_k^*, x_k - x \rangle \geq \liminf_{k \to \infty} \langle x^*, x_k - x \rangle = 0 \geq \limsup_{k \to \infty} \langle x_k^*, x_k - x \rangle$$

Thus

$$\lim_{k \to \infty} \langle x_k^*, x_k - x \rangle = 0.$$

Now let $x_t^* \in A(x + t(y - x))$, $t \in (0,1)$, where here $y$ is arbitrary. Then

$$\langle x_n^* - x_t^*, x_n - x + t(x - y) \rangle \geq 0$$
Hence
\[
\liminf_{n \to \infty} (x_n^*, x_n - x + t(x - y)) \geq \liminf_{n \to \infty} (x_n^*, x_n - x + t(x - y))
\]
and so from the above limit,
\[
t \liminf_{n \to \infty} (x_n^*, x - y) \geq t (x_n^*, x - y)
\]
 Cancel the t.
\[
\liminf_{n \to \infty} (x_n^*, x - y) = \liminf_{n \to \infty} (x_n^*, x_n - y) \geq (x_t^*, x - y)
\]
Now you have a fixed y and \(x_t^* \in A(x + t(y - x))\). The subspace determined by \(x, y\) is finite dimensional. Also it was shown above that \(A\) is locally bounded at \(x\) and so there is a subsequence, still denoted as \(x_t^*\) such that \(x_t^* \to x^*(y)\) weakly. Now from the upper semicontinuity on finite dimensional spaces shown above, for every \(S\) a finite subset of X and \(\varepsilon > 0\), it follows that for all \(t\) small enough,
\[
x_t^* \in Ax + BS(0, \varepsilon)
\]
Thus \(x^*(y) \in Ax\). Hence, there exists \(x^*(y) \in Ax\) such that
\[
\liminf_{n \to \infty} (x_n^*, x_n - y) \geq (x^*(y), x - y)
\]
Suppose \(T\) is a bounded pseudomonotone operator and \(S\) is a maximal monotone operator, both defined on a strictly convex reflexive Banach space. What of their sum? Is \((T + S)(x)\) convex and closed? Say \(t_i \in Tx\) and \(s_i \in Sx\) is it the case that \(\theta (s_i + t_i) + (1 - \theta)(s_2 + t_2) \in (T + S)(x)\) whenever \(\theta \in [0, 1]\)? Of course this is so. Thus \(T + S\) has convex values. Does it have closed values? Suppose \(\{s_n + t_n\}\) converges to \(z \in X', s_n \in Sx, t_n \in Tx\). Is \(z \in (T + S)(x)\)? Taking a subsequence, and using the assumption that \(T\) is bounded, it can be assumed that \(t_n \to t \in Tx\) weakly. Therefore, \(s_n\) must also converge weakly and so it converges to some \(s = z - t \in Sx\). Convex and closed implies weakly closed. Thus \(T + S\) has closed convex values. Is it upper semicontinuous on finite dimensional subspaces? Suppose \(x_n \to x\) in a finite dimensional subspace \(F\) does it follow that
\[
(T + S) x_n \subseteq (T + S) x + B(0, r)
\]
for all \(n\) sufficiently large? It is known that \(Sx_n \subseteq Sx + B(0, r/2)\) and \(Tx_n \subseteq Tx + B(0, r/2)\) whenever \(n\) is sufficiently large and so it follows that
\[
(T + S) x_n \subseteq (T + S) x + B(0, r/2) + B(0, r/2) \subseteq (T + S) x + B(0, r)
\]
whenever \(n\) is large enough.
What of the pseudomonotone condition? Suppose
\[
\limsup_{n \to \infty} \langle u_n^* + v_n^*, x_n - x \rangle \leq 0
\]
where \(u_n^* \in Sx_n\) and \(v_n^* \in Tx_n\) where \(x_n \to x\) weakly. Is it the case that for every \(y\), there exists \(u^* \in Sx\) and \(v^* \in Tx\) such that
\[
\liminf_{n \to \infty} \langle u_n^* + v_n^*, x_n - y \rangle \geq \langle u^* + v^*, x - y \rangle?
\]
By monotonicity,
\[
0 \geq \limsup_{n \to \infty} \langle u_n^* + v_n^*, x_n - x \rangle \geq \limsup_{n \to \infty} \langle u_n^* + v_n^*, x_n - x \rangle
\]
\[
= \limsup_{n \to \infty} \langle v_n^*, x_n - x \rangle
\]
Hence
\[
\limsup_{n \to \infty} \langle v_n^*, x_n - x \rangle \leq 0
\]
which implies
\[
\liminf_{n \to \infty} \langle v_n^*, x_n - x \rangle \geq \langle v^*, x - x \rangle = 0 \geq \limsup_{n \to \infty} \langle v_n^*, x_n - x \rangle
\]
showing that
\[
\lim_{n \to \infty} \langle v_n^*, x_n - x \rangle = 0
\] (27.6.42)
It follows that if \( y \) is given, there exists \( v^* \in T(x) \) such that
\[
\lim_{n \to \infty} \inf \langle v^*_n, x_n - y \rangle \geq \langle v^*, x - y \rangle
\]
Now let \( u^*_t \in S(x + t(y - x)) \) for \( t > 0 \). Thus
\[
\langle u^*_n - u^*_t, x_n - x + t(x - y) \rangle \geq 0
\]
\[
\langle u^*_n, x_n - x + t(x - y) \rangle \geq \langle u^*_t, x_n - x + t(x - y) \rangle
\]
Then using the above and the convergence in \( X \),
\[
\lim_{n \to \infty} \inf \langle u^*_n + v^*_n, x_n - y \rangle \geq \lim_{n \to \infty} \inf \langle u^*_t + v^*_n, x_n - y \rangle
\]
\[
= \langle u^*_t, x - y \rangle + \langle v^*, x - y \rangle
\]
Now as before where it was shown that maximal monotone and defined on \( X \) implied pseudomonotone, and the theorem which says that maximal monotone operators are locally bounded on the interior of their domains, it follows that there exists a sequence, still denoted as \( u^*_n \) which converges to something called \( u^* \). Then as before, the subspace spanned by \( x, y \) is finite dimensional and so from upper semicontinuity, for all \( t \) small enough,
\[
u^*_t \in S(x) + B(0, r)
\]
Note that weak convergence is the same as strong on finite dimensional spaces. Since this is true for all \( r \) and \( S(x) \) is closed, it follows that \( u^* \in S(x) \). Thus, passing to a limit as \( t \to 0 \) one gets \( u^* \in S(x) \), \( v^* \in T(x) \), and
\[
\lim_{n \to \infty} \inf \langle u^*_n + v^*_n, x_n - y \rangle \geq \langle u^* + v^*, x - y \rangle
\]
This proves the following generalization of Theorem 27.6.21.

**Theorem 27.6.21** Let \( T, S : X \to \mathcal{P}(X') \) where \( X \) is a strictly convex reflexive Banach space and suppose \( T \) is bounded and pseudomonotone while \( S \) is maximal monotone. Then \( T + S \) is pseudomonotone.

Also, there is an interesting result which is based on the obvious observation that if \( A \) is maximal monotone, then so is \( \hat{A}(x) \equiv A(x_0 + x) \).

**Lemma 27.6.22** Let \( A \) be maximal monotone. Then for each \( \lambda > 0 \),
\[
x \to \lambda F(x - x_0) + Ax
\]
is onto.

**Proof:** Let \( \hat{A}(x) \equiv A(x_0 + x) \) so as earlier, \( \hat{A} \) is maximal monotone. Then let \( y^* \in X' \). Then there exists \( y \) such that
\[
\hat{A}(y) + \lambda F(y) \ni y^*. \quad \text{Now define } x \equiv y + x_0, \text{ Then}
\]
\[
\hat{A}(y) + \lambda F(y) \ni y^*, \quad \hat{A}(x - x_0) + \lambda F(x - x_0) \ni y^*, \quad A(x) + \lambda F(x - x_0) \ni y^* \square
\]

**Definition 27.6.23** Let \( A : D(A) \to \mathcal{P}(X') \) be maximal monotone. Let \( A^{-1} : A(D(A)) \to \mathcal{P}(X') \) be defined as follows.
\[
x \in A^{-1} x^* \text{ if and only if } x^* \in Ax
\]

**Observation 27.6.24** \( A^{-1} \) is also maximal monotone. This is easily seen as follows. \([x, y] \in \mathcal{G}(A) \) if and only if \([y, x] \in \mathcal{G}(A^{-1}) \).

Earlier, it was shown that if \( B \) is monotone and hemicontinuous and coercive, then it was onto. It was not necessary to assume that \( B \) is bounded. The same thing holds for \( A \) maximal monotone. This will follow from the next result. Recall that a maximal monotone operator is locally bounded at every interior point of its domain which was shown above. Also it appears to not be possible to show that a maximal monotone operator is locally bounded at a limit point of \( D(A) \). The following result is in \( \square \) although he claims a better result than what I am proving here in which it is only necessary to verify \( A^{-1} \) is locally bounded at every point of \( A(D(A)) \). However, I was unable to follow the argument and so I am proving another theorem with the same argument he uses. It looks like a typo to me but I often have trouble following hard theorems so I am not sure. Anyway, the following is the best I can do. I think it is still a very interesting result.
Theorem 27.6.25 Suppose $A^{-1}$ is locally bounded at every point of $A(D(A))$. Then in fact $A(D(A)) = X'$ and in fact $A(D(A)) = A(D(A))$.

Proof: This is done by showing that $A(D(A))$ is both open and closed. Since it is nonempty, it must be all of $X'$ because $X'$ is connected. First it is shown that $A(D(A))$ is closed. Suppose $y_n \in Ax_n$ and $y_n \to y$. Does it follow that $y \in A(D(A))$? Since $y$ is a limit point of $A(D(A))$, it follows that $A^{-1}$ is locally bounded at $y$. Thus there is a subsequence still denoted by $y_n$ such that $y_n \to y$ and for $x_n \in A^{-1}y_n$ or in other words, $y_n \in Ax_n$, it follows that $x_n$ is bounded. Hence there exists a subsequence, still denoted with the subscript $n$ such that $x_n \to x$ weakly and $y_n \to y$ strongly. Hence if $[u,v] \in G(A)$,

$$\langle y - v, x - u \rangle = \lim_{n \to \infty} \langle y_n - v, x_n - x \rangle \geq 0$$

Since $[u,v]$ is arbitrary and $A$ is maximal monotone, it follows that $y \in Ax$ or in other words, $x \in A^{-1}y$ and $y \in A(D(A))$. Thus $A(D(A))$ is closed.

Next consider why $A(D(A))$ is open. Let $y_0 \in A(D(A))$. Then there exists $D_r = \overline{B(y_0,r)}$ centered at $y_0$ such that $A^{-1}$ is bounded on $D_r$. Since $A$ is maximal monotone, for each $y \in X'$ there is a solution $x_\varepsilon$ to the inclusion

$$y \in \varepsilon F(x_\varepsilon - x_0) + Ax_\varepsilon, \ y_\varepsilon \equiv y - \varepsilon F(x_\varepsilon - x_0) \in Ax_\varepsilon$$

Consider only $y \in B(y_0, \frac{r}{2})$.

$$\langle (y - \varepsilon F(x_\varepsilon - x_0)) - y_0, x_\varepsilon - x_0 \rangle \geq 0$$

Then using $\langle F\varepsilon, z \rangle = \|z\|^2$,

$$\|y - y_0\| \|x_\varepsilon - x_0\| \geq \langle y - y_0, x_\varepsilon - x_0 \rangle \geq \varepsilon \|x_\varepsilon - x_0\|^2$$

and so $\varepsilon \|x_\varepsilon - x_0\| = \varepsilon \|F(x_\varepsilon - x_0)\| \leq \|y - y_0\| < r/2$. Thus $y_\varepsilon$ stays in $B(y_0,r)$. This is because $y$ is closer to $y_0$ than $r/2$ while $y_\varepsilon$ is within $r/2$ of $y$. It follows that the $x_\varepsilon$ are bounded and so $x_\varepsilon - x_0$ is bounded and so $\varepsilon F(x_\varepsilon - x_0) \to 0$. Thus $y_\varepsilon \to y$ strongly. Since the $x_\varepsilon$ are bounded, there exists a further subsequence, still denoted as $x_\varepsilon$ such that $x_\varepsilon \to x$, some point of $X$. Then if $[u,v] \in G(A)$,

$$\langle y_\varepsilon - v, x_\varepsilon - u \rangle \geq 0$$

and letting $\varepsilon \to 0$ using the strong convergence of $y_\varepsilon$ one obtains

$$\langle y - v, x - u \rangle \geq 0$$

which shows that $y \in Ax$. Thus $B(y_0, \frac{r}{2}) \subseteq A(D(A)) \equiv D(A^{-1})$ and so $A(D(A))$ is open. ■

The proof featured the usual duality map.

Note that as part of the proof $A(D(A))$ was shown to be closed so although it was assumed at the outset that $A^{-1}$ was locally bounded on $A(D(A))$, this is the same as saying that $A^{-1}$ is locally bounded on $A(D(A))$.

Corollary 27.6.26 Suppose $A : D(A) \to P(X')$ is maximal monotone and coercive. Then $A$ is onto.

Proof: From Theorem 27.6.25 it suffices to show that $A^{-1}$ is locally bounded at $y^* \in A(D(A))$. The case of an interior point follows from Theorem 27.6.16. Assume then that $y^*$ is a limit point of $A(D(A))$. Of course this includes the case of interior points. Then there exists $y_n^* \to y^*$ where $y_n^* \in Ax_n$. Then

$$\frac{\langle y_n^*, x_n \rangle}{\|x_n\|} \leq \|y_n^*\|$$

and the right side is bounded. Hence by coercivity, so is $\|x_n\|$. Therefore, there is a further subsequence, still denoted as $x_n$ such that $x_n \to x$ weakly while $y_n^* \to y^*$ strongly. Then letting $[u,v] \in G(A)$,

$$\langle y^* - u, x - u \rangle = \lim_{n \to \infty} \langle y_n^* - v^*, x_n - u \rangle \geq 0$$

Hence $y^* \in Ax$ and $y^* \in A(D(A))$. Thus $A^{-1}$ is locally bounded on $A(D(A))$ and so $A$ is onto from the above theorem. ■
27.6.3 Approximation Theorems

This section continues following Barbu [6]. Always it is assumed that the situation is of a real reflexive Banach space \( X \) having strictly convex norm and its dual \( X' \). As observed earlier, there exists a solution \( x_\lambda \) to the inclusion

\[
0 \in F(x_\lambda - x) + \lambda A(x_\lambda)
\]

To see this, you consider \( \hat{A}(y) \equiv A(x + y) \). Then \( \hat{A} \) is also maximal monotone and so there exists a solution to

\[
0 \in F(\hat{x}) + \lambda \hat{A}(\hat{x}) = F(\hat{x}) + \lambda A(x + \hat{x})
\]

Now let \( x_\lambda = x + \hat{x} \) so \( \hat{x} = x_\lambda - x \). Hence

\[
0 \in F(x_\lambda - x) + \lambda Ax_\lambda
\]

Here you could have \( F \) the duality map for any given \( p > 1 \).

The symbol \( \limsup_{n,m \to \infty} a_{mn} \) means \( \limsup_{m,n \geq N} (\sup_{m \geq N, n \geq N} a_{mn}) \). Then here is a simple observation.

**Lemma 27.6.27** Suppose \( \limsup_{n,m \to \infty} a_{mn} \leq 0 \). Then \( \limsup_{m \to \infty} (\limsup_{n \to \infty} a_{mn}) \leq 0 \).

**Proof:** There exists \( N \) such that if both \( m, n \geq N, a_{mn} \leq \varepsilon \). Then

\[
\limsup_{n \to \infty} a_{mn} = \lim sup_{n \to \infty, n \geq N} a_{mn} \leq \varepsilon
\]

Thus also

\[
\limsup_{m \to \infty} \left( \limsup_{n \to \infty} a_{mn} \right) = \lim sup_{m \to \infty, m \geq N} \left( \limsup_{n \to \infty} a_{mn} \right) \leq \varepsilon.
\]

The argument will be based on the following lemma.

**Lemma 27.6.28** Let \( A : D(A) \to \mathcal{P}(X') \) be maximal monotone and let \( v_n \in Au_n \) and

\[
u_n \to u, \ v_n \to v \text{ weakly.}
\]

Also suppose that

\[
\limsup_{m,n \to \infty} \langle v_n - v_m, u_n - u_m \rangle \leq 0
\]

or

\[
\limsup_{n \to \infty} \langle v_n - v, u_n - u \rangle \leq 0
\]

Then \([u, v] \in \mathcal{G}(A) \) and \( \langle v_n, u_n \rangle \to \langle v, u \rangle \).

**Proof:** By monotonicity,

\[
\lim_{m,n \to \infty} \langle v_n - v_m, u_n - u_m \rangle = 0
\]

Suppose then that \( \langle v_n, u_n \rangle \) fails to converge to \( \langle v, u \rangle \). Then there is a subsequence, still denoted with subscript \( n \) such that \( \langle v_n, u_n \rangle \to \mu \neq \langle v, u \rangle \). Let \( \varepsilon > 0 \). Then there exists \( M \) such that if \( n, m > M \), then

\[
|\langle v_n, u_n \rangle - \mu| < \varepsilon, |\langle v_n - v_m, u_n - u_m \rangle| < \varepsilon
\]

Then if \( n, m > M \),

\[
|\langle v_n - v_m, u_n - u_m \rangle| = |\langle v_n, u_n \rangle + \langle v_m, u_m \rangle - \langle v_n, u_m \rangle - \langle v_m, u_n \rangle| < \varepsilon
\]

Hence it is also true that

\[
|\langle v_n, u_n \rangle + \langle v_m, u_m \rangle - \langle v_n, u_m \rangle - \langle v_m, u_n \rangle| \leq |2\mu - (\langle v_n, u_m \rangle + \langle v_m, u_n \rangle)| < 3\varepsilon
\]

Now take a limit first with respect to \( n \) and then with respect to \( m \) to obtain

\[
|2\mu - (\langle v, u \rangle + \langle v, u \rangle)| < 3\varepsilon
\]

Since \( \varepsilon \) is arbitrary, \( \mu = \langle v, u \rangle \) after all. Hence the claim that \( \langle v_n, u_n \rangle \to \langle v, u \rangle \) is verified. Next suppose \([x, y] \in \mathcal{G}(A) \) and consider

\[
\langle v - y, u - x \rangle = \langle v, u \rangle - \langle v, x \rangle - \langle y, u \rangle + \langle y, x \rangle
\]
Thus you are looking at
$$\text{Theorem 27.6.30}$$
for
$$p > 1$$
are in a single Hilbert space. This is just a generalization to mappings between Banach spaces and their duals. This is in the case where $$F$$

Also denote
$$\text{Recall how this}$$
$$x$$

This is for $$F$$

which shows that $$\lim \sup_{n \to \infty} \langle v_n - v, u_n - u \rangle \geq 0$$

and since $$[x, y]$$ is arbitrary, it follows that $$v \in Au$$.

Next suppose $$\lim \sup_{n \to \infty} \langle v_n - v, u_n - u \rangle \leq 0$$. It is not known that $$[u, v] \in G \left( A \right)$$. Formally, and to help remember what is going on, you are looking at a generalization of

Also denote $$\text{Recall how this}$$

Thus $$\lim \sup_{n \to \infty} \langle v_n, u_n \rangle \leq \langle v, u \rangle$$. Now let $$[x, y] \in G \left( A \right)$$

$$\langle v - y, u - x \rangle = \langle v, u \rangle - \langle v, x \rangle - \langle y, u \rangle + \langle y, x \rangle$$

$$\geq \lim \sup_{n \to \infty} \left( \langle v_n, u_n \rangle - \langle v_n, x \rangle - \langle y, u_n \rangle + \langle y, x \rangle \right)$$

$$\geq \lim \inf_{n \to \infty} \left( \langle v_n - y, u_n - x \rangle \right) \geq 0$$

Hence $$[u, v] \in G \left( A \right)$$. Now

$$\lim \sup_{n \to \infty} \langle v_n - v, u_n - u \rangle \leq 0 \leq \lim \inf_{n \to \infty} \langle v_n - v, u_n - u \rangle$$

the second coming from monotonicity and the fact that $$v \in Au$$. Therefore,

$$\lim_{n \to \infty} \langle v_n - v, u_n - u \rangle = 0$$

which shows that $$\lim_{n \to \infty} \langle v_n, u_n \rangle = \langle v, u \rangle$$.

\textbf{Definition 27.6.29} Let $$x_\lambda$$ just defined be denoted by $$J_\lambda x$$ and define also

$$A_\lambda \left( x \right) = -\lambda^{-1} \left( p - 1 \right) F \left( x_\lambda - x \right).$$

This is for $$F$$ a duality map with $$p > 1$$. Thus for the usual duality map, you would have

$$A_\lambda \left( x \right) = -\lambda^{-1} F \left( x_\lambda - x \right)$$

Recall how this $$x_\lambda$$ is defined. In general,

$$0 \in F \left( x_\lambda - x \right) + \lambda^{p-1} Ax_\lambda$$

Also denote $$J_\lambda x \equiv x_\lambda$$. Thus, from the definition,

$$A_\lambda \left( x \right) \in A \left( J_\lambda x \right)$$

Formally, and to help remember what is going on, you are looking at a generalization of

$$A_\lambda x = \frac{A}{1 + \lambda A} x = \frac{1}{\lambda} \left( x - \left( I + \lambda A \right)^{-1} x \right)$$

This is in the case where $$F = I$$ to keep things simpler. You have $$0 = x_\lambda - x + \lambda Ax_\lambda$$ and so formally $$x_\lambda = \left( I + \lambda A \right)^{-1} x$$. Thus you are looking at $$\frac{1}{\lambda} \left( x - x_\lambda \right) = \frac{1}{\lambda} \left( x - \left( I + \lambda A \right)^{-1} x \right) = A_\lambda x$$. In fact, this is exactly what you do when you are in a single Hilbert space. This is just a generalization to mappings between Banach spaces and their duals.

Then there are some things which can be said about these operators. It is presented for the general duality map for $$p > 1$$.

\textbf{Theorem 27.6.30} The following hold. Here $$X$$ is a reflexive Banach space with strictly convex norm. $$A : D \left( A \right) \to \mathcal{P} \left( X' \right)$$ is maximal monotone. Then

1. $$J_\lambda$$ and $$A_\lambda$$ are bounded single valued operators defined on $$X$$. Bounded means they take bounded sets to bounded sets. Also $$A_\lambda$$ is a monotone operator.
2. $A_\lambda, J_\lambda$ are demicontinuous. That is, strongly convergent sequences are mapped to weakly convergent sequences.

3. For every $x \in D(A), \|A_\lambda(x)\| \leq |Ax| \equiv \inf \{\|y^*\| : y^* \in Ax\}$. For every $x \in \text{conv}(D(A))$, it follows that $\lim_{\lambda \to 0} J_\lambda(x) = x$. The new symbol means the closure of the convex hull. It is the closure of the set of all convex combinations of points of $D(A)$.

**Proof:**

1. It is clear that these are single valued operators. What about the assertion that they are bounded? Let $y^* \in Ax_\lambda$ such that the inclusion defining $x_\lambda$ becomes an equality. Thus

$$F(x_\lambda - x) + \lambda^{p-1}y^* = 0$$

Then let $x_0 \in D(A)$ be given.

$$(F(x_\lambda - x), x_\lambda - x) + \lambda^{p-1}(y^*, x_\lambda - x_0) + \lambda^{p-1}(y^*, x_0 - x) = 0$$

Then by monotonicity of $A$,

$$\|x_\lambda - x\|^p + \lambda^{p-1}(y^*_0, x_\lambda - x_0) + \lambda^{p-1}(y^*, x_0 - x) \leq 0$$

It follows that

$$\|x_\lambda - x\|^p \leq \lambda^{p-1}\|y^*_0\|\|x_\lambda - x_0\| + \lambda^{p-1}\|y^*\|\|x_0 - x\|$$

Hence if $x$ is in a bounded set, it follows the resulting $x_\lambda = J_\lambda x$ remain in a bounded set. Now from the definition of $A_\lambda$, it follows that this is also a bounded operator.

Why is $A_\lambda$ monotone?

$$0 \leq \langle A_\lambda x - A_\lambda y, x - y \rangle = \langle A_\lambda x - A_\lambda y, x - J_\lambda x - (y - J_\lambda y) \rangle$$

$$+ \langle A_\lambda x - A_\lambda y, J_\lambda x - J_\lambda y \rangle$$

$$= \langle \lambda^{-(p-1)}F(J_\lambda x - x) - \lambda^{-(p-1)}F(J_\lambda y - y), J_\lambda x - x - (J_\lambda y - y) \rangle$$

$$+ \langle A_\lambda x - A_\lambda y, J_\lambda x - J_\lambda y \rangle$$

and both terms are nonnegative, the first because $F$ is monotone so indeed $A_\lambda$ is monotone.

2. What of the demicontinuity of $A_\lambda$? This one is really tricky. Suppose $x_n \to x$. Does it follow that $A_\lambda x_n \to A_\lambda x$ weakly? The proof will be based on a pair of equations. These are

$$\lim_{m,n \to \infty} \langle F(J_\lambda x_n - x_n) - F(J_\lambda x_m - x_m), J_\lambda x_n - x_n - (J_\lambda x_m - x_m) \rangle = 0$$

and

$$\lim_{m,n \to \infty} \langle A_\lambda(x_n) - A_\lambda(x_m), J_\lambda x_n - J_\lambda x_m \rangle = 0$$

When these have been established, Lemma 24.6.28 is used to get the desired result for a subsequence. It will be shown that every sequence has a subsequence which gives the right sort of weak convergence and from this the desired weak convergence of $A_\lambda x_n$ to $A_\lambda x$ follows.

$$0 \in F(J_\lambda x_n - x_n) + \lambda^{p-1}A(J_\lambda x_n)$$

$$0 \in F(J_\lambda x - x) + \lambda^{p-1}A(J_\lambda x)$$

$$-\lambda^{-(p-1)}F(J_\lambda x - x) \equiv A_\lambda(x) \in A(J_\lambda x)$$

$$-\lambda^{-(p-1)}F(J_\lambda x_n - x_n) \equiv A_\lambda(x_n) \in A(J_\lambda x_n)$$

Note also that for a given $x$ there is only one solution $J_\lambda x$ to $0 \in F(J_\lambda x - x) + \lambda^{p-1}A(J_\lambda x)$. By monotonicity of $F$,

$$0 \leq \langle F(J_\lambda x_n - x_n) - F(J_\lambda x_m - x_m), x_m - x_n + J_\lambda x_n - J_\lambda x_m \rangle$$

Then from the above,

$$\langle F(J_\lambda x_n - x_n) - F(J_\lambda x_m - x_m), x_n - x_m \rangle$$

$$\leq \langle F(J_\lambda x_n - x_n) - F(J_\lambda x_m - x_m), J_\lambda x_n - J_\lambda x_m \rangle$$
Now from the boundedness of these operators, the left side of the above inequality converges to 0 as \( n,m \to \infty \). Thus
\[
\lim_{n,m \to \infty} \langle F(J_\lambda x_n - x_n) - F(J_\lambda x_m - x_m), J_\lambda x_n - J_\lambda x_m \rangle \geq 0
\] (27.6.43)

\[
\lim_{n,m \to \infty} \langle -\lambda^{p-1} A_\lambda(x_n) - (-\lambda^{p-1} A_\lambda(x_m)), J_\lambda x_n - J_\lambda x_m \rangle \geq 0
\]

\[
\lim_{n,m \to \infty} \left( \lambda^{p-1} A_\lambda(x_m) - \lambda^{p-1} A_\lambda(x_n), J_\lambda x_n - J_\lambda x_m \right) \geq 0
\]

The expression on the left in the above is non positive. Multiplying by \(-1\),
\[
0 \geq \lim_{n,m \to \infty} \sup \langle A_\lambda(x_n) - A_\lambda(x_m), J_\lambda x_n - J_\lambda x_m \rangle
\]
\[
\geq \lim_{n,m \to \infty} \inf \langle A_\lambda(x_n) - A_\lambda(x_m), J_\lambda x_n - J_\lambda x_m \rangle \geq 0
\] (27.6.44)

Thus, in fact, the expression in 27.6.43 converges to 0. By boundedness considerations and the strong convergence given,
\[
\lim_{n,m \to \infty} \langle F(J_\lambda x_n - x_n) - F(J_\lambda x_m - x_m), J_\lambda x_n - x_n - (J_\lambda x_m - x_m) \rangle = 0
\] (27.6.45)

From boundedness again, there is a subsequence still denoted with the subscript \( n \) such that
\[ J_\lambda x_n \to a - x, \ F(J_\lambda x_n - x_n) \to b \] both weakly.

Since \( F \) is maximal monotone, (Theorem 27.6.43) it follows from Lemma 27.6.28 that \([a - x, b] \in \mathcal{G}(F)\) and so in fact \( F(a - x) = b \). Thus this has just shown that \( F(J_\lambda x_n - x_n) \to F(a - x) \). Next consider 27.6.43. We have \( J_\lambda x_n \to a \) weakly and \( A_\lambda(x_n) = -\lambda^{-(p-1)}F(J_\lambda x_n - x_n) \to -\lambda^{-(p-1)}b \) weakly. Then from Lemma 27.6.28 again, \([a, -\lambda^{-(p-1)}b] \in \mathcal{G}(A)\) so \(-\lambda^{-(p-1)}b \in A(a)\) so \( b \in -\lambda^{-(p-1)}A(a)\). But it was just shown that \( b = F(a - x) \) and so
\[ F(a - x) \in -\lambda^{-(p-1)}A(a) \] so \( 0 \in F(a - x) + \lambda^{-(p-1)}A(a) \), so \( a = J_\lambda x \).

As noted at the beginning, there is only one solution to this inclusion for a given \( x \) and it is \( a = J_\lambda x \). This has shown that in terms of weak convergence,
\[ A_\lambda(x_n) \to -\lambda^{-(p-1)}b = -\lambda^{-(p-1)}F(a - x) = -\lambda^{-(p-1)}F(J_\lambda x - x) \equiv A_\lambda(x) \]

This has shown that \( A_\lambda \) is demicontinuous. Also it has shown that \( J_\lambda \) is also demicontinuous. (This result is a lot nicer in Hilbert space. )

3.) Why is \( \|A_\lambda(x)\| \leq |Ax| \) whenever \( x \in D(A) \)?
\[ A_\lambda(x) = -\lambda^{-(p-1)}F(J_\lambda x - x) \]
where \( 0 \in F(J_\lambda x - x) + \lambda^{p-1}A(J_\lambda x) \). Therefore, \( A_\lambda(x) \in A(J_\lambda x) \). Then letting \([u, v] \in \mathcal{G}(A)\),
\[ 0 \leq \langle v - A_\lambda(x), u - J_\lambda x \rangle \]
In particular, if \( y \in Ax \)
\[ 0 \leq \langle y - A_\lambda(x), x - J_\lambda x \rangle = \left( y + \lambda^{-(p-1)}F(J_\lambda x - x), x - J_\lambda x \right) \]
Hence
\[ \lambda^{-(p-1)}\|J_\lambda x - x\|^p \leq \|y\| \|J_\lambda x - x\| \]
and so
\[ \lambda^{-(p-1)}\|J_\lambda x - x\|^p = \lambda^{-(p-1)}\|F(J_\lambda x - x)\| = \|A_\lambda(x)\| \leq \|y\| \]
and since \( y \in Ax \) is arbitrary, \( \|A_\lambda(x)\| \leq |Ax| \equiv \inf \{|y| : y \in Ax\} \).

Next consider the claim that for all \( x \in \text{conv}(D(A)) \), it follows that
\[ \lim_{\lambda \to 0} J_\lambda(x) = x. \]
Let \([u,v] \in G(A)\) and \(x\) is arbitrary.

\[
0 \leq \langle v - A\lambda (x), u - J\lambda x \rangle = \left( v + \lambda^{-(p-1)}F(J\lambda x - x), u - J\lambda x \right)
\]

Thus

\[
\|J\lambda x - x\|^p \leq \lambda^{p-1} \langle v, u - x \rangle + \langle F(J\lambda x - x), u - x \rangle + \lambda^{p-1} \langle v, x - J\lambda x \rangle
\] (27.6.46)

for \(x\) arbitrary and \(u\) anything in \(D(A)\). It follows that \(27.6.40\) holds for any \(u \in \text{conv}(D(A))\). Say \(u = x_n \in \text{conv}(D(A))\) where \(x_n \to x\). Then

\[
\|J\lambda x - x\|^p \leq \lambda^{p-1} \langle v, x_n - x \rangle + \langle F(J\lambda x - x), x_n - x \rangle + \lambda^{p-1} \langle v, x - J\lambda x \rangle
\]

\[
\leq \lambda^{p-1} \|v\| \|x_n - x\| + \|J\lambda x - x\|^{p-1} \|x_n - x\| + \lambda^{p-1} \|v\| \|J\lambda x - x\|
\]

You have something like this: \(y_\lambda = \|J\lambda x - x\|, \ a_n = \|x_n - x\|, \ y_\lambda^p \leq \lambda^{p-1} \|v\| a_n + y_\lambda^{p-1} a_n + \lambda^{p-1} \|v\| y_\lambda, \ y_\lambda \geq 0\)

where \(p > 1\) and \(a_n \to 0\). Then

\[
\limsup_{\lambda \to 0} y_\lambda^p \leq \limsup_{\lambda \to 0} y_\lambda^{p-1} a_n
\]

and so,

\[
\limsup_{\lambda \to 0} y_\lambda \leq a_n
\]

Hence

\[
\limsup_{\lambda \to 0} \|J\lambda x - x\| \leq \|x_n - x\|
\]

Since \(x_n\) is arbitrary, it follows that for every \(\varepsilon > 0\),

\[
\limsup_{\lambda \to 0} \|J\lambda x - x\| \leq \varepsilon
\]

and so in fact, \(\limsup_{\lambda \to \infty} \|J\lambda x - x\| = 0\).

Now here is an interesting corollary.

**Corollary 27.6.31** Let \(A\) be maximal monotone. \(A : X \to X'\) where \(X\) is a strictly convex reflexive Banach space. Then \(\overline{D(A)}\) is convex.

**Proof:** It is known that \(J\lambda : X \to D(A)\) for any \(\lambda\). Also, if \(x \in \overline{\text{conv}}(D(A))\), then it was shown that \(J\lambda x \to x\).

Clearly

\[
\overline{\text{conv}}(D(A)) \supseteq \overline{D(A)}
\]

Now if \(x\) is in the set on the left, \(J\lambda x \to x\) and so in fact, since \(J\lambda x \in D(A)\), it must be the case that \(x \in \overline{D(A)}\).

Thus the two sets are the same and so in fact, \(\overline{D(A)}\) is closed and convex. \(\blacksquare\)

Note that this implies that \(\overline{\mathcal{A}(D(A))}\) is also convex. This is because \(A^{-1}\) described above, is maximal monotone with domain \(A(D(A))\).

Next, is a useful generalization of some of the earlier material used to establish the above results on approximation. It will include the general case of \(F\) a duality mapping for \(p > 1\).

**Proposition 27.6.32** Suppose \(A : X \to \mathcal{P}(X')\) where \(X\) is a reflexive Banach space with strictly convex norm. Suppose also that \(A\) is maximal monotone. Then if \(\lambda_n \to 0\) and if \(x_n \to x\) weakly, \(A\lambda_n x_n \to x^\star\) weakly, and

\[
\lim_{n,m \to \infty} \sup_{m} \langle A\lambda_n x_n - A\lambda_m x_m, x_n - x_m \rangle \leq 0
\]

Then

\[
\lim_{n,m \to \infty} \langle A\lambda_n x_n - A\lambda_m x_m, x_n - x_m \rangle = 0,
\]

\([x,x^\star] \in G(A)\), and \(\langle A\lambda_n x_n, x_n \rangle \to \langle x^\star, x \rangle\).
Proof: Let \( \alpha = \limsup_{n \to \infty} \langle A_{\lambda_n} x_n, x_n \rangle \). It is finite because the expression is bounded independent of \( n \). Then

\[
\limsup_{m \to \infty} \left( \limsup_{n \to \infty} \left( \langle A_{\lambda_n} x_n, x_n \rangle + \langle A_{\lambda_n} x_m, x_m \rangle - \left( \langle A_{\lambda_n} x_n, x_m \rangle + \langle A_{\lambda_n} x_m, x_n \rangle \right) \right) \right) \leq 0
\]

Thus

\[
\limsup_{m \to \infty} \left( \alpha + \langle A_{\lambda_n} x_m, x_m \rangle - \left( \langle x^*, x_m \rangle + \langle A_{\lambda_n} x_m, x \rangle \right) \right) \leq 0
\]

and so

\[
2\alpha - 2 \langle x^*, x \rangle \leq 0
\]

The next simple observation is that

\[
\|A_{\lambda_n} x_n\| = \left\|\lambda_n^{-(p-1)} F(J_{\lambda_n} x_n - x_n)\right\| \leq C
\]

due to the weak convergence. Hence \( \lambda_n^{-(p-1)} \|J_{\lambda_n} x_n - x_n\|^{p-1} \leq C \) and so

\[
\|J_{\lambda_n} x_n - x_n\| \leq \lambda_n C^{1/(p-1)}.
\] (27.6.47)

Thus if \( [u, u^*] \in \mathcal{G}(A) \),

\[
\liminf_{n \to \infty} \langle A_{\lambda_n} x_n - u^*, x_n - u \rangle = \liminf_{n \to \infty} \langle A_{\lambda_n} x_n - u^*, J_{\lambda_n} x_n - u \rangle \geq 0
\]

because \( A_{\lambda} x \in AJ_{\lambda}x \). However, the left side satisfies

\[
0 \leq \liminf_{n \to \infty} \langle A_{\lambda_n} x_n - u^*, x_n - u \rangle \leq \limsup_{n \to \infty} \langle A_{\lambda_n} x_n - u^*, x_n - u \rangle
\]

\[
= \limsup_{n \to \infty} \left( \langle A_{\lambda_n} x_n, x_n \rangle - \langle A_{\lambda_n} x_n, u^* \rangle - \langle u^*, x_n \rangle + \langle u^*, u \rangle \right)
\]

\[
= \alpha - \langle x^*, u \rangle - \langle u^*, x \rangle + \langle u^*, u \rangle \leq \langle x^*, x \rangle - \langle x^*, u \rangle - \langle u^*, x \rangle + \langle u^*, u \rangle
\]

\[
= \langle x^* - u^*, x - u \rangle
\]

and this shows that \( [x, x^*] \in \mathcal{G}(A) \) since \( [u, u^*] \) was arbitrary.

Next let \( [u, u^*] \in \mathcal{G}(A) \). Then thanks to 27.6.64.

\[
0 \leq \liminf_{n \to \infty} \langle A_{\lambda_n} x_n - u^*, J_{\lambda_n} x_n - u \rangle = \liminf_{n \to \infty} \langle A_{\lambda_n} x_n - u^*, x_n - u \rangle
\]

\[
\leq \limsup_{n \to \infty} \langle A_{\lambda_n} x_n - u^*, x_n - u \rangle
\]

\[
= \limsup_{n \to \infty} \left( \langle A_{\lambda_n} x_n, x_n \rangle - \langle A_{\lambda_n} x_n, u^* \rangle - \langle u^*, x_n \rangle + \langle u^*, u \rangle \right)
\]

\[
= \limsup_{n \to \infty} \langle A_{\lambda_n} x_n, x_n \rangle - \langle x^*, u \rangle - \langle u^*, x \rangle + \langle u^*, u \rangle
\]

\[
\leq \langle x^*, x \rangle - \langle x^*, u \rangle - \langle u^*, x \rangle + \langle u^*, u \rangle = \langle x^* - u^*, x - u \rangle
\]

In particular, you could let \( [u, u^*] = [x, x^*] \) and conclude that

\[
\lim_{n \to \infty} \langle A_{\lambda_n} x_n - x^*, x_n - x \rangle = \lim_{n \to \infty} \left( \langle A_{\lambda_n} x_n, x_n \rangle - \langle A_{\lambda_n} x_n, x^* \rangle + \langle x^*, x \rangle - \langle x^*, x_n \rangle \right)
\]

\[
= \lim_{n \to \infty} \langle A_{\lambda_n} x_n, x_n \rangle - \langle x^*, x \rangle + \langle x^*, x \rangle - \langle x^*, x \rangle = 0
\]

which shows that \( \lim_{n \to \infty} \langle A_{\lambda_n} x_n, x_n \rangle = \langle x^*, x \rangle \). Then it follows from this that

\[
\lim_{n,m \to \infty} \langle A_{\lambda_n} x_n - A_{\lambda_m} x_m, x_n - x_m \rangle = 0 \]

For the rest of this, the usual duality map for \( p = 2 \) will be used. It may be that one could change this, but I don’t have a need to do it right now so from now on, \( F \) will be the usual thing.
27.6.4 Sum Of Maximal Monotone Operators

To begin with, here is a nice lemma.

**Lemma 27.6.33** Let \( 0 \in D(A) \) and let \( A \) be maximal monotone and let \( B : X \to X' \) be monotone hemicontinuous, bounded, and coercive. Then \( B + A \) is also maximal monotone. Also \( B + A \) is onto.

**Proof:** By Theorem 27.6.32 there exists \( x \in D(A) \) such that for all \([u,u^*] \in \mathcal{G}(A)\),

\[
\langle Bx + Fx - y^* + u^*, u - x \rangle \geq 0
\]

Hence for all \([u,u^*]\),

\[
\langle u^* - (y^* - (Bx + Fx)), u - x \rangle \geq 0
\]

It follows that

\[
y^* - (Bx + Fx) \in Ax
\]

and so \( y^* \in Bx + Ax + Fx \) showing that \( B + A \) is maximal monotone because it added to \( F \) is onto. As to the last claim, just don’t add in \( F \) in the argument. Thus for all \([u,u^*]\),

\[
\langle Bx - y^* + u^*, u - x \rangle \geq 0
\]

Then the rest is as before. You find that \( y^* - Bx \in Ax \). ■

**Corollary 27.6.34** Suppose instead of \( 0 \in D(A) \), it is known that \( x_0 \in D(A) \) and

\[
\lim_{\|x\| \to \infty} \frac{\langle B(x_0 + x), x \rangle}{\|x\|} = \infty
\]

Then if \( B \) is monotone and hemicontinuous and \( A \) is maximal monotone, then \( B + A \) is onto.

**Proof:** Let \( \hat{A}(x) \equiv A(x_0 + x) \) so in fact \( 0 \in D(\hat{A}) \). Then letting \( \hat{B} \) be defined similarly, it follows from the above lemma that if \( y^* \in X' \), there exists \( x \) such that

\[
y^* \in \hat{A}x + \hat{B}x \equiv A(x_0 + x) + B(x_0 + x)
\]

**Lemma 27.6.35** Let \( 0 \) be on the interior of \( D(A) \) and also in \( D(B) \). Also let \( 0 \in B(0) \) and \( 0 \in A(0) \). Then if \( A,B \) are maximal monotone, so is \( A + B \).

**Proof:** Note that, since \( 0 \in A(0) \), if \( x^* \in Ax \), then \( \langle x^*, x \rangle \geq 0 \). Also note that \( \|B(0)\| \leq \|B(0)\| = 0 \) and so also \( \langle B(0)x, x \rangle \geq 0 \). It is necessary to show that \( F + A + B \) is onto. However, \( B_\lambda \) is monotone hemicontinuous, bounded and coercive. Hence, by Lemma 27.6.33, \( B_\lambda + A \) is maximal monotone. If \( x^* \in X' \) is given, there exists a solution to

\[
x^* \in Fx_\lambda + B_\lambda x_\lambda + Ax_\lambda
\]

Do both sides to \( x_\lambda \) and let \( x^*_\lambda \in Ax_\lambda \) be such that equality holds in the above.

\[
x^* = Fx_\lambda + B_\lambda x_\lambda + x^*_\lambda
\]

(27.6.48)

Then

\[
\langle x^*, x_\lambda \rangle = \|x_\lambda \|^2 + \langle x^*_\lambda, x_\lambda \rangle
\]

It follows that

\[
\|x_\lambda \| \leq \|x^*\|, \quad \langle x^*_\lambda, x_\lambda \rangle \leq \langle x^*, x_\lambda \rangle \leq \|x^*\| \|x_\lambda \| \leq \|x^*\|^2
\]

(27.6.49)

Next, \( 0 \) is on the interior of \( D(A) \) and so from Theorem 27.6.31 there exists \( \rho > 0 \) such that if \( y^* \in Ax \) for \( \|x\| \leq \rho \), then \( \|y^*\| < M \) and in fact, all such \( x \) are in \( D(A) \). Now let

\[
y_\lambda = \frac{1}{2\|x^*_\lambda\|} F^{-1}(x^*_\lambda) \quad \text{so} \quad \|y_\lambda\| < \rho
\]

Thus \( y_\lambda \in D(A) \) and if \( y^*_\lambda \in Ay_\lambda \), then \( \|y^*_\lambda\| < M \). Then for such bounded \( y^*_\lambda \),

\[
0 \leq \langle y^*_\lambda - x^*_\lambda, y_\lambda - x_\lambda \rangle = \langle y^*_\lambda, y_\lambda \rangle - \langle x^*_\lambda, y_\lambda \rangle - \langle y^*_\lambda, x_\lambda \rangle + \langle x^*_\lambda, x_\lambda \rangle
\]
Then

\[ \frac{1}{2} \| x_\lambda \|^2 = \left\langle x_\lambda, \frac{1}{2} \| x_\lambda \|^2 F^{-1} (x_\lambda) \right\rangle = \langle x_\lambda, y \lambda \rangle \leq \langle y_\lambda^*, y \lambda \rangle - \langle y_\lambda^*, x_\lambda \rangle + \langle x_\lambda^*, x_\lambda \rangle \leq M \rho + M \| x_\lambda \| + \langle x_\lambda^*, x_\lambda \rangle \]

From \textbf{27.6.34.}

\[ \| x_\lambda \|^2 \leq 2 \left( M \rho + M \| x^* \| + \| x^* \|^2 \right) \]

Thus from \textbf{27.6.34.}, \( x_\lambda, x_\lambda^*, Fx_\lambda \) are all bounded. Hence it follows from \textbf{27.6.38.} that \( B_\lambda x_\lambda \) is also bounded. Therefore, there is a sequence, \( \lambda_n \to 0 \) such that

\[ x_{\lambda_n} \to z \text{ weakly} \]
\[ x_\lambda^* \to u^* \text{ weakly} \]
\[ Fx_\lambda \to u^* \text{ weakly} \]
\[ B_\lambda x_{\lambda_n} \to b^* \text{ weakly} \]

Using \textbf{27.6.38.} it follows that

\[ \langle Fx_{\lambda_n} + x_\lambda^* + B_\lambda x_{\lambda_n} - (Fx_{\lambda_m} + x_\lambda^* + B_\lambda x_{\lambda_m}), x_{\lambda_n} - x_{\lambda_m} \rangle = 0 \]

Thus

\[ \langle Fx_{\lambda_n} + x_\lambda^* - (Fx_{\lambda_m} + x_\lambda^*) , x_{\lambda_n} - x_{\lambda_m} \rangle + \langle B_\lambda x_{\lambda_m} - B_\lambda x_{\lambda_m}, x_{\lambda_n} - x_{\lambda_m} \rangle = 0 \]  
(27.6.50)

Now \( F + A \) is surely monotone and so

\[ \lim_{m,n \to \infty} \sup \langle B_\lambda x_{\lambda_m} - B_\lambda x_{\lambda_m}, x_{\lambda_n} - x_{\lambda_m} \rangle \leq 0 \]

By Proposition \textbf{27.6.32.}, \( b^* \in Bz \) and

\[ \lim_{m \to \infty} \langle B_\lambda x_{\lambda_m} - B_\lambda x_{\lambda_m}, x_{\lambda_n} - x_{\lambda_m} \rangle = 0 \]

Then returning to \textbf{27.6.31.}

\[ \lim_{m,n \to \infty} \sup \langle Fx_{\lambda_n} + x_\lambda^* - (Fx_{\lambda_m} + x_\lambda^*) , x_{\lambda_n} - x_{\lambda_m} \rangle \leq 0 \]

Now from Lemma \textbf{27.6.33.}, \( F + A \) is maximal monotone. Hence Proposition \textbf{27.6.32.} applies again and it follows that \( u^* + w^* \in Fz + Az \). Then passing to the limit as \( n \to \infty \) in

\[ x^* = Fx_{\lambda_n} + B_\lambda x_{\lambda_n} + x_\lambda^* \]

it follows that

\[ x^* = u^* + b^* + w^* = Fz + Az + Bz \]

and this shows that \( A + B \) is maximal monotone. You don’t need to assume all that stuff about \( 0 \in (A) \), \( 0 \in B \), \( 0 \) on interior of \( D(A) \) and so forth.

**Theorem 27.6.36** Suppose \( A, B \) are maximal monotone and the interior of \( D(A) \) has nonempty intersection with \( D(B) \). Then \( A + B \) is maximal monotone.

**Proof:** Let \( x_0 \) be on the interior of \( D(A) \) and also in \( D(B) \). Let \( \hat{A}(x) = A(x_0 + x) - x_0^* \) where \( x_0^* \in A(x_0) \). Thus \( 0 \in D(\hat{A}) \) and \( 0 \in \hat{A}(0) \). Do the same thing for \( B \) to get \( \hat{B} \) defined similarly. Are these still maximal monotone? Suppose for all \([u, u^*] \in G(\hat{A})\)

\[ \langle y^* - u^*, y - u \rangle \geq 0 \]

Does it follow that \( y^* \in \hat{A}y \)? It is given that \( u^* \in A(x_0 + u) \). The above implies for all \([u, u^*] \in G(\hat{A})\)

\[ \langle y^* + x_0^* - (u^* + x_0^*), (y + x_0) - (u + x_0) \rangle \geq 0 \]
and since \( u + x_0 \) is a generic element of \( D(A) \) for \( u \in D(A) \), the above implies \( y^* + x_0^* \in A(y + x_0) \) and so \( y \in A(y + x_0) - x_0^* \equiv A(y) \). Hence the graph is maximal. Similar for \( \hat{B} \). Thus the lemma can be applied to \( \hat{A}, \hat{B} \) to conclude that the sum of these is maximal monotone. Now a repeat of the above reasoning which shows that \( \hat{\lambda} \) follows from almost a repeat of the last part of the proof of the above theorem. Since 

\[
\frac{\langle F(x + x_0) + B_\lambda(x + x_0), x \rangle}{\|x\|} = \frac{\langle F(x + x_0), x \rangle}{\|x\|} + \frac{\langle B_\lambda(x + x_0) - B_\lambda(x_0), x \rangle}{\|x\|} + \frac{\langle B_\lambda(x_0), x \rangle}{\|x\|} \\
\geq \frac{1}{2} \frac{\langle F(x + x_0), x \rangle}{\|x + x_0\|} - \|B_\lambda(x_0)\| \\
\geq \frac{1}{2} \frac{\langle F(x + x_0), x + x_0 \rangle}{\|x + x_0\|} - \frac{1}{2} \frac{\langle F(x + x_0), x_0 \rangle}{\|x + x_0\|} - \|B_\lambda(x_0)\| \\
\geq \frac{1}{2} \frac{\langle F(x + x_0), x + x_0 \rangle}{\|x + x_0\|} - \frac{1}{2} \frac{\langle F(x + x_0), x_0 \rangle}{\|x_0\|} - \|B_\lambda(x_0)\| \\
= \frac{1}{2} \|x + x_0\|^2 - \frac{1}{2} \|x + x_0\| - \|B_\lambda(x_0)\| 
\]

which shows that 

\[
\lim_{\|x\| \to \infty} \frac{\langle F(x + x_0) + B_\lambda(x + x_0), x \rangle}{\|x\|} = \infty
\]

and so by Corollary 27.6.3, there exists a solution to 27.6.51. This shows half of the following interesting theorem which is another version of the above major result.

**Theorem 27.6.37** Suppose \( A, B \) are maximal monotone operators. Then for each \( x^* \in X' \), there exists a solution \( x_\lambda \) to

\[
x^* \in Fx_\lambda + B_\lambda x_\lambda + Ax_\lambda, \ \lambda > 0
\]

If for \( \lambda \in (0, \delta) \), \( \{B_\lambda x_\lambda\} \) is bounded, then there exists a solution \( x \) to

\[
x^* \in Fx + Bx + Ax
\]

**Proof:** The existence of a solution to the inclusion 27.6.3 comes from the above discussion. The last claim follows from almost a repeat of the last part of the proof of the above theorem. Since \( \{B_\lambda x_\lambda\} \) is given to be bounded for \( \lambda \in (0, \delta) \), there is a sequence, \( \lambda_n \to 0 \) such that

\[
x_{\lambda_n} \to z \text{ weakly} \\
x_{\lambda_n}^* \to u^* \text{ weakly} \\
F x_{\lambda_n} \to u^* \text{ weakly} \\
B_{\lambda_n} x_{\lambda_n} \to b^* \text{ weakly}
\]
Using \((\alpha < \phi (x))\), it follows that
\[
\langle Fx_{\lambda_n} + x^*_\lambda_n + B_{\lambda_n}x_{\lambda_n} - (Fx_{\lambda_m} + x^*_\lambda_m + B_{\lambda_m}x_{\lambda_m}), x_{\lambda_n} - x_{\lambda_m} \rangle = 0
\]
Thus
\[
\langle Fx_{\lambda_n} + x^*_\lambda_n - (Fx_{\lambda_m} + x^*_\lambda_m), x_{\lambda_n} - x_{\lambda_m} \rangle \\
+ \langle B_{\lambda_n}x_{\lambda_n} - B_{\lambda_m}x_{\lambda_m}, x_{\lambda_n} - x_{\lambda_m} \rangle = 0
\] (27.6.53)
Now \(F + A\) is surely monotone and so
\[
\lim \sup_{m,n \to \infty} \langle B_{\lambda_n}x_{\lambda_n} - B_{\lambda_m}x_{\lambda_m}, x_{\lambda_n} - x_{\lambda_m} \rangle \leq 0
\]
By Proposition 27.6.52, \(b^* \in Bz\) and
\[
\lim_{m,n \to \infty} \langle B_{\lambda_n}x_{\lambda_n} - B_{\lambda_m}x_{\lambda_m}, x_{\lambda_n} - x_{\lambda_m} \rangle = 0
\]
Then returning to 27.6.52.
\[
\lim_{m,n \to \infty} \langle Fx_{\lambda_n} + x^*_\lambda_n - (Fx_{\lambda_m} + x^*_\lambda_m), x_{\lambda_n} - x_{\lambda_m} \rangle \leq 0
\]
Now from Corollary 27.6.51, \(F + A\) is maximal monotone (In fact, \(F + A\) is onto). Hence Proposition 27.6.32 applies again and it follows that \(u^* + w^* \in Fz + Az\). Then passing to the limit as \(n \to \infty\) in
\[
x^* = Fx_{\lambda_n} + B_{\lambda_n}x_{\lambda_n} + x^*_n
\]
it follows that
\[
x^* = u^* + b^* + w^* = Fz + Az + Bz
\]

27.6.5 Convex Functions, An Example

As before, \(X\) will be a Banach space in what follows. Sometimes it will be a reflexive Banach space and in this case, it will be assumed that the norm is strictly convex.

**Definition 27.6.38** Let \(\phi : X \to (-\infty, \infty]\). Then \(\phi\) is convex if whenever \(t \in [0,1]\), \(x, y \in X\),
\[
\phi (tx + (1-t) y) \leq t\phi (x) + (1-t) \phi (y)
\]
The epigraph of \(\phi\) is defined by
\[
\text{epi} (\phi) \equiv \{(x,y) : y \geq \phi (x)\}
\]
When \(\text{epi} (\phi)\) is closed in \(X \times (-\infty, \infty]\), we say that \(\phi\) is lower semicontinuous, l.s.c. The function is called proper if \(\phi (x) < \infty\) for some \(x\). The collection of all such \(x\) is called \(D (\phi)\), the domain of \(\phi\).

This definition of lower semicontinuity is equivalent to the usual definition.

**Lemma 27.6.39** The above definition of lower semicontinuity is equivalent to the assertion that whenever \(x_n \to x\), it follows that \(\phi (x) \leq \lim \inf_{n \to \infty} \phi (x_n)\). In case that \(\phi\) is convex, lower semicontinuity is equivalent to weak lower semicontinuity. That is \(\text{epi} (\phi)\) is closed if and only if \(\text{epi} (\phi)\) is weakly closed. In this case, the limit condition: If \(x_n \to x\) weakly, then \(\phi (x) \leq \lim \inf_{n \to \infty} \phi (x_n)\) is valid.

**Proof:** Suppose the limit condition holds. Why is \(\text{epi} (\phi)\) closed? Why is \(X \times (-\infty, \infty]\) \(\setminus \text{epi} (\phi) \equiv \text{epi} (\phi)^C\) open? Let \((x, \alpha) \in \text{epi} (\phi)^C\). Then \(\alpha < \phi (x)\), \(\alpha + \delta < \phi (x)\). Consider \(B (x, r) \times (\alpha - \frac{\delta}{2}, \alpha + \frac{\delta}{2})\). If every such open set contains a point of \(\text{epi} (\phi)\), then there exists \(x_n \to x, y_n < \alpha + \frac{\delta}{2}, y_n \geq \phi (x_n)\). Hence, from the limit condition,
\[
\phi (x) \leq \lim \inf_{n \to \infty} \phi (x_n) \leq \lim \inf_{n \to \infty} y_n \leq \alpha + \frac{\delta}{2} < \alpha + \delta < \phi (x)
\]
a contradiction. It follows that there exists \(r > 0\) such that \(B (x, r) \times (\alpha - \frac{\delta}{2}, \alpha + \frac{\delta}{2}) \cap \text{epi} (\phi) = \emptyset\). Since \(\text{epi} (\phi)^C\) is open, it follows that \(\text{epi} (\phi)\) is closed.
Next suppose epi \((\phi)\) is closed. Why does the limit condition hold? Suppose \(x_n \to x\). Then \((x_n, \phi(x_n)) \in epi(\phi)\). There is a subsequence such that

\[
\alpha \equiv \lim_{n \to \infty} \inf_{k \to \infty} \phi(x_n) = \lim_{k \to \infty} \phi(x_{n_k})
\]

and so \((x_{n_k}, \phi(x_{n_k})) \to (x, \alpha)\). Since epi \((\phi)\) is closed, this means \((x, \alpha) \in epi(\phi)\). Hence

\[
\alpha \equiv \lim_{n \to \infty} \inf_{k \to \infty} \phi(x_n) \geq \phi(x).
\]

Consider the last claim. In this case, epi \((\phi)\) is convex. If it is closed, then it is weakly closed thanks to separation theorems: If \((x, \alpha) \in epi(\phi)^C\), then \(\alpha < \infty\) and so there exists \((x^*, \beta) \in (X \times \mathbb{R})^l\) and \(l\) such that for all \((t, \gamma) \in epi(\phi)\),

\[
x^*(t) + \beta \gamma > l > x^*(x) + \alpha \beta
\]

Then \(B(x^*,\beta)((x,\alpha),\delta)\) is a weakly open set containing \((x,\alpha)\). For \(\delta\) small enough, it does not intersect epi \((\phi)\) since if not so, there would exist \((t_n, \gamma_n) \in epi(\phi) \cap B(x^*, \beta)((x, \alpha), \frac{1}{n})\) and so

\[
x^*(t_n) + \beta \gamma_n \to x^*(x) + \alpha \beta
\]

contrary to the above inequality. Thus epi \((\phi)\) is weakly closed. Also, if epi \((\phi)\) is weakly closed, then it is obviously strongly closed.

What of the limit condition using weak convergence instead of strong convergence? Say \(x_n \to x\) weakly. Does it follow that if epi \((\phi)\) is weakly closed that \(\phi(x) \leq \liminf_{n \to \infty} \phi(x_n)\)? It is just as above. There is a subsequence such that

\[
\alpha \equiv \lim_{n \to \infty} \inf_{k \to \infty} \phi(x_n) = \lim_{k \to \infty} \phi(x_{n_k})
\]

and so \((x_{n_k}, \phi(x_{n_k})) \to (x, \alpha)\) weakly. Since epi \((\phi)\) is weakly closed, this means \((x, \alpha) \in epi(\phi)\). Hence

\[
\alpha \equiv \lim_{n \to \infty} \inf_{k \to \infty} \phi(x_n) \geq \phi(x).
\]

There is also another convenient characterization of what it means for a function to be lower semicontinuous.

**Lemma 27.6.40** Let \(\phi : X \to (-\infty, \infty]\). Then \(\phi\) is lower semicontinuous if and only if \(\phi^{-1}((a, \infty])\) is open for any \(a \in \mathbb{R}\).

**Proof:** Suppose first that epi \((\phi)\) is closed. Consider \(x \in \phi^{-1}((a, \infty])\). Thus \(\phi(x) > a\). Thus \((x, a) \in epi(\phi)^C\) because \(a < \phi(x)\). Since epi \((\phi)\) is closed, there exists \(r, \varepsilon > 0\) such that

\[
B(x, r) \times (a - \varepsilon, a + \varepsilon) \subseteq epi(\phi)^C
\]

Hence if \(y \in B(x, r)\), it follows that \(\phi(y) \geq a + \varepsilon\) since otherwise there would be a point of epi \((\phi)^C\) in this open set

\[
B(x, r) \times (a - \varepsilon, a + \varepsilon).
\]

Hence \(B(x, r) \subseteq \phi^{-1}((a, \infty])\).

Conversely, suppose \(\phi^{-1}((a, \infty])\) is open for any \(a\) and let \((x, b) \in epi(\phi)^C\). Then \(\phi(x) > b\). Thus there exists \(B(x, r)\) such that for \(y \in B(x, r)\), it follows that \(\phi(y) > b\). That is, \(y \in \phi^{-1}((b, \infty])\). So consider \(B(x, r) \times (-\infty, b)\), then since \(\phi(y) > b\), \(a < \phi(y)\) and so there is no point of intersection between epi \((\phi)\) and this open set \(B(x, r) \times (-\infty, b)\).

Of course one can define upper semicontinuous the same way that \(\phi^{-1}(-\infty, a)\) is open. Thus a function is continuous if and only if it is both upper and lower semicontinuous.

In case \(X\) is reflexive, the limit condition implies that epi \((\phi)\) is weakly closed. Suppose \((x, \alpha)\) is a weak limit point of epi \((\phi)\). Then by the Eberlein Smulian theorem, there is a subsequence of points of \(X, (x_n, \alpha_n)\) which converges weakly to \((x, \alpha)\). Thus if the limit condition holds,

\[
\phi(x) \leq \lim_{n \to \infty} \inf_{k \to \infty} \phi(x_n) \leq \lim_{n \to \infty} \inf \alpha_n = \alpha
\]

and so \((x, \alpha) \in epi(\phi)\). If \(X\) is not reflexive, this isn’t all that clear because it is not clear that a limit point is the limit of a sequence. However, one could consider a limit condition involving nets and get a similar result.

**Definition 27.6.41** Let \(\phi : X \to (-\infty, \infty]\) be convex lower semicontinuous, and proper. Then

\[
\partial \phi(x) \equiv \{x^* : \phi(y) - \phi(x) \geq \langle x^*, y - x \rangle \text{ for all } y\}
\]

The domain of \(\partial \phi\), denoted as \(D(\partial \phi)\) is just the set of all \(x\) for which \(\partial \phi(x) \neq \emptyset\). Note that \(D(\partial \phi) \subseteq D(\phi)\) since if \(x \notin D(\phi)\), the defining inequality could not hold for all \(y\) because the left side would be \(-\infty\) for some \(y\).
Theorem 27.6.42 For \( X \) a real Banach space, let \( \phi(x) \equiv \frac{1}{2} \|x\|^2 \). Then \( F(x) = \partial \phi(x) \). Here \( F \) was the set valued map satisfying \( x^* \in Fx \) means
\[
\|x^*\| = \|Fx\|, (Fx, x) = \|x\|^2.
\]

Proof: Let \( x^* \in F(x) \). Then
\[
\langle x^*, y - x \rangle = \langle x^*, y \rangle - \langle x^*, x \rangle \\
\leq \|x\| \|y\| - \|x\|^2 \leq \frac{1}{2} \|y\|^2 - \frac{1}{2} \|x\|^2.
\]
This shows \( F(x) \subseteq \partial \phi(x) \).
Next suppose \( x^* \in \partial \phi(x) \). Then for all \( t \in \mathbb{R} \),
\[
\langle x^*, ty \rangle = \langle x^*, (ty + x) - x \rangle \leq \frac{1}{2} \left( \|ty + x\|^2 - \|x\|^2 \right).
\] (27.6.54)
Now if \( t > 0 \), divide both sides by \( t \). This yields
\[
\langle x^*, y \rangle \leq \frac{1}{2t} \left( \|x\| + t \|y\| \right)^2 - \|x\|^2 \\
= \frac{1}{2t} \left( 2t \|x\| \|y\| + t^2 \|y\|^2 \right)
\]
Letting \( t \to 0 \),
\[
\langle x^*, y \rangle \leq \|x\| \|y\|. \quad (27.6.55)
\]
Next suppose \( t = -s \), where \( s > 0 \) in 27.6.54. Then, since when you divide by a negative, you reverse the inequality, for \( s > 0 \)
\[
\langle x^*, y \rangle \geq \frac{1}{2s} \left[ \|x\|^2 - \|x - sy\|^2 \right] \\
= \frac{1}{2s} \left[ \|x - sy\|^2 - \|x\|^2 \right] \\
= \frac{1}{2s} \left[ \|x - sy\|^2 \|sy\| + \|sy\|^2 \right].
\] (27.6.56)
Taking a limit as \( s \to 0 \) yields
\[
\langle x^*, y \rangle \geq - \|x\| \|y\|. \quad (27.6.58)
\]
It follows from 27.6.53 and 27.6.58 that
\[
\|x^*, y\| \leq \|x\| \|y\|
\]
and that, therefore, \( \|x^*\| \leq \|x\| \) and \( \langle x^*, x \rangle \leq \|x\|^2 \). Now return to 27.6.54 and let \( y = x \). Then
\[
\langle x^*, x \rangle \geq \frac{1}{2s} \left[ -2 \|x - sx\| \|sx\| + \|sx\|^2 \right] \\
= - \|x\|^2 (1 - s) + s \|x\|^2
\]
Letting \( s \to 1 \),
\[
\langle x^*, x \rangle \geq \|x\|^2.
\]
Since it was already shown that \( \|x^*, x\| \leq \|x\|^2 \), this shows \( \langle x^*, x \rangle = \|x\|^2 \) and also \( \|x^*\| \leq \|x\| \). Thus
\[
\|x^*\| \geq \left( x^* \frac{x}{\|x\|} \right) = \|x\|
\]
so in fact \( x^* \in F(x) \). \( \square \)

The next result gives conditions under which the subgradient is onto. This means that if \( y^* \in X' \), then there exists \( x \in X \) such that \( y^* \in \partial \phi(x) \).

Theorem 27.6.43 Suppose \( X \) is a reflexive Banach space and suppose \( \phi : X \to (-\infty, \infty) \) is convex, proper, l.s.c., and for all \( y^* \in X' \), \( x \to \phi(x) - \langle y^*, x \rangle \) is coercive. Then \( \partial \phi \) is onto.
**Proof:** The function \( x \to \phi(x) - y^*(x) \equiv \psi(x) \) is convex, proper, l.s.c., and coercive. Let
\[
\lambda \equiv \inf \{ \phi(x) - \langle y^*, x \rangle : x \in X \}
\]
and let \( \{x_n\} \) be a minimizing sequence satisfying
\[
\lambda = \lim_{n \to \infty} \phi (x_n) - \langle y^*, x_n \rangle
\]
By coercivity,
\[
\lim_{||x|| \to \infty} \phi (x) - \langle y^*, x \rangle = \infty
\]
and so this minimizing sequence is bounded. By the Eberlein Smulian theorem, Theorem 27.6.44, there is a weakly convergent subsequence \( x_{n_k} \to x \). By Lemma 27.6.45,
\[
\lambda = \phi (x) - \langle y^*, x \rangle \leq \lim \inf_{k \to \infty} \phi (x_{n_k}) - \langle y^*, x_{n_k} \rangle = \lambda
\]
so there exists \( x \) which minimizes \( x \to \phi (x) - \langle y^*, x \rangle \equiv \psi(x) \). Therefore, \( 0 \in \partial \psi(x) \) because
\[
\psi(y) - \psi(x) \geq 0 = \langle 0, y - x \rangle
\]
Thus, \( 0 \in \partial \psi(x) = \partial \phi(x) - y^* \).

Now let \( \phi \) be a convex proper lower semicontinuous function defined on \( X \) where \( X \) is a reflexive Banach space with strictly convex norm. Consider \( \partial \phi \). Is it maximal monotone? Is it the case that \( F + \partial \phi \) is onto? First of all, is \( \partial \phi \) monotone? Let \( x^* \in \partial \phi(x), y^* \in \partial \phi(y) \). Then
\[
\phi(y) - \phi(x) \geq \langle x^*, y - x \rangle
\]
\[
\phi(x) - \phi(y) \geq \langle y^*, x - y \rangle
\]
Hence adding these yields
\[
\langle y^* - x^*, x - y \rangle \leq 0, \quad \langle y^* - x^*, y - x \rangle \geq 0.
\]
Yes, \( \partial \phi \) is certainly monotone. Is it maximal monotone?

**Theorem 27.6.44** Let \( \phi \) be convex, proper, and lower semicontinuous on \( X \) where \( X \) is a reflexive Banach space having strictly convex norm. Then \( \partial \phi \) is maximal monotone.

**Proof:** It is necessary to show that \( F + \partial \phi \) is onto. To do this, let
\[
\psi(x) \equiv \frac{1}{2} \| x \|^2 + \phi(x) - \langle y^*, x \rangle
\]
where \( y^* \) is a given element of \( X' \) and the idea is to show that \( y^* \in F(x) + \partial \phi(x) \) for some \( x \). Then by separation theorems, \( \phi(x) \geq b + \langle z^*, x \rangle \) for some \( b, z^* \). Hence it is clear that \( \psi \) is convex, lower semicontinuous and coercive in the sense that
\[
\lim_{||x|| \to \infty} \psi (x) = \infty
\]
It follows that any minimizing sequence for \( \psi \) is bounded. Hence by the weak lower semicontinuity, this function has a minimum at \( x_0 \) say. Thus
\[
\frac{1}{2} \| x_0 \|^2 + \phi(x_0) - \langle y^*, x_0 \rangle \leq \frac{1}{2} \| x \|^2 + \phi(x) - \langle y^*, x \rangle
\]
for all \( x \). Then
\[
\frac{1}{2} \| x_0 \|^2 - \frac{1}{2} \| x \|^2 + \langle y^*, x - x_0 \rangle \leq \phi(x) - \phi(x_0)
\]
Now from Theorem 27.6.42,
\[
\langle F(x), x_0 - x \rangle \leq \frac{1}{2} \| x_0 \|^2 - \frac{1}{2} \| x \|^2
\]
and so, the above reduces to
\[
\langle F(x), x_0 - x \rangle + \langle y^*, x - x_0 \rangle \leq \phi(x) - \phi(x_0)
\]
Next let \( x = x_0 + t (z - x_0), t \in (0, 1) \), where \( z \) is arbitrary. Then

\[
-t \langle F(x_0 + t(z - x_0)), z - x_0 \rangle + t \langle y^*, z - x_0 \rangle \leq \phi(x_0 + t(z - x_0)) - \phi(x_0)
\]

and so, by convexity,

\[
-t \langle F(x_0 + t(z - x_0)), z - x_0 \rangle + t \langle y^*, z - x_0 \rangle \leq (1 - t) \phi(x_0) + t \phi(z) - \phi(x_0)
\]

\[
t \langle y^*, z - x_0 \rangle \leq t (\phi(z) - \phi(x_0)) + t \langle F(x_0 + t(z - x_0)), z - x_0 \rangle
\]

Now cancel the \( t \) on both sides to obtain

\[
\langle y^*, z - x_0 \rangle \leq (\phi(z) - \phi(x_0)) + \langle F(x_0 + t(z - x_0)), z - x_0 \rangle
\]

By the fact that \( F \) is hemicontinuous, actually demicontinuous, one can let \( t \downarrow 0 \) and obtain

\[
\langle y^*, z - x_0 \rangle \leq (\phi(z) - \phi(x_0)) + \langle F(x_0), z - x_0 \rangle
\]

This says that \( y^* - F(x_0) \in \partial \phi(x_0) \) from the definition of what \( \partial \phi(x_0) \) means. \( \blacksquare \)

There is a much harder approach to this theorem which is based on a theorem about when the subgradient of a sum equals the sum of the subgradients. This major theorem is given next. Much of the above is in [113] but I don’t remember where I found the following proof.

**Theorem 27.6.45** Let \( \phi_1 \) and \( \phi_2 \) be convex, l.s.c. and proper having values in \((-\infty, \infty] \). Then

\[
\partial(\lambda \phi_1) (x) = \lambda \partial \phi_1 (x), \quad \partial(\phi_1 + \phi_2) (x) \supseteq \partial \phi_1 (x) + \partial \phi_2 (x)
\]

(27.6.59)

if \( \lambda > 0 \). If there exists \( \mathfrak{F} \in \text{dom}(\phi_1) \cap \text{dom}(\phi_2) \) and \( \phi_1 \) is continuous at \( \mathfrak{F} \) then for all \( x \in X \),

\[
\partial(\phi_1 + \phi_2) (x) = \partial \phi_1 (x) + \partial \phi_2 (x).
\]

(27.6.60)

**Proof**: \( \blacksquare \) is obvious so we only need to show \( \blacksquare \). Suppose \( \mathfrak{F} \) is as described. It is clear \( \blacksquare \) holds whenever \( x \notin \text{dom}(\phi_1) \cap \text{dom}(\phi_2) \) since then \( \partial(\phi_1 + \phi_2) = \emptyset \). Therefore, assume

\[
x \in \text{dom}(\phi_1) \cap \text{dom}(\phi_2)
\]

in what follows. Let \( x^* \in \partial(\phi_1 + \phi_2) (x) \). Is \( x^* \) is the sum of an element of \( \partial \phi_1 (x) \) and \( \partial \phi_2 (x) \)? Does there exist \( x_1^* \) and \( x_2^* \) such that for every \( y \),

\[
x^*(y - x) = x_1^*(y - x) + x_2^*(y - x)
\]

\[
\leq \phi_1(y) - \phi_1(x) + \phi_2(y) - \phi_2(x)?
\]

If so, then

\[
\phi_1(y) - \phi_1(x) - x^*(y - x) \geq \phi_2(x) - \phi_2(y).
\]

Define

\[
C_1 \equiv \{(y, a) \in X \times \mathbb{R} : \phi_1(y) - \phi_1(x) - x^*(y - x) \leq a \}
\]

\[
C_2 \equiv \{(y, a) \in X \times \mathbb{R} : a \leq \phi_2(x) - \phi_2(y) \}.
\]

I will show \( \text{int}(C_1) \cap C_2 = \emptyset \) and then by Theorem \( \blacksquare \) there exists an element of \( X' \) which does something interesting.

Both \( C_1 \) and \( C_2 \) are convex and nonempty. Say \( y_1, y_2 \in C_1 \) and \( t \in [0, 1] \). Then

\[
\phi_1((ty_1) + (1 - t)y_2) - \phi_1(x) - x^*((ty_1) + (1 - t)y_2) - x)
\]

\[
\leq t\phi(y_1) + (1 - t)\phi(y_2) - (t\phi_1(x) + (1 - t)\phi(x))
\]

\[
- (tx^*(y_1 - x) + (1 - t)x^*(y_2 - x))
\]

\[
\leq ta + (1 - t)a = a
\]

so \( C_1 \) is indeed convex. The case of \( C_2 \) is similar.
$C_1$ is nonempty because it contains $(x, \phi_1(x) - \phi_1(x) - x^*(x - x))$ since
\[ \phi_1(x) - \phi_1(x) - x^*(x - x) \leq \phi_1(x) - \phi_1(x) - x^*(x - x) \]

$C_2$ is also nonempty because it contains $(x, \phi_2(x) - \phi_2(x))$ since
\[ \phi_2(x) - \phi_2(x) \leq \phi_2(x) - \phi_2(x) \]

In addition to this,
\[ (x, \phi_1(x) - x^*(x - x) - \phi_1(x) + 1) \in \text{int}(C_1) \]
due to the assumed continuity of \( \phi_1 \) at \( x \) and so \( \text{int}(C_1) \neq \emptyset \). If \((y, a) \in \text{int}(C_1)\) then
\[ \phi_1(y) - x^*(y - x) - \phi_1(x) \leq a - \varepsilon \]
whenever \( \varepsilon \) is small enough. Therefore, if \((y, a)\) is also in \( C_2 \), the assumption that \( x^* \in \partial (\phi_1 + \phi_2)(x) \) implies
\[ a - \varepsilon \geq \phi_1(y) - x^*(y - x) - \phi_1(x) \geq \phi_2(x) - \phi_2(y) \geq a, \]
a contradiction. Therefore \( \text{int}(C_1) \cap C_2 = \emptyset \) and so by Theorem \ref{thm:27.6.61} there exists \((w^*, \beta) \in X' \times \mathbb{R} \) with
\[ (w^*, \beta) \neq (0, 0), \quad (27.6.61) \]
and
\[ w^*(y) + \beta a \geq w^*(y_1) + \beta a_1, \quad (27.6.62) \]
whenever \((y, a) \in C_1 \) and \((y_1, a_1) \in C_2 \).

**Claim:** \( \beta > 0 \).

**Proof of claim:** If \( \beta < 0 \) let
\[ a = \phi_1(x) - x^*(x - x) - \phi_1(x) + 1, \]
\[ a_1 = \phi_2(x) - \phi_2(x), \quad \text{and} \quad y = y_1 = x. \]

Then from \ref{27.6.61}
\[ \beta (\phi_1(x) - x^*(x - x) - \phi_1(x) + 1) \geq \beta (\phi_2(x) - \phi_2(x)). \]
Dividing by \( \beta \) yields
\[ \phi_1(x) - x^*(x - x) - \phi_1(x) + 1 \leq \phi_2(x) - \phi_2(x) \]
and so
\[ \phi_1(x) + \phi_2(x) - (\phi_1(x) + \phi_2(x)) + 1 \leq x^*(x - x) \]
\[ \phi_1(x) + \phi_2(x) - (\phi_1(x) + \phi_2(x)) \leq \phi_1(x) + \phi_2(x) - (\phi_1(x) + \phi_2(x)), \]
a contradiction. Therefore, \( \beta \geq 0 \).

Now suppose \( \beta = 0 \). Letting
\[ a = \phi_1(x) - x^*(x - x) - \phi_1(x) + 1, \]
and so there exists an open set \( U \) containing 0 and \( \eta > 0 \) such that
\[ x + U \times (a - \eta, a + \eta) \subseteq C_1. \]

Therefore, \ref{27.6.61} applied to \((x + z, a) \in C_1 \) and \((x, \phi_2(x) - \phi_2(x)) \in C_2 \) for \( z \in U \) yields
\[ w^*(x + z) \geq w^*(x) \]
for all \( z \in U \). Hence \( w^*(z) = 0 \) on \( U \) which implies \( w^* = 0 \), contradicting \ref{27.6.61}. This proves the claim.

Now with the claim, it follows \( \beta > 0 \) and so, letting \( z^* = w^*/\beta \), \ref{27.6.61} and Lemma \ref{lem:27.6.61} implies
\[ z^*(y) + a \geq z^*(y_1) + a_1 \quad (27.6.63) \]
whenever \((y, a) \in C_1 \) and \((y_1, a_1) \in C_2 \). In particular,
\[ (y, \phi_1(y) - \phi_1(x) - x^*(y - x)) \in C_1 \quad (27.6.64) \]
because
\[ \phi_1(y) - \phi_1(x) - x^*(y - x) \leq \phi_1(y) - x^*(y - x) - \phi_1(x) \]
and
\[ (y_1, \phi_2(x) - \phi_2(y_1)) \in C_2. \]  
(27.6.65)
by similar reasoning so letting \( y = x \),
\[ z^*(x) + \left( \phi_1(x) - x^*(x - x) - \phi_1(x) \right) \geq z^*(y_1) + \phi_2(x) - \phi_2(y_1). \]
Therefore,
\[ z^*(y_1 - x) \leq \phi_2(y_1) - \phi_2(x) \]
for all \( y_1 \) and so \( z^* \in \partial \phi_2(x) \). Now let \( y_1 = x \) in 27.6.65 and using 27.6.46 and 27.6.65, it follows
\[ z^*(y) + \phi_1(y) - x^*(y - x) - \phi_1(x) \geq z^*(x) \]
\[ \phi_1(y) - \phi_1(x) \geq x^*(y - x) - z^*(y - x) \]
and so \( x^* - z^* \in \partial \phi_1(x) \) so \( x^* = z^* + (x^* - z^*) \in \partial \phi_2(x) + \partial \phi_1(x) \).

**Corollary 27.6.46** Let \( \phi : X \to (-\infty, \infty] \) be convex, proper, and lower semicontinuous. Here \( X \) is a Banach space. Then \( \partial \phi \) is maximal monotone.

**Proof:** Let \( \psi(x) = \frac{1}{2} \| x \|^2 \). There exists \( x^* \) and some number \( b \) such that \( \phi(x) \geq b + \langle x^*, x \rangle \). Therefore, \( \psi + \phi \) is convex, lower semicontinuous, and bounded. It follows \( \partial (\psi + \phi) \) is onto by Theorem 27.6.43. However, \( \psi \) is continuous everywhere, in particular at every point of the domain of \( \phi \). Therefore, \( \partial \psi + \partial \phi = \partial (\phi + \psi) \) and by Theorem 27.6.42, this shows that \( F + \partial \phi \) is onto. \( \blacksquare \)

It seems to me that the above are the most important results about convex proper lower semicontinuous functions. However, there are many other very interesting properties known.

**Proposition 27.6.47** Let \( \phi : X \to (-\infty, \infty] \) be convex proper and lower semicontinuous. Then \( D(\partial \phi) \) is dense in \( D(\phi) \) and so \( \overline{D(\partial \phi)} = \overline{D(\phi)} \).

**Proof:** Let \( x_\lambda \) be the solution to \( 0 \in F(x_\lambda - x) + \lambda \partial \phi(x_\lambda) \). Here \( x \in D(\phi) \). Say \( u^*_x \in \partial \phi(x_\lambda) \) such that the inclusion becomes an equality. Then
\[ 0 = \langle F(x_\lambda - x) + \lambda u^*_x, x_\lambda - x \rangle = \| x_\lambda - x \|^2 - \lambda \langle u^*_x, x - x_\lambda \rangle \]
\[ \geq \| x_\lambda - x \|^2 - \lambda (\phi(x) - \phi(x_\lambda)) \]

Hence, letting \( z^*, b \) be such that \( \phi(y) \geq b + \langle z^*, y - x \rangle \),
\[ \lambda (\phi(x) - [b + \langle z^*, x_\lambda - x \rangle]) \geq \lambda (\phi(x) - \phi(x_\lambda)) \geq \| x_\lambda - x \|^2 \]
\[ \lambda \phi(x) - \lambda b \geq \| x_\lambda - x \|^2 - \lambda \| z^* \| \| x_\lambda - x \| \]
\[ \geq \| x_\lambda - x \|^2 - \lambda \left( \frac{\| z^* \|^2}{2} + \frac{\| x_\lambda - x \|^2}{2} \right) \]

Thus
\[ \lambda \phi(x) - \lambda b + \lambda \frac{\| z^* \|^2}{2} \geq \left( 1 - \frac{\lambda}{2} \right) \| x_\lambda - x \|^2 \]
It follows that \( x_\lambda \to x \). This shows that \( D(\phi) \subseteq D(\partial \phi) \) and so \( \overline{D(\phi)} \subseteq \overline{D(\partial \phi)} \subseteq D(\phi) \). \( \blacksquare \)

There is a really amazing theorem, Moreau’s theorem. It is in [13], [14] and [15]. It involves approximating a convex function with one which is differentiable, at least in the case where you have a Hilbert space. In the general case considered in this chapter, the function is continuous.
Theorem 27.6.48 Let \( \phi \) be a convex lower semicontinuous proper function defined on \( X \). Define \( A = \partial \phi, A_\lambda = (\partial \phi)_\lambda \)

\[
\phi_\lambda(x) = \min_{y \in X} \left( \frac{1}{2\lambda} \|x - y\|^2 + \phi(y) \right)
\]

Then the function is well defined, convex, Gateaux differentiable,

\[
D_z \phi_\lambda(x) = \lim_{t \downarrow 0} \frac{\phi_\lambda(x+tz) - \phi_\lambda(x)}{t} = \langle A_\lambda x, z \rangle
\]

so the Gateaux derivative is just \( A_\lambda x \) and for all \( x \in X \),

\[
\lim_{\lambda \to 0} \phi_\lambda(x) = \phi(x),
\]

In addition,

\[
\phi_\lambda(x) = \frac{1}{2\lambda} \|x - J_\lambda x\|^2 + \phi(J_\lambda(x))
\]

(27.6.66)

where \( J_\lambda x \) is as before, the solution to

\[
0 \in F(J_\lambda x - x) + \lambda \partial \phi(J_\lambda x)
\]

**Proof:** First of all, why does the minimum take place? By the convexity, closed epigraph, and assumption that \( \phi \) is proper, separation theorems apply and one can say that there exists \( z^* \) such that for all \( y \in H \),

\[
\frac{1}{2\lambda} \|x - y\|^2 + \phi(y) \geq \frac{1}{2\lambda} \|x - y\|^2 + (z^*, y) + c
\]

(27.6.67)

It follows easily that a minimizing sequence is bounded and so from lower semicontinuity which implies weak lower semicontinuity due to convexity, there exists \( y_x \) such that

\[
\min_{y \in H} \left( \frac{1}{2\lambda} \|x - y\|^2 + \phi(y) \right) = \left( \frac{1}{2\lambda} \|x - y_x\|^2 + \phi(y_x) \right)
\]

Why is \( \phi_\lambda \) convex? For \( \theta \in [0, 1] \),

\[
\phi_\lambda(\theta x + (1 - \theta) z) = \frac{1}{2\lambda} \|\theta x + (1 - \theta) z - y(\theta x + (1 - \theta) z)\|^2 + \phi(y(\theta x + (1 - \theta) z))
\]

\[
\leq \frac{1}{2\lambda} \|\theta x + (1 - \theta) z - (\theta y_x + (1 - \theta) y_z)\|^2 + \phi(\theta y_x + (1 - \theta) y_z)
\]

\[
\leq \frac{\theta}{2\lambda} \|x - y_x\|^2 + \frac{1 - \theta}{2\lambda} \|z - y_z\|^2 + \theta \phi(y_x) + (1 - \theta) \phi(y_z)
\]

\[
= \theta \phi_\lambda(x) + (1 - \theta) \phi_\lambda(z)
\]

So is there a formula for \( y_x \)? Since it involves minimization of the functional, it follows that

\[
0 \in -\frac{1}{\lambda} F(x - y_x) + \partial \phi(y_x) = \frac{1}{\lambda} F(y_x - x) + \partial \phi(y_x)
\]

Recall that if \( \psi(x) = \frac{1}{2} \|x\|^2 \), then \( \partial \psi(x) = F(x) \). Thus

\[
y_x = J_\lambda x
\]

because this was how \( J_\lambda x \) was defined. Therefore,

\[
\phi_\lambda(x) = \frac{1}{2\lambda} \|x - J_\lambda x\|^2 + \phi(J_\lambda(x)) = \frac{\lambda}{2} \|A_\lambda x\|^2 + \phi(J_\lambda x), \; A = \partial \phi
\]

It follows from this equation that

\[
\phi(J_\lambda x) \leq \phi_\lambda(x) \leq \phi(x),
\]

(27.6.68)

the second inequality following from taking \( y = x \) in the definition of \( \phi_\lambda \).
Next consider the claim about \( \phi_\lambda (x) \uparrow \phi (x) \). First suppose that \( x \in D (\phi) \). Then from Proposition 27.6.50, \( x \in D (\partial \phi) \) and so from the material on approximations, Theorem 27.6.53, it follows that \( J_\lambda x \to x \). Hence from 27.6.68 and lower semicontinuity of \( \phi \),

\[
\phi (x) \leq \liminf_{\lambda \to 0} \phi (J_\lambda x) \leq \liminf_{\lambda \to 0} \phi_\lambda (x) \leq \limsup_{\lambda \to 0} \phi_\lambda (x) \leq \phi (x)
\]

showing that in this case, \( \lim_{\lambda \to 0} \phi_\lambda (x) = \phi (x) \). Next suppose \( x \notin D (\phi) \) so that \( \phi (x) = \infty \). Why does \( \phi_\lambda (x) \to \infty \)? Suppose not. Then from the description of \( \phi_\lambda \) given above and using the fact that the epigraph is closed and convex, there would exist a subsequence, still denoted as \( \lambda \) such that

\[
C \geq \phi_\lambda (x) = \frac{1}{2\lambda} \| x - J_\lambda x \|^2 + \phi (J_\lambda (x)) \geq \frac{1}{2\lambda} \| x - J_\lambda x \|^2 + \langle z^*, x - J_\lambda x \rangle + b
\]

Then multiplying by \( \lambda \), it follows that for a suitable constant \( M \),

\[
\| x - J_\lambda x \|^2 \leq M\lambda + \lambda M \| x - J_\lambda x \|
\]

and so a use of the quadratic formula implies

\[
\| x - J_\lambda x \| \leq \frac{M}{2} \left( 1 + \sqrt{5} \right) \lambda
\]

Hence \( J_\lambda x \to x \) and so in 27.6.68 it follows from lower semicontinuity again that

\[
\infty = \phi (x) \leq \liminf_{\lambda \to 0} \phi (J_\lambda x) \leq \liminf_{\lambda \to 0} \phi_\lambda (x) \leq \limsup_{\lambda \to 0} \phi_\lambda (x) \leq \phi (x)
\]

and so again, \( \lim_{\lambda \to 0} \phi_\lambda (x) = \infty \). Also note that if \( \lambda > \mu \), then

\[
\min_{y \in X} \left( \frac{1}{2\lambda} \| x - y \|^2 + \phi (y) \right) \leq \min_{y \in X} \left( \frac{1}{2\mu} \| x - y \|^2 + \phi (y) \right)
\]

because for a given \( y \), \( \frac{1}{2\lambda} \| x - y \|^2 + \phi (y) \leq \frac{1}{2\mu} \| x - y \|^2 + \phi (y) \). Thus \( \phi_\lambda (x) \uparrow \phi (x) \).

Next consider the claim about the Gateaux differentiability. Using the description 27.6.69, \( \phi_\lambda (y) - \phi_\lambda (x) = \)

\[
\frac{1}{2\lambda} \| y - J_\lambda y \|^2 + \phi (J_\lambda (y)) - \left( \frac{1}{2\lambda} \| x - J_\lambda x \|^2 + \phi (J_\lambda (x)) \right)
\]

(27.6.69)

Using the fact that if \( \psi (x) = \| x \|^2 \), then \( \partial \psi (x) = Fx \), and that \( A_\lambda x \in \partial \phi (J_\lambda x) \),

\[
\begin{align*}
\geq \quad & \lambda^{-1} \langle F (x - J_\lambda (x)), (y - J_\lambda y) - (x - J_\lambda x) \rangle + \langle A_\lambda x, J_\lambda (y) - J_\lambda (x) \rangle \\
= \quad & \langle A_\lambda (x), (y - J_\lambda y) - (x - J_\lambda x) \rangle + \langle A_\lambda x, J_\lambda (y) - J_\lambda (x) \rangle = \langle A_\lambda x, y - x \rangle
\end{align*}
\]

Hence

\[
(\phi_\lambda (y) - \phi_\lambda (x)) - \langle A_\lambda x, y - x \rangle \geq 0
\]

Also from 27.6.69,

\[
\frac{1}{2\lambda} \| y - J_\lambda y \|^2 - \frac{1}{2\lambda} \| x - J_\lambda x \|^2 = - \left( \frac{1}{2\lambda} \| x - J_\lambda x \|^2 - \frac{1}{2\lambda} \| y - J_\lambda y \|^2 \right)
\]

\[
\leq - \frac{1}{\lambda} \langle F (y - J_\lambda y), (x - J_\lambda x) - (y - J_\lambda y) \rangle = \langle A_\lambda y, (y - J_\lambda y) - (x - J_\lambda x) \rangle
\]

Similarly, from 27.6.69,

\[
\phi (J_\lambda (y)) - \phi (J_\lambda (x)) = -(\phi (J_\lambda (x)) - \phi (J_\lambda (y)))
\]

\[
\leq - \langle A_\lambda (y), J_\lambda (x) - J_\lambda (y) \rangle = \langle A_\lambda (y), J_\lambda (y) - J_\lambda (x) \rangle
\]

It follows that

\[
\langle A_\lambda (y), J_\lambda (y) - J_\lambda (x) \rangle + \langle A_\lambda y, (y - J_\lambda y) - (x - J_\lambda x) \rangle \geq (\phi_\lambda (y) - \phi_\lambda (x)) \geq \langle A_\lambda x, y - x \rangle
\]
and so
\[ \langle A_\lambda(y), y - x \rangle \geq (\phi_\lambda(y) - \phi_\lambda(x)) \geq \langle A_\lambda x, y - x \rangle \]

Therefore,
\[ \langle A_\lambda(y) - A_\lambda(x), y - x \rangle \geq (\phi_\lambda(y) - \phi_\lambda(x)) - \langle A_\lambda x, y - x \rangle \geq 0 \]

Next let \( y = x + tz \) for \( t > 0 \). Then
\[ t \langle A_\lambda(x + tz) - A_\lambda(x), z \rangle \geq (\phi_\lambda(x + tz) - \phi_\lambda(x)) - t \langle A_\lambda x, z \rangle \geq 0 \]

Using the demicontinuity of \( A_\lambda \), you can divide by \( t \) and pass to a limit to obtain
\[
\lim_{t \to 0} \frac{\phi_\lambda(x + tz) - \phi_\lambda(x)}{t} = \langle A_\lambda x, z \rangle. \]

A much better theorem is available in case \( X = X' = H \) a Hilbert space. In this case \( \phi_\lambda \) is also Frechet differentiable. Everything is much nicer in the Hilbert space setting because \( F \) is just replaced with the identity and the approximations are defined more easily.

\[
0 \in J_\lambda x - x + \lambda AJ_\lambda x,
\]
\[
x \in J_\lambda x + \lambda AJ_\lambda x = (I + \lambda A)J_\lambda x
\]
\[
J_\lambda x = (I + \lambda A)^{-1} x
\]

Then one can show that \( J_\lambda \) is Lipschitz continuous and many other nice things happen.

Next is an interesting result about when the sum of a maximal monotone operator and a subgradient is also maximal monotone. A version of this is well known in the case of a single Hilbert space. In the case of a single Hilbert space, this result can be used to produce very regular solutions to evolution equations for functions which have values in the Hilbert space. You would get this by letting \( X = X' = H \) a space of Hilbert space valued functions which are square integrable. Then you could take \( Lu = u' \) with domain equal to those functions in \( X \) which are equal to 0 at the left end of the interval for example. This is done more generally later. In this case the duality map is just the identity. The next theorem includes the case of two different spaces. I am not sure whether this is a useful result at this time, in terms of evolution equations. However, it is good to have conditions which show that the sum of two maximal monotone operators is maximal monotone.

**Theorem 27.6.49** Let \( X \) be a reflexive Banach space with strictly convex norm and let \( \Phi \) be non negative, convex, proper, and lower semicontinuous. Suppose also that \( A : D(A) \to P(X') \) is a maximal monotone operator and there exists \( \xi \in D(A) \cap D(\Phi) \).

Suppose also that
\[ \Phi(J_\lambda x) \leq \Phi(x) + C\lambda \]

Then \( A + \partial \Phi \) is maximal monotone.

**Proof:** Recall that
\[ A_\lambda x = -\lambda^{-1}F(J_\lambda x - x), \text{ where } 0 \in F(J_\lambda x - x) + \lambda \partial A(J_\lambda x) \]

Let \( y^* \in X' \). From Theorem 27.6.31 there exists \( x_\lambda \in H \) such that
\[ y^* \in Fx_\lambda + A_\lambda x_\lambda + \partial \Phi(x_\lambda). \]

It is desired to show that \( A_\lambda x_\lambda \) is bounded. From the above,
\[ y^* - Fx_\lambda - A_\lambda x_\lambda \in \partial \Phi(x_\lambda) \]

and so
\[ \langle y^* - Fx_\lambda - A_\lambda x_\lambda, J_\lambda x_\lambda - x_\lambda \rangle \leq \Phi(J_\lambda x_\lambda) - \Phi(x_\lambda) \leq C\lambda \]

which implies
\[ \langle y^* - Fx_\lambda - A_\lambda x_\lambda, (-\lambda) F^{-1}(A_\lambda x) \rangle \leq \Phi(J_\lambda x_\lambda) - \Phi(x_\lambda) \leq C\lambda \]
Thus there exists a unique element of 

\[ \langle y^* - Fx_\lambda, A\lambda x_\lambda, -F^{-1}(A\lambda x) \rangle \leq C \]

Hence

\[ \langle y^* - Fx_\lambda, -F^{-1}(A\lambda x_\lambda) \rangle + \|A\lambda x_\lambda\|^2 \leq C \]

By Definition 27.7.1 strictly convex norm. It is therefore, assumed that the norm for the reflexive Banach space is strictly convex.

Let \( y^\ast \in V \), \( \langle x, y \rangle \) and \( (\langle x, y \rangle, (u, v)_\lambda) \equiv \langle y^\ast, u \rangle + \langle x, v^\ast \rangle \). It is known that for a reflexive Banach space, there is always an equivalent strictly convex norm. It is therefore, assumed that the norm for the reflexive Banach space is strictly convex.

**Definition 27.7.1** Let \( L : D(L) \subseteq V \rightarrow V' \) be a linear map where we always assume \( D(L) \) is dense in \( V \). Then

\[ D(L^*) \equiv \{ u \in V : \| L^* u \| \leq C \| v \| \text{ for all } v \in D(L) \} \]

For such \( u \), it follows that on a dense subset of \( V \), namely \( D(L), v \rightarrow \langle Lz, u \rangle \) is a continuous linear map. Hence there exists a unique element of \( V^\ast \), denoted as \( L^* u \) such that for all \( v \in D(L) \),

\[ \langle Lv, u \rangle_{V', V} = \langle L^* u, v \rangle_{V^*, V} \]

Thus

\[ L : D(L) \subseteq V \rightarrow V' \]

\[ L^* : D(L^*) \subseteq V \rightarrow V' \]
There is an interesting description of $L^*$ in terms of $L$ which will be quite useful.

**Proposition 27.7.2** Let $\tau : V \times V' \to V' \times V$ be given by $\tau (a, b) \equiv (-b, a)$. Also for $S \subseteq X$ a reflexive Banach space,

$$S^\perp \equiv \{ z^* \in X' : \langle z^*, s \rangle = 0 \text{ for all } s \in S \}$$

Also denote by $G(L) \equiv \{(x, Lx) : x \in D(L)\}$. Then

$$G(L^*) = (\tau G(L))^\perp$$

**Proof:** Let $(x, L^*x) \in G(L^*)$. This means that

$$|(Ly, x)| \leq C \| y \| \text{ for all } y \in D(L)$$

and $(Ly, x) = (L^*x, y)$ for all $y \in D(L)$. Let $(y, Ly) \in G(L)$, then $\tau(y, Ly) = (-Ly, y)$. Then

$$\langle (x, L^*x), (-Ly, y) \rangle = \langle x, -Ly \rangle + \langle L^*x, y \rangle = -\langle x, Ly \rangle + \langle x, L^*y \rangle = 0$$

Thus $G(L^*) \subseteq (\tau G(L))^\perp$. Next suppose $(x, y^*) \in (\tau G(L))^\perp$. This means that if $(u, Lu) \in G(L)$, then

$$\langle (x, y^*), (-Lu, u) \rangle \equiv \langle x, -Lu \rangle + \langle y^*, u \rangle = 0$$

and so for all $u \in D(L)$,

$$\langle y^*, u \rangle = \langle x, Lu \rangle$$

and so $x \in D(L^*)$. Hence for all $u \in D(L)$,

$$\langle y^*, u \rangle = \langle x, Lu \rangle = \langle L^*x, u \rangle$$

Then, since $D(L)$ is dense, it follows that $y^* = L^*x$ and so $(x, y) \in G(L^*)$. Thus these are the same. ■

Theorem 27.3.11 is a very nice surjectivity result for set valued pseudomonotone operators. We recall what it said here. Recall the meaning of coercive.

$$\lim_{\| v \| \to \infty} \inf \left\{ \frac{\langle z^*, v \rangle}{\| v \|} : z^* \in Tv \right\}$$

In this section, we use the convenient notation $\langle z^*, x \rangle_{V', V} \equiv z^*(x)$.

**Theorem 27.7.3** Let $V$ be a reflexive Banach space and let $T : V \to \mathcal{P} (V')$ be pseudomonotone, bounded and coercive. Then $T$ is onto. More generally, this continues to hold if $T$ is modified bounded pseudomonotone.

Recall the definition of pseudomonotone.

**Definition 27.7.4** For $X$ a reflexive Banach space, we say $A : X \to \mathcal{P} (X')$ is pseudomonotone if the following hold.

1. The set $Au$ is nonempty, closed and convex for all $u \in X$.

2. If $F$ is a finite dimensional subspace of $X$, $u \in F$, and if $U$ is a weakly open set in $V'$ such that $Au \subseteq U$, then there exists a $\delta > 0$ such that if $v \in B_\delta (u) \cap F$ then $Av \subseteq U$. (Weakly upper semicontinuous on finite dimensional subspaces.)

3. If $u_i \rightharpoonup u$ weakly in $X$ and $u^*_i \in Au_i$ is such that

$$\limsup_{i \to \infty} \langle u^*_i, u_i - u \rangle \leq 0,$$

then, for each $v \in X$, there exists $u^*(v) \in Au$ such that

$$\liminf_{i \to \infty} \langle u^*_i, u_i - v \rangle \geq \langle u^*(v), u - v \rangle.$$
Also recall the definition of modified bounded pseudomonotone. It is just the above except that the limit condition is replaced with the following condition: If \( u_i \to u \) weakly in \( X \) then there exists a subsequence, still denoted as \( \{u_i\} \) such that if

\[
\limsup_{i \to \infty} \langle u_i^*, u_i - u \rangle \leq 0,
\]

then, for each \( v \in X \), there exists \( u^*(v) \in A u \) such that

\[
\liminf_{i \to \infty} \langle u_i^*, u_i - v \rangle \geq \langle u^*(v), u - v \rangle.
\]

Also recall that this more general limit condition along with the assumption \( \| \) and the assumption that \( A \) is bounded is sufficient to obtain condition \( \| \). This was Lemma 27.4.1 proved earlier and stated here for convenience.

**Lemma 27.7.5** Let \( A : X \to \mathcal{P}(X') \) satisfy conditions \( \| \) and \( \| \) above and suppose \( A \) is bounded. Then if whenever \( x_n \to x \) in \( X \), there is a subsequence \( \{x_{n_k}\} \) such that for this subsequence, the pseudomonotone limit condition

\[
\limsup_{n \to \infty} \langle z_{n_k}, x_{n_k} - x \rangle \leq 0 \implies \liminf_{n \to \infty} \langle z_{n_k}, x_{n_k} - y \rangle \geq \langle z(y), x - y \rangle
\]

for \( z_{n_k} \in Ax_{n_k} \) and \( z(y) \) is some element of \( Ax \). Then if \( U \) is a weakly open set containing \( Ax \), then \( Ax_n \subseteq U \) for all \( n \) large enough.

**Definition 27.7.6** Now let \( L : D(L) \subseteq V \to V' \) such that \( L \) is linear, monotone, \( D(L) \) is dense in \( V \), \( L \) is closed, and \( L^* \) is monotone. Let \( A : V \to \mathcal{P}(V') \) be a bounded operator. Then \( A \) is called \( L \) pseudomonotone if for any sequence \( \{u_n\} \subseteq D(L) \) such that \( u_n \to u \) weakly in \( V \) and \( Lu_n \to Lu \) weakly in \( V' \), and for \( z_n^* \in A u_n \),

\[
\limsup_{n \to \infty} \langle z_n^*, u_n - u \rangle \leq 0
\]

then for every \( v \in V \), there exists \( z^*(v) \in A u \) such that

\[
\liminf_{n \to \infty} \langle z_n^*, u_n - v \rangle \geq \langle z^*(v), u - v \rangle
\]

It is called \( L \) modified bounded pseudomonotone if the above limit condition holds for some subsequence whenever \( u_n \to u \) weakly and \( Lu_n \to Lu \) weakly.

**Lemma 27.7.7** Suppose \( X \) is the Banach space

\[
X = D(L), \| u \|_X \equiv \| u \|_V + \| Lu \|_{V'},
\]

where \( L \) is as described in the above definition. Also assume that \( A \) is bounded. Then if \( A \) is \( L \) pseudomonotone, it follows that \( A \) is pseudomonotone as a map from \( X \) to \( \mathcal{P}(X') \). If \( A \) is \( L \) modified bounded pseudomonotone, then \( A \) is modified bounded pseudomonotone as a map from \( X \) to \( \mathcal{P}(X') \).

**Proof:** Is \( A \) bounded? Of course because the norm of \( X \) is stronger than the norm on \( V \). Is \( Au \) convex and closed? This also follows because \( X \subseteq V \). It is clear that \( Au \) is convex. If \( \{z_n\} \subseteq Au \) and \( z_n \to z \) in \( X' \), then does it follow that \( z \in Au \)? Since \( A \) is bounded, there is a further subsequence which converges weakly to \( w \) in \( V' \). However, \( Au \) is convex and closed so it is weakly closed. Hence \( w \in Au \) and also \( w = z \). It only remains to verify the pseudomonotone limit condition. Suppose then that \( u_n \to u \) weakly in \( X \) and for \( z_n^* \in A u_n \),

\[
\limsup_{n \to \infty} \langle z_n^*, u_n - u \rangle \leq 0
\]

Then it follows that \( Lu_n \to Lu \) weakly in \( V' \) and \( u_n \to u \) weakly in \( V \) so \( u \in X \). Hence the assumption that \( A \) is \( L \) pseudomonotone implies that for every \( v \in X \), and for every \( v \in X \), there exists \( z^*(v) \in Au \subseteq V' \subseteq X' \) such that

\[
\liminf_{n \to \infty} \langle z_n^*, u_n - v \rangle \geq \langle z^*(v), u - v \rangle
\]

The last claim goes the same way. You just have to take a subsequence. ■

Then we have the following major surjectivity result. In this theorem, we will assume for simplicity that all spaces are real spaces. Versions of this appear to be due to Brezis \([12]\) and Lions \([13]\). Of course the theorem holds for complex spaces as well. You just need to use \( \text{Re} \langle \rangle \) instead of \( \langle \rangle \).
Theorem 27.7.8 Let $L : D(L) \subset V \to V'$ where $D(L)$ is dense, $L$ is monotone, $L$ is closed, and $L^*$ is monotone, $L$ a linear map. Let $A : V \to \mathcal{P}(V')$ be $L$ pseudomonotone, bounded, coercive. Then $L + A$ is onto. Here $V$ is a reflexive Banach space such that the norms for $V$ and $V'$ are strictly convex. In case that $A$ is strictly monotone ($\langle Au - Av, u - v \rangle > 0$ implies $u \neq v$) the solution $u$ to $f \in Lu + Au$ is unique. If, in addition to this, $\langle Au - Av, u - v \rangle \geq r (\|u - v\|_V)$ where $U$ is some Banach space containing $V$, and $r$ is a positive strictly increasing function for which $\lim_{t \to 0^+} r(t) = 0$, then the map $f \to u$ where $f \in Lu + Au$ is continuous as a map from $V'$ to $U$.

The conclusion holds if $A$ is only $L$ modified bounded pseudomonotone.

Proof: Let $F$ be the duality map for $p = 2$. Consider the Banach space $X$ given by

$$X = D(L), \quad \|u\|_X = \|u\|_V + \|Lu\|_V'. $$

This is isometric with the graph of $L$ with the graph norm and so $X$ is reflexive. Now define a set valued map $G_e$ on $X$ as follows. $z^* \in G_e (u)$ means there exists $w^* \in Au$ such that

$$\langle z^*, v \rangle_{X',X} = \varepsilon \langle Lw, F^{-1} (Lu) \rangle_{V',V} + \langle Lu, v \rangle_{V',V} + \langle w^*, v \rangle_{V',V}$$

It follows from Lemma 27.7.4 that $G_e$ is the sum of a set valued $L$ modified bounded pseudomonotone operator with an operator which is demicontinuous, bounded, and monotone, hence pseudomonotone. Thus by Lemma 27.7.11 it is $L$ modified bounded pseudomonotone. Is it coercive?

$$\lim_{\|u\|_X \to \infty} \inf \left\{ \frac{\langle z^*, u \rangle + \varepsilon \langle Lu, F^{-1} (Lu) \rangle_{V',V} + \langle Lu, u \rangle_{V',V}}{\|u\|_X} : z^* \in Au \right\} = \infty?$$

It equals

$$\lim_{\|u\|_X \to \infty} \inf \left\{ \frac{\langle z^*, u \rangle + \varepsilon \langle FF^{-1} (Lu), F^{-1} (Lu) \rangle_{V',V} + \langle Lu, u \rangle_{V',V}}{\|u\|_X} : z^* \in Au \right\}$$

and this is

$$\geq \lim_{\|u\|_X \to \infty} \inf \left\{ \frac{\langle z^*, u \rangle + \varepsilon \|F^{-1} (Lu)\|_V^2}{\|u\|_X} : z^* \in Au \right\}$$

$$= \lim_{\|u\|_X \to \infty} \inf \left\{ \frac{\langle z^*, u \rangle + \varepsilon \|Lu\|_V^2}{\|u\|_V + \|Lu\|_V'} : z^* \in Au \right\}$$

because $L$ is monotone. Now let $M$ be an arbitrary positive number. By assumption, there exists $R$ such that if $\|u\|_V > R$, then

$$\inf \left\{ \frac{\langle z^*, u \rangle}{\|u\|_V} : z^* \in Au \right\} > M$$

and so for every $z^* \in Au$,

$$\frac{\langle z^*, u \rangle}{\|u\|_V} > M, \quad \langle z^*, u \rangle > M \|u\|_V$$

Thus if $\|u\|_V > R$, then

$$\inf \left\{ \frac{\langle z^*, u \rangle + \varepsilon \|Lu\|_V^2}{\|u\|_V + \|Lu\|_V'} : z^* \in Au \right\} \geq \frac{M \|u\|_V + \varepsilon \|Lu\|_V^2}{\|u\|_V + \|Lu\|_V'}$$

I claim that if $\|u\|_V$ is large enough, the above is larger than $M/2$. If not, then there exists $\{u_n\}$ such that $\|u_n\|_X \to \infty$ but the right side is less than $M/2$. First say $\|Lu_n\|_V$ is bounded. Then there is an obvious contradiction since the right hand side then converges to $M$. Thus it can be assumed that $\|Lu_n\|_V \to \infty$. Hence, for all $n$ large enough, $\varepsilon \|Lu_n\|_V > M \|Lu_n\|_V$, for all $n$ large enough. However, this implies the right side is larger than

$$M \|u_n\|_V + \varepsilon \|Lu_n\|_V = M > M/2$$

This is a contradiction. Hence the right side is larger than $M/2$ for all $n$ large enough. It follows since $M$ is arbitrary, that

$$\lim_{\|u\|_X \to \infty} \inf \left\{ \frac{\langle z^*, u \rangle + \varepsilon \|Lu\|_V^2}{\|u\|_V + \|Lu\|_V'} : z^* \in Au \right\} = \infty.$$
It follows from Theorem 27.7.11 that if $f \in V'$, there exists $u_\varepsilon$ such that for all $v \in D(L) = X$,

$$\varepsilon \langle Lv, F^{-1}(Lu_\varepsilon) \rangle_{V',V} + \langle Lu_\varepsilon, v \rangle_{V',V} + \langle w^*_\varepsilon, v \rangle_{V',V} = (f, v), \ w^*_\varepsilon \in Au_\varepsilon$$

(27.7.79)

First we get an estimate.

$$\varepsilon \langle Lu_\varepsilon, F^{-1}(Lu_\varepsilon) \rangle_{V',V} + \langle Lu_\varepsilon, u_\varepsilon \rangle_{V',V} + \langle w^*_\varepsilon, u_\varepsilon \rangle_{V',V} = (f, u_\varepsilon)$$

Hence it follows from the coercivity of $A$ that $\|u_\varepsilon\|_V$ is bounded independent of $\varepsilon$. Thus the $w^*_\varepsilon$ are also bounded in $V'$ because it is assumed that $A$ is bounded. Now from the equation solved 27.7.79, it follows that $F^{-1}(Lu_\varepsilon) \in D(L^*)$. Thus the first term is just $\varepsilon \langle L^* (F^{-1}(Lu_\varepsilon)), v \rangle_{V',V}$. It follows, since $D(L) = X$ is dense in $V$ that

$$\varepsilon L^* (F^{-1}(Lu_\varepsilon)) + Lu_\varepsilon + w^*_\varepsilon = f$$

(27.7.80)

Then act on $F^{-1}(Lu_\varepsilon)$ on both sides. From monotonicity of $L^*$, this yields $\|Lu_\varepsilon\|_{V'}$ is bounded independent of $\varepsilon > 0$. Thus there is a subsequence still denoted with a subscript of $\varepsilon$ such that

$$u_\varepsilon \rightharpoonup u \text{ in } V$$

$$Lu_\varepsilon \rightharpoonup Lu \text{ in } V'$$

This because of the fact that the graph of $L$ is closed, hence weakly closed. Thus $u \in X$. Also

$$w^*_\varepsilon \rightharpoonup w^* \text{ in } V'.$$

It follows that we can pass to a limit in 27.7.80 and obtain

$$Lu + w^* = f$$

(27.7.81)

Now by assumption on $A$, it is $L$ modified bounded pseudomonotone and so there is a subsequence, still denoted as $u_\varepsilon$ such that the pseudomonotone limit condition holds. This will be what is referred to in what follows. Then

$$\langle \varepsilon L^* (F^{-1}(Lu_\varepsilon)), u_\varepsilon - u \rangle + \langle Lu_\varepsilon, u_\varepsilon - u \rangle + \langle w^*_\varepsilon, u_\varepsilon - u \rangle = (f, u_\varepsilon - u)$$

and so,

$$\varepsilon \langle F^{-1}(Lu_\varepsilon), Lu_\varepsilon - Lu \rangle + \langle Lu_\varepsilon, u_\varepsilon - u \rangle + \langle w^*_\varepsilon, u_\varepsilon - u \rangle = (f, u_\varepsilon - u)$$

using the monotonicity of $L$,

$$\varepsilon \langle Lu_\varepsilon - Lu, F^{-1}(Lu_\varepsilon) - F^{-1}(Lu) \rangle + \varepsilon \langle Lu_\varepsilon - Lu, F^{-1}(Lu) \rangle$$

$$+ \langle Lu, u_\varepsilon - u \rangle + \langle w^*_\varepsilon, u_\varepsilon - u \rangle \leq (f, u_\varepsilon - u)$$

Now using monotonicity of $F^{-1}$,

$$\varepsilon \langle Lu_\varepsilon - Lu, F^{-1}(Lu) \rangle + \langle Lu, u_\varepsilon - u \rangle + \langle w^*_\varepsilon, u_\varepsilon - u \rangle \leq (f, u_\varepsilon - u)$$

and so, passing to a limit as $\varepsilon \to 0$,

$$\limsup_{\varepsilon \to 0} \langle w^*_\varepsilon, u_\varepsilon - u \rangle \leq 0$$

It follows that for all $v \in X = D(L)$ there exists $w^*(v) \in Au$

$$\liminf_{\varepsilon \to 0} \langle w^*_\varepsilon, u_\varepsilon - v \rangle \geq \langle w^*(v), u - v \rangle$$

But the left side equals

$$\liminf_{\varepsilon \to 0} [\langle w^*_\varepsilon, u_\varepsilon - u \rangle + \langle w^*_\varepsilon, u - v \rangle]$$

$$\leq \limsup_{\varepsilon \to 0} (w^*_\varepsilon, u_\varepsilon - u) + \langle w^*, u - v \rangle \leq \langle w^*, u - v \rangle$$

and so

$$\langle w^*, u - v \rangle \geq \langle w^*(v), u - v \rangle$$
for all $v$.

Is $w^* \in Au$? Suppose not. Then $Au$ is a closed convex set and $w^*$ is not in it. Hence, since $V$ is reflexive, there exists $z \in V$ such that whenever $y^* \in Au, \langle w^*, z \rangle < \langle y^*, z \rangle$. Now simply choose $v$ such that $u - v = z$ and it follows that

$$\langle w^*(v), u - v \rangle > \langle w^*, u - v \rangle \geq \langle w^*(v), u - v \rangle \geq \langle w^*(v), u - v \rangle$$

which is clearly a contradiction. Hence $w^* \in Au$. Thus from Lemma 27.7.8, this has shown that $L + A$ is onto.

Consider the claim about uniqueness and continuous dependence. Say you have $f_i \in Lu_i + Au_i, i = 1, 2$. Let $z_i^* \in Au_i$ be such that equality holds in the two inclusions. Then

$$f_1 - f_2 = z_1^* - z_2^* + Lu_1 - Lu_2$$

It follows that

$$\langle f_1 - f_2, u_1 - u_2 \rangle = \langle z_1^* - z_2^* + Lu_1 - Lu_2, u_1 - u_2 \rangle \geq r(\|u_1 - u_2\|)$$

Thus if $f_1 = f_2$, then $u_1 = u_2$. If $f_n \to f$ in $V'$, then $r(\|u - u_n\|) \to 0$ where $u_n$ goes with $f_n$ and $u$ with $f$ as just described, and so $u_n \to u$ because the coercivity estimate given above shows that the $u_n$ and $u$ are all bounded. Thus the map just described is continuous. ■

The following lemma is interesting in terms of the hypotheses of the above theorem. [28]

**Lemma 27.7.9** Let $L : D(L) \to X'$ where $D(L)$ is dense and $L$ is a closed operator. Then $L$ is maximal monotone if and only if both $L, L^*$ are monotone.

**Proof:** Suppose both $L, L^*$ are monotone. One must show that $\lambda F + L$ is onto. However, $F$ is monotone and hemicontinuous (actually demicontinuous) and coercive. Hence the fact that $\lambda F + L$ is onto follows from Theorem 27.6.2. Next suppose $L$ is maximal monotone. If $L$ is maximal monotone, then for every $\varepsilon > 0$ there exists a solution $u_\varepsilon$ such that $\varepsilon Lu_\varepsilon + F(u_\varepsilon - u) = 0$. Here $u \in D(L^*)$. This is from Lemma 27.6.2. It is originally due to Browder [21]. Then

$$\varepsilon \langle Lu_\varepsilon, u_\varepsilon \rangle + \langle F(u_\varepsilon - u), u_\varepsilon \rangle = 0$$

and so $\langle F(u_\varepsilon - u), u_\varepsilon \rangle \leq 0$. Then

$$\langle F(u_\varepsilon - u), u_\varepsilon - u \rangle \leq \langle F(u_\varepsilon - u), u \rangle$$

so $\|u_\varepsilon - u\|^2 \leq \|u_\varepsilon - u\| \|u\|$ and so

$$\|u_\varepsilon - u\| \leq \|u\|$$

Thus the $u_\varepsilon$ are bounded.

Next let $v \in D(L)$.

$$\|u_\varepsilon - u\|^2 = \langle F(u_\varepsilon - u), u_\varepsilon - u \rangle = \langle F(u_\varepsilon - u), u_\varepsilon - v \rangle + \langle F(u_\varepsilon - u), v - u \rangle$$

Hence

$$\limsup_{\varepsilon \to 0} \|u_\varepsilon - u\|^2 \leq \limsup_{\varepsilon \to 0} (\varepsilon \langle Lv, v - u_\varepsilon \rangle + \langle F(u_\varepsilon - u), v - u \rangle)$$

and so $u_\varepsilon \to u$ strongly. Also

$$\langle F(u_\varepsilon - u), u_\varepsilon \rangle = -\varepsilon \langle Lu_\varepsilon, u_\varepsilon \rangle \leq 0$$

Then

$$\langle L^*u, u \rangle = \lim_{\varepsilon \to 0} \langle L^*u, u_\varepsilon \rangle = \lim_{\varepsilon \to 0} \langle Lu_\varepsilon, u \rangle = \lim_{\varepsilon \to 0} (\frac{1}{\varepsilon} \langle F(u_\varepsilon - u), u \rangle)$$

Both of these last terms are nonnegative, the first obviously and the second from the above where it was shown that $\langle F(u_\varepsilon - u), u_\varepsilon \rangle \leq 0$. ■

In the hypotheses of Theorem 27.7.8, one could have simply said that $L$ is closed, linear, densely defined and maximal monotone. One can also show that if $L$ is maximal monotone, then it must be densely defined. This is done in [18].
27.7. PERTURBATION THEOREMS

One can go further in obtaining a perturbation theorem like the above. Let linear $L$ be densely defined with $L$ closed and $L, L^*$ monotone. In short, $L$ is densely defined and maximal monotone, $L : X \to X'$. Let $A$ be a set valued $L$ pseudomonotone operator which is coercive and bounded. Also let $B : D(B) \to \mathcal{P}(X)$ be maximal monotone. It is of interest to consider whether $L + A + B$ is onto $X'$. In considering this, I will add further assumptions as needed. First note that $\langle Lx, x \rangle = \langle Lx - L0, x - 0 \rangle \geq 0$.

**Definition 27.7.10** Define $\limsup_{m,n \to \infty} a_{m,n} = \lim_{k \to \infty} \sup \{a_{m,n} : \min(m, n) \geq k\}

Then $\limsup_{m,n \to \infty} a_{m,n} \geq \limsup_{n \to \infty} \left(\limsup_{m \to \infty} a_{m,n}\right)$. To see this, suppose $a > \limsup_{m,n \to \infty} a_{m,n}$. Then there exist $k$ such that whenever $m, n > k$,

$$a_{m,n} < a$$

It follows that for $m \geq k$,

$$\lim_{n \to \infty} \sup a_{m,n} \leq a$$

Hence

$$\lim_{m \to \infty} \left(\limsup_{n \to \infty} a_{m,n}\right) \leq a$$

Since $a > \limsup_{m,n \to \infty} a_{m,n}$ is arbitrary, it follows that

$$\lim_{m \to \infty} \left(\limsup_{n \to \infty} a_{m,n}\right) \leq \limsup_{m,n \to \infty} a_{m,n}.$$

Then the following lemma is useful. I found this result in a paper by Gasinski, Migorski and Ochal [4]. They begin with the following interesting lemma or something like it which is similar to some of the ideas used in the section on approximation of maximal monotone operators.

**Lemma 27.7.11** Suppose $A$ is a set valued operator, $A : X \to \mathcal{P}(X)$ and $u^*_n \in Au_n$. Suppose also that $u_n \to u$ weakly and $u^*_n \to u^*$ weakly. Suppose also that

$$\lim_{m,n \to \infty} \langle u^*_n - u^*_m, u_n - u_m \rangle \leq 0$$

Then one can conclude that

$$\limsup_{n \to \infty} \langle u^*_n, u_n - u \rangle \leq 0$$

**Proof:** Let $\alpha = \limsup_{n \to \infty} \langle u^*_n, u_n \rangle$. It is a finite number because these sequences are bounded. Then using the weak convergence,

\[
0 \geq \limsup_{m \to \infty} \left(\limsup_{n \to \infty} \langle u^*_n - u^*_m, u_n - u_m \rangle\right)
\]

\[
= \limsup_{m \to \infty} \left(\limsup_{n \to \infty} \langle u^*_n, u_n \rangle + \langle u^*_m, u_n \rangle - \langle u^*_m, u_m \rangle - \langle u^*_m, u_n \rangle\right)
\]

\[
= \limsup_{m \to \infty} \left(\alpha + \langle u^*_m, u_m \rangle - \langle u^*_m, u_m \rangle - \langle u^*_m, u \rangle\right)
\]

\[
= \alpha + \alpha - \langle u^*, u \rangle - \langle u^*, u \rangle = 2\alpha - 2\langle u^*, u \rangle
\]

Now

$$\limsup_{n \to \infty} \langle u^*_n, u_n - u \rangle = \alpha - \langle u^*, u \rangle \leq 0.\]$$

To begin with, consider the approximate problem which is to determine whether $L + A + B$ is onto. Here $B_\lambda x = -\lambda^{-1} F(x_\lambda - x)$ where $0 \in F(x_\lambda - x) + \lambda Bx$. In the notation given above, $B_\lambda x = -\lambda^{-1} F(J_\lambda x - x)$. Then by Theorem 27.7.31, $B_\lambda$ is monotone, demicontinuous, and bounded. In addition, we assume $0 \in D(B)$. Then

$$\langle B_\lambda x, x \rangle \geq \langle B_0, 0 \rangle \geq -|B(0)||x|| \quad (27.7.82)$$

**Lemma 27.7.12** Let $A$ be pseudomonotone, bounded and coercive and let $0 \in D(B)$. Then if $y^* \in X'$, there exists a solution $x_\lambda$ to

$$y^* \in Lx_\lambda + Ax_\lambda + B_\lambda x_\lambda$$
Proof: From the inequality, $A + B$ is coercive. It is also bounded and pseudomonotone. It is pseudomonotone from Theorem 27.6.24. Therefore, there exists a solution $x_\lambda$ by Theorem 27.7.8. ■

Acting on $x_\lambda$ and using the inequality, it follows that these solutions $x_\lambda$ lie in a bounded set. The details follow. Letting $z^*_\lambda \in Ax_\lambda$ be such that equality holds in the above inclusion,

$$y^* = Lx_\lambda + z^*_\lambda + Bx_\lambda \quad (27.7.83)$$

$$\|y^*\| \geq \frac{\langle y^*, x_\lambda \rangle}{\|x_\lambda\|} = \frac{\langle Lx_\lambda, x_\lambda \rangle + \langle z^*_\lambda, x_\lambda \rangle + \langle Bx_\lambda, x_\lambda \rangle}{\|x_\lambda\|} \geq \frac{\langle Lx_\lambda, x_\lambda \rangle + \langle z^*_\lambda, x_\lambda \rangle - |B(0)|\|x_\lambda\|}{\|x_\lambda\|} \geq \frac{\langle z^*_\lambda, x_\lambda \rangle}{\|x_\lambda\|} - |B(0)| \quad (27.7.84)$$

Thus, from coercivity, $\|x_\lambda\|$ are bounded. Then since $A$ is bounded, the $z^*_\lambda$ are all bounded also independent of $\lambda$. The top line shows also that

$$\langle y^*, x_\lambda \rangle = \langle Lx_\lambda, x_\lambda \rangle + \langle z^*_\lambda, x_\lambda \rangle + \langle Bx_\lambda, x_\lambda \rangle \geq \langle z^*_\lambda, x_\lambda \rangle + \langle Bx_\lambda, x_\lambda \rangle \geq \langle Bx_\lambda, x_\lambda \rangle - M \geq - |B(0)|\|x_\lambda\| - \hat{M}$$

where $|\langle z^*_\lambda, x_\lambda \rangle| \leq \hat{M}$ for all $\lambda$. Hence there is a constant $M$ such that

$$|\langle Bx_\lambda, x_\lambda \rangle| \leq M$$

Definition 27.7.13 A set valued operator $B$ is quasi-bounded if whenever $x \in D(B)$ and $x^* \in Bx$ are such that

$$|\langle x^*, x \rangle|, \|x\| \leq M,$$

it follows that $\|x^*\| \leq K_M$. Bounded would mean that if $\|x\| \leq M$, then $\|x^*\| \leq K_M$. Here you only know this if there is another condition.

By Proposition 27.6.18 an example of a quasi-bounded operator is a maximal monotone operator $G$ for which $0 \in \text{int}(D(G))$.

Lemma 27.7.14 In the above situation, suppose the maximal monotone operator $B$ is quasi-bounded and $|\langle Bx_\lambda, x_\lambda \rangle| \leq M$. Then the $Bx_\lambda$ are bounded. Also

$$\|J_\lambda x_\lambda - x_\lambda\|^2 \leq M \lambda$$

Proof: Now $Bx_\lambda \in BJ_\lambda x_\lambda$

$$- |B(0)|\|x_\lambda\| \leq \langle Bx_\lambda, x_\lambda \rangle = \langle Bx_\lambda, J_\lambda x_\lambda \rangle + \langle Bx_\lambda, x_\lambda - J_\lambda x_\lambda \rangle \leq \langle Bx_\lambda, J_\lambda x_\lambda \rangle + \langle \lambda^{-1} F(J_\lambda x_\lambda - x_\lambda), J_\lambda x_\lambda - x_\lambda \rangle \leq \langle Bx_\lambda, J_\lambda x_\lambda \rangle + \lambda^{-1} \|J_\lambda x_\lambda - x_\lambda\|^2 \leq M$$

This inequality shows that $J_\lambda x_\lambda - x_\lambda \rightarrow 0$ and so $J_\lambda x_\lambda$ is bounded as is $x_\lambda$. Also $Bx_\lambda \in BJ_\lambda x_\lambda$ and since $B$ is quasi-bounded, it follows that $Bx_\lambda$ is bounded. ■

Assume from now on that $B$ is quasi-bounded. Then the estimate and this lemma shows that $Bx_\lambda$ is also bounded independent of $\lambda$. Thus, adjusting the constants, there exists an estimate of the form

$$\|x_\lambda\| + \|J_\lambda x_\lambda\| + \|Bx_\lambda\| + \|z^*_\lambda\| + \|Lx_\lambda\| \leq C, \|x_\lambda - J_\lambda x_\lambda\| \leq \sqrt{\lambda}$$

(27.7.85)

Let $\lambda = 1/n$. Also denote by $J_n$ the the operator $J_{1/n}$ to save notation. There exists a subsequence

$$x_n \rightarrow x \text{ weakly},$$

$$J_n x_n \rightarrow x \text{ weakly},$$

$$B_n x_n \rightarrow g^* \text{ weakly},$$

$$z^*_n \rightarrow z^* \text{ weakly},$$

$$z^*_n \rightarrow z^* \text{ weakly},$$
\[ Lx_n \rightarrow Le \text{ weakly} \]

Now from the inclusion satisfied,
\[
0 = \langle z^*_n - z^*_m, x_n - x_m \rangle + \langle B_n x_n - B_m x_m, x_n - x_m \rangle \tag{27.7.86}
\]

Consider that last term. \( B_n x_n \in BJ_n x_n \) similar for \( B_m x_m \). Hence this term is of the form
\[
\langle B_n x_n - B_m x_m, x_n - x_m \rangle = (B_n x_n - B_m x_m, J_n x_n - J_m x_m)
\]

From the estimate \[\text{Lemma}\]
\[
\langle B_n x_n - B_m x_m, x_n - x_m \rangle \geq \langle B_n x_n - B_m x_m, (x_n - J_n x_n) - (x_m - J_m x_m) \rangle
\]
and
\[
|\langle B_n x_n - B_m x_m, (x_n - J_n x_n) - (x_m - J_m x_m) \rangle| \leq 2C \left( \sqrt{\frac{1}{n}} + \sqrt{\frac{1}{m}} \right)
\]

Then from \[\text{Lemma}\]
\[
0 \geq \langle z^*_n - z^*_m, x_n - x_m \rangle + e_{n,m}
\]
where \( e_{n,m} \rightarrow 0 \) as \( n, m \rightarrow \infty \). Hence
\[
\lim_{m,n \rightarrow \infty} \sup\ (z^*_n - z^*_m, x_n - x_m) \leq 0
\]

From Lemma \[\text{Lemma}\]
\[
\lim_{n \rightarrow \infty} \sup\ (z^*_n, x_n - x) \leq 0
\]
Hence, since \( A \) is pseudomonotone, for every \( y \), there exists \( z^* (y) \in Ax \) such that
\[
\lim_{n \rightarrow \infty} \inf\ (z^*_n, x_n - y) \geq (z^* (y), x - y)
\]

In particular, if \( x = y \), this shows that
\[
\lim_{n \rightarrow \infty} \inf\ (z^*_n, x_n - x) \geq 0 \geq \lim_{n \rightarrow \infty} \sup\ (z^*_n, x_n - x)
\]
showing that
\[
\lim_{n \rightarrow \infty} (z^*_n, x_n) = (z^*, x)
\]

Next, returning to the inclusion solved,
\[
0 = Lx_n + z^*_n + B_n x_n
\]
Act on \((x_n - x)\). Then from monotonicity of \( L \),
\[
0 \geq (Lx_n, x_n - x) + (z^*_n, x_n - x) + (B_n x_n, x_n - x)
\]
Thus, taking \( \limsup \) of both sides,
\[
\lim_{n \rightarrow \infty} \sup\ (B_n x_n, x_n - x) = \lim_{n \rightarrow \infty} \sup\ (B_n x_n, J_n x_n - x) \leq 0
\]
Hence
\[
\lim_{n \rightarrow \infty} \sup\ (B_n x_n, J_n x_n) \leq \langle g^*, x \rangle
\]
Letting \([a, b^*] \in G (B)\),
\[
(B_n x_n - b^*, J_n x_n - a) = (B_n x_n, J_n x_n) - (B_n x_n, a) - \langle b^*, J_n x_n \rangle + \langle b^*, a \rangle
\]
Then taking $\lim sup$

$$0 \leq \lim sup_{n \to \infty} \langle B_n x_n - b^*, J_n x_n - a \rangle$$

$$\leq \langle g^*, x \rangle - \langle g^*, a \rangle - \langle b^*, x \rangle + \langle b^*, a \rangle = \langle g^* - b^*, x - a \rangle$$

It follows that $g^* \in B(x)$ and $x \in D(B)$.

Thus, passing to the limit in the equation 27.7.83 where, as explained $\lambda = 1/n$, one obtains

$$y^* = Lu + z^* + g^*$$

where $z^* \in Ax$ and $g^* \in Bx$. This proves the following nice generalization of the above perturbation theorem.

**Theorem 27.7.15** Let $B$ be maximal monotone from $X$ to $P(X')$, $0 \in D(B)$, and $B$ is quasi-bounded as explained above. Let $A : X \to P(X')$ be pseudomonotone, bounded, and coercive. Also let $L$ be a densely defined linear operator such that both $L$ and $L^*$ are monotone. (That is, $L$ is linear and maximal monotone.) Then $L + A + B$ is onto $X'$. 
Chapter 28

Implicit Stochastic Equations

28.1 Introduction

In this chapter, implicit evolution equations are considered. These are of the form

\[ Bu(t, \omega) - Bu_0(\omega) + \int_0^t A(s, u(t, \omega), \omega) \, ds = \int_0^t f(s) \, ds + B \int_0^t \Phi \, dW \]

the term on the end being a stochastic integral. The novelty is in allowing \( B \) to be an operator which could vanish or have other interesting features. Thus the integral equation could degenerate to a non stochastic elliptic equation. This generalization of evolution equations has proven useful in the study of deterministic evolution equations and we give some interesting examples which indicate that this may be true in the case of stochastic equations also. In any case, it is an interesting generalization and equations of the usual form are recovered by using a Gelfand triple in which \( B = I \).

Like deterministic equations, there are many ways to consider stochastic equations. Here it is based on an approach due to Bardos and Brezis \[12\] which avoids the consideration of finite dimensional problems. A generalized Ito formula is summarized in the next section. It is Theorem 28.2.3.

28.2 Preliminary Results

Let \( X \) have values in \( W \) and satisfy the following

\[ BX(t) = BX_0 + \int_0^t Y(s) \, ds + B \int_0^t Z(s) \, dW(s), \quad (28.2.1) \]

\( X_0 \in L^2(\Omega; W) \) and is \( \mathcal{F}_0 \) measurable, where \( Z \) is \( L_2(Q^{1/2}U, W) \) progressively measurable and

\[ \|Z\|_{L^2([0,T] \times \Omega, \mathcal{L}_2(Q^{1/2}U, W))} < \infty. \]

This is what is needed to define the stochastic integral in the above formula. Here \( Q \) is a nonnegative self adjoint operator defined on a separable real Hilbert space \( U \). In what follows, \( J \) will denote a one to one Hilbert Schmidt operator mapping \( Q^{1/2}U \) into another separable Hilbert space \( U_1 \). For more explanation on this situation see \[72\].

Assume \( X,Y \) satisfy

\[ X \in K \equiv L^p([0,T] \times \Omega; V), \quad Y \in K' = L^{p'}([0,T] \times \Omega; V') \]

where \( 1/p' + 1/p = 1, p > 1 \), and \( X,Y \) are progressively measurable into \( V \) and \( V' \) respectively.

The sense in which the equation \[28.2.1\] holds is as follows. For a.e. \( \omega \), the equation holds in \( V' \) for all \( t \in [0,T] \). Assume that

\[ X \in L^2([0,T] \times \Omega, W), \]

\[ BX \in L^2([0,T] \times \Omega, \mathcal{B}([0,T]) \times \mathcal{F}, W'), \quad X \in L^p([0,T] \times \Omega, \mathcal{B}([0,T]) \times \mathcal{F}, V) \]
Note that, since $X$ is progressively measurable into $V$, this implies that $BX$ is progressively measurable into $W'$. Also $W(t)$ is a $JJ^*$ Wiener process on $U_1$ in the following diagram. ($W$ is a cylindrical Wiener process.)

\[
\begin{array}{c}
U \\
\downarrow Q^{1/2} \\
J_{1-1} \\
\downarrow Q^{1/2}U \\
\downarrow \Phi \\
W
\end{array}
\]

We will also make use of the following generalization of familiar concepts from Hilbert space.

**Lemma 28.2.1** Suppose $V, W$ are separable Banach spaces, $W$ also a Hilbert space such that $V$ is dense in $W$ and $B \in \mathcal{L}(W, W')$ satisfies

\[
\langle Bx, x \rangle \geq 0, \quad \langle Bx, y \rangle = \langle By, x \rangle, \quad B \neq 0.
\]

Then there exists a countable set \{e$_i$\} of vectors in $V$ such that

\[
\langle Be_i, e_j \rangle = \delta_{ij}
\]

and for each $x \in W$,

\[
\langle Bx, x \rangle = \sum_{i=1}^{\infty} |\langle Bx, e_i \rangle|^2,
\]

and also

\[
Bx = \sum_{i=1}^{\infty} \langle Bx, e_i \rangle Be_i,
\]

the series converging in $W'$.

Then in the above situation, we have the following fundamental estimate.

**Lemma 28.2.2** In the above situation where, off a set of measure zero, 28.2.1 holds for all $t \in [0, T]$, and $X$ is progressively measurable into $V$,

\[
E \left( \sup_{t \in [0, T]} \langle BX, X \rangle (t) \right) < C \left( ||Y||_{K'}, ||X||_K, ||Z||_{J'}, ||\langle BX_0, X_0 \rangle||_{L^1(\Omega)} \right) < \infty.
\]

where $\langle BX, X \rangle (t) = \langle B(X(t)), X(t) \rangle$ a.e. and $\langle BX, X \rangle$ is progressively measurable and continuous in $t$.

\[
J = L^2 \left( [0, T] \times \Omega; \mathcal{L}_2 \left( Q^{1/2}U; W' \right) \right), K \equiv L^p ([0, T] \times \Omega; V), \quad K' \equiv L^{p'} ([0, T] \times \Omega; V').
\]

Also, $C$ is a continuous function of its arguments and $C (0, 0, 0, 0) = 0$. Thus for a.e. $\omega$,

\[
\sup_{t \in [0, T]} \langle BX, X \rangle (t) \leq C (\omega) < \infty.
\]

For a.e. $\omega, t \to BX(t, \omega)$ is weakly continuous with values in $W'$ for $t$ off a set of measure zero. Also $t \to \langle BX(t), X(t) \rangle$ is lower semicontinuous off a set of measure zero.

Then from this fundamental lemma, the following Itô formula is valid. The proof of this theorem follows the same methods used for a similar result in 72.

**Theorem 28.2.3** Off a set of measure zero, for every $t \in [0, T]$,

\[
\langle BX, X \rangle (t) = \langle BX_0, X_0 \rangle + \int_0^t \left( 2 \langle Y(s), X(s) \rangle + \langle BZ, Z \rangle \right) ds.
\]
28.2. PRELIMINARY RESULTS

\[ + 2 \int_0^t (Z \circ J^{-1})^* BX \circ JdW \]  

(28.2.2)

Also

\[ E \langle BX, X \rangle (t) = E \langle BX_0, X_0 \rangle + E \left( \int_0^t (2 \langle Y(s), X(s) \rangle + \langle BZ, Z \rangle_{L^2}) \, ds \right) \]  

(28.2.3)

The quadratic variation of the stochastic integral is dominated by

\[ C \int_0^t \|Z\|^2_{L^2} \|BX\|^2_{W'} \, ds \]  

(28.2.4)

for a suitable constant C. Also \( t \to BX(t) \) is continuous with values in \( W' \) for \( t \in N_C \).

We will often abuse the notation and write \( \langle BX(t), X(t) \rangle \) instead of the more precise \( \langle BX, X \rangle (t) \). No harm is done because these two are equal a.e.

In addition to the above, we will use the following basic theorems about nonlinear operators. This is Proposition 27.1.8 above.

**Proposition 28.2.4** Suppose \( A : V \to V' \) is type M, see \([73]\), and suppose \( L : V \to V' \) is monotone, bounded and linear. Here \( V \) is a separable reflexive Banach space. Then \( L + A \) is type M.

As an important example, we give the following definition.

**Definition 28.2.5** Let \( f : [0,T] \times \Omega \to V \)

\[ \tau_h f (t, \omega) = \begin{cases} f(t-h, \omega) & \text{if } t \geq h \\ 0 & \text{if } t < h \end{cases} \]

Then letting \( B \) be a monotone nonnegative, self adjoint operator, \( B : W \to W' \) for \( W \) a separable Hilbert space, consider the linear operator \( L : L^2(0,T,W) \equiv W \to L^2(0,T,W') \equiv W' \) given as

\[ Lu \equiv \left( \frac{I - \tau_h}{h} \right) Bu. \]

Is it the case that \( L \) is monotone? Clearly it is linear and so it suffices to consider \( \langle Lu, u \rangle_{W', W} \) which equals

\[ \frac{1}{h} \int_0^T \langle Bu(t), u(t) \rangle \, dt - \frac{1}{h} \int_h^T \langle Bu(t-h), u(t) \rangle \, dt \]

\[ = \frac{1}{h} \int_0^T \langle Bu(t), u(t) \rangle \, dt - \frac{1}{h} \int_0^{T-h} \langle Bu(t), u(t+h) \rangle \, dt \]

\[ \geq \frac{1}{h} \int_0^T \langle Bu(t), u(t) \rangle \, dt \]

\[ - \frac{1}{h} \int_0^{T-h} \left( \frac{1}{2} \langle Bu(t), u(t) \rangle + \frac{1}{2} \langle Bu(t+h), u(t+h) \rangle \right) \, dt \]

\[ = \frac{1}{2h} \int_0^{T-h} \langle Bu(t), u(t) \rangle \, dt + \frac{1}{h} \int_T^{T-h} \langle Bu(t), u(t) \rangle \, dt \]

\[ - \frac{1}{2h} \int_0^{T-h} \langle Bu(t+h), u(t+h) \rangle \, dt \]

\[ = \frac{1}{2h} \int_0^{T-h} \langle Bu(t), u(t) \rangle \, dt + \frac{1}{h} \int_{T-h}^{T} \langle Bu(t), u(t) \rangle \, dt \]

\[ - \frac{1}{2h} \int_h^{T} \langle Bu(t), u(t) \rangle \, dt \]
\[
\frac{1}{2h} \int_{T}^{T-h} \langle B(t), u(t) \rangle \, dt + \frac{1}{2h} \int_{0}^{h} \langle B(t), u(t) \rangle \, dt
\]

\[
+ \frac{1}{h} \int_{T-h}^{T} \langle B(t), u(t) \rangle \, dt - \frac{1}{2h} \int_{0}^{h} \langle B(t), u(t) \rangle \, dt
\]

\[
= \frac{1}{2h} \int_{T}^{T-h} \langle B(t), u(t) \rangle \, dt + \frac{1}{h} \int_{T-h}^{T} \langle B(t), u(t) \rangle \, dt
\]

\[
- \frac{1}{2h} \int_{h}^{T-h} \langle B(t), u(t) \rangle \, dt
\]

\[
+ \frac{1}{2h} \int_{0}^{h} \langle B(t), u(t) \rangle \, dt - \frac{1}{2h} \int_{h}^{T-h} \langle B(t), u(t) \rangle \, dt
\]

\[
= \frac{1}{2h} \int_{T-h}^{T} \langle B(t), u(t) \rangle \, dt + \frac{1}{2h} \int_{0}^{h} \langle B(t), u(t) \rangle \, dt \geq 0 \tag{28.2.5}
\]

The following is a restatement of Theorem 27.1.13.

**Theorem 28.2.6** Let \( A : V \to V' \) be type \( M \), bounded, and coercive

\[
\lim_{\|u\| \to \infty} \frac{\langle A(u + u_0), u \rangle}{\|u\|} = \infty,
\]

for some \( u_0 \in V \), where \( V \) is a separable reflexive Banach space. Then \( A \) is surjective.

In addition, there is a fundamental definition and theorem about weak derivatives which will be used.

**Definition 28.2.7** Let \( f \in L^1(a,b, V') \) where \( V' \) is the dual of a Banach space \( V \). Let \( D^* (a,b) \) linear mappings from \( C^\infty_c (a,b) \) to \( V' \). Then we can consider \( f \in D^* (a,b) \), the linear transformations defined on \( C^\infty_c (a,b) \) as follows.

\[
f(\phi) \equiv \int_{a}^{b} f \phi ds
\]

This is well defined due to regularity considerations for Lebesgue measure. Then define \( Df \in D^* (a,b) \) by

\[
Df(\phi) \equiv - \int_{a}^{b} f \phi' ds
\]

To say that \( Df \in L^1(a,b,V') \) is to say that there exists \( g \in L^1(a,b,V') \) such that

\[
Df(\phi) \equiv - \int_{a}^{b} f \phi' ds = \int_{a}^{b} g \phi ds
\]

for all \( \phi \in C^\infty_c (a,b) \). Note that regularity considerations imply that \( g \) is unique if it exists.

The following is Theorem 21.2.9.

**Theorem 28.2.8** Suppose that \( f \) and \( Df \) are both in \( L^1(a,b,V') \). Then \( f \) is equal to a continuous function a.e., still denoted by \( f \) and

\[
f(x) = f(a) + \int_{a}^{x} Df(t) \, dt.
\]

In the next section are theorems about how shifts in time relate to progressive measurability.
28.3 The Existence Of Approximate Solutions

The situation is as follows. There are spaces \( V \subseteq W \) where \( V \) is a reflexive separable Banach space and \( W \) is a separable Hilbert space. It is assumed that \( V \) is dense in \( W \). Define the spaces

\[
V = L^p([0, T] \times \Omega, V), \quad W = L^2([0, T] \times \Omega, W)
\]

where in each case, the \( \sigma \) algebra of measurable sets will be the progressively measurable sets. Thus, from the Riesz representation theorem,

\[
V' = L^{p'}([0, T] \times \Omega, V'), \quad W' = L^2([0, T] \times \Omega, W')
\]

It will be assumed for the sake of convenience that \( p \geq 2 \). It follows that

\[
V \subseteq W; \quad W' \subseteq V'
\]

The entire presentation will be based on the following lemma.

**Lemma 28.3.1** Let \( V = L^p([0, T] \times \Omega, V) \) where \( V \) is a separable Banach space and the \( \sigma \) algebra of measurable sets consists of those which are progressively measurable. Then for \( h \in (0, T) \), \( \tau_h : V \rightarrow V \).

**Proof:** First consider \( Q \) which is a progressively measurable set. Is it the case that \( \tau_h X_Q \) is also progressively measurable? Define \( Q + h \) as

\[
Q + h = \{(t + h, \omega) : (t, \omega) \in Q\}
\]

Then

\[
\tau_h X_Q(t, \omega) = \begin{cases} 
X_{Q + h}(t, \omega) & \text{if } t \geq h \\
0 & \text{if } t < h
\end{cases}
\]

Is this function progressively measurable? For \((s, \omega) \in [0, t] \times \Omega\), we have the following

\[
0 < \alpha \leq 1, [(s, \omega) : \tau_h X_Q(s, \omega) \geq \alpha] = [h, t] \times \Omega \cap (Q + h)
\]

\[
\alpha > 1, [(s, \omega) : \tau_h X_Q(s, \omega) \geq \alpha] = \emptyset \in B([0, t]) \times F_t
\]

\[
\alpha \leq 0, [(s, \omega) : \tau_h X_Q(s, \omega) \geq \alpha] = [0, t] \times \Omega \cap \emptyset = \emptyset \in B([0, t]) \times F_t
\]

It suffices to show that for \( t \geq h \), \([h, t] \times \Omega \cap (Q + h) \) is \( B([0, t]) \times F_t \) measurable. It is known that \([0, t] \times \Omega \cap Q \) is \( B([0, t]) \times F_t \) measurable and also that \([0, t - h] \times \Omega \cap Q \) is \( B([0, t - h]) \times F_{t-h} \) measurable. Let

\[
\mathcal{G} = \{Q \in B([0, t - h]) \times F_{t-h} : [h, t] \times \Omega \cap Q + h \in B([0, t]) \times F_t\}
\]

First consider \( I \times B \) where \( I \) is an interval in \( B([0, t - h]) \) and \( B \in F_{t-h} \). Then

\[
[h, t] \times \Omega \cap (I + h) \times B = I' \times B
\]

where \( I' \) is in \( B([0, t]) \). Thus the sets of this form, are in \( \mathcal{G} \). Next suppose \( Q \in \mathcal{G} \). Is \( Q^C \in \mathcal{G} \)?

\[
[h, t] \times \Omega \cap (Q^C + h) \cup [h, t] \times \Omega \cap (Q + h) \cup [0, h] \times \Omega = [0, t] \times \Omega
\]

Then all of these disjoint sets but the first are in \( B([0, t - h]) \times F_{t-h} \). It follows that the first is also in \( B([0, t]) \times F_t \). It is clear that \( \mathcal{G} \) is also closed with respect to countable disjoint unions. Therefore, \( \mathcal{G} \) contains the \( \pi \) system of sets of the form \( I \times B \) just described. It follows that \( \mathcal{G} = B([0, t - h]) \times F_{t-h} \). Now if \( Q \) is progressively measurable, then \([0, t - h] \times \Omega \cap Q \) is \( B([0, t - h]) \times F_{t-h} \) measurable and so from what was just shown, \([h, t] \times \Omega \cap Q + h \in B([0, t]) \times F_t \). Thus \( \tau_h X_Q \) is progressively measurable. It follows that if \( f \in V \), you could consider \( \phi(f) \) for \( \phi \in V' \) and the positive and negative parts of this function. Each of these is the limit of a sequence of simple functions involving combinations of indicator functions of the form \( X_Q \). Thus \( \tau_h \phi(f) = \phi(\tau_h f) \) is the limit of simple functions involving combinations of functions \( \tau_h X_Q \) and, as just shown, these simple functions are progressively measurable. Thus \( \tau_h f \) is also progressively measurable by the Pettis theorem.

This Lemma states that you can do \( \tau_h \) to progressively measurable functions and end up with one which is progressively measurable. Let

\[
B \in \mathcal{L}(W, W')
\]

satisfy

\[
\langle Bx, y \rangle = \langle By, x \rangle, \quad \langle Bx, x \rangle \geq 0 \quad (28.3.7)
\]
Also suppose that

\[ A \text{ is monotone and hemicontinuous from } \mathcal{V} \text{ to } \mathcal{V}' \quad (28.3.8) \]

This means the operator is monotone:

\[ \langle Au - Au, u - v \rangle_{\mathcal{V}', \mathcal{V}} \geq 0 \]

and hemicontinuous:

\[ \lim_{t \to 0} \langle A(u + tv), w \rangle_{\mathcal{V}', \mathcal{V}} = \langle Au, w \rangle_{\mathcal{V}', \mathcal{V}} \]

Also we assume that \( A \) is bounded and takes the form

\[ Au(t, \omega) = A(t, u(t, \omega), \omega) \]

for \( u \in \mathcal{V} \). Such an operator is type \( M \) and this is what we use. Such an operator is defined by:

\[ \text{If } u_n \to u \text{ weakly in } \mathcal{V} \text{ and } Au_n \to \xi \text{ weakly in } \mathcal{V}' \text{ and } \limsup_{n \to \infty} \langle Au_n, u_n \rangle \leq \langle \xi, u \rangle \]

Then the above implies

\[ Au = \xi. \]

We define \( \mathcal{V}_\omega \) as \( L^p(0, T, \mathcal{V}) \) with the definition of \( \mathcal{V}'_\omega \) similar, the subscript denoting that \( \omega \) is fixed, the \( \sigma \) algebra of measurable sets being the Borel sets, \( B([0, T]) \). Also,

\[ (t, u, \omega) \to A(t, u, \omega) \quad (28.3.9) \]

is progressively measurable.

Suppose \( A(\omega) \) is monotone and hemicontinuous and bounded from \( \mathcal{V}_\omega \) to \( \mathcal{V}'_\omega \). Thus

\[ A(\omega) \text{ is type } M \text{ from } \mathcal{V}_\omega \text{ to } \mathcal{V}'_\omega \quad (28.3.10) \]

where

\[ A(\omega) u \equiv A(t, u, \omega) \]

We assume the estimates found in the next lemma.

**Lemma 28.3.2** If \( p \geq 2 \) and

\[ \langle A(t, u, \omega), u \rangle_{\mathcal{V}} \geq \delta \|u\|^p_{\mathcal{V}} - c(t, \omega) \quad (28.3.11) \]

\[ \|A(t, u, \omega)\|_{\mathcal{V}'} \leq k \|u\|^{p-1}_{\mathcal{V}} + c^1 (t, \omega) \quad (28.3.12) \]

where \( c \geq 0, c \in L^1([0, T] \times \Omega) \), then if \( (t, \omega) \to q(t, \omega) \) is in \( \mathcal{V}_\omega \), it follows that for a.e. \( \omega \), similar inequalities hold for \( \bar{A} \) given by

\[ \bar{A}(t, u, \omega) \equiv A(t, u + q(t, \omega), \omega) \quad (28.3.13) \]

**Proof:** Letting \( q \) be progressively measurable, \( q(t, \omega) \in \mathcal{V} \) only consider \( \omega \) such that \( t \to q(t, \omega) \) is in \( L^p(0, T, \mathcal{V}) \).

\[ \langle \bar{A}(t, u, \omega), u \rangle = \]

\[ = \langle A(t, u + q(t, \omega), \omega), u + q(t, \omega) \rangle - \langle A(t, u + q(t, \omega), \omega), q(t, \omega) \rangle \]

\[ \geq \delta \|u + q(t, \omega)\|^p_{\mathcal{V}} - k \|u + q(t, \omega)\|^p_{\mathcal{V}} - c^1 (t, \omega) \|q(t, \omega)\|_{\mathcal{V}} \]

\[ \geq \delta \|u + q(t, \omega)\|^p_{\mathcal{V}} - k \|u + q(t, \omega)\|^p_{\mathcal{V}} - c(t, \omega) \]

\[ \geq \delta \frac{1}{2} \|u + q(t, \omega)\|^p_{\mathcal{V}} - C(k, \delta, T) \|q(t, \omega)\|^p_{\mathcal{V}} - 2c(t, \omega) \]

Now

\[ \|u + q(t, \omega)\| \geq \|u\| - \|q(t, \omega)\| \]

and so by convexity,

\[ \frac{\|u + q(t, \omega)\|^p + \|q(t, \omega)\|^p}{2} \geq \left( \frac{\|u + q(t, \omega)\| + \|q(t, \omega)\|}{2} \right)^p \geq \left( \frac{\|u\|}{2} \right)^p \]
This implies
\[ \|u + q(t, \omega)\|^p \geq 2 \left( \frac{\|u\|^p}{2^p} - \frac{\|q(t, \omega)\|^p}{2} \right) \]

Therefore,
\[ \langle \bar{A}(t, u, \omega), u \rangle = \]
\[ \langle A(t, u + q(t, \omega), u) \rangle \geq \frac{\delta}{2} \left( 2 \left( \frac{\|u\|^p}{2^p} - \frac{\|q(t, \omega)\|^p}{2} \right) \right) - C(k, \delta, T) \|q(t, \omega)\|^p_V - 2c(t, \omega) \]
\[ \geq \frac{\delta}{2^p} \|u\|^p - c'(t, \omega) \]
where \( c' \in L^1([0, T] \times \Omega) \).

Consider the other inequality. Let \( \|z\|_V \leq 1 \). Then
\[ |\langle A(t, u + q(t, \omega), z) \rangle| \leq k \|u + q(t, \omega)\|^{p-1} + c^{1/p'}(t, \omega) \]

Since \( p \geq 2 \), a convexity argument shows that
\[ \langle A(t, u + q(t, \omega), z) \rangle \leq k \left( 2^{p-2} \|u\|^{p-1} + 2^{p-2} \|q(t, \omega)\|^{p-1} \right) + c^{1/p'}(t, \omega) \]
\[ = 2^{-p} k \|u\|^{p-1} + (\tilde{c}(t, \omega))^{1/p'} \]
where \( \tilde{c} \in L^1([0, T] \times \Omega) \). Thus the same two inequalities continue to hold.

In what follows, \( c \geq 0 \) and is in \( L^1([0, T] \times \Omega) \), the \( \sigma \) algebra being \( \mathcal{B}([0, T]) \times \mathcal{F}_T \).

\[ \langle A(t, u, \omega), u \rangle \rangle \geq \frac{\delta}{2^p} \|u\|^p - c(t, \omega) \quad (28.3.14) \]
\[ \|A(t, u, \omega)\|_V \leq k \|u\|^{p-1} + c^{1/p'}(t, \omega) \quad (28.3.15) \]

Letting \( \bar{A} \) be defined above in 28.3.13,
\[ \bar{A}(t, u, \omega) \equiv A(t, u + q, \omega) \equiv \bar{A}(\omega)(t, u) \]

Assume the following pathwise uniqueness condition which is the hypothesis of the following lemma.

**Lemma 28.3.3** Suppose it is true that whenever \( u, v \in \mathcal{V}_\omega \) and
\[ Bu(t) - Bv(t) + \int_0^t A(u) - A(v) = 0 \quad (28.3.16) \]

it follows that \( u = v \). Then if
\[ (Bu)' + \bar{A}(\omega) u = f \quad \text{in} \quad \mathcal{V}'_\omega, \quad Bu(0) = Bu_0 \]
\[ (Bv)' + \bar{A}(\omega) v = f \quad \text{in} \quad \mathcal{V}'_\omega, \quad Bv(0) = Bu_0 \quad (28.3.17) \]
it follows that \( u = v \) in \( \mathcal{V}_\omega \). Here \( u_0 \in W \).

**Proof:** If \( (Bu)' + \bar{A}(\omega) u = f \) and \( (Bv)' + \bar{A}(\omega) v = f \), then
\[ Bu(t) - Bv(t) + \int_0^t A(u + q) - A(v + q) \, ds = 0 \]
Hence
\[ B(u(t) + q(t)) - B(v(t) + q(t)) + \int_0^t A(u + q) - A(v + q) \, ds = 0 \]
and so \( u + q = v + q \) showing that \( u = v \). ■

We give the following measurability lemma.
Lemma 28.3.4 Suppose $f_n$ is progressively measurable and converges weakly to $\bar{f}$ in 

$$L^\alpha ([0, T] \times \Omega, X, \mathcal{B}([0, T]) \times \mathcal{F}_T), \quad \alpha > 1$$

where $X$ is a reflexive separable Banach space. Also suppose that for each $\omega \notin N$ a set of measure zero,

$$f_n (\cdot, \omega) \to f (\cdot, \omega) \quad \text{weakly in} \quad L^\alpha (0, T, X)$$

Then there exists $N$ such that for $\omega \notin N$,

$$\bar{f} (\cdot, \omega) = f (\cdot, \omega) \quad \text{in} \quad L^\alpha (0, T, X).$$

Also $\bar{f}$ is progressively measurable.

Proof: By the Pettis theorem, $\bar{f}$ is progressively measurable. Letting $\phi \in L^{\alpha'} ([0, T] \times \Omega, X', \mathcal{B}([0, T]) \times \mathcal{F}_T)$, it is known that for a.e. $\omega$,

$$\int_0^T \langle \phi (t, \omega), f_n (t, \omega) \rangle \, dt \to \int_0^T \langle \phi (t, \omega), f (t, \omega) \rangle \, dt$$

Therefore, the function of $\omega$ on the right is at least $\mathcal{F}_T$ measurable. Now let $g \in L^\infty (\Omega, X', \mathcal{F}_T)$ and let $\psi \in C ([0, T])$. Then for $1 < p \leq \alpha$,

$$\int_\Omega \left| \int_0^T \langle g (\omega) \psi (t), f_n (t, \omega) \rangle \, dt \right|^p \, dP \leq C (T) \int_\Omega \| g \|_{L^\infty (\Omega, X')} \int_0^T |\psi (t)|^p \| f_n (t, \omega) \|_X^p \, dt \, dP$$

for some $C$. Since $\int_0^T \langle g (\omega) \psi (t), f_n (t, \omega) \rangle \, dt$ is bounded in $L^p (\Omega)$ independent of $n$ because $\int_\Omega \int_0^T \| f_n (t, \omega) \|_X^p \, dt \, dP$ is given to be bounded, it follows that the functions

$$\omega \to \int_0^T \langle g (\omega) \psi (t), f_n (t, \omega) \rangle \, dt$$

are uniformly integrable and so it follows from the Vitali convergence theorem that

$$\int_\Omega \int_0^T \langle g (\omega) \psi (t), f_n (t, \omega) \rangle \, dt \, dP \to \int_\Omega \int_0^T \langle g (\omega) \psi (t), f (t, \omega) \rangle \, dt \, dP$$

But also from the assumed weak convergence to $\bar{f}$

$$\int_\Omega \int_0^T \langle g (\omega) \psi (t), f_n (t, \omega) \rangle \, dt \, dP \to \int_\Omega \int_0^T \langle g (\omega) \psi (t), \bar{f} (t, \omega) \rangle \, dt \, dP$$

It follows that

$$\int_\Omega \left\langle g (\omega), \int_0^T (f - \bar{f}) \psi (t) \, dt \right\rangle \, dP = 0$$

This is true for every such $g \in L^\infty (\Omega, X')$, and so for a fixed $\psi \in C ([0, T])$ and the Riesz representation theorem,

$$\int_\Omega \left\| \int_0^T (f - \bar{f}) \psi (t) \, dt \right\|_X \, dP = 0$$

Therefore, there exists $N_\psi$ such that if $\omega \notin N_\psi$, then

$$\int_0^T (f - \bar{f}) \psi (t) \, dt = 0$$

Enlarge $N$, the exceptional set to also include $\cup_{\psi \in D} N_\psi$ where $D$ is a countable dense subset of $C ([0, T])$. Therefore, if $\omega \notin N$, then the above holds for all $\psi \in C ([0, T])$. It follows that for such $\omega$, $f (t, \omega) = \bar{f} (t, \omega)$ for a.e. $t$. Therefore, $f (\cdot, \omega) = \bar{f} (\cdot, \omega)$ in $L^\alpha (0, T, X)$ for all $\omega \notin N$. ■

Then one can obtain the following existence theorem using a technique of Bardos and Brezis [12].
Lemma 28.3.5 Let $q \in \mathcal{V}$ and let the conditions \[ (28.3.7) \] - \[ (28.3.14) \] be valid. Let $f \in \mathcal{V}'$ be given. Then for each $\omega$ off a set of measure zero, there exists $u (\cdot, \omega) \in \mathcal{V}_\omega$ such that $(Bu)' (\cdot, \omega) \in \mathcal{V}'_\omega$ and

$$Bu (0, \omega) = 0$$

and also the following equation holds in $\mathcal{V}'_\omega$ for a.e. $\omega$

$$(Bu)' (\cdot, \omega) + \bar{A} (\omega) (\cdot, u (\cdot, \omega)) = f (\cdot, \omega)$$

In addition to this, it can be assumed that $(t, \omega) \to u (t, \omega)$ is progressively measurable into $\mathcal{V}$. That is, for each $\omega$ off a set of measure zero, $t \to u (t, \omega)$ can be modified on a set of measure zero in $[0, T]$ such that the resulting $u$ is progressively measurable.

**Proof:** Consider the equation

$$L_h Bu + \bar{A} u = \frac{1}{h} (I - \tau_h) (Bu) + \bar{A} u = f \text{ in } \mathcal{V}'$$  \hspace{1cm} (28.3.18)

By Proposition \[ (28.3.4) \] and Theorem \[ (28.3.10) \], there exists a solution to the above equation if the left side is coercive. However, it was shown above in the computations leading to \[ (28.3.7) \] that $L_h \circ B$ is monotone. Hence the coercivity follows right away from Lemma \[ (28.3.7) \].

Thus \[ (28.3.18) \] holds in $\mathcal{V}'$. It follows that, indexing the solution by $h$,

$$\int_0^T \left\| \frac{1}{h} (I - \tau_h) (Bu_h) + \bar{A} u_h - f \right\|_{\mathcal{V}'}^p \, dt dP = 0$$

and so there exists a set of measure zero $N_h$ such that for $\omega \notin N_h$, the following equation holds in $\mathcal{V}'_\omega$

$$\frac{1}{h} (I - \tau_h) (Bu_h (\cdot, \omega)) + \bar{A} (\omega) (u_h (\cdot, \omega)) = f (\cdot, \omega)$$

Let $h$ denote a sequence converging to $0$ and let $N$ be a set of measure zero which includes $\cup_h N_h$.

Letting $u_h \in \mathcal{V}$ be the above solution to \[ (28.3.18) \], it also follows from the above estimates \[ (28.3.7) \] - \[ (28.3.14) \] that for $\omega$ off $N$, $\|u_h (\cdot, \omega)\|_{\mathcal{V}}$ is bounded independent of $h$. Thus, for such $\omega$ off this set, there exists a subsequence still called $u_h$ such that the following convergences hold.

$$u_h \rightharpoonup u \text{ in } \mathcal{V}_\omega$$

$$\bar{A} (\omega) u_h \rightharpoonup \xi \text{ in } \mathcal{V}'_\omega$$

$$\frac{1}{h} (I - \tau_h) (Bu_h) \rightharpoonup \zeta \text{ in } \mathcal{V}'_\omega$$

First we need to identify $\zeta$. Let $\phi \in C^\infty ([0, T])$ where $\phi = 0$ near $T$ and let $w \in \mathcal{V}$. Then

$$\left\langle \int_0^T \zeta \phi, w \right\rangle = \lim_{h \to 0} \left\langle \int_0^T \frac{1}{h} (I - \tau_h) (Bu_h), w \phi \right\rangle$$

$$= \lim_{h \to 0} \left\langle \int_0^T Bu_h (t) \frac{\phi (t)}{h} - \int_h^T Bu_h (t - h) \frac{\phi (t)}{h}, w \right\rangle$$

$$= \lim_{h \to 0} \left\langle \int_0^T Bu_h (t) \frac{\phi (t)}{h} - \int_h^T \frac{Bu_h (t) + Bu_h (t + h)}{h} \phi (t), w \right\rangle$$

$$= \lim_{h \to 0} \left( \left\langle \int_0^{T-h} Bu_h \frac{\phi (t) - \phi (t + h)}{h}, w \right\rangle + \int_0^T \frac{Bu_h (t)}{h} \phi (t) \right)$$

$$= \left\langle - \int_0^T Bu (t) \phi' (t), w \right\rangle$$
Since this holds for all \( \phi \in C_c^\infty (0, T) \), it follows that \( \zeta = (Bu)' \). Hence letting \( \phi \) be an arbitrary function in \( C^\infty ([0, T]) \) which equals zero near \( T \), this implies from the above that

\[
- \int_0^T \zeta \phi \, dt = \left\langle \int_0^T Bu(t) \phi'(t), w \right\rangle
\]

\[
= \left\langle \int_0^T (Bu(0) + \int_0^t (Bu)'(s) \, ds) \phi'(t), w \right\rangle
\]

\[
= \int_0^T (Bu(0), w) \phi'(t) \, dt + \left\langle \int_0^T \int_0^t (Bu)'(s) \, ds \phi'(t), w \right\rangle
\]

\[
= - (Bu(0), w) \phi(0) + \left\langle \int_0^T (Bu)'(s) \int_s^T \phi'(t) \, dt \, ds, w \right\rangle
\]

\[
= - (Bu(0), w) \phi(0) - \left\langle \int_0^T (Bu)'(s) \phi(s) \, ds, w \right\rangle
\]

Hence, since \( \zeta = (Bu)' \),

\[
0 = - (Bu(0), w) \phi(0)
\]

Then it follows that \( (Bu(0), w) = 0 \). Since \( w \) was arbitrary, \( Bu(0) = 0 \) and \( \zeta = (Bu)' \).

Thus, passing to a limit in \( 28.3.18 \),

\[
(Bu)' + \xi = f \text{ in } V'_\omega,
\]

\( Bu(0) = 0 \)

It is desired to identify \( \xi \) with \( \bar{A}(\omega) u \). First let

\[
L_h \equiv \frac{I - \tau_h}{h}
\]

Then

\[
L_h (Bu)(t) = \begin{cases} 
\frac{1}{h} \int_t^{t+h} (Bu)'(s) \, ds & \text{if } t \geq h \\
\frac{1}{h} \int_0^t (Bu)'(s) \, ds & \text{if } t < h
\end{cases}
\]

Then from standard considerations involving approximate identities,

\[
\lim_{h \to 0} L_h (Bu) = (Bu)' \text{ strongly in } V'_\omega \quad (28.3.19)
\]

Thus

\[
\left\langle L_h (Bu_h) - (Bu)', u_h - u \right\rangle =
\]

\[
\left\langle L_h (Bu_h) - L_h (Bu) , u_h - u \right\rangle + \left\langle L_h (Bu) - (Bu)', u_h - u \right\rangle
\]

\[
\geq \left\langle L_h (Bu) - (Bu)', u_h - u \right\rangle
\]

and the above strong convergence implies that this converges to 0. Therefore, from \( 28.3.18 \),

\[
L_h Bu_h + \bar{A}u_h = f \text{ in } V'_\omega
\]

and so

\[
\left\langle L_h Bu_h, u_h - u \right\rangle + \left\langle \bar{A}u_h, u_h - u \right\rangle = \left\langle f, u_h - u \right\rangle
\]

From the above,

\[
\left\langle (Bu)', u_h - u \right\rangle + \left\langle L_h (Bu) - (Bu)', u_h - u \right\rangle + \left\langle \bar{A}u_h, u_h - u \right\rangle \leq \left\langle f, u_h - u \right\rangle
\]

and so, taking \( \limsup_{h \to 0} \) of both sides, it follows from \( 28.3.18 \) that

\[
\limsup_{h \to 0} \left\langle \bar{A}u_h, u_h - u \right\rangle \leq 0,
\]

\[
\limsup_{h \to 0} \left\langle \bar{A}u_h, u_h \right\rangle \leq \left\langle \xi, u \right\rangle
\]
Since $A$ is monotone and hemicontinuous, the same is true of $\bar{A}$ and so

$$\bar{A}u = \xi$$

Thus

$$((Bu)' (\cdot, \omega)) + \bar{A} (\cdot, \omega) u (\cdot, \omega) = f (\cdot, \omega) \text{ in } V'_\omega, \quad Bu (0, \omega) = 0 \quad (28.3.20)$$

It follows from the uniqueness assumption \[28.3.14\] that for each $\omega$ off a set of measure zero, there exists a unique solution to

$$(Bu)' (\cdot, \omega) + \bar{A} (\cdot, u (\cdot, \omega), \omega) = f (\cdot, \omega) \text{ in } V'_\omega,$$

$$B (u (\cdot, \omega)) (0) = 0$$

You can consider the function of two variables $u (t, \omega)$. Is this function progressively measurable? Right now, this is not clear because we have done nothing more than solve a problem for each $\omega$.

However, we can at least say that $u_h$ is progressively measurable because $u_h \in V$. Recall also that

$$\frac{1}{h} (f - \tau_h) Bu_h + \bar{A} (\omega) u_h = f, \quad u_h \in V$$

Next we show that because of uniqueness, one can assume that $u$ is progressively measurable. To do this, we show that the sequence for which convergence holds in the above can be chosen independent of $\omega$.

Claim: A single sequence $h \to 0$ works for all $\omega$ off a set of measure zero.

Proof: Since there is only one solution to the above initial value problem for $\omega \notin N$, then letting $h \to 0$ be a single sequence, one can conclude that $u_h (\cdot, \omega) \to u (\cdot, \omega)$ in $V_\omega = L^p (0, T, V)$. Otherwise, from the above argument, one could obtain another subsequence which converges to a solution different than $u (\cdot, \omega)$ which would violate uniqueness.

From the coercivity condition, it follows that there exists a constant $C (f)$ depending on $f$ such that for all $h$,

$$\|u_h\|_V \leq C (f)$$

Therefore, there is a further subsequence still denoted by $h$ such that

$$u_h \to \bar{u} \text{ in } L^p ([0, T] \times \Omega; V) \quad (28.3.21)$$

where the measurable sets are just the product measurable sets $B ([0, T]) \times F_T$. Then it follows from Lemma \[28.3.14\] that $u (\cdot, \omega) = \bar{u} (\cdot, \omega)$ in $V_\omega$ for all $\omega$ off a set of measure zero. It follows that in all of the above, we could substitute $\bar{u}$ for $u$ at least for $\omega$ off a single set of measure zero. Thus $u$ can be assumed progressively measurable. ■

Note the importance of path uniqueness in obtaining the result on progressive measurability of the solutions.

We will write $u$ rather than $\bar{u}$ to save notation. Now with this lemma, it is easy to obtain the following proposition.

**Proposition 28.3.6** Let $q \in V$ such that $t \to q (t, \omega)$ is continuous and $q (0, \omega) = 0$, and let the conditions \[28.3.14\] - \[28.3.17\] be valid. Also let $u_0 \in L^2 (\Omega, V)$ such that $u_0$ is $F_0$ measurable. Let $f \in V'$ be given. Then for each $\omega$ off a set of measure zero, there exists $u (\cdot, \omega) \in V_\omega$ such that $(Bu)' (\cdot, \omega) \in V'_\omega$ and

$$Bu (0, \omega) = Bu_0$$

and also the following equation holds in $V'_\omega$ for a.e. $\omega$

$$(Bu - Bq)' (\cdot, \omega) + A (\cdot, u (\cdot, \omega), \omega) = f (\cdot, \omega)$$

In addition to this, it can be assumed that $t \to u (t, \omega)$ is progressively measurable into $V$. That is, for each $\omega$ off a set of measure zero, $t \to u (t, \omega)$ can be modified on a set of measure zero in $[0, T]$ such that the resulting $u$ is progressively measurable. Then one also obtains that $u$ is the unique solution to the integral equation which holds for a.e. $\omega$

$$Bu (t, \omega) - Bu_0 + \int_0^t A (s, u (s, \omega), \omega) \, ds = \int_0^t f (s, \omega) \, ds + Bq (t, \omega) \quad (28.3.22)$$

Proof: Recall

$$\bar{A} (\omega) (t, u) \equiv A (t, u + q (t, \omega), \omega)$$

where $q$ was in $V$. Therefore, replace this definition of $A$ with

$$\bar{A} (\omega) (t, u) \equiv A (t, u + q (t, \omega) + u_0, \omega)$$
Then from Lemma 28.3.4, there exists \( w \in \mathcal{V} \) such that
\[
(Bw)'(\cdot, \omega) + A(\cdot, w(\cdot, \omega) + q(\cdot, \omega) + u_0(\omega), \omega) = f(\cdot, \omega), \quad Bw(0) = 0
\]
Let \( u(t, \omega) = w(t, \omega) + q(t, \omega) + u_0(\omega) \). Then for fixed \( \omega \), \( Bu(0) = Bw(0) + Bu_0 = Bu_0 \). Also
\[
(B(u - q))' + A(\cdot, u, \omega) = f(\cdot, \omega), \quad Bu(0) = Bu_0
\]
Then an integration yields
\[
(0) = Bu_0
\]
One can easily generalize this using an exponential shift argument.

**Corollary 28.3.7** Suppose the situation of the above proposition but that all that is known is that \( \lambda B + A \) is monotone and hemicontinuous on \( \mathcal{V} \) and \( \mathcal{V} \) for all \( \lambda \) sufficiently large. Then defining
\[
\langle A_\lambda(t, w, \omega), v \rangle_{\mathcal{V}} \equiv \langle e^{-\lambda t} A(t, e^{\lambda t} w, \omega), v \rangle_{\mathcal{V}}
\]
for such \( \lambda \), it follows that \( \lambda B + A_\lambda \) is also monotone and hemicontinuous. Then replace the coercivity, and boundedness conditions above with the following weaker conditions
\[
\lambda \langle Bu, u \rangle + \langle A(t, u, \omega), u \rangle_{\mathcal{V}} \geq \delta \|u\|_{\mathcal{V}}^p - c(t, \omega)
\]  
(28.3.23)
for all \( \lambda \) large enough.
\[
\|A(t, u, \omega)\|_{\mathcal{V}} \leq k \|u\|_{\mathcal{V}}^{p-1} + e^{c_1/(p-1)} (t, \omega)
\]  
(28.3.24)
where \( c \in L^1([0, T] \times \Omega) \), \( c \geq 0 \). Then the conclusion of Proposition 28.3.17 is still valid. There exists a unique \( u \in \mathcal{V} \) such that for a.e. \( \omega \),
\[
(Bu - Bq)'(\cdot, \omega) + A(\omega)(u(\cdot, \omega)) = f(\cdot, \omega), \quad Bu(0) = Bu_0
\]  
(28.3.25)
**Proof:** That \( \lambda B + A_\lambda \) is monotone and hemicontinuous follows from the definition. Also, from the above estimates,
\[
\lambda \langle Bu, u \rangle + \langle A_\lambda(t, u, \omega), u \rangle_{\mathcal{V}} \geq e^{-2\lambda t} \left( \delta \|e^{\lambda t} u\|_{\mathcal{V}}^p - c(t, \omega) \right) \geq e^{-2\lambda t} \left( \delta \|e^{\lambda t} u\|_{\mathcal{V}}^p - e^{\lambda pt} e^{-\lambda pt} c(t, \omega) \right) \geq e^{-2\lambda t} e^{\lambda pt} \left( \delta \|u\|_{\mathcal{V}}^p - e^{-\lambda pt} c(t, \omega) \right) \geq \delta \|u\|_{\mathcal{V}}^p - e^{-\lambda pt} c(t, \omega)
\]
which is of the right form.
Similarly
\[
\|\lambda Bw + A_\lambda(t, w, \omega)\|_{\mathcal{V}} \leq \|\lambda Bw\|_{\mathcal{V}} + \|e^{-\lambda t} A(t, e^{\lambda t} w, \omega)\|_{\mathcal{V}} \leq \|\lambda B\| \|w\|_{\mathcal{V}} + \|e^{-\lambda t} A(t, e^{\lambda t} w, \omega)\|_{\mathcal{V}} \leq \|\lambda B\| \|w\|_{\mathcal{V}} + e^{-\lambda t} \|e^{\lambda t} w\|_{\mathcal{V}}^{p-1} + e^{-\lambda t} c_1/(p-1) (t, \omega)
\]
Since \( p \geq 2 \), this is no larger than
\[
\lambda \|B\|^{p/(p-1)} \|w\|_{\mathcal{V}}^{p-1} + e^{(p-1)\lambda t} e^{-\lambda t} \|w\|_{\mathcal{V}}^{p-1} + e^{-\lambda t} c_1/(p-1) (t, \omega)
\]
\[
\leq \left( e^{\lambda t} + 1 \right) \|w\|_{\mathcal{V}}^{p-1} + e^{-\lambda t} c_1/(p-1) (t, \omega) + (\lambda \|B\|)^{p/(p-1)}
\]
\[
\leq \bar{k} \|w\|_{\mathcal{V}}^{p-1} + \bar{c}(t, \omega)^{1/(p-1)}
\]
Now note that \( w \) is a solution to
\[
B \left( w - e^{-\lambda(t)} q \right)' + \lambda Bw + e^{-\lambda(t)} A(\cdot, e^{\lambda(t)} w, \omega) = e^{-\lambda(t)} f(\cdot, \omega) + \lambda B e^{-\lambda(t)} q(\cdot, \omega) \quad \text{in} \, \mathcal{V}_\omega
\]
\[
B \left( w - e^{-\lambda(t)} q \right)(0) = Bu_0
\]
28.3. **THE EXISTENCE OF APPROXIMATE SOLUTIONS**

if and only if \( u(t) \equiv e^{\lambda t} w(t) \) is a solution to

\[
(B(u-q))' + A(t, u, \omega) = f(\cdot, \omega), \quad B(u-q)(0) = Bu_0
\]

Thus the necessary uniqueness condition holds for the initial value problem for \( w \) and hence it follows from Proposition 28.3.17 that there exists a unique progressively measurable solution to the initial value problem for \( w \) and hence a unique progressively measurable solution to the above one for \( u \). ■

Now suppose the situation of the above corollary and let \( E \) be a separable Hilbert space which is dense in \( V \) and let

\[ \Phi \in L^2 \left( [0, T] \times \Omega, \mathcal{L}_2 \left( Q^{1/2} U, E \right) \right), \quad \Phi \text{ being progressively measurable}, \]

so that one can consider the stochastic integral \( \int_0^t \Phi dW \). Let

\[
\tau_n \equiv \inf \left\{ t : \left\| \int_0^t \Phi dW \right\|_E > 2^n \right\}
\]

Thus

\[
\left\| \int_0^{t \wedge \tau_n} \Phi dW \right\|_E \leq 2^n
\]

Then you could pick \( u_0 \in L^p(\Omega, V) \), \( u_0 \) being \( \mathcal{F}_0 \) measurable, and let

\[
q(t, \omega) = \int_0^{t \wedge \tau_n} \Phi dW.
\]

The result is clearly in \( V \) and is continuous in \( t \). Therefore, from Corollary 28.3.17 there exists a unique solution \( u \in V \) to the initial value problem

\[
\left( Bu - B \int_0^{t \wedge \tau_n} \Phi dW \right)'(\cdot, \omega) + A(\omega)(u(\cdot, \omega)) = f(\cdot, \omega), \quad Bu(0) = Bu_0
\]

Integrating, one obtains a unique solution \( u_n \in V \) to the integral equation

\[
Bu_n(t, \omega) - Bu_0(\omega) + \int_0^t A(s, u_n, \omega) ds = \int_0^t f(s, \omega) ds + B \int_0^{t \wedge \tau_n} \Phi dW
\]

This holds in \( \mathcal{V}'_\omega \) and is so for all \( \omega \) off a set of measure zero \( N_n \). Let \( N = \cup_n N_n \). For \( \omega \notin N, \ t \rightarrow \int_0^t \Phi dW \) is continuous and so for all \( n \) large enough, \( \tau_n = \infty \). Thus for a fixed \( \omega \), it follows that for all \( n \) large enough \( \tau_n = \infty \) and so one obtains

\[
Bu_n(t, \omega) - Bu_0(\omega) + \int_0^t A(s, u_n, \omega) ds = \int_0^t f(s, \omega) ds + B \int_0^t \Phi dW
\]

Then for \( k \) some other index sufficiently large, the same holds for \( u_k \). By the uniqueness assumption 28.3.17, \( u_k(t, \omega) = u_n(t, \omega) \) and so it follows that \( \lim_{n \to \infty} u_n(t, \omega) \) exists because for each \( \omega \) off a set of measure zero, there is eventually no change in \( u_n \). Defining \( u(t, \omega) \equiv \lim_{n \to \infty} u_n(t, \omega) \equiv u_n(t, \omega) \) for all \( n \) large enough, it follows that \( u \) is progressively measurable since it is the pointwise limit of progressively measurable functions and

\[
Bu(t, \omega) - Bu_0(\omega) + \int_0^t A(s, u, \omega) ds = \int_0^t f(s, \omega) ds + B \int_0^t \Phi dW
\]

This has shown the following lemma.

**Lemma 28.3.8** Let \( (t, u, \omega) \to A(t, u, \omega) \) be progressively measurable into \( V' \) and suppose for some \( \lambda \),

\[
\lambda B + A(\omega) : \mathcal{V'}_\omega \to \mathcal{V'}_\omega,
\]

\[
\lambda B + A : \mathcal{V} \to \mathcal{V'}
\]

are both monotone bounded and hemicontinuous. Also suppose the two estimates giving boundedness and coercivity 28.3.22 - 28.3.24 of Corollary 28.3.7 above. Here \( V, W \) are as described above \( V \subseteq W, W' \subseteq V' \), \( W \) is a separable Hilbert space and \( V \) is a separable reflexive Banach space. \( B : W \to W' \) is nonnegative and self adjoint. Let \( f \in \mathcal{V}' \)
and let $u_0 \in L^p(\Omega, V)$ where $u_0$ is $F_0$ measurable. Then if $\Phi \in L^2 \left( [0, T] \times \Omega, \mathcal{L}_2 \left( Q^{1/2}U, E \right) \right)$, $\Phi$ being progressively measurable, into $E$, where $E$ is a Hilbert space dense in $V$ with $\|u\|_E \geq \|u\|_V$, then there exists a unique solution to the integral equation

$$Bu(t) - Bu_0 + \int_0^t A(s,u,\omega) \, ds = \int_0^t f(s,\omega) \, ds + B \int_0^t \Phi \, dW$$

in the sense that $u$ is in $V$ and there exists a set of measure zero $N$ such that if $\omega \notin N$, then the above integral equation holds for all $t$.

### 28.4 The General Case

Suppose $\lambda B + A(\omega), \lambda B + A$ are both monotone bounded and hemicontinuous on $\mathcal{V}_\omega$ and $V$ respectively for $\lambda$ sufficiently large. Also suppose the two estimates giving boundedness and coercivity of Corollary 28.4.1 above. We strengthen the assumption that $\lambda B + A(\omega)$ is monotone as follows. In the usual case where $B$ is the identity, this conclusion is obvious, but here we need to assume it.

$$\langle (\lambda B + A(\omega))(u) - (\lambda B + A(\omega))(v), u - v \rangle \geq \delta \|u - v\|^\alpha_L, \quad \alpha \geq 1$$

(28.4.26)

where here $U$ is a reflexive Banach space such that $V \subseteq U$ and the inclusion map is continuous, $V$ being dense in $U$.

In regards to this monotonicity condition, here is a simple lemma which will be used later.

**Lemma 28.4.1** Suppose $u_n \to w$ weakly in $\mathcal{V}_\omega$ and that for a.e.$t$, $u_n(t) \to u(t)$ in $U$. Then $w(t) = u(t)$ a.e.

**Proof:** You know that $\|u_n\|_{L^p([0,T],V)}$ is bounded. Now consider $\phi \in U'$ and $\psi \in C \left( [0,T] \right)$. Then the weak convergence implies

$$\lim_{n \to \infty} \int_0^T \langle \phi, u_n \rangle_{U',U} \psi \, dt = \int_0^T \langle \phi, w \rangle_{U',U} \psi \, dt$$

because it is also the case that $u_n \to w$ weakly in $L^p \left( [0,T], U \right)$. However, the fact that $\|u_n\|_{L^p([0,T],V)}$ is bounded means that, by the assumed pointwise convergence,

$$\lim_{n \to \infty} \int_0^T \langle \phi, u_n \rangle_{U',U} \psi \, dt = \int_0^T \langle \phi, u \rangle_{U',U} \psi \, dt$$

It follows that

$$\int_0^T \langle \phi, u - w \rangle \psi \, dt = 0$$

Since this is true for all $\psi \in C \left( [0,T] \right)$, there exists a set of measure zero $Q_\phi$ such that for $t \notin Q_\phi$,

$$\langle \phi, u(t) - w(t) \rangle = 0$$

Letting $Q = \cup_{\phi \in D} Q_\phi$, where $D$ is a countable dense subset of $U'$, it follows that for $t \notin Q$, the above holds for all $\phi \in U'$. Hence $u(t) = w(t)$ for $t \notin Q$ and $m(Q) = 0$. $\blacksquare$

Typically $\alpha = 2$ and $U = W$.

Recall that

$$V \subseteq W, \quad W' \subseteq V'$$

each space dense in the one to its right and the inclusion maps are continuous.

Assume only

$$\Phi \in L^2 \left( [0, T] \times \Omega, \mathcal{L}_2 \left( Q^{1/2}U, W \right) \right).$$

By density of $E$ into $W$, there exists a sequence

$$\Phi_n \in L^2 \left( [0, T] \times \Omega, \mathcal{L}_2 \left( Q^{1/2}U, E \right) \right)$$

such that

$$\|\Phi_n - \Phi\|_{L^2([0,T] \times \Omega, \mathcal{L}_2(Q^{1/2}U,W))} \to 0,$$

$$\|\Phi_n\|_{\mathcal{L}_2(Q^{1/2}U,W)} \leq \|\Phi\|_{\mathcal{L}_2(Q^{1/2}U,W)}.$$
28.4. THE GENERAL CASE

Also let \( u_{0n} \in L^p(\Omega, V) \) where \( u_{0n} \) is \( \mathcal{F}_0 \) measurable and such that \( u_{0n} \in L^p(\Omega, V) \) and
\[
\|u_{0n}(\omega) - u_0(\omega)\|_W \to 0, \quad \langle Bu_{0n}, u_{0n} \rangle \leq 2\langle Bu_0, u_0 \rangle
\]
for each \( \omega \). The existence of such an approximating sequence follows from density considerations of \( E \) into \( V \) and of \( V \) into \( W \).

By Lemma 28.3.8 there is a solution \( u_n \) to the integral equation
\[
Bu_n(t) - Bu_{0n} + \int_0^t A(s, u_n, \omega) \, ds = \int_0^t f(s, \omega) \, ds + B \int_0^t \Phi_n dW
\]
Then by the Implicit Ito formula there is a set of measure zero such that for all \( n, m \)
\[
\frac{1}{2} \langle B(u_n - u_m), u_n - u_m \rangle(t) - \frac{1}{2} \langle Bu_{0n} - Bu_{0m}, u_{0n} - u_{0m} \rangle
+ \delta \int_0^t \|u_n - u_m\|_U^\alpha \, ds
\leq \lambda \int_0^t \langle B(u_n - u_m), u_n - u_m \rangle(s) \, ds
+ \frac{1}{2} \int_0^t \langle B(\Phi_n - \Phi_m), \Phi_n - \Phi_m \rangle_{\mathcal{L}_2} \, ds + M_{nn}(t)
\]
Also the last term is a martingale whose quadratic variation satisfies
\[
[M_{mn}](t) \leq C \int_0^t \|\Phi_n - \Phi_m\|_{\mathcal{L}_2(Q^{1/2}U,W)}^2 \|B(u_n - u_m)\|_W^2 \, ds
\leq C \int_0^t \|\Phi_n - \Phi_m\|_{\mathcal{L}_2(Q^{1/2}U,W)}^2 \langle Bu_n - Bu_m, u_n - u_m \rangle \, ds
\]
Then from Gronwall’s inequality, and adjusting the constants,
\[
\langle B(u_n - u_m), u_n - u_m \rangle(t) + \int_0^t \|u_n - u_m\|_U^\alpha \, ds
\leq C (u_{0n} - u_{0m}, \Phi_n - \Phi_m) + C(T) M_{nn}^*(t)
\]
where the expectation of the first constant on the right converges to 0 as \( m, n \to \infty \). Here
\[
M_{nm}^*(t) = \sup_{s \in [0,t]} [M_{nm}(t)]
\]
Since \( M^* \) is increasing, this implies that after adjusting constants,
\[
\sup_{s \in [0,t]} \left( \langle B(u_n - u_m), u_n - u_m \rangle(s) + \int_0^t \|u_n - u_m\|_U^\alpha \, ds \right)
\leq C (u_{0n} - u_{0m}, \Phi_n - \Phi_m) + C(T) M_{mn}^*(t)
\]
Then taking expectations and using the Burkholder Davis Gundy inequality,
\[
E \left( \sup_{s \in [0,t]} \left( \langle B(u_n - u_m), u_n - u_m \rangle(s) + \int_0^t \|u_n - u_m\|_U^\alpha \, ds \right) \right)
\leq C (u_{0n} - u_{0m}, \Phi_n - \Phi_m) +
C(T) \int_\Omega \left( \int_0^t \|\Phi_n - \Phi_m\|_{\mathcal{L}_2(Q^{1/2}U,W)}^2 \langle B(u_n - u_m), u_n - u_m \rangle \, ds \right)^{1/2} \, dP
\leq C_{n,m} + 2C \int_\Omega \sup_{s \in [0,t]} \langle B(u_n - u_m), u_n - u_m \rangle^{1/2}(s) \cdot
\left( \int_0^t \|\Phi_n - \Phi_m\|_{\mathcal{L}_2(Q^{1/2}U,W)}^2 \right)^{1/2} \, dP
Then adjusting the constants,
\[
E \left( \sup_{s \in [0,t]} \left( B(u_n - u_m), u_n - u_m \right)(s) + \int_0^t \|u_n - u_m\|^\alpha_U ds \right) \leq C_{n,m} + C \int_0^T \|\Phi_n - \Phi_m\|^2_{Z_2(Q^{1/2}_U,W)} dt dP = C_{n,m}
\]
where \(C_{n,m} \to 0\) as \(n, m \to \infty\). In particular, it is true for \(t = T\)
\[
E \left( \sup_{s \in [0,T]} \left( B(u_n - u_m), u_n - u_m \right)(s) + \int_0^T \|u_n - u_m\|^\alpha_U ds \right) \leq C_{n,m}
\]
Then
\[
P \left( \sup_{s \in [0,T]} \left( B(u_n - u_m), u_n - u_m \right)(s) + \int_0^T \|u_n - u_m\|^\alpha_U ds \geq \lambda \right) \leq \frac{C_{n,m}}{\lambda}
\]
Now take a subsequence such that if \(m > n_k, C_{n,k,m} < 4^{-k}\). Then the above inequality implies that
\[
P \left( \sup_{s \in [0,T]} \left( B(u_n - u_m), u_n - u_m \right)(s) + \int_0^T \|u_n - u_{n+k}\|^\alpha_U ds \geq 2^{-k} \right) \leq \frac{4^{-k}}{2^{-k}} = 2^{-k}
\]
and so, by the Borel Cantelli lemma, there is a set of measure zero \(N\) including all earlier exceptional sets of measure zero such that for \(\omega \notin N\),
\[
\sup_{s \in [0,T]} \left( B(u_n - u_m), u_n - u_m \right)(s) + \int_0^T \|u_{n,k} - u_{n,k+1}\|^\alpha_U ds \leq 2^{-k}
\]
for all \(k\) large enough. We will denote this new subsequence by \(\{u_n\}\). Thus for such \(\omega\), it follows that \(\{Bu_n\}\) is a Cauchy sequence in \(C(\Omega; W')\) for \(N\omega\) an exceptional set of measure zero where \(B(u_n - u_m)(t) = B(u_n(t) - u_m(t))\) and also \(\{u_n\}\) is a Cauchy sequence in \(L^\alpha(0,T,U)\). It follows
\[
Bu_n \to z \text{ strongly in } C(\Omega; W') \text{ with uniform norm } \tag{28.4.28}
\]
\[
\lim_{m,n \to \infty} \sup_{s \in [0,T]} \left( B(u_n - u_m), u_n - u_m \right)(s) = 0 \tag{28.4.29}
\]
There exists \(u \in L^\alpha(0,T,U)\) such that for \(\omega \notin N\),
\[
\|u_n - u\|_{L^\alpha(0,T,U)} \to 0, \quad u_n(t,\omega) \to u(t,\omega) \text{ for a.e.t in } U \tag{28.4.30}
\]
Of course a technical issue is the fact that \(B\) is a degenerate operator which might not be invertible. In the above limit, we do not know that \(z = Bu\) for some \(u\). We resolve this issue by obtaining pointwise estimates for a given \(\omega\) and then pass to a limit. After this, a time integration will give the desired result. There are easier ways to do this if \(B\) is not degenerate.

From now on, this or a subsequence of this one will be the sequence of interest. Return to \(\Phi_n\) and use the Ito formula again. Thus using the estimates,
\[
\frac{1}{2} \langle Bu_n, u_n \rangle(t) - \frac{1}{2} \langle Bu_0, u_0 \rangle + \delta \int_0^t \|u_n\|^p_U ds - \lambda \int_0^t \langle Bu_n, u_n \rangle ds
\]
\[
= \frac{1}{2} \int_0^t \langle B\Phi_n, \Phi_n \rangle ds + \int_0^t c(s,\omega) ds + \int_0^t \langle f, u_n \rangle ds + M_n(t)
\]
where \(M_n(t)\) is a local martingale whose quadratic variation satisfies
\[
\left[M_n\right](t) \leq C \int_0^t \|\Phi_n\|^2_{Z_2} \|Bu_n\|^2_W ds
\]
Then adjusting the constants,
\[
\langle Bu_n, u_n \rangle(t) + \int_0^t \|u_n\|^p_U ds \leq C \langle u_0, \Phi_n, f, c \rangle + CM^*_n(t)
\]
where the expectation of the first constant on the right is no larger than a constant \( C \) which is independent of \( n \).
Since the right term is increasing in \( t \),
\[
\sup_{s \in [0,t]} \langle Bu_n, u_n \rangle (s) + \int_0^t \|u_n\|_V^p \, ds \leq C (u_{qn}, \Phi_n, f, c) + CM_n^* (t)
\]  
(28.4.31)

Now using the Burkholder Davis Gundy inequality as before and taking the expectation,
\[
E \left( \sup_{s \in [0,t]} \langle Bu_n, u_n \rangle (s) \right) + E \int_0^t \|u_n\|_V^p \, ds \\
\leq C + C \int_\Omega \left( \int_0^t \|\Phi_n\|_{L_2}^2 \|Bu_n\|_W^2 \, ds \right)^{1/2} \, dP
\]
\[
\leq C + C \int_\Omega \left( \int_0^t \|\Phi_n\|_{L_2}^2 \langle Bu_n, u_n \rangle \, ds \right)^{1/2} \, dP
\]
\[
\leq C + C \int_\Omega \sup_{s \in [0,t]} \langle Bu_n, u_n \rangle^{1/2} (s) \left( \int_0^t \|\Phi_n\|_{L_2}^2 \, ds \right)^{1/2} \, dP
\]

Then adjusting the constants and using the approximation properties of \( \Phi_n \) given above, there is a constant \( C \) independent of \( n, t \leq T \) such that
\[
E \left( \sup_{s \in [0,t]} \langle Bu_n, u_n \rangle (s) \right) + E \int_0^t \|u_n\|_V^p \, ds \leq C
\]

In particular
\[
E \left( \sup_{s \in [0,T]} \langle Bu_n, u_n \rangle (s) \right) + E \int_0^T \|u_n\|_V^p \, ds \leq C
\]  
(28.4.32)

Next use monotonicity to obtain
\[
\frac{1}{2} \langle Bu_r - Bu_q, u_r - u_q \rangle (t) \leq \frac{1}{2} \int_0^t ((\Phi_r - \Phi_q) \circ J^{-1})^* B (u_r - u_q) \circ J \, dW \\
+ C \int_0^t (Bu_r - Bu_q, u_r - u_q) \, ds + \int_0^t \|\Phi_r - \Phi_q\|^2 \, ds
\]

and so, from Gronwall’s inequality, there is a constant \( C \) which is independent of \( r, q \) such that
\[
\langle Bu_r - Bu_q, u_r - u_q \rangle (t) \leq CM_{rq} (t) \leq CM_{rq}^* (T) + C \int_0^t \|\Phi_r - \Phi_q\|^2 \, ds
\]

where \( M_{rq} \) refers to that local martingale on the right. Thus also
\[
\sup_{t \in [0,T]} \langle Bu_r - Bu_q, u_r - u_q \rangle (t) \leq CM_{rq} (t) \leq CM_{rq}^* (T) + C \int_0^T \|\Phi_r - \Phi_q\|^2 \, ds
\]  
(28.4.33)

Taking the expectation and using the Burkholder Davis Gundy inequality again, and similar estimates to the above, using appropriate stopping times as needed, we obtain
\[
E \left( \sup_{t \in [0,T]} \langle Bu_r - Bu_q, u_r - u_q \rangle (t) \right) \leq C \int_\Omega \int_0^T \|\Phi_r - \Phi_q\|^2 \, dt \, dP
\]

Now the right side converges to 0 as \( r, q \to \infty \) and so there is a subsequence, denoted with the index \( k \) such that if \( p > k \),
\[
E \left( \sup_{t \in [0,T]} \langle Bu_k - Bu_p, u_k - u_p \rangle (t) \right) \leq \frac{1}{2^k}
\]  
(28.4.34)
Then consider the earlier local martingales. One of these is of the form
\[ M_k = \int_0^t (\Phi_k \circ J^{-1})^* B u_k \circ J dW \]
Then by the Burkholder Davis Gundy inequality and modifying constants as appropriate,
\[ E \left( (M_k - M_{k+1})^* \right) \]
\[ \leq C \int_\Omega \left( \int_0^T \| \Phi_k \circ J^{-1} \|^2 (Bu_k) + \| \Phi_{k+1} \|^2 (Bu_k - Bu_{k+1}, u_k - u_{k+1}) \right)^{1/2} dP \]
\[ \leq C \int_\Omega \left( \int_0^T \| \Phi_k - \Phi_{k+1} \|^2 (Bu_k, u_k) \right)^{1/2} dP \]
\[ + C \int_\Omega \left( \int_0^T \| \Phi_{k+1} \|^2 (Bu_k - Bu_{k+1}, u_k - u_{k+1}) \right)^{1/2} dP \]
\[ \leq C \int_\Omega \sup_t (Bu_k, u_k)^{1/2} \left( \int_0^T \| \Phi_k - \Phi_{k+1} \|^2 dt \right)^{1/2} dP \]
\[ + C \int_\Omega \sup_t (Bu_k - Bu_{k+1}, u_k - u_{k+1})^{1/2} \left( \int_0^T \| \Phi_{k+1} \|^2 dt \right)^{1/2} dP \]
\[ \leq C \left( \int_\Omega \sup_t (Bu_k, u_k) dP \right)^{1/2} \left( \int_\Omega \int_0^T \| \Phi_k - \Phi_{k+1} \|^2 dt dP \right)^{1/2} \]
\[ + C \left( \int_\Omega \sup_t (Bu_k - Bu_{k+1}, u_k - u_{k+1}) dP \right)^{1/2} \left( \int_\Omega \int_0^T \| \Phi_{k+1} \|^2 dt dP \right)^{1/2} \]
From the above inequalities, after adjusting the constants, the above is no larger than an expression of the form
\[ C \left( \frac{1}{2} \right)^{k/2} \]
which is a summable sequence. Then
\[ \sum_k \int_\Omega \sup_{t \in [0,T]} |M_k(t) - M_{k+1}(t)| dP < \infty \]
Then \( \{M_k\} \) is a Cauchy sequence in \( M_T^2 \) and so there is a continuous martingale \( M \) such that
\[ \lim_{k \to \infty} E \left( \sup_t |M_k(t) - M(t)| \right) = 0 \] (28.4.35)
Taking a further subsequence if needed, one can also have
\[ P \left( \sup_t |M_k(t) - M(t)| > \frac{1}{k} \right) \leq \frac{1}{2^k} \]
and so by the Borel Cantelli lemma, there is a set of measure zero such that off this set, \( \sup_t |M_k(t) - M(t)| \) converges to 0. Hence for such \( \omega \), \( M_k^2(T) \) is bounded independent of \( k \). Thus for \( \omega \) off a set of measure zero, \( 28.3.34 \) implies that for such \( \omega \),
\[ \sup_{s \in [0,T]} \langle Bu_r, u_r \rangle(s) + \int_0^T \| u_r(s) \|^p \eta(ds) \leq C(\omega) \]
where \( C(\omega) \) does not depend on the index \( r \), this for the subsequence just described which will be the sequence of interest in what follows. Using the boundedness assumption for \( A \), one also obtains an estimate of the form
\[ \sup_{s \in [0,T]} \langle Bu_r, u_r \rangle(s) + \int_0^T \| u_r(s) \|^p \eta(ds) + \int_0^T \| z_r \|^p \eta(ds) \leq C(\omega) \] (28.4.36)
Lemma 28.4.2 There is a subsequence, still indexed by \( n \) and a set of measure zero \( N \), containing all the preceding sets of measure zero such that for \( \omega \notin N \),

\[
\sup_{s \in [0, T]} (Bu_n, u_n) (s) + \int_0^T \| u_n \|_V^p ds \leq C (\omega) < \infty
\]

From the theory of the stochastic integral, there is a further subsequence of the above such that

\[
\int_0^t \Phi_n dW \rightarrow \int_0^t \Phi dW \text{ strongly in } C ([0, T], W)
\]

for all \( \omega \) off a set of measure zero. Enlarge the exceptional set \( N \) and only use subsequences of this one so that both the above estimate in the lemma and the above convergence hold for \( \omega \notin N \). Recall the integral equation solved.

\[
Bu_n(t) - Bu_0n + \int_0^t A (s, u_n, \omega) ds = \int_0^t f (s, \omega) ds + B \int_0^t \Phi_n dW
\]

(28.4.37)

Thus

\[
\left( Bu_n - B \int_0^t \Phi_n dW - Bu_0n \right) ' + Au_n = f
\]

Then for \( \omega \notin N \), a subsequence of the one for which the above lemma holds, still denoted as \( \{ u_n \} \) yields the following convergences,

\[
u_n \rightarrow u \text{ weakly in } V_{\omega}
\]

(28.4.38)

\[Au_n \rightharpoonup \xi \text{ weakly in } V_{\omega}'
\]

(28.4.39)

\[
\left( Bu_n - B \int_0^t \Phi_n dW - Bu_0n \right) ' \rightarrow \zeta \text{ weakly in } V_{\omega}'
\]

(28.4.40)

By the earlier convergence, this \( u \) is the same as the one in the above.

Consider \( \zeta \). Let \( \psi \) be infinitely differentiable and equal to 0 near \( T \) and let \( g \in V \). Then since \( Bu_n (0) = Bu_0n \),

\[
\int_0^T \langle \zeta, \psi g \rangle dt = \lim_{n \to \infty} \int_0^T \left( \left( Bu_n - B \int_0^t \Phi_n dW - Bu_0n \right) ', \psi g \right) dt
\]

\[= - \lim_{n \to \infty} \int_0^T \left( \left( Bu_n - B \int_0^t \Phi_n dW - Bu_0n \right), \psi' g \right) dt
\]

\[= - \int_0^T \left( \psi' Bg, u - \int_0^t \Phi dW - u_0 \right) dt
\]

\[= - \int_0^T B \left( u - \int_0^t \Phi dW - u_0 \right), \psi' g \right) dt
\]

which shows that

\[
\zeta = \left( B \left( u - \int_0^t \Phi dW - u_0 \right) \right) '
\]

in the sense of \( V' \) valued distributions. Also from the above,

\[
\int_0^T \langle \zeta, \psi g \rangle dt = \langle Bu (0) - Bu_0, \psi (0) g \rangle
\]

\[+ \int_0^T \left( \left( Bu - B \int_0^t \Phi dW - Bu_0 \right), \psi' g \right) dt
\]

\[= \langle Bu (0) - Bu_0, \psi (0) g \rangle + \int_0^T \langle \zeta, \psi g \rangle dt
\]
Hence \( B (u (0, \omega)) = Bu_0 \). Thus this has shown that

\[
\left( B \left( u - \int_0^t \Phi dW - u_0 \right) \right)^t + \xi (\cdot, \omega) = f (\cdot, \omega) \text{ in } \mathcal{V}_\omega, \quad Bu (0) = Bu_0.
\]

Thus integrating this, we get

\[
Bu (t, \omega) = Bu_0 (\omega) + \int_0^t \xi (s, \omega) ds = \int_0^t f (s, \omega) ds + B \int_0^t \Phi dW \quad (28.4.41)
\]

**Lemma 28.4.3** The above sequence does not depend on \( \omega \notin N \). In fact, it is not necessary to take a further subsequence.

**Proof:** In fact, it is not necessary to take a subsequence to get the convergences. This is because of the pointwise convergence of the sequences and Lemma 28.4.1. If the original sequence did not converge, then there would be two subsequences converging weakly to two different functions in \( \mathcal{V}_\omega v, w \) which is impossible because of the lemma since it would require \( v (t) = w (t) \text{ a.e.} \)

The question at this point is whether \( u \) is progressively measurable. From the assumed estimates, the Ito formula, and the same kind of estimates used earlier show that there exists an estimate of the form

\[
\|u_n\|_{\mathcal{V}} \leq C
\]

Therefore, there exists a further subsequence such that

\[
u_n \rightarrow \bar{u} \text{ weakly in } \mathcal{V}
\]

It follows from Lemma 28.4.3 that off an enlarged exceptional set of measure zero, still denoted as \( N \),

\[
\bar{u} (\cdot, \omega) = u (\cdot, \omega) \text{ in } \mathcal{V}_\omega
\]

Hence we can assume that \( u \) is progressively measurable into \( V \). It follows that \( Bu \) is progressively measurable into \( W' \).

Thus also

\[
(t, \omega) \rightarrow \int_0^t \xi (s, \omega) ds
\]

is progressively measurable into \( V' \).

Of course the next task is to identify \( \xi \). This is always a problem even in the non stochastic case. Here it is especially difficult because in order to identify \( \xi \) we need to use the implicit Ito formula which only holds if \( \xi \) is sufficiently measurable. However, we have obtained \( \xi \) as a weak limit for fixed \( \omega \). Therefore, this is a significant issue. In stochastic evolution problems where \( B = I \) this is not as difficult because one gets \( \xi \) as a weak limit in \( V \) and then \( \xi \) is progressively measurable. We cannot do it this way and still get the best results in which there is a solution to the integral equation which holds for all \( t \) off a set of measure zero because of the degenerate nature of the operator \( B \). However, \( \xi \) is only an equivalence class of functions. We show in the next lemma that there exists a representative of this equivalence class for each \( \omega \) off an exceptional set of measure zero such that the resulting \( \xi \) is progressively measurable. This will enable us to use the implicit Ito formula and indentify \( \xi \).

The following lemma will allow the use of the Ito formula and eventually identify \( \xi \).

**Lemma 28.4.4** Enlarging the exceptional set, one can assume that \( \xi \) is also progressively measurable. In fact, if

\[
\xi_n \equiv \int_{t-(1/n)}^t \xi ds
\]

is known to be progressively measurable, \( \xi (t, \omega) \equiv 0 \text{ for } t < 0 \), then there exists a set of measure zero \( N \) such that for \( \omega \notin N, \xi (t, \omega) = \xi (t, \omega) \text{ for all } t \text{ off a set of measure zero and } \xi \text{ is progressively measurable}.

**Proof:** Define

\[
\xi_n \equiv n \int_{t-(1/n)}^t \xi ds
\]
where \( \xi \) is defined to be zero for \( t \leq 0 \). Then by what was just shown, this is progressively measurable. Also, standard approximate identity arguments verify that for each \( \omega, \xi_n \rightarrow \xi \) in \( \mathcal{V}_\omega \). Next note that the set where \( \xi_n \) is not a Cauchy sequence is a progressively measurable set. It equals

\[
\bigcup_n \bigcap_m \bigcup_{k,l \geq m} \left\{ (t, \omega) : \| \xi_l(t, \omega) - \xi_k(t, \omega) \| > \frac{1}{n} \right\} \equiv S
\]

Now for \( p > 0 \)

\[
\lim_{m \to \infty} P \left( \sup_{p > 0} \| \xi_{m+p} - \xi_m \|_{\mathcal{V}_\omega} > \varepsilon \right) = 0
\]

This is because of the convergence of \( \xi_n \) to \( \xi \) in \( \mathcal{V}_\omega \). Therefore, there is a subsequence still called \( \xi_n \) such that

\[
P \left( \sup_{p > 0} \| \xi_{n+p} - \xi_n \|_{\mathcal{V}_\omega} > 2^{-n} \right) < 2^{-n}
\]

and so there is an enlarged set of measure zero, still denoted as \( N \) such that all of the above considerations hold for \( \omega \not\in N \) and also for \( \omega \not\in N \),

\[
\sup_{p > 0} \| \xi_{n+p} - \xi_n \|_{\mathcal{V}_\omega} \leq 2^{-n}
\]

for all \( n \) large enough. Now let \( S \) defined above, correspond to this particular subsequence. Let \( S(\omega) \) be those \( t \) such that \( (t, \omega) \in S \). Then \( S(\omega) \) is a set of measure zero for each \( \omega \not\in N \) because the above inequality implies that \( t \rightarrow \xi_n(t, \omega) \) is a Cauchy sequence off a set of measure zero which by definition is \( S(\omega) \). Then consider \( \{ \xi_n(t, \omega) \mathcal{X}_{S(\omega)}(t, \omega) \} \). For each \( \omega \) off \( N \), this converges for all \( t \). Thus it converges pointwise to a function \( \xi \) which must be progressively measurable. However, \( t \rightarrow \xi(t, \omega) \) must also equal \( t \rightarrow \xi(t, \omega) \) in \( \mathcal{V}_\omega \) by the above construction. Therefore, we can assume without loss of generality that \( \xi \) is itself progressively measurable. ■

From the weak convergence of \( u_n \) to \( u \) in \( \mathcal{V}_\omega \),

\[
Bu_n \rightarrow Bu \text{ weakly in } \mathcal{V}_\omega'
\]

and so

\[
(\lambda B + A(\omega))u_n \rightarrow \lambda Bu + \xi \text{ weakly in } \mathcal{V}_\omega'
\]

Now the above convergences and the integral equation imply that off the exceptional set \( N \), for each \( t \)

\[
Bu_n(t) \rightarrow Bu(t) \text{ weakly in } V'
\]

From a generalization of standard theorems in Hilbert space, stated in Lemma \([28.4.31]\) there exist vectors \( \{ e_i \} \subseteq V \) such that

\[
\langle Bu_n(t), u_n(t) \rangle = \sum_{i=1}^{\infty} |\langle Bu_n(t), e_i \rangle|^2
\]

Hence

\[
\lim_{n \to \infty} \inf \sum_{i=1}^{\infty} \lim_{n \to \infty} \inf \| \langle Bu_n(t), e_i \rangle \|^2
\]

\[
= \sum_{i=1}^{\infty} \| \langle Bu(t), e_i \rangle \|^2 = \langle Bu(t), u(t) \rangle
\]

(28.4.42)

Thus the above inequalities and formulas hold for a.e. \( t \).

Return to the equation \([28.4.31]\). Define the stopping time

\[
\tau_p \equiv \inf \left\{ t \in [0, T] : \langle Bu, u \rangle(t) + \int_0^t \| \xi \|_{\mathcal{V}}', ds > p \right\}
\]

From \([28.3.28]\) and the fact that \( \xi \in \mathcal{V}_\omega' \), it follows that \( \tau_p = \infty \) for all \( p \) large enough. Then stop the equation using this stopping time.

\[
Bu_{\tau_p}(t, \omega) - Bu_0(\omega) + \int_0^t \mathcal{X}_{[0, \tau_p]} \xi_{\tau_p}(s, \omega) ds
\]

\[
= \int_0^t \mathcal{X}_{[0, \tau_p]} f(s, \omega) ds + B \int_0^t \mathcal{X}_{[0, \tau_p]} \Phi dW
\]
From the implicit Ito formula Theorem \textbf{28.4.3}, for a.e. $t$,
\[
\frac{1}{2} \langle Bu^p \rangle (t), u^p (t) \rangle - \frac{1}{2} \langle Bu_0, u_0 \rangle + \int_0^t \lambdaBu_0 + \xi, u \rangle ds = \frac{1}{2} \int_0^t \langle B\Phi, \Phi \rangle ds \\
\]

\[
+ \int_0^t \lambdaBu_0 \rangle (f, u^p ) ds + \int_0^t \lambdaBu_0 \rangle (\Phi \circ J^{-1}) Bfu \circ JdW \\
\]

Then letting $p \to \infty$ this yields the following formula for a.e. $t$
\[
\frac{1}{2} \langle Bu \rangle (t), u (t) \rangle - \frac{1}{2} \langle Bu_0, u_0 \rangle + \int_0^t \lambdaBu + \xi, u \rangle ds = \frac{1}{2} \int_0^t \langle B\Phi, \Phi \rangle ds \\
\]

\[
+ \int_0^t \langle f, u \rangle ds + \int_0^t \lambdaBu \rangle (\Phi \circ J^{-1}) Bfu \circ JdW + \int_0^t \lambdaBu, u \rangle ds \\
\]

\begin{equation}
(28.4.43)
\end{equation}

**Lemma 28.4.5** It is true that
\[
\lim_{n \to \infty} \int_0^T \langle Bu_n, u_n \rangle dt = \int_0^T \langle Bu, u \rangle dt \\
\]

**Proof:** From \textbf{28.4.3} $Bu_n \to z$ strongly in $C (N^C, W')$. But also, for each $t$, $Bu_n (t) \to Bu (t)$ weakly in $V'$ and so $z (t) = Bu (t)$. This strong convergence in $C (N^C, W')$ along with the uniform norm with the weak convergence of $u_n$ to $u$ in $V_o$ is sufficient to obtain the above limit.

You might think that
\[
\int_0^T \lambdaBu_n \rangle (\Phi \circ J^{-1}) Bfu \circ JdW = \int_0^T \lambdaBu, u \rangle (\Phi \circ J^{-1}) Bfu \circ JdW \\
\]

but this is not entirely clear. It will be true in the case that in \textbf{28.4.2} $\alpha = 2$ and $U = W$ and this is shown later. However, it is not clearly true here unless it is also the case that $\Phi \in L^2 (\Omega, L^\infty ([0, T], \mathcal{L}_2 (Q^{1/2}U, W)))$.

**Lemma 28.4.6** If $\Phi \in L^2 (\Omega, L^\infty ([0, T], \mathcal{L}_2 (Q^{1/2}U, W)))$ then
\[
\int_0^T \lambdaBu_n \rangle (\Phi \circ J^{-1}) Bfu \circ JdW \\
\]

**Proof:**
\[
E \left( \int_0^T \lambdaBu_n \rangle (\Phi \circ J^{-1}) Bfu \circ JdW - \int_0^T \lambdaBu, u \rangle (\Phi \circ J^{-1}) Bfu \circ JdW \right) \\
\]

\[
\leq E \left( \int_0^T \lambdaBu_n \rangle (\Phi \circ J^{-1}) Bfu \circ JdW - \int_0^T \lambdaBu, u \rangle (\Phi \circ J^{-1}) Bfu \circ JdW \right) \\
+ E \left( \int_0^T \lambdaBu_n \rangle (\Phi \circ J^{-1}) Bfu \circ JdW - \int_0^T \lambdaBu, u \rangle (\Phi \circ J^{-1}) Bfu \circ JdW \right) \\
\]

\[
\leq \int_\Omega \left( \left( \int_0^T \|\Phi_n - \Phi\|_{\mathcal{L}_2}^2 \langle Bu_n, u_n \rangle \right)^{1/2} \right) dP \\
+ \int_\Omega \left( \left( \int_0^T \|\Phi\|_{\mathcal{L}_2}^2 \langle Bu_n - Bu, u_n - u \rangle dt \right)^{1/2} \right) dP \\
\]

\begin{equation}
(28.4.44)
\end{equation}
Consider that second term. It is no larger than
\[
\int_{\Omega} \left( \int_0^T \left( Bu_n - Bu_u, u_n - u \right) dt \right)^{1/2} dP \\
\leq \left( \int_{\Omega} \left( \int_0^T \left( Bu_n - Bu_u, u_n - u \right) dt \right)^2 dP \right)^{1/2}.
\]

Now consider the following. Letting the \( e_i \) be the special vectors of Theorem 28.4.37, it follows,
\[
\int_{\Omega} \int_0^T \left( Bu_n - Bu, u_n - u \right) dt dP = \int_{\Omega} \int_0^T \sum_{i=1}^\infty \left( Bu_n - Bu, e_i \right)^2 dt dP
\]
\[
= \int_{\Omega} \int_0^T \lim_{p \to \infty} \sum_{i=1}^\infty \left( Bu_n - Bu, e_i \right)^2 dt dP \\
\leq \lim_{p \to \infty} \int_{\Omega} \int_0^T \sum_{i=1}^\infty \left( Bu_n - Bu, e_i \right)^2 dt dP \\
= \lim_{p \to \infty} \int_{\Omega} \int_0^T \left( Bu_n - Bu, u_n - u_p \right) dt dP \leq \frac{T}{2^n}
\]
The last inequality follows from 28.4.34. Therefore, the second term in 28.4.35 is no larger than \( C \left( T, \Phi \right) / 2^n \)^{1/2} which converges to 0 as \( n \to \infty \). Now consider the first term in 28.4.35,
\[
\int_{\Omega} \left( \left( \int_0^T \left( \Phi_n - \Phi \right)^2 dt \right)^{1/2} \right) dP
\]
\[
\leq \sup_{t \in [0,T]} \left( Bu_n, u_n \right)^{1/2} \left( \left( \int_0^T \left( \Phi_n - \Phi \right)^2 dt \right)^{1/2} \right) dP \\
\leq \left( \left( \sup_{t \in [0,T]} \left( Bu_n, u_n \right) \right)^{1/2} \left( \int_0^T \left( \Phi_n - \Phi \right)^2 dt \right)^{1/2} \right)
\]
From 28.4.32
\[
\leq C \left( \int_0^T \left( \Phi_n - \Phi \right)^2 dt \right)^{1/2}
\]
which converges to 0.  

Return now to the equation solved by \( u_n \) in 28.4.31. Apply the Ito formula to this one. This yields for a.e. \( t \),
\[
\frac{1}{2} \left( Bu_n \left( t, u_n \left( t \right) \right) \right) - \frac{1}{2} \left( Bu_{0n}, u_{0n} \right) + \int_0^t \left( A \left( \omega \right) u_n, u_n \right) ds = \frac{1}{2} \int_0^t \left( B \Phi_n, \Phi_n \right) ds \\
+ \int_0^t \left( f, u_n \right) ds + \int_0^t \left( \Phi_n \circ J^{-1} \right)^* Bu_n \circ J dW
\]
Assume without loss of generality that \( T \) is not in the exceptional set. If not, consider all \( T' \) close to \( T \) such that \( T' \) is not in the exceptional set.
\[
\int_0^T \left( \left( \lambda B + A \left( \omega \right) \right) u_n, u_n \right) ds
\]
\[
= \frac{1}{2} \left( Bu_{0n}, u_{0n} \right) - \frac{1}{2} \left( Bu_n \left( T, u_n \left( T \right) \right) \right) + \int_0^T \left( f, u_n \right) ds \\
+ \int_0^T \left( \Phi_n \circ J^{-1} \right)^* Bu_n \circ J dW + \frac{1}{2} \int_0^T \left( B \Phi_n, \Phi_n \right) ds + \int_0^T \left( \lambda Bu_n, u_n \right) ds
\]
Now it follows from applied to \( t = T \) and the above lemma that

\[
\limsup_{n \to \infty} \int_0^T \langle (\lambda B + A (\omega)) u_n, u_n \rangle \, ds \\
\leq \frac{1}{2} \langle Bu_0, u_0 \rangle - \frac{1}{2} \langle Bu (T), u (T) \rangle + \int_0^T \langle f, u \rangle \, ds \\
+ \int_0^T \left( \Phi \circ J^{-1} \right)^* Bu \circ J dW + \frac{1}{2} \int_0^T \langle B \Phi, \Phi \rangle \, ds + \int_0^T \langle \lambda Bu, u \rangle \, ds
\]

and from the expression on the right equals \( \int_0^T \langle \lambda Bu + \xi, u \rangle \, ds \). Hence

\[
\limsup_{n \to \infty} \int_0^T \langle (\lambda B + A (\omega)) u_n, u_n \rangle \, ds \leq \int_0^T \langle \lambda Bu + \xi, u \rangle \, ds
\]

Then since \( \lambda B + A (\omega) \) is monotone and hemicontinuous, it is type \( M \) and so this requires \( A (\omega) u = \xi \).

Hence we obtain

\[
Bu (t) - Bu_0 (\omega) + \int_0^t A (\omega) (u) \, ds = \int_0^t f (s, \omega) \, ds + B \int_0^t \Phi dW
\]

This is a solution for a given \( \omega \notin \mathbb{N} \). Also, a stopping time argument like the above and the coercivity estimates for \( A \) along with the implicit Ito formula show that \( u \in \mathcal{V} \). This yields the existence part of the following existence and uniqueness theorem.

**Theorem 28.4.7** Suppose \( \mathcal{V} \equiv L^p ([0, T] \times \Omega, \mathcal{V}) \) where \( p \geq 2 \), with the \( \sigma \) algebra of progressively measurable sets and \( \mathcal{V}_\omega = L^p ([0, T] \times \Omega) \).

\[
\Phi \in L^2 \left( [0, T] \times \Omega, \mathcal{L}_2 \left( Q^{1/2} U, W \right) \right) \cap L^2 \left( \Omega, L^\infty \left( [0, T], \mathcal{L}_2 \left( Q^{1/2} U, W \right) \right) \right), \\
f \in \mathcal{V}' \equiv L^{p'} ([0, T] \times \Omega, \mathcal{V}')
\]

and both are progressively measurable. Suppose that

\[
\lambda B + A (\omega) : \mathcal{V}_\omega \to \mathcal{V}_\omega, \quad \lambda B + A : \mathcal{V} \to \mathcal{V}'
\]

are monotone hemicontinuous and bounded where

\[
A (\omega) u (t) \equiv A (t, u (t), \omega)
\]

and \( (t, u, \omega) \to A (t, u, \omega) \) is progressively measurable. Also suppose for \( p \geq 2 \), the coercivity, and the boundedness conditions

\[
\lambda \langle Bu, u \rangle + \langle A (t, u, \omega), u \rangle \geq \delta \| u \|^p_{\mathcal{V}} - c (t, \omega) \quad (28.4.46)
\]

where \( c \in L^1 ([0, T] \times \Omega) \) for all \( \lambda \) large enough. Also,

\[
\| A (t, u, \omega) \|_{\mathcal{V}'} \leq k \| u \|_{\mathcal{V}}^{p-1} + c^{1/p'} (t, \omega) \quad (28.4.47)
\]

also suppose the monotonicity condition for all \( \lambda \) large enough.

\[
\langle (\lambda B + A (\omega)) (u) - (\lambda B + A (\omega)) (v), u - v \rangle \geq \delta \| u - v \|^2_{\mathcal{V}} \quad (28.4.48)
\]

Then if \( u_0 \in L^2 (\Omega, W) \) with \( u_0 \mathcal{F}_0 \) measurable, there exists a unique solution \( u (\cdot, \omega) \in \mathcal{V}_\omega \) with \( u \in \mathcal{V} \) \((L^p ([0, T] \times \Omega, \mathcal{V}) \) and progressively measurable) such that for \( \omega \) off a set of measure zero,

\[
Bu (t, \omega) - Bu_0 (\omega) + \int_0^t A (s, u (s, \omega), \omega) \, ds = \int_0^t f (s) \, ds + B \int_0^t \Phi dW.
\]

It is also assumed that \( \mathcal{V} \) is a reflexive separable real Banach space.
Proof: The uniqueness assertion follows easily from the monotonicity condition. 

Now we remove the assumption that \( \Phi \in L^2 (\Omega, L^\infty ([0, T], L_2 (Q^{1/2}U, W))) \). Everything is the same except for the need for a different argument to show that \( \int_0^T (\Phi_n \circ J^{-1})^* Bu_n \circ JdW \rightarrow \int_0^T (\Phi \circ J^{-1})^* Bu \circ JdW \). In this case we assume

\[
(\langle \lambda B + A(\omega) \rangle (u) - (\lambda B + A(\omega)) (v), u - v) \geq \delta \|u - v\|^2_W
\]

Then repeating the above argument with this change yields a set of measure zero, still denoted as \( N \) such that for \( \omega \notin N \)

\[
\int_0^T \|u_n - u_{n+1}\|^2_W ds \leq 2^{-n}
\]

for all \( n \) large enough. Hence for such \( \omega, u_n (\cdot, \omega) \) is Cauchy in \( L^2 ([0, T], W) \) and in fact \( u_n (t, \omega) \) is a Cauchy sequence in \( W \). Thus \( \{u_n (\cdot, \omega)\} \) converges in \( L^2 ([0, T], W) \) to \( u (\cdot, \omega) \in L^2 ([0, T], W) \) and by the above considerations involving continuous dependence of \( V \) into \( W \), it follows that \( u (\cdot, \omega) \) will be the same as the \( u \) from the above convergences. Now this convergence implies that in addition, for a.e. \( t \),

\[
\lim_{n \to \infty} \langle Bu_n (t, \omega) - Bu (t, \omega), u_n (t, \omega) - u (t, \omega) \rangle = 0
\]

\[
\lim_{n \to \infty} \int_0^T \langle Bu_n (t, \omega) - Bu (t, \omega), u_n (t, \omega) - u (t, \omega) \rangle dt = 0
\]

What is known from (28.4.35) is that for

\[
M_n = \int_0^t (\Phi_n \circ J^{-1})^* Bu_n \circ JdW
\]

there is a continuous martingale \( M \in M^1_T \) such that

\[
\lim_{n \to \infty} E \left( \sup_{t \in [0, T]} |M_n (t) - M (t)| \right) = 0
\]

Define a stopping time

\[
\tau_p = \inf \left\{ t : \langle Bu, u \rangle (t) + \sup_n \langle Bu_n, u_n \rangle (t) > p \right\}
\]

This is a good enough stopping time because the function used to define it as a hitting time is lower semicontinuous.

Lemma 28.4.8 \( \int_0^T (\Phi_n \circ J^{-1})^* Bu_n \circ JdW \rightarrow \int_0^T (\Phi \circ J^{-1})^* Bu \circ JdW \) in probability. Also there is a further subsequence and set of measure zero such that off this set,

\[
\lim_{n \to \infty} \left( \sup_{t \in [0, T]} \left| \int_0^t (\Phi \circ J^{-1})^* Bu \circ JdW - \int_0^t (\Phi_n \circ J^{-1})^* Bu_n \circ JdW \right| \right) = 0.
\]

In particular, what is needed here is valid,

\[
\lim_{n \to \infty} \int_0^T (\Phi_n \circ J^{-1})^* Bu_n \circ JdW = \int_0^T (\Phi \circ J^{-1})^* Bu \circ JdW
\]

Proof: Let \( \varepsilon > 0 \). Then define

\[
A_n = \left\{ \omega : \left| \int_0^T (\Phi_n \circ J^{-1})^* Bu_n \circ JdW - \int_0^T (\Phi \circ J^{-1})^* Bu \circ JdW \right| > \varepsilon \right\}
\]

It was shown earlier that \( \lim_{n \to \infty} \int_0^T (\Phi_n \circ J^{-1})^* Bu_n \circ JdW \) exists. Then

\[
A_n = \cup_{p=1}^\infty A_n \cap ([\tau_p = \infty] \setminus [\tau_{p-1} < \infty]), [\tau_0 < \infty] \equiv \emptyset
\]

the sets in the union being disjoint. Then \( A \cap ([\tau_p = \infty]) \subseteq \left\{ \omega : \left| \int_0^T \chi_{[0, \tau_p]} (\Phi_n \circ J^{-1})^* Bu_n \circ JdW - \int_0^T \chi_{[0, \tau_p]} (\Phi \circ J^{-1})^* Bu \circ JdW \right| > \varepsilon \right\} \)
Then as before,
\[
E \left( \int_0^T \mathcal{X}_{[0,\tau_p]} (\Phi_n \circ J^{-1})^* B u_n \circ J dW - \int_0^T \mathcal{X}_{[0,\tau_p]} (\Phi \circ J^{-1})^* B u \circ J dW \right)
\]
\[
\leq \int_\Omega \left( \left( \int_0^T \|\Phi_n - \Phi\|_{L^2}^2 \mathcal{X}_{[0,\tau_p]} \langle B u_n, u_n \rangle \right)^{1/2} \right) dP
\]
\[
+ \int_\Omega \left( \int_0^T \mathcal{X}_{[0,\tau_p]} \|\Phi\|_{L^2}^2 \langle B u_n - B u, u_n - u \rangle dt \right)^{1/2} dP
\]
(28.4.52)

Consider the second term. It is no larger than
\[
\left( \int\int_\Omega \mathcal{X}_{[0,\tau_p]} \|\Phi\|_{L^2}^2 \langle B u_n - B u, u_n - u \rangle dtdP \right)^{1/2}
\]
Now \( t \to \langle B u_n, u_n \rangle \) is continuous and so \( \mathcal{X}_{[0,\tau_p]} \langle B u_n, u_n \rangle (t) \leq p \). If not, then you would have \( \langle B u_n, u_n \rangle (t) > p \) for some \( t \leq \tau_p \) and so, by continuity, there would be \( s < t \leq \tau_p \) for which \( \langle B u_n, u_n \rangle (s) > p \) contrary to the definition of \( \tau_p \). Then \( \mathcal{X}_{[0,\tau_p]} (B u_n - B u, u_n - u) \) is bounded a.e. and also converges to 0 for a.e. \( t \leq \tau_p \) as \( n \to \infty \). Therefore, off a set of measure zero, including the set where \( t \to \|\Phi\|_{L^2}^2 \) is not in \( L^1 \), the double integral converges to 0 by the dominated convergence theorem. As to the first integral in (28.4.32) it is dominated by
\[
\int_\Omega \mathcal{X}_{[0,\tau_p]} \sup_{t \in [0,\tau_p]} \langle B u_n, u_n \rangle^{1/2} (t) \left( \int_0^T \|\Phi_n - \Phi\|_{L^2}^2 \right)^{1/2} dP
\]
\[
\leq \left( \int \sup_{t \in [0,\tau]} \langle B u_n, u_n \rangle dP \right)^{1/2} \left( \int \int_\Omega \|\Phi_n - \Phi\|_{L^2}^2 dtdP \right)^{1/2}
\]
From the estimate (28.3.32),
\[
\leq C \left( \int \int_0^T \|\Phi_n - \Phi\|_{L^2}^2 dtdP \right)^{1/2}
\]
for a constant \( C \) independent of \( n \). Therefore,
\[
\lim_{n \to \infty} E \left( \int_0^T \mathcal{X}_{[0,\tau_p]} (\Phi_n \circ J^{-1})^* B u_n \circ J dW - \int_0^T \mathcal{X}_{[0,\tau_p]} (\Phi \circ J^{-1})^* B u \circ J dW \right) = 0
\]
Hence
\[
P \left( A_n \cap [\tau_p = \infty] \right) \leq \frac{1}{\varepsilon} E \left( \left| \int_0^T \mathcal{X}_{[0,\tau_p]} (\Phi_n \circ J^{-1})^* B u_n \circ J dW - \int_0^T \mathcal{X}_{[0,\tau_p]} (\Phi \circ J^{-1})^* B u \circ J dW \right| \right)
\]
and so
\[
\lim_{n \to \infty} P \left( A_n \cap [\tau_p = \infty] \right) = 0
\]
Then
\[
P \left( A_n \right) = \sum_{p=1}^{\infty} P \left( A_n \cap ([\tau_p = \infty] \setminus [\tau_{p-1} < \infty]) \right)
\]
and so from the dominated convergence theorem,
\[
\lim_{n \to \infty} P \left( A_n \right) = \sum_{p=1}^{\infty} \lim_{n \to \infty} P \left( A_n \cap ([\tau_p = \infty] \setminus [\tau_{p-1} < \infty]) \right) = \sum_{p} 0 = 0.
\]
There was nothing special about $T$. The same argument holds for all $t$ and so $M(t)$ mentioned above has been identified as $\int_0^t (\Phi \circ J^{-1})^* Bu \circ JdW$. Then from (28.4.51)

$$
\lim_{n \to \infty} E \left( \sup_{t \in [0,T]} \left| \int_0^t (\Phi \circ J^{-1})^* Bu \circ JdW - \int_0^t (\Phi_n \circ J^{-1})^* Bu_n \circ JdW \right| \right) = 0
$$

It follows from the usual Borel Cantelli argument that there is a set of measure zero and a further subsequence such that off this set, all the above convergences happen and also

$$
\int_0^t (\Phi_n \circ J^{-1})^* Bu_n \circ JdW \to \int_0^t (\Phi \circ J^{-1})^* Bu \circ JdW
$$

uniformly on $[0,T]$. ■

The rest of the argument is identical. This yields the following theorem.

**Theorem 28.4.9** Suppose $V \equiv L^p ([0,T] \times \Omega, V)$ where $p \geq 2$, with the $\sigma$ algebra of progressively measurable sets and $V_\omega = L^p ([0,T], V)$.

$$
\Phi \in L^2 \left( [0,T] \times \Omega, L_2 \left( Q^{1/2} U, W \right) \right),
$$

$$
f \in V' \equiv L^{p'} ([0,T] \times \Omega, V')
$$

and both are progressively measurable. Suppose that

$$
\lambda B + A(\omega) : V_\omega \to V'_\omega, \quad \lambda B + A : V \to V'
$$

are monotone hemicontinuous and bounded where

$$
A(\omega) u(t) = A(t,u(t),\omega)
$$

and $(t,u,\omega) \to A(t,u,\omega)$ is progressively measurable. Also suppose for $p \geq 2$, the coercivity, and the boundedness conditions

$$
\lambda \langle Bu, u \rangle + \langle A(t,u,\omega), u \rangle_V \geq \delta \|u\|_V^p - c(t,\omega)
$$

(28.4.53)

where $c \in L^1 ([0,T] \times \Omega)$ for all $\lambda$ large enough. Also,

$$
\|A(t,u,\omega)\|_{V'} \leq k \|u\|_{V}^{p-1} + e^{1/p'}(t,\omega)
$$

(28.4.54)

also suppose the monotonicity condition for all $\lambda$ large enough.

$$
\langle (\lambda B + A(\omega))(u) - (\lambda B + A(\omega))(v), u-v \rangle \geq \delta \|u-v\|_W^2
$$

(28.4.55)

Then if $u_0 \in L^2 (\Omega, W)$ with $u_0 F_0$ measurable, there exists a unique solution $u(\cdot,\omega) \in V_\omega$ with $u \in V (L^p ([0,T] \times \Omega, V)$ and progressively measurable) such that for $\omega$ off a set of measure zero,

$$
Bu(t,\omega) - Bu_0(\omega) + \int_0^t A(s,u(s,\omega),\omega) ds = \int_0^t f ds + B \int_0^t \Phi dW.
$$

It is also assumed that $V$ is a reflexive separable real Banach space.

### 28.5 Replacing $\Phi$ With $\sigma (u)$

It is not hard to include the case where $\Phi$ is replaced with a function $\sigma (u)$. We make the following assumptions. **For each** $r > 0$ there exists $\lambda$ large enough that

$$
(\lambda B(u) + A(u) - (\lambda B(\dot{u}) + A(\dot{u})), u-\dot{u}) \geq r \|u-\dot{u}\|_W^2
$$

Note that in the case where $B = I$ and there is a conventional Gelfand triple, $V,H,V'$, this kind of condition is obvious if $\lambda I + A$ is monotone for some $\lambda$. Thus this is not an unreasonable assumption to make although it is stronger than some of the assumptions used above with the integral given by $\int_0^t \Phi dW$. 

As to \( \sigma \) we make the following assumptions,
\[
(t, u, \omega) \in [0, T] \times W \times \Omega \rightarrow \sigma (t, u, \omega) \text{ is progressively measurable into } W
\]
\[
\| \sigma (t, u, \omega) \|_W \leq C + C \| u \|_W
\]
\[
\| \sigma (t, u, \omega) - \sigma (t, \tilde{u}, \omega) \|_{L^2(Q/2U,W)} \leq K \| u - \tilde{u} \|_W
\]
That is, it has linear growth and is Lipschitz.

Let \( \lambda \) correspond to \( r \) where \( r - \| B \| K^2 > 4 \). Also let \( T \) be such that
\[
\hat{C}e^{\lambda T}K^2 < 3
\]
where \( \hat{C} \) is a constant used in the Burkholder Davis Gundy inequality. This is a restriction on the size of \( K \). Thus we only give a solution if \( K \) is small enough. Later, this will be removed in the most interesting case. This will give a local solution valid for a fixed \( T > 0 \) and then the global solution can be obtained by applying this result on the succession of intervals \([0, T], [T, 2T], [3T, 4T], \) and so forth.

From Theorem 28.3.4 if \( w \in W \), there exists a unique solution \( u \) to
\[
Bu (t, \omega) - Bu_0 (\omega) + \int_0^t A(s, u(s, \omega), \omega) ds = \int_0^t f ds + B \int_0^t \sigma(w) dW.
\]
holding in the sense described there. Let \( u_i \) result from \( w_i \). Then from the implicit Itô formula and the above monotonicity estimate,
\[
\frac{1}{2} \langle B (u_1 - u_2), u_1 - u_2 \rangle (t) + r \int_0^t \| u_1 - u_2 \|^2_W ds
\]
\[
- \lambda \int_0^t \langle B (u_1 - u_2), u_1 - u_2 \rangle ds
\]
\[
- \int_0^t \langle B \sigma (u_1) - B \sigma (u_2), \sigma (u_1) - \sigma (u_2) \rangle_{L^2} ds \leq M^* (t)
\]
where the right side is of the form \( \sup_{s \in [0,t]} |M(s)| \) where \( M(t) \) is a local martingale having quadratic variation dominated by
\[
C \int_0^t \| \sigma (w_1) - \sigma (w_2) \|^2 \langle B (u_1 - u_2), u_1 - u_2 \rangle ds
\]
(28.5.56)

Therefore, since \( M^* \) is increasing in \( t \), it follows from the Lipschitz condition on \( \sigma \) that
\[
\frac{1}{2} \langle B (u_1 - u_2), u_1 - u_2 \rangle (t) + r \int_0^t \| u_1 - u_2 \|^2_W ds
\]
\[
- \lambda \int_0^t \langle B (u_1 - u_2), u_1 - u_2 \rangle ds - \| B \| K^2 \int_0^t \| u_1 - u_2 \|^2_W ds \leq M^* (t)
\]
Thus, from the assumption about \( r \),
\[
\sup_{s \in [0,t]} \langle B (u_1 - u_2), u_1 - u_2 \rangle (s) + 4 \int_0^t \| u_1 - u_2 \|^2_W ds
\]
\[
\leq \lambda \int_0^t \langle B (u_1 - u_2), u_1 - u_2 \rangle ds + 2M^* (t)
\]
Then applying Gronwall’s inequality,
\[
\sup_{s \in [0,t]} \langle B (u_1 - u_2), u_1 - u_2 \rangle (s) + 4 \int_0^t \| u_1 - u_2 \|^2_W ds \leq 2e^{\lambda T}M^* (t)
\]
Now take expectations and use the Burkholder Davis Gundy inequality. The expectation of the right side is then dominated by
\[
2\hat{C}e^{\lambda T} \int \left( \int_0^t \| \sigma (w_1) - \sigma (w_2) \|^2 \langle B (u_1 - u_2), u_1 - u_2 \rangle ds \right)^{1/2} dP
\]
\[
\begin{align*}
&\leq \left[ \int_{\Omega} \sup_{s \in [0,t]} \langle B(u_1 - u_2), u_1 - u_2 \rangle^{1/2} \right.
\left. \cdot 2\bar{C} e^{\lambda T} \left( \int_0^t K^2 \| w_1 - w_2 \|_W \ dt \right)^{1/2} \ dP \right] \\
&\leq E \left( \frac{1}{2} \sup_{s \in [0,t]} \langle B(u_1 - u_2), u_1 - u_2 \rangle (s) \right)
+ \bar{C} e^{\lambda T} E \left( \int_0^t K^2 \| w_1 - w_2 \|_W \ dt \right)
\end{align*}
\]

It follows that, after adjusting constants as needed, one gets an inequality of the following form.
\[
\frac{1}{2} E \left( \sup_{s \in [0,t]} \langle B(u_1 - u_2), u_1 - u_2 \rangle (s) \right) + 4 \int_0^t \| u_1 - u_2 \|_W \ ds dP \\
\leq \bar{C} e^{\lambda T} E \left( \int_0^t K^2 \| w_1 - w_2 \|_W \ dt \right)
\]

This holds for every \( t \leq T \) and so, from the estimate on the size of \( T \), it follows that
\[
\int_0^T \int_{\Omega} \| u_1 - u_2 \|_W^2 \ ds dP \leq \frac{3}{4} \int_0^T \int_{\Omega} \| w_1 - w_2 \|_W^2 \ dt dP
\]

Therefore, there is a unique fixed point to this mapping which takes \( w \in W \) to \( u \) the solution to the integral equation. We denote it as \( u \). Thus \( u \) is progressively measurable and for \( \omega \) off a set of measure zero, we have a solution to the integral equation
\[
Bu(t,\omega) - Bu_0(\omega) + \int_0^t A(s, u(s, \omega), \omega) \ ds \\
= \int_0^t f ds + B \int_0^t \sigma (u) dW, t \in [0, T]
\]

Now the same argument can be repeated for the succession of intervals mentioned above. However, you need to be careful that at \( T \), you have \( Bu(T, \omega) = B(u(T, \omega)) \) for \( \omega \) off a set of measure zero. If this is not so, you locate \( T' \) close to \( T \) for which it is so as in Lemma 28.5.1 and use this \( T' \) instead, but these are mainly technical issues. This proves the following existence and uniqueness theorem.

**Theorem 28.5.1** Suppose \( f \in V' \) is progressively measurable and that \((t, \omega) \to \sigma (t, \omega, u(t, \omega)) \) is progressively measurable whenever \( u \) is. Suppose that
\[
\lambda B + A(\omega) : V_\omega \to V_\omega', \ \lambda B + A : V \to V'
\]
are monotone hemicontinuous and bounded where
\[
A(\omega) u(t) \equiv A(t, u(t), \omega)
\]
and \((t, u, \omega) \to A(t, u, \omega) \) is progressively measurable. Also suppose for \( p \geq 2 \), the coercivity, and the boundedness conditions
\[
\lambda \langle B(u), u \rangle \geq \delta \| u \|_{V'}^2 - c(t, \omega) \quad (28.5.57)
\]
for all \( \lambda \) large enough.
\[
\| A(t, u, \omega) \|_{V'} \leq k \| u \|_{V'}^{-1} + c^{1/p'}(t, \omega) \quad (28.5.58)
\]
where \( c \in L^1 ([0,T] \times \Omega) \). Also suppose the monotonicity condition that for all \( r > 0 \) there exists \( \lambda \) such that
\[
\langle (\lambda B + A(\omega)) (u) - (\lambda B + A(\omega)) (v), u - v \rangle \geq r \| u - v \|_W^2
\]
Also suppose that
\[
(t, u, \omega) \in [0,T] \times W \times \Omega \to \sigma (t, u, \omega) \) is progressively measurable into \( W \\
\| \sigma (t, u, \omega) \|_W \leq C + C \| u \|_W
\]
\[ \| \sigma(t, u, \omega) - \sigma(t, \hat{u}, \omega) \|_{L_2(Q^{1/2}U, W)} \leq K \| u - \hat{u} \|_W \]

Then if \( u_0 \in L^2(\Omega, W) \) with \( u_0 \mathcal{F}_0 \) measurable, there exists a unique solution \( u(\cdot, \omega) \in \mathcal{V}_\omega \) with \( u \in \mathcal{V} (L^p([0, T] \times \Omega, V)) \) and progressively measurable) such that for \( \omega \) off a set of measure zero,

\[ Bu(t, \omega) - Bu_0(\omega) + \int_0^t A(s, u(s, \omega), \omega) \, ds = \int_0^t f(s, \omega) \, ds + B \int_0^t \sigma(u) \, dW. \]

In case \( B \) is the Riesz map, you do not have to make any assumption on the size of \( K \). Thus

\[ \langle Bu, u \rangle = \| u \|_W^2 \]

The case of most interest is the usual one where \( V \subseteq W = W' \subseteq V' \), the case of a Gelfand triple in which \( B \) is the identity. As to \( \sigma \), the assumption is made that

\[ \| \sigma(t, u, \omega) \|_W \leq C + C \| u \|_W \]

\[ \| \sigma(t, \omega, u_1) - \sigma(t, \omega, u_2) \|_{L_2(Q^{1/2}U, W)} \leq K \| u_1 - u_2 \|_W \]

Of course it is also assumed that whenever \( u \) has values in \( W \) and is progressively measurable, \( (t, \omega) \rightarrow \sigma(t, \omega, u(t, \omega)) \) is also progressively measurable into \( L_2(Q^{1/2}U, W) \).

Letting \( w_i \in L^2([0, T] \times \Omega, W) \) each \( w_i \) being progressively measurable, the above assumptions imply that there exists a solution \( u_i \) to the integral equation

\[ Bu_i(t, \omega) - Bu_0(\omega) + \int_0^t A(u_i) \, ds = \int_0^t f(s, \omega) \, ds + B \int_0^t \sigma(w_i) \, dW \]

different equations.

In particular, consider

\[ w \in L^2([0, T] \times \Omega, W) \cap L^\infty([0, T], L^2(\Omega, W)) \]

and let \( u \) be the solution which results from placing \( w \) in \( \sigma \). Then from the estimates,

\[ \langle Bu, u \rangle(t) - \langle Bu_0, u_0 \rangle + \delta \int_0^t \| u \|_V^p \, ds = 2 \int_0^t \langle f, u \rangle \, ds + C(b_3, b_4, b_5) + \lambda \int_0^t \langle Bu, u \rangle \, ds + \int_0^t \langle B\sigma(w), \sigma(w) \rangle \, ds + 2M^* (t) \]

\[ \leq 2 \int_0^t \langle f, u \rangle \, ds + C(b_3, b_4, b_5) + \lambda \int_0^t \langle Bu, u \rangle \, ds + \int_0^t \left( C + C \| u \|_W^2 \right) \, ds + 2M^* (t) \]

where \( M^*(t) = \sup_{s \in [0, t]} |M(s)| \) and the quadratic variation of \( M \) is no larger than

\[ \int_0^t \| \sigma(w) \|^2 (B(u), u) \, ds \]

Then using Gronwall’s inequality, one obtains an inequality of the form

\[ \sup_{s \in [0, T]} \langle Bu, u \rangle(s) \leq C + C \left( M^*(t) + \int_0^t \| w \|_W^2 \, ds \right) \]

where \( C = C(u_0, f, \delta, \lambda, b_3, b_4, b_5, T) \) and is integrable. Then take expectation. By Burkholder Davis Gundy inequality and adjusting constants as needed,

\[ E \left( \sup_{s \in [0, T]} \langle Bu, u \rangle(s) \right) \]

\[ \leq C + C \int_0^T \| w \|_W^2 \, ds dP + C \int_\Omega \left( \int_0^T \| \sigma(w) \|^2 (B(u), u) \, ds \right)^{1/2} \, dP \]

\[ \leq C + C \int_0^T \| w \|_W^2 \, ds dP + C \int_\Omega \sup_{s \in [0, T]} \langle Bu, u \rangle^{1/2} (s) \left( \int_0^T \| \sigma(w) \|^2 \, ds \right)^{1/2} \, dP \]
Thus
\[ E(⟨Bu, u⟩(t)) \leq E \left( \sup_{s \in [0, T]} ⟨Bu, u⟩(s) \right) \leq C + C \int_0^T \|w\|^2_W \, ds \, dP \]
and so
\[ \|u\|^2_{L^∞([0, T], L^2(Ω, W))} \leq C + C \int_0^T \|w\|^2_W \, ds \, dP \]
which implies \( u \in L^∞ \left( [0, T], L^2(Ω, W) \right) \) and is progressively measurable.

Using the monotonicity assumption, there is a suitable \( \lambda \) such that
\[ \frac{1}{2} \langle B(u_1 - u_2), u_1 - u_2 \rangle (t) + \lambda \int_0^t \|u_1 - u_2\|^2_W \, ds \]
\[ -\lambda \int_0^t \langle B(u_1 - u_2), u_1 - u_2 \rangle \, ds \]
\[ - \int_0^t \langle Bσ(u_1) - Bσ(u_2), σ(u_1) - σ(u_2)⟩_{L^2} \, ds \leq M^* (t) \]
where the right side is of the form \( \sup_{s \in [0, t]} |M(s)| \) where \( M(t) \) is a local martingale having quadratic variation dominated by
\[ C \int_0^t \|σ(w_1) - σ(w_2)\|^2 \langle B(u_1 - u_2), u_1 - u_2 \rangle \, ds \]  
(28.5.59)

Then by assumption and using Gronwall’s inequality, there is a constant \( C = C(λ, K, T) \) such that
\[ \langle B(u_1 - u_2), u_1 - u_2 \rangle (t) \leq CM^* (t) \]

Then also, since \( M^* \) is increasing,
\[ \sup_{s \in [0, t]} \langle B(u_1 - u_2), u_1 - u_2 \rangle (s) \leq CM^* (t) \]

Taking expectations and from the Burkholder Davis Gundy inequality,
\[ E \left( \sup_{s \in [0, t]} \langle B(u_1 - u_2), u_1 - u_2 \rangle (s) \right) \leq C \int_0^t \left( \int_0^s \|σ(w_1) - σ(w_2)\|^2 \langle B(u_1 - u_2), u_1 - u_2 \rangle \right)^{1/2} \, dP \]
\[ \leq C \int_0^t \sup_{s \in [0, t]} \langle B(u_1 - u_2), u_1 - u_2 \rangle^{1/2} \left( \int_0^s \|σ(w_1) - σ(w_2)\|^2 \right)^{1/2} \, dP \]

Then it follows after adjusting constants that there exists an inequality of the form
\[ E \left( \sup_{s \in [0, t]} \langle B(u_1 - u_2), u_1 - u_2 \rangle (s) \right) \leq CE \left( \int_0^t \|σ(w_1) - σ(w_2)\|^2_{L^2} \, ds \right) \]

Hence
\[ E \left( \sup_{s \in [0, t]} \langle B(u_1 - u_2), u_1 - u_2 \rangle (t) \right) \leq CK^2E \left( \int_0^t \|w_1 - w_2\|^2_W \, ds \right) \]

Thus, for each \( t \leq T \)
\[ \int_0^t \langle B(u_1 - u_2), u_1 - u_2 \rangle (t) \, dP \leq CK^2E \left( \int_0^t \|w_1 - w_2\|^2_W \, ds \right) \]

one can consider the map \( ψ(w) ≡ u \) as described above. Then the above inequality implies
\[ E(⟨B(ψ^n w_1 - ψ^n w_2), ψ^n w_1 - ψ^n w_2⟩ (t)) \leq CK^2E \left( \int_0^t \|ψ^{n-1} w_1 - ψ^{n-1} w_2\|^2_W \, dt \right) \]
\[
CK^2 E \left( \int_0^t \langle B (\psi^{n-1}w_1 - \psi^{n-1}w_2), \psi^{n-1}w_1 - \psi^{n-1}w_2 \rangle (t) \, dt \right)
\]
\[
\leq (CK^2)^2 E \left( \int_0^t \int_0^{t_1} \langle B (\psi^{n-2}w_1 - \psi^{n-2}w_2), \psi^{n-2}w_1 - \psi^{n-2}w_2 \rangle (t_2) \, dt_2 \, dt_1 \right)
\]
\[
\cdots \leq (CK^2)^n E \left( \int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} \langle B (w_1 - w_2), w_1 - w_2 \rangle (t_n) \, dt_n \cdots dt_1 \right)
\]
\[
= (CK^2)^n \int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} E (\langle B (w_1 - w_2), w_1 - w_2 \rangle (t_n) \, dt_n \cdots dt_1
\]
\[
\leq (CK^2)^n \sup_t E (\langle B (w_1 - w_2), w_1 - w_2 \rangle (t)) \frac{T^n}{n!} < \frac{1}{2} \|w_1 - w_2\|^2_{L^\infty([0,T],L^2(\Omega,W))}
\]
provided \( n \) is sufficiently large. It follows that
\[
\|\psi^n w_1 - \psi^n w_2\|^2_{L^\infty([0,T],L^2(\Omega,W))} \leq \frac{1}{2} \|w_1 - w_2\|^2_{L^\infty([0,T],L^2(\Omega,W))}
\]
for all \( n \) sufficiently large. Hence, if one begins with \( w \in L^\infty ([0,T],L^2 (\Omega,W)) \cap L^2 ([0,T] \times \Omega, W) \), the sequence of iterates \( \{\psi^n w\}_{n=1}^\infty \) must converge to some fixed point \( u \) in \( L^\infty ([0,T],L^2 (\Omega, W)) \). This \( u \) is automatically in \( L^2 ([0,T] \times \Omega, W) \) and is progressively measurable since each of the iterates is progressively measurable. This proves the following theorem.

**Theorem 28.5.2**  **Suppose** \( f \in \mathcal{V}' \)  **is progressively measurable and that** \( (t, \omega) \to \sigma (t, u, \omega) \) **is progressively measurable whenever** \( u \) **is. Suppose that** \( B : W \to W' \) **is a Riesz map.**

\[
\lambda B + A (\omega) : \mathcal{V}_\omega \to \mathcal{V}_\omega', \lambda B + A : \mathcal{V} \to \mathcal{V}'
\]

**are monotone hemicontinuous and bounded where**

\[
A (\omega) u (t) \equiv A (t, u (t), \omega)
\]

and \( (t, u, \omega) \to A (t, u, \omega) \) **is progressively measurable. Also suppose for** \( p \geq 2 \), **the coercivity, and the boundedness conditions**

\[
\lambda \langle Bu, u \rangle + \langle A (t, u, \omega), u \rangle_V \geq \delta \|u\|_V^p - c (t, \omega)
\]

for all \( \lambda \) large enough.

\[
\|A (t, u, \omega)\|_{\mathcal{V}'} \leq k \|u\|_{\mathcal{V}}^{p-1} + c^{1/p'} (t, \omega)
\]

where \( c \in L^1 ([0,T] \times \Omega) \). Also suppose that

\[
\|\sigma (t, u, \omega)\|_W \leq C + C \|u\|_W
\]

\[
\|\sigma (t, u, \omega) - \sigma (t, \tilde{u}, \omega)\|_{L^2 (\Omega/2, W)} \leq K \|u - \tilde{u}\|_W
\]

Then if \( u_0 \in L^2 (\Omega, W) \) with \( u_0 \mathcal{F}_0 \) measurable, there exists a unique solution \( u (\cdot, \omega) \in \mathcal{V}_\omega \) with \( u \in \mathcal{V} (L^p ([0,T] \times \Omega, V) \) and progressively measurable) such that for \( \omega \) off a set of measure zero,

\[
Bu (t, \omega) - Bu_0 (\omega) + \int_0^t A (s, u (s, \omega), \omega) \, ds = \int_0^t f \, ds + B \int_0^t \sigma (u) \, dW.
\]

### 28.6  Examples

Here we give some examples. The first is a standard example, the porous media equation, which is discussed well in [E]. For stochastic versions of this example, see [M]. The generalization to stochastic equations does not require the theory developed here. We will show, however, that it can be considered in terms of the theory of this paper without much difficulty using an approach proposed in [M]. These examples involve operators which are not monotone, in the usual way but they can be transformed into equations which do fit the above theory.
Example 28.6.1 The stochastic porous media equation is
\[ u_t - \Delta \left( u |u|^{p-2} \right) = f, \quad u(0) = u_0, \quad u = 0 \text{ on } \partial U \]
where here \( U \) is a bounded open set in \( \mathbb{R}^n, n \leq 3 \) having Lipschitz boundary. One can consider a stochastic version of this as a solution to the following integral equation
\[ u(t) - u_0 + \int_0^t \left( -\Delta \left( u |u|^{p-2} \right) \right) ds = \int_0^t \Phi dW + \int_0^t f ds \quad (28.6.62) \]
where here \( \Phi \in L^2 \left( [0,T] \times \Omega, \mathcal{L}_2 \left( Q^{1/2} U, H \right) \right) \cap L^2 \left( \Omega, L^\infty \left( [0,T], \mathcal{L}_2 \left( Q^{1/2} U, H \right) \right) \right), \)
\( H = L^2(U) \) and the equation holds in the manner described above in \( H^{-1}(U) \). Assume \( p \geq 2 \) and \( f \in L^2((0,T) \times \Omega,H) \).

One can consider this as an implicit integral equation of the form
\[ (\Delta)^{-1} u(t) - (\Delta)^{-1} u_0 + \int_0^t u |u|^{p-2} ds = (\Delta)^{-1} \int_0^t \Phi dW + (\Delta)^{-1} \int_0^t f ds \quad (28.6.63) \]
where \( -\Delta \) is the Riesz map of \( H_0^1(U) \) to \( H^{-1}(U) \). Then we can also consider \( (\Delta)^{-1} \) as a map from \( L^2(U) \) to \( L^2(U) \) as follows.
\[ (\Delta)^{-1} \Delta = f \text{ where } -\Delta u = f, \quad u = 0 \text{ on } \partial U. \]
Thus we let \( W = L^2(U) = H \) and \( V = L^p(U) \). Let \( B \equiv (\Delta)^{-1} \) on \( L^2(U) \) as just described. Let \( A(u) = u |u|^{p-2} \).

It is obvious that the necessary coercivity condition holds. In addition, there is a strong monotonicity condition which holds. Therefore, if \( u_0 \in L^2(\Omega,L^2(U)) \) and \( F_0 \) measurable, Theorem 28.4.7 applies and we can conclude that there exists a unique solution to the integral equation \( (28.6.63) \) in the sense described in this theorem. Here \( u \in L^p((0,T) \times \Omega,L^p(U)) \) and is progressively measurable, the integral equation holding for all \( t \) for \( \omega \) off a set of measure zero. Since \( A \) satisfies for some \( \delta > 0 \) an inequality of the form
\[ (Au - Av,u-v) \geq \delta \|u-v\|_{L^p(U)}^2 \]
it follows easily from the above methods that the solution is also unique. In fact, this follows right away from Theorem 28.4.7 because \( (\Delta^{-1}u,u) = \|u\|_{H^{-1}}^2 \).

Also note that from the integral equation,
\[ (\Delta)^{-1} \left( u(t) - u_0 - \int_0^t \Phi dW \right) + \int_0^t u |u|^{p-2} ds = (\Delta)^{-1} \int_0^t f ds \]
and so, since \( (\Delta) \) is the Riesz map on \( H_0^1(U) \), the integral equation above shows that off a set of measure zero,
\[ \int_0^t u |u|^{p-2} ds = (\Delta)^{-1} \left( \int_0^t f ds - \left( u(t) - u_0 - \int_0^t \Phi dW \right) \right) \in L^2(0,T,H_0^1(U) \cap H^2(U)) \]
by elliptic regularity results. If it were not for that stochastic integral, one could assert that \( |u|^\frac{p-2}{2} u \in L^2(0,T,H_0^1(U)) \).

This is shown in [28.6.6]. However, it appears that no such condition can be obtained here because of the nowhere differentiability of the stochastic integral, even if more is assumed on \( u_0 \) and \( \Phi \).

Also in this reference is a treatment of the Stefan problem. The Stefan problem involves a partial differential equation
\[ u_t - \sum_i \frac{\partial}{\partial x_i} \left( k(u) \frac{\partial u}{\partial x_i} \right) = f, \quad \text{on } U \times [0,T] \equiv Q \]
for \( (x,t) \notin S \) where \( u \) is the temperature and \( k(u) \) has a jump at \( \sigma \) and \( S \) is given by \( u(x,t) = \sigma \). It is assumed that \( 0 < k_1 \leq k(r) \leq k_2 < \infty \) for all \( r \in \mathbb{R} \). For example, its graph could be of the form
\[
\begin{array}{c|c}
& \\
\hline
k(u) & \sigma \\
\hline
\end{array}
\]
\[ u \]
On $S$ there is a jump condition which is assumed to hold. Namely

$$bn_t - (k (u^+) u,i (+) - k (u^-) u,i (-)) n_i = 0$$

where the sum is taken over repeated indices and $b > 0$. $u (+)$ is the “limit” as $(x', t') \to (x, t) \in S$ where $(x', t') \in S_+$, $u (-)$ defined similarly. Also $n$ will denote the unit normal which goes from $S_+ \equiv \{(x, t) : u(x, t) > \sigma\}$ toward $S_- \equiv \{(x, t) : u(x, t) < \sigma\}$.

In addition, there is an initial condition and boundary conditions

$$u(x, 0) = u_0(x) \notin S, \quad u(x, t) = 0 \text{ on } \partial U.$$

The idea is to obtain a variational formulation of this thing. To do this, let $K(r) \equiv \int_0^r k(s) \, ds$. Thus in the case of the above picture, the graph of $K(r)$ would look like

$$\begin{array}{c}
K(u) \\
\sigma \\
\end{array}$$

Now let $\beta(t)$ be a function which satisfies

$$\beta'(t) = \frac{1}{k(K^{-1}(t))} \text{ for } t \neq \tau \equiv K(\sigma)$$

and it has a jump equal to $b$ at $\tau$.

$$\begin{array}{c}
\beta(v) \\
\tau \\
\end{array}$$

Let $v = K(u)$. Then for $u \neq \sigma$, equivalently $v \neq \tau$,

$$v_t = K'(u) u_t = k(K^{-1}(v)) u_t$$

and so

$$u_t = \frac{1}{k(K^{-1}(v))} v_t = \frac{d}{dt} (\beta(v))$$

Also,

$$v,i = K'(u) u,i = k(u) u,i$$

and so

$$u,i = \frac{1}{k(u)} v,i$$

Hence

$$(k(u) u,i)_i = \left(k(u) \frac{1}{k(u)} v,i\right)_i = \Delta v$$

Thus, off the set $S$,

$$\beta(v)_t - \Delta v = f$$

Now let $\phi \in L^2(0, T, H^1_0(U))$ with $\phi(x, T) = 0$. Then assume $S$ is sufficiently smooth that things like divergence theorem apply. Also note that $u = \sigma$ is the same as $v = \tau$.

$$\int_Q (\beta(v)_t - \Delta v) \phi = \int_{S_+} (\beta(v)_t - \Delta v) \phi + \int_{S_-} (\beta(v)_t - \Delta v) \phi$$

$$= \int_{S_+} (\beta(v) \phi)_t - \beta(v) \phi_t - (v_i \phi)_i + v,i \phi,i$$

$$+ \int_{S_-} (\beta(v) \phi)_t - \beta(v) \phi_t - (v_i \phi)_i + v,i \phi,i$$
Now using the divergence theorem, and continuing these formal manipulations, the above reduces to
\[
\int_{S} \beta(v(+)) \phi n_t - (v,i(+)) \phi n_i + \int_{S_+} -\beta(v) \phi_t + v_i \phi_i - \int_{U \cap S_+} \beta(v(x,0)) \phi(x,0) \\
+ \int_{S} -\beta(v(-)) \phi n_t + (v,i(-)) \phi n_i + \int_{S_-} -\beta(v) \phi_t + v_i \phi_i - \int_{U \cap S_-} \beta(v(x,0)) \phi(x,0)
\]
Combining the two integrals over \(S\) yields
\[
\int_{S} (bn_t - (v,i(+)) - v_i(-)) n_i \phi = \int_{S} (bn_t - (k(u) u,i(+)) - k(u) u,i(-)) n_i \phi = 0
\]
by assumption. Therefore, including \(f\), we obtain
\[
\int_{Q} -\beta(v) \phi_t + v_i \phi_i - \int_{U} \beta(v(x,0)) \phi(x,0) = \int_{Q} f \phi
\]
which implies, using the initial condition
\[
\int_{U} \beta(v_0) \phi(x,0) + \int_{Q} ((\beta(v))^t \phi + v_i \phi_i) - \int_{U} \beta(v(x,0)) \phi(x,0) = \int_{Q} f \phi
\]
Regard \(\beta\) as a maximal monotone graph and let \(\alpha(t) \equiv \beta^{-1}(t)\). Thus \(\alpha\) is single valued. It just has a horizontal place corresponding to the jump in \(\beta\). Then let \(w = \beta(v)\) so that \(v = \alpha(w)\).

\[
\alpha(w)
\]

Then in terms of \(w\), the above equals
\[
\int_{U} w_0 \phi(x,0) + \int_{Q} (w' \phi + \alpha(w)i \phi_i) - \int_{U} w(x,0) \phi(x,0) = \int_{Q} f \phi
\]
and so this simplifies to
\[
w' - \Delta(\alpha(w)) = f, \ w(0) = w_0
\]
where \(\alpha\) maps onto \(\mathbb{R}\) and is monotone and satisfies
\[
(\alpha(r_1) - \alpha(r_2)) (r_1 - r_2) \geq 0, \ |\alpha(r)| \leq m |r|, \\
|\alpha(r_1) - \alpha(r_2)| \leq m |r_1 - r_2|, \alpha(r) r \geq \delta |r|^2
\]
for some \(\delta, m > 0\). Then \(K^{-1}(\alpha(w)) = u\) where \(u\) is the original dependent variable. Obviously, the original function \(k\) could have had more than one jump and you would handle it the same way by defining \(\beta\) to be like \(K^{-1}\) except for having appropriate jumps at the values of \(K(u)\) corresponding to the jumps in \(k\). This explains the following example.

**Example 28.6.2** It can be shown that the Stefan problem can be reduced to the consideration of an equation of the form
\[
w_t - \Delta(\alpha(w)) = f, \ w(0) = w_0
\]
where \(\alpha : L^2(0,T,L^2(U)) \rightarrow L^2(0,T,L^2(U))\) is monotone hemicontinuous and coercive, \(\alpha\) being a single valued function. Here \(f\) is the same which occurred in the original partial differential equation
\[
u_t - \sum_i \frac{\partial}{\partial x_i} (k(u) \frac{\partial u}{\partial x_i}) = f
\]
Thus a stochastic Stefan problem could be considered in the form
\[
(-\Delta^{-1}) w(t) - (-\Delta)^{-1} w_0 + \int_0^t \alpha(w) ds = (-\Delta^{-1}) \int_0^t f ds + (-\Delta^{-1}) \int_0^t \Phi dW
\]
This example can be included in the above general theory because
\[
\left(-\Delta^{-1}\right) u - \left(-\Delta^{-1}\right) v, u - v \geq \|u - v\|_{V'}, \quad \langle y, R^{-1}y \rangle_{L^2(U)} = \langle RR^{-1}y, R^{-1}y \rangle_{V', V} = \|R^{-1}y\|^2_V = \|y\|^2_{V'}
\]

This is seen as follows. \( V, L^2(U), V' \) is a Gelfand triple. Then \(-\Delta\) is the Riesz map \( R \) from \( H^1_0(U) \) to \( H^{-1}(U) \).
then you have
\[
\left(\begin{array}{c}
y(x) - \frac{1}{2} \int_0^T \nabla \cdot \left( |\nabla u|^{p-2} \nabla u \right) dx + \frac{1}{2} \int_0^T \nabla \cdot \left( |\nabla v|^{p-2} \nabla v \right) dx = 0,
\end{array}\right)
\]

Next we give a simple example which is a singular and degenerate equation. This is a model problem which illustrates how the theory can be used. This problem is mixed parabolic and stochastic and nonlinear elliptic. It is a singular equation because the coefficient \( b \) can be unbounded. The existence of a solution is easy to obtain from the above theory but it does not follow readily from other methods. If \( p = 2 \) it is an abstract version of stochastic heat equation which could model a material in which the density becomes vanishingly small in some regions and very large in other regions.

**Example 28.6.3** Suppose \( U \) is a bounded open set in \( \mathbb{R}^3 \) and \( b(x) \geq 0, b \in L^p(U), p \geq 4 \) for simplicity. Consider the degenerate stochastic initial boundary value problem

\[
\begin{align*}
&b(\cdot)(t, \cdot) - b(\cdot) u_0(\cdot) - \int_0^t \nabla \cdot \left( |\nabla u|^{p-2} \nabla u \right) = b \int_0^t \Phi dW,
&u = 0 \quad \text{on } \partial U,
\end{align*}
\]

where \( \Phi \in L^2([0, T] \times \Omega, L^2(Q^{1/2}U, W)) \) for \( W = H^1_0(U) \).

To consider this equation and initial condition, it suffices to let \( W = H^1_0(U), V = W^{3,p}_0(U) \),

\[
A : V \rightarrow V', \quad (Au, v) = \int_U |\nabla u|^{p-2} \nabla u \cdot \nabla v dx,
\]

\[
B : W \rightarrow W', \quad (Bu, v) = \int_U b(x) u(x) v(x) dx
\]

Then by the Sobolev embedding theorem, \( B \) is obviously self adjoint, bounded and nonnegative. This follows from a short computation:

\[
\left| \int_U b(x) u(x) v(x) dx \right| \leq \|v\|_{L^4(U)} \left( \int_U |b(x)|^{4/3} |u(x)|^{4/3} dx \right)^{3/4}
\]

\[
\leq \|v\|_{H^1_0(U)} \left( \int_U |b(x)| dx \right)^{1/3} \left( \int_U |u(x)|^2 dx \right)^{2/3} \left( \int_U |u(x)|^4 dx \right)^{1/3}
\]

\[
= \|v\|_{H^1_0(U)} \|b\|_{L^4(U)} \|u\|_{L^2(U)} \leq \|b\|_{L^4} \|u\|_{H^1_0} \|v\|_{H^1_0}
\]

Also for some \( \delta > 0 \)

\[
\langle Au - Av, u - v \rangle \geq \delta \|u - v\|^p_{V'}
\]

The nonlinear operator is obviously monotone and hemicontinuous. As for \( u_0 \), it is only necessary to assume \( u_0 \in L^2(\Omega, W) \) and \( f \) measurable. Then Theorem 28.4.4 gives the existence of a solution in the sense that for a.e. \( \omega \) the integral equation holds for all \( t \). Note that \( b \) can be unbounded and may also vanish. Thus the equation can degenerate to the case of a non stochastic nonlinear elliptic equation.

The existence theorems can easily be extended to include the situation where \( \Phi \) is replaced with a function of the unknown function \( u \). This is done by splitting the time interval into small sub intervals of length \( h \) and retarding the function in the stochastic integral, like a standard proof of the Peano existence theorem. Then the Ito formula is applied to obtain estimates and a limit is taken.

Other examples of the usefulness of this theory will result when one considers stochastic versions of systems of partial differential equations in which there is a nonlinear coupling between a parabolic equation and a nonlinear elliptic equation. These kinds of problems occur, for example as quasistatic damage problems in which the damage parameter satisfies a parabolic equation and the balance of momentum is a nonlinear elliptic equation and the two equations are coupled in a nonlinear way.
28.7 Other Examples, Inclusions

The above general result can also be used as a starting point for evolution inclusions or other situations where one does not have hemicontinuous operators. Assume here that

$$\Phi \in L^\infty ([0,T] \times \Omega, L_2 (Q^{1/2}U, H)).$$

We will use the following simple observation. Let $\alpha > 2$. Let $\|\Phi\|_\infty$ denote the norm in $L^\infty ([0,T] \times \Omega, L_2 (Q^{1/2}U, H))$.

By the Burkholder Davis Gundy inequality,

$$\int_\Omega \left( \int_s^t \Phi dW \right)^\alpha dP \leq C \int_\Omega \left( \int_s^t \|\Phi\|^2 d\tau \right)^{\alpha/2} dP \leq C \|\Phi\|_\infty^\alpha \int_\Omega \left( \int_s^t d\tau \right)^{\alpha/2} dP = C \|\Phi\|_\infty^\alpha |t-s|^{\alpha/2}$$

By the Kolmogorov Čentsov theorem, this shows that $t \to \int_0^t \Phi dW$ is Holder continuous with exponent

$$\gamma < \frac{(\alpha/2) - 1}{\alpha} = \frac{1}{2} - \frac{1}{\alpha}$$

Since $\alpha > 2$ is arbitrary, this shows that for any $\gamma < 1/2$, the stochastic integral is Holder continuous with exponent $\gamma$. This is exactly the same kind of continuity possessed by the Wiener process. We state this as the following lemma.

**Lemma 28.7.1** Let $\Phi \in L^\infty ([0,T] \times \Omega, L_2 (Q^{1/2}U, H))$ then for any $\gamma < 1/2$, the stochastic integral $\int_0^t \Phi dW$ is Holder continuous with exponent $\gamma$.

To begin with, we consider a stochastic inclusion. Suppose, in the context of Theorem 28.4.7, that $V$ is a closed subspace of $W^{\sigma,p} (U)$, $\sigma > 1$ which contains $C_c^\infty (U)$ where $U$ is an open bounded set in $\mathbb{R}^n$, different than the Hilbert space $U$. (In case the matrix $A$ which follows equals 0, it suffices to take $\sigma \geq 1$.) Let

$$\sum_{i,j} a_{i,j} (x) \xi_i \xi_j \geq 0, \quad a_{ij} = a_{ji}$$

where the $a_{i,j} \in C^1 (\bar{U})$. Denoting by $A$ the matrix whose $ij^{th}$ entry is $a_{ij}$, let

$$W \equiv \left\{ u \in L^2 (U) : ( u, \quad A^{1/2} \nabla u ) \in L^2 (U)^{n+1} \right\}$$

with a norm given by

$$\|u\|_W \equiv \left( \int_U \left( uv + \sum_{i,j} a_{ij} (x) \partial_i u \partial_j \right) dx \right)^{1/2}$$

$B : W \to W'$ be given by

$$\langle Bu, v \rangle \equiv \int_U \left( uv + \sum_{i,j} a_{ij} (x) \partial_i u \partial_j \right) dx$$

so that $B$ is the Riesz map for this space. The case where the $a_{ij}$ could vanish is allowed. Thus $B$ is a positive self adjoint operator and is therefore, included in the above discussion. In this example, it will be significant that $B$ is one to one and does not vanish.

This operator maps onto $L^2 (U)$ because of basic considerations concerning maximal monotone operators. This is because

$$\langle Du, v \rangle \equiv \int_U \sum_{i,j} a_{ij} (x) \partial_i u \partial_j v dx$$
can be obtained as a subgradient of a convex lower semicontinuous and proper functional defined on $L^2(U)$. Therefore, the operator is maximal monotone on $L^2(U)$ which means that $I + D$ is onto. The domain of $D$ consists of all $u \in L^2(U)$ such that

$$Du = -\sum_{i,j} \partial_j (a_{ij}(x) \partial_i u) \in L^2(U)$$

along with suitable boundary conditions determined by the choice of $V$. It follows that if $u + Du = Bu = f \in H = L^2(U)$, then

$$u - \sum_{i,j} \partial_j (a_{ij} \partial_i u) = f$$

Therefore,

$$\|u\|_{L^2(U)}^2 + \int_U \sum_{i,j} a_{ij}(x) \partial_i u \partial_j u = \|u\|_W^2 = (f,u) \leq \|f\|_{L^2(U)} \|u\|_{L^2(U)} \leq \|f\|_{L^2(U)} \|u\|_W$$

which shows that the map $B^{-1}: H = L^2(U) \to W$ is continuous.

Next suppose that $\Phi \in L^\infty([0,T] \times \Omega; L_2(Q^{1/2}U,H))$. Then by continuity of the mapping $B^{-1}$, it follows that $\Psi \equiv B^{-1}\Phi$ satisfies $\Psi \in L^\infty([0,T] \times \Omega; L_2(Q^{1/2}U,W))$. Thus $\Phi = B\Psi$. In addition to this, to simplify the presentation, assume in addition that

$$\langle A(t,u,\omega) - A(t,v,\omega),u-v \rangle \geq \delta^2 \|u-v\|^p_V$$

$$\langle A(t,u,\omega),u \rangle \geq \delta^2 \|u\|^p_V$$

Also assume the uniqueness condition of Lemma 28.8.10 is satisfied. Consider the following graph.

There is a monotone Lipschitz function $J_n$ which is approximating a function with the indicated jump. For a convex function $\phi$, we denote by $\partial \phi$ its subgradient. Thus for $y \in \partial \phi(x)$

$$(y,u) \leq \phi(x+u) - \phi(x).$$

Denote the Lipschitz function as $J_n$ and the maximal monotone graph which it is approximating as $J$. Thus $J$ denotes the ordered pairs $(x,y)$ which are of the form $(0,y)$ for $|y| \leq 1$ along with ordered pairs $(x,1), x > 0$ and ordered pairs $(x,-1)$ for $x < 0$. The graph of $J$ is illustrated in the above picture and is a maximal monotone graph. Thus $J = \partial \phi$ where $\phi(r) = |r|$. As illustrated in the graph, $J_n$ is piecewise linear.

Let $\phi_n(r) = \int_0^r J_n(s) \, ds$. It follows easily that $\phi_n(r) \to \phi(r)$ uniformly on $\mathbb{R}$. Also let $h \geq 0$ be progressively measurable and uniformly bounded by $M$ and let $u_0 \in L^2(\Omega,W), u_0 \mathcal{F}_0$ measurable. Then from the above theorems, there exists a unique solution to the integral equation

$$Bu_n(t) - Bu_0 + \int_0^t A(s,u_n,s) \, ds + \int_0^t h(s) J_n(u_n) \, ds = B \int_0^t \Psi dW,$$

the last term equaling $\int_0^t \Phi dW$. The integral equation holds off a set of measure zero and is progressively measurable.

Then from the Ito formula, one obtains, using the monotonicity of $J_n$ an estimate in which $C$ does not depend on $n$

$$\frac{1}{2} \mathbb{E} \langle Bu_n(t),u_n(t) \rangle - \frac{1}{2} \mathbb{E} \langle Bu_0,u_0 \rangle + E \int_0^t \langle Au_n,u_n \rangle \, ds \leq C$$

In particular, this holds for $n = 1$. Therefore, adjusting the constant, it follows that

$$\int_\Omega \langle Bu_1(t),u_1(t) \rangle + \int_\Omega \int_0^T \|u_1\|^p_V \, dt \, ds \leq C$$
Consequently, there exists a set of measure zero $N$ such that for $\omega \notin N$,

$$
\langle Bu_1(t), u_1(t) \rangle + \int_0^t \|u_t\|^p_V \, dt \leq C(\omega)
$$

(28.7.64)

From the integral equation, it follows that, enlarging $N$ by including countably many sets of measure zero, for $\omega \notin N$

$$
Bu_n(t) - Bu_1(t) + \int_0^t A(s, u_n, \omega) - A(s, u_1, \omega) \, ds + \int_0^t h(s, \omega) J_n(u_n) - h(s, \omega) J_1(u_1) \, ds = 0
$$

Now it is certainly true that $|J_n(u_n) - J_1(u_n)| \leq 2$. Thus

$$
\int_0^t \langle h(s, \omega) J_n(u_n) - h(s, \omega) J_1(u_1), u_n - u_1 \rangle \, ds
$$

$$
= \int_0^t \langle h(s, \omega) J_n(u_n) - h(s, \omega) J_1(u_1), u_n - u_1 \rangle \, ds
$$

$$
+ \int_0^t \langle h(s, \omega) (J_n(u_n) - J_1(u_1)), u_n - u_1 \rangle \, ds
$$

$$
\geq -2M \int_0^t |u_n - u_1| \, ds
$$

Therefore, from the Ito formula and for $\omega \notin N$,

$$
\frac{1}{2} \langle Bu_n(t) - Bu_1(t), u_n(t) - u_1(t) \rangle + \delta^2 \int_0^t \|u_n - u_1\|^p_V \, ds
$$

$$
\leq \int_0^t 2M |u_n - u_1| \, ds \leq \left( 2 + \frac{1}{2} \int_0^t |u_n - u_1|^2 \, ds \right) M
$$

$$
\leq \left( 2 + \frac{1}{2} \int_0^t (Bu_n - Bu_1, u_n - u_1) \, ds \right) M
$$

where $M$ is an upper bound to $h$. Then by Gronwall’s inequality

$$
\frac{1}{2} \langle Bu_n(t) - Bu_1(t), u_n(t) - u_1(t) \rangle \leq 2Me^{MT}
$$

Hence

$$
\frac{1}{2} \langle Bu_n(t) - Bu_1(t), u_n(t) - u_1(t) \rangle + \delta^2 \int_0^t \|u_n - u_1\|^p_V \, ds \leq 2M + TM2e^{TM}
$$

It follows from (28.7.65) that for all $\omega \notin N$ and adjusting the constant,

$$
\langle Bu_n(t), u_n(t) \rangle + \int_0^T \|u_n\|^p_V \, ds \leq C(\omega)
$$

(28.7.65)

for all $n$, where $C(\omega)$ depends only on $\omega$.

For $\omega \notin N$, the above estimate implies there exists a further subsequence, still called $n$ such that

$$
Bu_n \rightarrow Bu \text{ weak } \ast \text{ in } L^\infty(0, T, W')
$$

$$
u_n \rightarrow u \text{ weak } \ast \text{ in } L^\infty(0, T, H)
$$

$$
u_n \rightarrow u \text{ weakly in } V_\omega
$$

From the integral equation solved and the assumption that $A$ is bounded, it can also be assumed that

$$
\left( B \left( u_n - \int_0^t \Psi dW \right) \right)' \rightarrow \left( B \left( u - \int_0^t \Psi dW \right) \right)'
$$

weakly in $V'_\omega$

$$
Bu(0) = Bu_0
$$

$Au_n \rightarrow \xi$ weakly in $V'_\omega$
It is known that \( u_n \) is bounded in \( \mathcal{V}_\omega \). Also it is known that \( \left( B \left( u_n - \int_0^t \Psi dW \right) \right)' \) is bounded in \( \mathcal{V}_\omega' \). Therefore, \( B \left( u_n - \int_0^t \Psi dW \right) \) satisfies a Holder condition into \( V' \). Since \( \Psi \) is in \( L^\infty \), \( \int_0^t \Psi dW \) satisfies a Holder condition, and so \( Bu_n \) satisfies a Holder condition into \( V' \) while \( Bu_n \) is bounded in \( \mathcal{V}_\omega' \). By compactness of the embedding of \( V \) into \( W \), it follows that \( W' \) embeds compactly into \( V' \). This is sufficient to conclude that \( \{ Bu_n \} \) is precompact in \( \mathcal{W}_\omega' \). The proof is similar to one given by Lions. See Theorem 24.6.2. Since \( B \) is the Riesz map, this implies that \( \{ u_n \} \) is precompact in \( \mathcal{W}_\omega \) and hence in \( \mathcal{H}_\omega \).

Therefore, one can take a further subsequence and conclude that

\[
 u_n \to u \text{ strongly in } \mathcal{H}_\omega \equiv L^2 ([0,T], L^2(U))
\]

Therefore, a further subsequence, still denoted by \( n \) satisfies

\[
 u_n(t) \to u(t) \quad \text{in } L^2(U) \quad \text{for a.e. } t
\]

We can also assume that

\[
 J_n(u_n) \to \zeta \quad \text{weak * in } L^\infty (0,T,L^\infty(U))
\]

From the integral equation solved,

\[
 \left\langle \left( B \left( u_n - \int_0^t \Psi dW \right) \right)', u_n - u \right\rangle_{\mathcal{V}_\omega} + \langle A(t,u_n) + h(t,\omega) J_n(u_n), u_n - u \rangle = 0 \quad (28.7.67)
\]

We claim that

\[
 \int_0^t \left\langle \left( B \left( u_n - \int_0^s \Psi dW \right) \right) - \left( B \left( u - \int_0^s \Psi dW \right) \right)' \right\rangle, u_n - u \rangle ds \geq 0
\]

The difficulty is that \( \int_0^t \Psi dW \) is only in \( W \). To see that the conclusion is so, note that it is clear from a computation that

\[
 \int_0^t \left( 1 - \frac{\tau(h)}{h} \right) \left( Bu_n - B \int_0^s \Psi dW \right) - \frac{1 - \tau(h)}{h} \left( B u - B \int_0^s \Psi dW \right), u_n - u \right) ds \geq 0 \quad (28.7.68)
\]

**Claim:** The above is indeed nonnegative.

**Proof:** Denote by \( q(t) \) the stochastic integral, \( u_n \) as \( u \) and \( u \) as \( v \) to save notation. Then the left side of the above equals

\[
 \frac{1}{h} \int_0^t \langle B(u - q), u - v \rangle ds
\]

\[
 - \frac{1}{h} \int_0^t \langle B(u(s - h) - q(s - h)) - B(v(s - h) - q(s - h)), u(s) - v(s) \rangle ds
\]

\[
 \geq \frac{1}{h} \int_0^t \langle B(u - q) - B(v - q), u - v \rangle ds
\]

\[
 - \frac{1}{2h} \int_0^t \langle B(u(s - h) - q(s - h)) - B(v(s - h) - q(s - h)), (u(s) - h - v(s)) \rangle ds
\]

\[
 - \frac{1}{2h} \int_0^t \langle B(u - q) - B(v - q), u - v \rangle ds
\]

\[
 \geq \frac{1}{h} \int_0^t \langle B(u - q) - B(v - q), u - v \rangle ds
\]

\[
 - \frac{1}{2h} \int_0^t \langle B(u(s) - q(s)) - B(v(s) - q(s)), (u(s) - v(s)) \rangle ds
\]

\[
 - \frac{1}{2h} \int_0^t \langle B(u - q) - B(v - q), u - v \rangle ds
\]
\[
\frac{1}{h} \int_{t-h}^{t} \langle B(u - q) - B(v - q), u - v \rangle \, ds + \frac{1}{h} \int_{0}^{t-h} \langle B(u - q) - B(v - q), u - v \rangle \, ds
\]

\[
- \frac{1}{2h} \int_{0}^{t-h} \langle B(u - q) - B(v - q), (u - v) \rangle \, ds
\]

\[
- \frac{1}{2h} \int_{h}^{t} \langle B(u - q) - B(v - q), u - v \rangle \, ds
\]

\[
= \frac{1}{h} \int_{t-h}^{t} \langle B(u - q) - B(v - q), u - v \rangle \, ds + \frac{1}{2h} \int_{0}^{t-h} \langle B(u - q) - B(v - q), (u - v) \rangle \, ds
\]

\[
- \frac{1}{2h} \int_{0}^{t-h} \langle B(u - q) - B(v - q), (u - v) \rangle \, ds - \frac{1}{2h} \int_{h}^{t} \langle B(u - q) - B(v - q), u - v \rangle \, ds
\]

\[
= \frac{1}{2h} \int_{h}^{t} \langle B(u - q) - B(v - q), u - v \rangle \, ds + \frac{1}{2h} \int_{0}^{h} \langle B(u - q) - B(v - q), (u - v) \rangle \, ds
\]

which is nonnegative as can be seen by replacing \( u - v \) with \( (u - q) - (v - q) \) and using monotonicity of \( B \).

Now pass to a limit in (28.7.68) as \( h \to 0 \) to get the desired inequality. Therefore, from (28.7.67),

\[
\lim \sup_{n \to \infty} \int_{0}^{T} \langle A(t, u_n) + h(t, \omega) J_n(u_n), u_n - u \rangle \, dt \leq 0
\]

From the above strong convergence, the left side equals

\[
\lim \sup_{n \to \infty} \int_{0}^{T} \langle A(t, u_n), u_n - u \rangle \, dt \leq 0
\]

It follows that for all \( v \in \mathcal{V}_\omega \),

\[
\int_{0}^{T} \langle A(t, u), u - v \rangle \, dt \leq \lim \inf_{n \to \infty} \int_{0}^{T} \langle A(t, u_n), u_n - v \rangle \, dt
\]

\[
= \lim \sup_{n \to \infty} \left[ \int_{0}^{T} \langle A(t, u_n), u_n - u \rangle \, dt + \int_{0}^{T} \langle A(t, u_n), u - v \rangle \, dt \right]
\]

\[
\leq \int_{0}^{T} \langle \xi, u - v \rangle \, dt
\]

Since \( v \) is arbitrary, \( A(\cdot, u) = \xi \in \mathcal{V}_\omega' \). Passing to the limit in the integral equation yields

\[
Bu(t) - Bu_0 + \int_{0}^{t} A(s, u) \, ds + \int_{0}^{t} h(s, \omega) \zeta(s, \omega) \, ds = \int_{0}^{t} \Phi dW
\]

What is \( h(s, \omega) \zeta(s, \omega) \)?

\[
\int_{0}^{T} \langle h(s, \omega) J_n(u_n(s)), v(s) - u_n(s) \rangle \, ds \leq \int_{0}^{T} h(s, \omega) (\phi_n(v) - \phi_n(u_n)) \, ds
\]

Passing to the limit and using the strong convergence described above along with the uniform convergence of \( \phi_n \) to \( \phi \),

\[
\int_{0}^{T} (h(s, \omega) \zeta(s), v(s) - u(s))_H \, ds \leq \int_{0}^{T} h(s, \omega) (\phi(v(s)) - \phi(u(s))) \, ds
\]

Hence,

\[
\int_{0}^{T} (h(s, \omega) \phi(v(s)) - h(s, \omega) \phi(u(s))) - (h(s, \omega) \zeta(s), v(s) - u(s))_H \, ds \geq 0
\]

for any choice of \( v \in \mathcal{H}_\omega \). It follows that for a.e. \( s, h(s, \omega) \zeta(s) \in \partial_n (h(s, \omega) \phi(u(s))) \).

This has shown that for each \( \omega \notin N \), there exists a solution to the integral equation

\[
Bu(t) - Bu_0 + \int_{0}^{t} A(s, u) \, ds + \int_{0}^{t} h(s, \omega) \zeta(s, \omega) \, ds = \int_{0}^{t} \Phi dW
\]  

(28.7.69)
Then from the above example, there exists a solution to the integral equation of a set of measure zero in terms of inclusions, there exists a set of measure zero such that off this set, where for a.e. $\lambda$ for all $t$

\[Bu_1(t) - Bu_2(t) + \int_0^t A(s, u_1) - A(s, u_2) \, ds + \int_0^t h(s, \omega) (\zeta_1(s, \omega) - \zeta_2(s, \omega)) \, ds = 0\]

Then from monotonicity of the subgradient it follows that $u_1 = u_2$. Then the two integral equations yield that for a.e. $t$

\[\left(B \left( u_1 - \int_0^t \Psi dW \right) \right)'(t) + A(s, u_1(t)) + h(t, \omega) \zeta_1(t, \omega) = 0\]

\[\left(B \left( u_2 - \int_0^t \Psi dW \right) \right)'(t) + A(s, u_2(t)) + h(t, \omega) \zeta_2(t, \omega) = 0\]

Therefore, for a.e. $t$, $h(t, \omega) \zeta_1(t, \omega) - h(t, \omega) \zeta_2(t, \omega)$. Thus the solution to the integral equation for each $\omega$ off a set of measure zero is unique.

At this point it is not clear that $(t, \omega) \rightarrow u(t, \omega)$ is progressively measurable. We claim that for $\omega \notin N$ it is not necessary to take a subsequence in the above. This is because the above argument shows that if $u_n$ fails to converge weakly, then there would exist two subsequences converging weakly to two different solutions to the integral equation which would contradict uniqueness.

Therefore, for $\omega \notin N$, $u_n(t, \omega) \rightarrow u(t, \omega)$ weakly in $V_\omega^\prime$ for a single sequence. Using the estimate \[\|[u_n - \bar{u}] \in L^p([0, T] \times \Omega; V)\]

where the measurable sets are just the product measurable sets $B([0, T]) \times \mathcal{F}_T$. By Lemma \[\|[u_n - \bar{u}] \in L^p([0, T] \times \Omega; V)\]

for all $u_n \in L^2(\Omega, W)$, $u_0 \in \mathcal{F}_0$ measurable. Suppose $\lambda I + A(t, u, \omega)$ satisfies

\[(\lambda I + A(t, u, \omega)) - (\lambda I + A(t, v, \omega)) + u - v \geq \delta^2 \|u - v\|^p_V\]

for all $\lambda$ large enough. Also assume $\Phi \in L^\infty([0, T] \times \Omega, L_2(Q^{1/2}U, H))$ with $\Phi = B\Psi$ where $\Psi \in L^\infty([0, T] \times \Omega, L_2(Q^{1/2}U, V))$

and progressively measurable. Then there exists a unique solution to the integral equation

\[Bu(t) - Bu_0 + \int_0^t A(s, u) \, ds + \int_0^t h(s, \omega) \zeta(s, \omega) \, ds = \int_0^t \Phi dW\] (28.7.70)

where for a.e. $s$, $h(s, \omega) \zeta(s, \omega) \in \partial_u (h(s, \omega) \phi(u(s)))$ where $\phi(r) \equiv |r|$. The symbol $\partial_u$ is the subgradient of $\phi(u)$.

Written in terms of inclusions, there exists a set of measure zero such that off this set,

\[\left(B \left( u - \int_0^t \Phi dW \right) \right)' + A(t, u) \in \partial_u (h(t, \omega) \phi(u(s))) \text{ a.e. } t\]

\[u(0) = u_0\]

Note that one can replace $\Phi \in L^\infty([0, T] \times \Omega, L_2(Q^{1/2}U, H))$ with $\Phi \in L^2([0, T] \times \Omega, L_2(Q^{1/2}U, H))$ along with an assumption that $t \rightarrow \Phi(t, \omega)$ is continuous. This can be done by defining a stopping time

\[\tau_n \equiv \inf \{ t : \| \Phi(t) \| > n \}\]

Then from the above example, there exists a solution to the integral equation off a set of measure zero

\[Bu_n(t) - Bu_0 + \int_0^t A(s, u_n) \, ds + \int_0^t h(s, \omega) \zeta_n(s, \omega) \, ds = \int_0^{t \wedge \tau_n} \Phi dW\]
Since $\Phi$ is a continuous process, $\tau_n = \infty$ for all $n$ large enough. Hence, one can replace the above with the desired integral equation. Of course the size of $n$ depends on $\omega$, but we can define

$$u(t, \omega) = \lim_{n \to \infty} u_n(t, \omega)$$

because by uniqueness which comes from monotonicity, if for a particular $\omega$, both $n, k$ are sufficiently large, then $u_n = u_k$. Thus $u$ is progressively measurable and is the desired solution.

Next we show that the above theory can also be used as a starting point for some second order in time problems. Consider a beam which has a point mass of mass $m$ attached to one end. Suppose for sake of illustration that the left end is clamped, $u(0, t) = u_x(0, t) = 0$, while the right end which has the attached mass is free to move, $u_x(1, t) = 0$, and the beam occupies the interval $[0, 1]$ in material coordinates. Then the stress is $\sigma = -u_{xxx}$ and balance of momentum is

$$u_{tt} = \sigma_x + f$$

where $f$ is a body force. Thus, letting $w \in V \equiv \{w \in H^2(0, 1) : w(0) = w_x(0) = 0, w_x(1) = 0\}$ be a test function,

$$\int_0^1 u_{tt}w dx = \sigma w|_0^1 + \int_0^1 (-\sigma) w_x dx + \int_0^1 f w dx = -mu_{tt} (1, t) w(1, t) + \int_0^1 u_{xxx} w_x dx + \int_0^1 f w dx$$

Doing another integration by parts and using the boundary conditions, it follows that an appropriate variational formulation for this problem is

$$\int_0^1 u_{tt}w dx + m\gamma_1 u_{tt}\gamma_1 w + \int_0^1 u_{xx}w_{xx} dx = \int_0^1 f w dx$$

where here $\gamma_1$ is the trace map on the right end.

Letting

$$u(t) = u_0 + \int_0^t v(s) ds,$$

where $u(0, t) = u_0$, we can write the above variational equation in the form

$$(Bv)' + Au = f, \quad Bv(0) = Bv_0$$

where we assume that $v_0 \in W$ where $W$ is the closure of $V$ in $H^1(0, 1)$ and the operators are given by

$$B : V' \to W', \quad \langle Bu, w \rangle \equiv \int_0^1 uw dx + m\gamma_1 u\gamma_1 w$$

$$A : V \to V', \quad \langle Au, w \rangle \equiv \int_0^1 u_{xx}w_{xx} dx$$

Thus in terms of an integral equation, this would be of the form

$$Bv(t) - Bv_0 + \int_0^t A(u) ds = \int_0^t f ds$$

This suggests a stochastic version of the form

$$Bv(t) - Bv_0 + \int_0^t A(u) ds = \int_0^t f ds + \int_0^t \Phi dW$$

where $\Phi \in L^\infty((0, T) \times \Omega, L_2(Q^{1/2}U, H))$ for $H = L^2(0, 1)$. As in the previous example, simple considerations involving maximal monotone operators imply that there exists $\Psi \in L^\infty((0, T) \times \Omega, L_2(Q^{1/2}U, W))$ such that $\Phi = B\Psi$. We also assume that $f \in V'$ and $u_0, v_0$ are in $L^2(\Omega, V)$ and $L^2(\Omega, W)$ respectively, both being $\mathcal{F}_0$ measurable. The above equation does not fit the general theory developed earlier because it is second order in time and is a stochastic version of a hyperbolic equation rather than a parabolic one. We consider it using a parabolic regularization which can be studied with the above general theory along with a simple fixed point argument.
Consider the approximate problem which is to find a solution to
\[ Bv(t) - Bv_0 + \varepsilon \int_0^t Av ds + \int_0^t A(u) \, ds = \int_0^t f ds + \int_0^t \Phi dW \] (28.7.71)
where \( u \) is given above as an integral of \( v \). First we argue that there exists a unique solution to the above integral equation and then we pass to a limit as \( \varepsilon \to 0 \). Let \( u \in \mathcal{V} \) be given.

From Corollary (28.4.9) there exists a unique solution \( v \) to (28.4.9). Now suppose \( u_1, u_2 \) are two given in \( \mathcal{V} \) and denote by \( v_i \) the corresponding \( v \) which solves the above. Then from the Ito formula or standard considerations,
\[
\frac{1}{2} E \langle B (v_1 (t) - v_2 (t)), v_1 (t) - v_2 (t) \rangle + \varepsilon E \int_0^t \| v_1 - v_2 \|_V^2 \, ds
\]
\[
\leq \frac{\varepsilon}{2} E \int_0^t \| v_1 - v_2 \|_V^2 \, ds + C_\varepsilon E \int_0^t \| u_1 - u_2 \|_V^2 \, ds
\]
Now define a mapping \( \theta \) from \( \mathcal{V} \) to \( \mathcal{V} \) as follows. Begin with \( v \) then obtain
\[ u (t) = u_0 + \int_0^t v (s) \, ds \] (28.7.72)
Use this \( u \) in (28.7.71). Then \( \theta v \) is the solution to (28.7.71) which corresponds to \( u \). Then the above inequality shows that
\[
\int_0^t \int_\Omega \| \theta v_1 (s) - \theta v_2 (s) \|^2 \, dP ds \leq \frac{C_\varepsilon}{\varepsilon} \int_0^t \int_\Omega \| u_1 - u_2 \|_V^2 \, dP ds
\]
\[
\leq \frac{C_\varepsilon}{\varepsilon} C_T \int_0^t \int_\Omega \| v_1 (r) - v_2 (r) \|^2 \, dP dr ds
\]
It follows that a high enough power of \( \varepsilon \) is a contraction map on \( L^2 (0, T, L^2 (\Omega, V)) \) and so there exists a unique fixed point. This yields a unique solution to the above approximate problem (28.7.71) in which \( u, v \) are related by (28.7.72).

Next we let \( \varepsilon \to 0 \). Index the above solution with \( \varepsilon \). By the Ito formula again,
\[
\frac{1}{2} E \langle B v_\varepsilon (t), v_\varepsilon (t) \rangle - \frac{1}{2} E \langle B v_0, v_0 \rangle + \varepsilon \int_0^t E \| v_\varepsilon \|_V^2 \, ds
\]
\[
+ \frac{1}{2} E \| u_\varepsilon (t) \|_V^2 - \frac{1}{2} E \| u_0 \|_V^2 = \int_0^t E (f, v_\varepsilon) \, ds
\]
Then one can obtain an estimate and pass to a limit as \( \varepsilon \to 0 \) obtaining the following convergences.
\[ \varepsilon v_\varepsilon \to 0 \text{ strongly in } \mathcal{V} \]
\[ B v_\varepsilon \to B v \text{ weak * in } L^\infty (0, T, L^2 (\Omega, W')) \]
\[ u_\varepsilon (t) \to u (t) \text{ weak * in } L^\infty (0, T, L^2 (\Omega, V)) \]
Then one can simply pass to a limit in the approximate integral equation and obtain, thanks to linearity considerations, that
\[ Bv(t) - Bv_0 + \int_0^t A(u) \, ds = \int_0^t f \, ds + \int_0^t \Phi dW, \quad u(t) = u_0 + \int_0^t v(s) \, ds \] (28.7.73)
the equation holding in \( \mathcal{V}' \). Thus for a.e. \( \omega \), the above holds for a.e. \( t \). It is possible to work harder and have the equation holding for all \( t \). This involves using the other form of the Ito formula, estimating for a fixed \( \omega \) as done above and then arguing that by uniqueness one can use a single subsequence which works independent of \( \omega \).

**Example 28.7.3** Let \( u_0 \in L^2 (\Omega, V) \) where \( V \) is described above and let \( v_0 \in L^2 (\Omega, W) \) for \( W \) described above. Let both of these initial conditions be progressively measurable. Also let \( f \in \mathcal{V}' \) and \( \Phi \in L^\infty ((0, T) \times \Omega, L_2 (Q^{1/2} U, H)) \). Then there exists a unique solution to the the integral equation (28.7.73) which can be written in the form
\[ Bu_t(t) - Bu_0 + \int_0^t A \left( u_0 + \int_0^t u_t (r) \, dr \right) \, ds = \int_0^t f \, ds + \int_0^t \Phi dW \]

Note that a more standard model involves no point mass at the tip of the beam. This would be done in the same way but it would not require the generalized Ito formula presented earlier. A more standard version would work.

One can find many other examples where this generalized Ito formula is a useful tool to study various kinds of stochastic partial differential equations. We have presented five examples above in which it was helpful to have the extra generality.
Chapter 29

The Yankov von Neumann Aumann theorem

The Yankov von Neumann Aumann theorem deals with the projection of a product measurable set. It is a very difficult but interesting theorem. The material of this chapter is taken from 23, 24, 8, and 48. We use the standard notation that for $S$ and $F$ $\sigma$ algebras, $S \times F$ is the $\sigma$ algebra generated by the measurable rectangles, the product measure $\sigma$ algebra. The next result is fairly easy and the proof is left for the reader.

Lemma 29.0.4 Let $(X,d)$ be a metric space. Then if $d_1(x,y) = \frac{d(x,y)}{1 + d(x,y)}$, it follows that $d_1$ is a metric on $X$ and the basis of open balls taken with respect to $d_1$ yields the same topology as the basis of open balls taken with respect to $d$.

Theorem 29.0.5 Let $(X_i,d_i)$ denote a complete metric space and let $X \equiv \prod_{i=1}^\infty X_i$. Then $X$ is also a complete metric space with the metric

$$
\rho(x,y) \equiv \sum_{i=1}^\infty 2^{-i} \frac{d_i(x_i,y_i)}{1 + d_i(x_i,y_i)}.
$$

Also, if $X_i$ is separable for each $i$ then so is $X$.

Proof: It is clear from the above lemma that $\rho$ is a metric on $X$. We need to verify $X$ is complete with this metric. Let $\{x^n\}$ be a Cauchy sequence in $X$. Then it is clear from the definition that $\{x^n_i\}$ is a Cauchy sequence for each $i$ and converges to $x_i \in X_i$. Therefore, letting $\epsilon > 0$ be given, we choose $N$ such that

$$
\sum_{k=N}^\infty 2^{-k} < \frac{\epsilon}{2},
$$

we choose $M$ large enough that for $n > M$,

$$
2^{-i} \frac{d_i(x^n_i,x_i)}{1 + d_i(x^n_i,x_i)} < \frac{\epsilon}{2(N+1)}
$$

for all $i = 1, 2, \ldots, N$. Then letting $x = \{x_i\}$,

$$
\rho(x,x^n) \leq \frac{\epsilon N}{2(N+1)} + \sum_{k=N}^\infty 2^{-k} \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
$$

We need to verify that $X$ is separable. Let $D_i$ denote a countable dense set in $X_i, D_i \equiv \{r_k\}_{k=1}^\infty$. Then let

$$
D_k \equiv D_1 \times \cdots \times D_k \times \{r_{k+1}^1\} \times \{r_{k+2}^1\} \times \cdots
$$

Thus $D_k$ is a countable subset of $X$. Let $D \equiv \bigcup_{k=1}^\infty D_k$. Then $D$ is countable and we can see $D$ is dense in $X$ as follows. The projection of $D_k$ onto the first $k$ entries is dense in $\prod_{i=1}^k X_i$ and for $k$ large enough the remaining component’s contribution to the metric, $\rho$ is very small. Therefore, obtaining $d \in D$ close to $x \in X$ may be accomplished by finding $d \in D$ such that $d$ is close to $x$ in the first $k$ components for $k$ large enough. Note that we do not use $\prod_{k=1}^\infty D_k$!
Definition 29.0.6 A complete separable metric space is called a polish space.

Theorem 29.0.7 Let X be a polish space. Then there exists \( f : \mathbb{N}^\infty \to X \) which is onto and continuous. Here \( \mathbb{N}^\infty = \prod_{i=1}^{\infty} \mathbb{N} \) and a metric is given according to the above theorem. Thus for \( n, m \in \mathbb{N}^\infty \),

\[
\rho(n, m) = \sum_{i=1}^{\infty} 2^{-i} \frac{|n_i - m_i|}{1 + |n_i - m_i|}
\]

Proof: Since X is polish, there exists a countable covering of X by closed sets having diameters no larger than \( 2^{-i} \), \( \{ B(i) \}_{i=1}^{\infty} \). Each of these closed sets is also a polish space and so there exists a countable covering of \( B(i) \) by a countable collection of closed sets, \( \{ B(i, j) \}_{j=1}^{\infty} \) each having diameter no larger than \( 2^{-2j} \) where \( B(i, j) \subseteq B(i) \neq \emptyset \) for all \( j \). Continue this way. Thus

\[
B(n_1, n_2, \ldots, n_m) = \bigcup_{i=1}^{\infty} B(n_1, n_2, \ldots, n_m, i)
\]

and each of \( B(n_1, n_2, \ldots, n_m, i) \) is a closed set contained in \( B(n_1, n_2, \ldots, n_m) \) whose diameter is at most half of the diameter of \( B(n_1, n_2, \ldots, n_m) \). Now we define our mapping from \( \mathbb{N}^\infty \) to X. If \( n = \{ n_k \}_{k=1}^{\infty} \in \mathbb{N}^\infty \), we let \( f(n) = \bigcap_{m=1}^{\infty} B(n_1, n_2, \ldots, n_m) \). Since the diameters of these sets converge to 0, there exists a unique point in this countable intersection and this is \( f(n) \).

We need to verify \( f \) is continuous. Let \( n \in \mathbb{N}^\infty \) be given and suppose \( m \) is very close to \( n \). The only way this can occur is for \( n_k \) to coincide with \( m_k \) for many \( k \). Therefore, both \( f(n) \) and \( f(m) \) must be contained in \( B(n_1, n_2, \ldots, n_m) \) for some fairly large \( m \). This implies, from the above construction that \( f(m) \) is close to \( f(n) \) as \( 2^{-m} \), proving \( f \) is continuous. To see that \( f \) is onto, note that from the construction, if \( x \in X \), then \( x \in B(n_1, n_2, \ldots, n_m) \) for some choice of \( n_1, \ldots, n_m \) for each \( m \). Note nothing is said about \( f \) being one to one. It probably is not one to one.

Definition 29.0.8 We call a topological space, \( X \) a Suslin space if \( X \) is a Hausdorff space and there exists a polish space, \( Z \) and a continuous function \( f \) which maps \( Z \) onto \( X \). These are also called analytic sets in some contexts but we will use the term Suslin space in referring to them.

Corollary 29.0.9 \( X \) is a Suslin space, if and only if there exists a continuous mapping from \( \mathbb{N}^\infty \) onto \( X \).

Proof: We know there exists a polish space and a continuous function, \( h : Z \to X \) which is onto. By the above theorem there exists a continuous map, \( g : \mathbb{N}^\infty \to Z \) which is onto. Then \( h \circ g \) is a continuous map from \( \mathbb{N}^\infty \) onto \( X \). The “if” part of this theorem is accomplished by noting that \( \mathbb{N}^\infty \) is a polish space.

Lemma 29.0.10 Let \( X \) be a Suslin space and suppose \( X_i \) is a subspace of \( X \) which is also a Suslin space. Then \( \bigcup_{i=1}^{\infty} X_i \) and \( \bigcap_{i=1}^{\infty} X_i \) are also Suslin spaces. Also every Borel set in \( X \) is a Suslin space.

Proof: Let \( f_i : Z_i \to X_i \) where \( Z_i \) is a polish space and \( f_i \) is continuous and onto. Without loss of generality we may assume the spaces \( Z_i \) are disjoint because if not, we could replace \( Z_i \) with \( Z_i \times \{ i \} \). Now we define a metric, \( \rho \), for \( Z = \bigcup_{i=1}^{\infty} Z_i \) as follows.

\[
\rho(x, y) \equiv 1 \text{ if } x \in Z_i, y \in Z_k, i \neq k
\]

\[
\rho(x, y) = \frac{d_i(x, y)}{1 + d_i(x, y)} \text{ if } x, y \in Z_i.
\]

Here \( d_i \) is the metric on \( Z_i \). It is easy to verify \( \rho \) is a metric and that \((Z, \rho)\) is a polish space. Now we define \( f : Z \to \bigcup_{i=1}^{\infty} X_i \) as follows. For \( x \in Z_i \), \( f(x) = f_i(x) \). This is well defined because the \( Z_i \) are disjoint. If \( y \) is very close to \( x \) it must be that \( x \) and \( y \) are in the same \( Z_i \); otherwise this could not happen. Therefore, continuity of \( f \) follows from continuity of \( f_i \). This shows countable unions of Suslin subspaces of a Suslin space are Suslin spaces.

If \( H \subseteq X \) is a closed subset, then, letting \( f : Z \to X \) be onto and continuous, it follows \( f : f^{-1}(H) \to H \) is onto and continuous. Since \( f^{-1}(H) \) is closed, it follows \( f^{-1}(H) \) is a polish space. Therefore, \( H \) is a Suslin space.

Now we show countable interections of Suslin spaces are Suslin. It is clear that \( \theta : \prod_{i=1}^{\infty} Z_i \to \prod_{i=1}^{\infty} X_i \) given by \( \theta(z) \equiv x = \{ x_i \} \) where \( x_i = f_i(z_i) \) is continuous and onto. Therefore, \( \prod_{i=1}^{\infty} X_i \) is a Suslin space. Now let \( P \equiv \{ y \in \prod_{i=1}^{\infty} f_i(Z_i) : y_i = y_j \text{ for all } i, j \} \). Then \( P \) is a closed subspace of a Suslin space and so it is Suslin. Then we define \( h : P \to \bigcap_{i=1}^{\infty} X_i \) by \( h(y) \equiv f_i(y_i) \). This shows \( \bigcap_{i=1}^{\infty} X_i \) is Suslin because \( h \) is continuous and onto. \((h \circ \theta : \theta^{-1}(P) \to \bigcap_{i=1}^{\infty} X_i \) is continuous and \( \theta^{-1}(P) \) being a closed subset of a polish space is polish.)
Next let $U$ be an open subset of $X$. Then $f^{-1}(U)$, being an open subset of a polish space, can be obtained as an increasing limit of closed sets, $K_n$. Therefore, $U = \bigcup_{n=1}^{\infty} f(K_n)$. Each $f(K_n)$ is a Suslin space because it is the continuous image of a polish space, $K_n$. Therefore, by the first part of the lemma, $U$ is a Suslin space. Now let

$$\mathcal{F} \equiv \{ E \subseteq X : \text{ both } E^C \text{ and } E \text{ are Suslin} \}.$$  

We see that $\mathcal{F}$ is closed with respect to taking complements. The first part of this lemma shows $\mathcal{F}$ is closed with respect to countable unions. Therefore, $\mathcal{F}$ is a $\sigma$ algebra and so, since it contains the open sets, must contain the Borel sets.

It turns out that Suslin spaces tend to be measurable sets. In order to develop this idea, we need a technical lemma.

**Lemma 29.0.11** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and denote by $\mu^*$ the outer measure generated by $\mu$. Thus

$$\mu^*(S) \equiv \inf \{ \mu(E) : E \supseteq S, E \in \mathcal{F} \}.$$  

Then $\mu^*$ is regular, meaning that for every $S$, there exists $E \in \mathcal{F}$ such that $E \supseteq S$ and $\mu(E) = \mu^*(S)$. If $S_n \uparrow S$, it follows that $\mu^*(S_n) \uparrow \mu^*(S)$. Also if $\mu(\Omega) < \infty$, then a set, $E$ is measurable if and only if

$$\mu^*(\Omega) \geq \mu^*(E) + \mu^*(\Omega \setminus E).$$

**Proof:** First we verify that $\mu^*$ is regular. If $\mu^*(S) = \infty$, let $E = \Omega$. Then $\mu^*(S) = \mu(E)$ and $E \supseteq S$. On the other hand, if $\mu^*(S) < \infty$, then we can obtain $E_n \in \mathcal{F}$ such that $\mu^*(S) + \frac{1}{n} \geq \mu(E_n)$ and $E_n \supseteq S$. Now let $F_n = \bigcap_{i=1}^{n} \bigcap_{i=1}^{n} E_i$. Then $F_n \supseteq S$ and so $\mu^*(S) + \frac{1}{n} \geq \mu(F_n) \geq \mu^*(S)$. Therefore, letting $F = \bigcap_{n=1}^{\infty} F_n \in \mathcal{F}$ it follows $\mu(F) = \lim_{n \to \infty} \mu(F_n) = \mu^*(S)$.

Now let $E_n \supseteq S_n$ be such that $E_n \in \mathcal{F}$ and $\mu(E_n) = \mu^*(S_n)$. Also let $E_\infty \supseteq S$ such that $\mu(E_\infty) = \mu^*(S)$ and $E_\infty \in \mathcal{F}$. Now consider $B_n \equiv \bigcup_{k=1}^{n} E_k$. We claim

$$\mu(B_n) = \mu(S_n).$$

Here is why:

$$\mu(E_1 \setminus E_2) = \mu(E_1) - \mu(E_1 \cap E_2) = \mu^*(S_1) - \mu^*(S_1) = 0.$$  

Therefore,

$$\mu(B_2) = \mu(E_1 \cup E_2) = \mu(E_1 \setminus E_2) + \mu(E_2) = \mu(S_2).$$

Continuing in this way we see that (29.0.1) holds. Now let $B_n \cap E_\infty \equiv C_n$. Then $C_n \uparrow C \equiv \bigcup_{n=1}^{\infty} C_n \in \mathcal{F}$ and $\mu(C_n) = \mu^*(S_n)$. Since $S_n \uparrow S$ and each $C_n \supseteq S_n$, it follows $C \supseteq S$ and therefore,

$$\mu^*(S) \leq \mu(C) = \lim_{n \to \infty} \mu(C_n) = \lim_{n \to \infty} \mu^*(S_n) \leq \mu^*(S).$$

Now we verify the second claim of the lemma. It is clear the formula holds whenever $E$ is measurable. Suppose now that the formula holds. Let $S$ be an arbitrary set. We need to verify that

$$\mu^*(S) \geq \mu^*(S \cap E) + \mu^*(S \setminus E).$$

Let $F \supseteq S, F \in \mathcal{F}$, and $\mu(F) = \mu^*(S)$. Then since $\mu^*$ is subadditive,

$$\mu^*(\Omega \setminus F) \leq \mu^*(E \setminus F) + \mu^*(E \cap F^C).$$

Since $F$ is measurable,

$$\mu^*(E) = \mu^*(E \cap F) + \mu^*(E \setminus F).$$

and

$$\mu^*(\Omega \setminus E) = \mu^*(F \setminus E) + \mu^*(\Omega \cap E^C \cap F^C)$$

and by the hypothesis,

$$\mu^*(\Omega) \geq \mu^*(E) + \mu^*(\Omega \setminus E).$$

Therefore,

$$\mu(\Omega) \geq \mu^*(E) + \mu^*(\Omega \setminus E) = \mu^*(E \cap F) + \mu^*(E \setminus F) + \mu^*(\Omega \setminus E) = \mu^*(E \cap F) + \mu^*(E \setminus F) + \mu^*(F \setminus E) + \mu^*(\Omega \cap E^C \cap F^C) \geq \mu^*(\Omega \setminus F) + \mu^*(F \setminus E) + \mu^*(E \cap F) \geq \mu^*(\Omega \setminus F) + \mu^*(F \setminus E) = \mu(\Omega).$$
showing that all the inequalities must be equal signs. Hence, referring to the top and fourth lines above,
\[ \mu(\Omega) = \mu^*(\Omega \setminus F) + \mu^*(F \setminus E) + \mu^*(E \cap F). \]
Subtracting \( \mu^*(\Omega \setminus F) = \mu(\Omega \setminus F) \) from both sides gives
\[ \mu^*(S) = \mu(F) = \mu^*(F \setminus E) + \mu^*(E \cap F) \geq \mu^*(S \setminus E) + \mu^*(E \cap S). \]
This proves the lemma.

The next theorem is a major result. It states that the Suslin subsets are measurable under appropriate conditions.

**Theorem 29.0.12** Let \( \Omega \) be a metric space and let \((\Omega, \mathcal{F}, \mu)\) be a complete Borel measure space with \( \mu(\Omega) < \infty \). Denote by \( \mu^* \) the outer measure generated by \( \mu \). Then if \( A \) is a Suslin subset of \( \Omega \), it follows that \( A \) is \( \mu^* \) measurable.

**Proof:** We need to verify that
\[ \mu^*(\Omega) \geq \mu^*(A) + \mu^*(\Omega \setminus A). \]
We know from Corollary 29.0.9, there exists a continuous map, \( f : \mathbb{N}^\mathbb{N} \to A \) which is onto. Let
\[ E(k) = \{ n \in \mathbb{N}^\mathbb{N} : n_1 \leq k \}. \]
Then \( E(k) \uparrow \mathbb{N}^\mathbb{N} \) and so from Lemma 29.0.11 we know \( \mu^*(f(E(k))) \uparrow \mu^*(A) \). Therefore, there exists \( m_1 \) such that
\[ \mu^*(f(E(m_1))) > \mu^*(A) - \frac{\varepsilon}{2}. \]
Now \( E(k) \) is clearly not compact but it is trying to be as far as the first component is concerned. Now we let
\[ E(m_1, k) = \{ n \in \mathbb{N}^\mathbb{N} : n_1 \leq m_1 \text{ and } n_2 \leq k \}. \]
Thus \( E(m_1, k) \uparrow E(m_1) \) and so we can pick \( m_2 \) such that
\[ \mu^*(f(E(m_1, m_2))) > \mu^*(f(E(m_1))) - \frac{\varepsilon}{2^2}. \]
We continue in this way obtaining a decreasing list of sets, \( f(E(m_1, m_2, \ldots, m_{k-1}, m_k)) \), such that
\[ \mu^*(f(E(m_1, m_2, \ldots, m_{k-1}, m_k))) > \mu^*(f(E(m_1, m_2, \ldots, m_{k-1}))) - \frac{\varepsilon}{2^k}. \]
Therefore,
\[ \mu^*(f(E(m_1, m_2, \ldots, m_{k-1}, m_k))) - \mu^*(A) > \sum_{i=1}^{k} \left( \frac{\varepsilon}{2^i} \right) > -\varepsilon. \]
Now define a closed set,
\[ C = \cap_{k=1}^{\infty} f(E(m_1, m_2, \ldots, m_{k-1}, m_k)). \]
The sets \( f(E(m_1, m_2, \ldots, m_{k-1}, m_k)) \) are decreasing as \( k \to \infty \) and so
\[ \mu^*(C) = \lim_{k \to \infty} \mu^*(f(E(m_1, m_2, \ldots, m_{k-1}, m_k))) \geq \mu^*(A) - \varepsilon. \]
We wish to verify that \( C \subseteq A \). If we can do this we will be done because \( C \), being a closed set, is measurable and so
\[ \mu^*(\Omega) = \mu^*(C) + \mu^*(\Omega \setminus C) \geq \mu^*(A) - \varepsilon + \mu^*(\Omega \setminus A). \]
Since \( \varepsilon \) is arbitrary, this will conclude the proof. Therefore, we only need to verify that \( C \subseteq A \).

What we know is that each \( f(E(m_1, m_2, \ldots, m_{k-1}, m_k)) \) is contained in \( A \). We do not know their closures are contained in \( A \). We let \( m \equiv \{ m_i \}_{i=1}^{\infty} \) where the \( m_i \) are defined above. Then letting
\[ K = \{ n \in \mathbb{N}^\mathbb{N} : n_i \leq m_i \text{ for all } i \}, \]
we see that \( K \) is a closed, hence complete subset of \( \mathbb{N}^\mathbb{N} \) which is also totally bounded due to the definition of the distance. Therefore, \( K \) is compact and so \( f(K) \) is also compact, hence closed due to the assumption that \( \Omega \) is a Hausdorff space and we know that \( f(K) \subseteq A \). We verify that \( C = f(K) \). We know \( f(K) \subseteq C \). Suppose therefore,
$p \in C$. From the definition of $C$, we know there exists $r^k \in E(m_1, m_2, \cdots, m_{k-1}, m_k)$ such that $d\left(f\left(r^k\right), p\right) < \frac{1}{k}$. Denote by $\tilde{r}^k$ the element of $\mathbb{N}^\mathbb{N}$ which consists of modifying $r^k$ by taking all components after the $k^{th}$ equal to one. Thus $\tilde{r}^k \in K$. Now $\left\{\tilde{r}^k\right\}$ is in a compact set and so taking a subsequence we can have $\tilde{r}^k \to r \in K$. But from the metric on $\mathbb{N}^\mathbb{N}$, it follows that $\rho\left(\tilde{r}^k, r^k\right) < \frac{1}{2^k}$. Therefore, $r^k \to r$ also and so $f\left(r^k\right) \to f\left(r\right) = p$. Therefore, $p \in f\left(K\right)$ and this proves the theorem.

Note we could have proved this under weaker assumptions. If we had assumed only that every point has a countable basis (first axiom of countability) and $\Omega$ is Hausdorff, the same argument would work. We will need the following definition.

**Definition 29.0.13** Let $\mathcal{F}$ be a $\sigma$ algebra of sets from $\Omega$ and let $\mu$ denote a finite measure defined on $\mathcal{F}$. We let $\mathcal{F}_\mu$ denote the completion of $\mathcal{F}$ with respect to $\mu$. Thus we let $\mu^*$ be the outer measure determined by $\mu$ and $\mathcal{F}_\mu$ will be the $\sigma$ algebra of $\mu^*$ measurable subsets of $\Omega$. We also define $\widehat{\mathcal{F}}$ by

$$\widehat{\mathcal{F}} \equiv \cap \{\mathcal{F}_\mu : \mu \text{ is a finite measure defined on } \mathcal{F}\}.$$

Also, if $X$ is a topological space, we will denote by $B\left(X\right)$ the Borel sets of $X$.

With this notation, we can give the following simple corollary of Theorem 29.0.12

**Corollary 29.0.14** Let $\Omega$ be a compact metric space and let $A$ be a Suslin subset of $\Omega$. Then $A \in \overline{B\left(\Omega\right)}$.

**Proof:** Let $\mu$ be a finite measure defined on $B\left(\Omega\right)$. By Theorem 29.0.12 $A \in \overline{B\left(\Omega\right)}$. Since this is true for every finite measure, $\mu$, it follows $A \in \overline{B\left(\Omega\right)}$ as claimed. This proves the corollary.

We give another technical lemma about the completion of measure spaces.

**Lemma 29.0.15** Let $\mu$ be a finite measure on a $\sigma$ algebra, $\Sigma$. Then $A \in \Sigma_\mu$ if and only if there exists $A_1 \in \Sigma$ and $N_1$ such that $A = A_1 \cup N_1$ where there exists $N \in \Sigma$ such that $\mu\left(N\right) = 0$ and $N_1 \subseteq N$.

**Proof:** Suppose first $A = A_1 \cup N_1$ where these sets are as described. Let $S \in \mathcal{P}\left(\Omega\right)$ and let $\mu^*$ denote the outer measure determined by $\mu$. Then since $A_1 \in \Sigma \subseteq \Sigma_\mu$

$$\mu^*(S) \leq \mu^*(S \setminus A) + \mu^*(S \cap A)$$

$$\leq \mu^*(S \setminus A_1) + \mu^*(S \cap A_1) + \mu^*(N_1)$$

$$= \mu^*(S \setminus A_1) + \mu^*(S \cap A_1) = \mu^*(S)$$

showing that $A \in \Sigma_\mu$.

Now suppose $A \in \Sigma_\mu$. Then there exists $B_1 \supseteq A$ such that $\mu^*(B_1) \leq \mu^*(A)$, and $B_1 \in \Sigma$. Also there exists $A_1^c \in \Sigma$ with $A_1^c \supseteq A^C$ and $\mu\left(A_1^c\right) = \mu\left(A^C\right)$. Then $A_1 \subseteq A \subseteq B_1$

$$A \subseteq A_1 \cup (B_1 \setminus A_1)$$

Now

$$\mu\left(A_1\right) + \mu^*(A^C) = \mu\left(A_1\right) + \mu\left(A_1^c\right) = \mu\left(\Omega\right)$$

and so

$$\mu\left(B_1 \setminus A_1\right) = \mu^*(B_1 \setminus A_1)$$

$$= \mu^*(B_1 \setminus A) + \mu^*(A \setminus A_1)$$

$$= \mu^*(B_1) - \mu^*(A) + \mu^*(A) - \mu^*(A_1)$$

$$= \mu^*(A) - (\mu\left(\Omega\right) - \mu^*(A^C)) = 0$$

because $A \in \Sigma_\mu$ implying $A = A_1 \cup (B_1 \setminus A_1) \cap A$ and $N_1 \subseteq N \equiv (B_1 \setminus A_1) \in \Sigma$ with $\mu\left(N\right) = 0$. This proves the lemma.

Next we need another definition.
Definition 29.0.16 We say \((\Omega, \Sigma)\), where \(\Sigma\) is a \(\sigma\) algebra of subsets of \(\Omega\), is separable if there exists a sequence \(\{A_n\}_{n=1}^{\infty} \subseteq \Sigma\) such that \(\sigma(\{A_n\}) = \Sigma\) and if \(w \neq w'\), then there exists \(A \in \Sigma\) such that \(\mathcal{X}_A(\omega) \neq \mathcal{X}_A(\omega')\). This last condition is referred to by saying \(\{A_n\}\) separates the points of \(\Omega\). Given two measure spaces, \((\Omega, \Sigma)\) and \((\Omega', \Sigma')\), we say they are isomorphic if there exists a function, \(f : \Omega \to \Omega'\) which is one to one and \(f(E) \in \Sigma'\) whenever \(E \in \Sigma\) and \(f^{-1}(F) \in \Sigma\) whenever \(F \in \Sigma'\).

The interesting thing about separable measure spaces is that they are isomorphic to a very simple sort of measure space in which topology plays a significant role.

Lemma 29.0.17 Let \((\Omega, \Sigma)\) be separable. Then there exists \(E \in \{0,1\}^\mathbb{N}\) such that \((\Omega, \Sigma)\) and \((E, \mathcal{B}(E))\) are isomorphic.

Proof: First we show \(\{A_n\}\) separates the points. We already know \(\Sigma\) separates the points. If this is not so, there exists \(\omega, \omega_1 \in \Omega\) such that for all \(n\), \(\mathcal{X}_{A_n}(\omega) = \mathcal{X}_{A_n}(\omega_1)\). Then let

\[ \mathcal{F} \equiv \{ F \in \Sigma : \mathcal{X}_F(\omega) = \mathcal{X}_F(\omega_1) \} \]

Thus \(A_n \in \mathcal{F}\) for all \(n\). It is also clear that \(\mathcal{F}\) is a \(\sigma\) algebra and so \(\mathcal{F} = \Sigma\) contradicting the assumption that \(\Sigma\) separates points. Now we define a function from \(\Omega\) to \(\{0,1\}^\mathbb{N}\) as follows.

\[ f(\omega) \equiv \{ \mathcal{X}_{A_n}(\omega) \}_{n=1}^{\infty} \]

We also let \(E \equiv f(\Omega)\). Since the \(\{A_n\}\) separate the points, we see that \(f\) is one to one. A subbasis for the topology of \(\{0,1\}^\mathbb{N}\) consists of sets of the form \(\prod_{i=1}^\infty H_i\) where \(H_i = \{0,1\}\) for all \(i\) except one, when \(i = j\) and \(H_j\) equals either \(\{0\}\) or \(\{1\}\). Therefore, \(f^{-1}(\text{subbasic open set}) \in \Sigma\) because if \(H_j\) is the exceptional set then this equals \(A_j\) if \(H_j = \{1\}\) and \(A_j^c\) if \(H_j = \{0\}\). Intersections of these subbasic sets with \(E\) gives a countable subbasis for \(E\) and so the inverse image of all sets in a countable subbasis for \(E\) are in \(\Sigma\), showing that \(f^{-1}(\text{open set}) \in \Sigma\). Now we consider \(f(A_n)\).

\[ f(A_n) \equiv \{ \lambda_n \}_{n=1}^{\infty} : \lambda_n = 1 \} \cap E, \]

an open set in \(E\). Hence \(f(A_n) \in \mathcal{B}(E)\). Now letting

\[ \mathcal{F} \equiv \{ G \subseteq \Omega : f(G) \in \mathcal{B}(E) \}, \]

we see that \(\mathcal{F}\) is a \(\sigma\) algebra which contains \(\{A_n\}_{n=1}^{\infty}\) and so \(\mathcal{F} \supseteq \sigma(\{A_n\}) = \Sigma\). Thus \(f(F) \in \mathcal{B}(E)\) for all \(A \in \Sigma\). This proves the lemma.

Lemma 29.0.18 Let \(\phi : (\Omega_1, \Sigma_1) \to (\Omega_2, \Sigma_2)\) where \(\phi^{-1}(U) \in \Sigma_1\) for all \(U \in \Sigma_2\). Then if \(F \in \hat{\Sigma}_2\), it follows \(\phi^{-1}(F) \in \hat{\Sigma}_1\).

Proof: Let \(\mu\) be a finite measure on \(\Sigma_1\) and define a measure \(\phi(\mu)\) on \(\Sigma_2\) by the rule

\[ \phi(\mu)(F) \equiv \mu(\phi^{-1}(F)) . \]

Now let \(A \in \Sigma_{2(\mu)}\). Then by Lemma 29.0.13, \(A = A_1 \cup N_1\) where there exists \(N \in \Sigma_2\) with \(\phi(\mu)(N) = 0\) and \(A_1 \in \Sigma_2\). Therefore, from the definition of \(\phi(\mu)\), we have \(\mu(\phi^{-1}(N)) = 0\) and therefore, \(\phi^{-1}(A) = \phi^{-1}(A_1) \cup \phi^{-1}(N_1)\) where \(\phi^{-1}(N_1) \subseteq \phi^{-1}(N) \in \Sigma_1\) and \(\mu(\phi^{-1}(N)) = 0\). Therefore, \(\phi^{-1}(A) \in \Sigma_{1(\mu)}\) and so if \(F \in \hat{\Sigma}_2\), then

\[ F \in \cap \{ \Sigma_{2(\nu) : \nu \text{ is a finite measure on } \Sigma_2} \} \subseteq \cap \{ \Sigma_{2(\mu) : \mu \text{ is a finite measure on } \Sigma_1} \} , \]

and so \(\phi^{-1}(F) \in \Sigma_{1(\mu)}\). Since \(\mu\) is arbitrary, this shows \(\phi^{-1}(F) \in \hat{\Sigma}_1\).

The next lemma is a special case of the Yankov von Neumann Aumann projection theorem. It contains the main idea of the proof of the more general theorem.

Lemma 29.0.19 Let \((\Omega, \Sigma)\) be separable and let \(X\) be a Suslin space. Let \(G \in \Sigma \times B(X)\). (Recall \(\Sigma \times B(X)\) is the \(\sigma\) algebra of product measurable sets, the smallest \(\sigma\) algebra containing the measurable rectangles.) Then

\[ \text{proj}_\Omega(G) \in \hat{\Sigma} . \]
**Proof:** Let \( f : (\Omega, \Sigma) \to (E, B(E)) \) be the isomorphism of Lemma 29.0.14. We have the following claim.

**Claim:** \( f \times \text{id}_X \) maps \( \Sigma \times B(X) \) to \( B(E) \times B(X) \).

**Proof of the claim:** First of all, assume \( A \times B \) is a measurable rectangle where \( A \in \Sigma \) and \( B \in B(X) \). Then by the assumption that \( f \) is an isomorphism, \( f(A) \in B(E) \) and so

\[
f \times \text{id}_X (A \times B) \in B(E) \times B(X).
\]

Now let

\[
\mathcal{F} \equiv \{ P \in \Sigma \times B(X) : f \times \text{id}_X (P) \in B(E) \times B(X) \}.
\]

Then we see that \( \mathcal{F} \) is a \( \sigma \) algebra and contains the elementary sets. (\( \mathcal{F} \) is closed with respect to complements because \( f \) is one to one.) Therefore, \( \mathcal{F} = \Sigma \times B(X) \) and this proves the claim.

Therefore, since \( G \in \Sigma \times B(X) \), we see

\[
f \times \text{id}_X (G) \in B(E) \times B(X) \subseteq B(E \times X).
\]

The set inclusion follows from the observation that if \( A \in B(E) \) and \( B \in B(X) \) then \( A \times B \in B(E \times X) \) and the collection of sets in \( B(E) \times B(X) \) which are in \( B(E \times X) \) is a \( \sigma \) algebra.

Therefore, there exists \( D \), a Borel set in \( E \times X \) such that \( f \times \text{id}_X (G) = D \cap (E \times X) \). Now from this it follows from Lemma 29.0.14 that \( D \) is a Suslin space. Letting \( Y \) be \( \{0,1\}^\mathbb{N} \), it follows that \( \text{proj} \), \( Y \) is a Suslin space in \( Y \).

**Proof:** First suppose \( G \) is a measurable rectangle, \( G = A \times B \) where \( A \in \Sigma \) and \( B \in B(X) \). Letting \( \Sigma_0 \) be the finite \( \sigma \) algebra, \( \{\emptyset, A^C, \Omega\} \), we see that \( G \in \Sigma_0 \times B(X) \). Similarly, if \( G \) equals an elementary set, then the conclusion of the lemma holds for \( G \).

Let

\[
\mathcal{F} \equiv \{ H \in \Sigma \times B(X) : H \in \Sigma_0 \times B(X) \}
\]

for some countably generated \( \sigma \) algebra, \( \Sigma_0 \). We just saw that \( \mathcal{F} \) contains the elementary sets. If \( H \in \mathcal{F} \), then \( H^C \in \Sigma_0 \times B(X) \) for the same \( \Sigma_0 \) and so \( \mathcal{F} \) is closed with respect to complements. Now suppose \( H_n \in \mathcal{F} \). Then for each \( n \), there exists a countably generated \( \sigma \) algebra, \( \Sigma_{0n} \) such that \( H_n \in \Sigma_{0n} \times B(X) \). Then \( \bigcup_{n=1}^\infty H_n \in \sigma (\{\Sigma_{0n} \times B(X)\}) \).

We will be done when we show

\[
\sigma (\{\Sigma_{0n} \times B(X)\}_{n=1}^\infty) \subseteq \sigma (\{\Sigma_{0n} \times B(X)\}_{n=1}^\infty) \times B(X)
\]

because it is clear that \( \sigma (\{\Sigma_{0n} \times B(X)\}_{n=1}^\infty) \) is countably generated. We see that

\[
\sigma (\{\Sigma_{0n} \times B(X)\}_{n=1}^\infty)
\]

is generated by sets of the form \( A \times B \) where \( A \in \Sigma_{0n} \) and \( B \in B(X) \). But each such set is also contained in \( \sigma (\{\Sigma_{0n} \times B(X)\}_{n=1}^\infty) \times B(X) \) and so the desired inclusion is obtained. Therefore, \( \mathcal{F} \) is a \( \sigma \) algebra and so since \( \mathcal{F} \) was shown to contain the measurable rectangles, this verifies \( \mathcal{F} = \Sigma \times B(X) \) and this proves the lemma.

**Theorem 29.0.21** Let \( (\Omega, \Sigma) \) be a measure space and let \( G \in \widehat{\Sigma} \times B(X) \) where \( X \) is a Suslin space. Then

\[
\text{proj}_\Omega (G) \in \widehat{\Sigma}.
\]
**Proof:** By the previous lemma, \( G \in \Sigma_0 \times B (X) \) where \( \Sigma_0 \) is countably generated. If \((\Omega, \Sigma_0)\) were separable, we could then apply Lemma 29.0.14 and be done. Unfortunately, we don’t know \( \Sigma_0 \) separates the points of \( \Omega \). Therefore, we define an equivalence class on the points of \( \Omega \) as follows. We say \( \omega \sim \omega_1 \) if and only if \( \mathcal{X}_A (\omega) = \mathcal{X}_A (\omega_1) \) for all \( A \in \Sigma_0 \). Now the nice thing to notice about this equivalence relation is that if \( \omega \in A \in \Sigma_0 \), and if \( \omega \sim \omega_1 \), then \( 1 = \mathcal{X}_A (\omega) = \mathcal{X}_A (\omega_1) \) implying \( \omega_1 \in A \) also. Therefore, every set of \( \Sigma_0 \) is the union of equivalence classes. It follows that for \( A \in \Sigma_0 \), and \( \pi \) the map given by \( \pi \omega = [\omega] \) where \([\omega] \) is the equivalence class determined by \( \omega \),

\[
\pi (A) \cap \pi (\Omega \setminus A) = \emptyset.
\]

Suppose now that \( H_n \in \Sigma_0 \times B (X) \). If \( ([\omega], x) \in \bigcap_{n=1}^{\infty} \pi \times \text{id}_X (H_n) \), then for each \( n \),

\[
([\omega], x) = (\pi w_n, x)
\]

for some \((\omega_n, x) \in H_n \). But this implies \( \omega \sim \omega_n \), and so from the above observation that the sets of \( \Sigma_0 \) are unions of equivalence classes, it follows that \( (\omega, x) \in H_n \). Therefore, \( (\omega, x) \in \bigcap_{n=1}^{\infty} H_n \) and so \( ([\omega], x) = \pi \times \text{id}_X (\omega, x) \) where \( (\omega, x) \in \bigcap_{n=1}^{\infty} H_n \). This shows that

\[
\pi \times \text{id}_X (\bigcap_{n=1}^{\infty} H_n) \supseteq \bigcap_{n=1}^{\infty} \pi \times \text{id}_X (H_n) .
\]

In fact these two sets are equal because the other inclusion is obvious. We will denote by \( \Omega_1 \) the set of equivalence classes and \( \Sigma_1 \) will be the subsets, \( S_1 \), of \( \Omega_1 \) such that \( S_1 = \{ [\omega] : \omega \in S \in \Sigma_0 \} \). Then \( (\Omega_1, \Sigma_1) \) is clearly a measure space which is separable. Let

\[
\mathcal{F} \equiv \{ H \in \Sigma_0 \times B (X) : \pi \times \text{id}_X (H), \pi \times \text{id}_X (H^C) \in \Sigma_1 \times B (X) \} .
\]

We see that the measurable rectangles, \( A \times B \) where \( A \in \Sigma_0 \) and \( B \in B (X) \) are in \( \mathcal{F} \), that from the above observation on countable intersections, \( \mathcal{F} \) is closed with respect to countable unions and closed with respect to complements. Therefore, \( \mathcal{F} \) is a \( \sigma \) algebra and so \( \mathcal{F} = \Sigma_0 \times B (X) \). By Lemma 29.0.14 \((\Omega_1, \Sigma_1)\) is isomorphic to \((E, B (E))\) where \( E \) is a subspace of \( \{0,1\}^\Omega \). Denoting the isomorphism by \( h \), it follows as in Lemma 29.0.14 that \( h \times \text{id}_X \) maps \( \Sigma_1 \times B (X) \) to \( B (E) \times B (X) \). Therefore, we see \( f = h \circ \pi \) is a mapping from \( \Omega \) to \( E \) which has the property that \( f \times \text{id}_X \) maps \( \Sigma_0 \times B (X) \) to \( B (E) \times B (X) \). Now from the proof of Lemma 29.0.14 starting with the claim, we see that \( G \in \Sigma_0 \).

However, if \( \mu \) is a finite measure on \( \widehat{\Sigma} \), then \( \widehat{\Sigma} = \Sigma_\mu \) and so \( \Sigma_0 \subseteq \widehat{\Sigma} \subseteq \widehat{\Sigma} \). This proves the theorem.

### 29.1 Multifunctions And Their Measurability

#### 29.1.1 The General Case

Let \( X \) be a separable complete metric space and let \((\Omega, \mathcal{C}, \mu)\) be a set, a \( \sigma \) algebra of subsets of \( \Omega \), and a measure \( \mu \) such that this is a complete \( \sigma \) finite measure space. Also let \( \Gamma : \Omega \to \mathcal{P}_F (X) \), the closed subsets of \( X \).

**Definition 29.1.1** We define \( \Gamma^{-} (S) \equiv \{ \omega \in \Omega : \Gamma (\omega) \cap S \neq \emptyset \} \)

We will consider a theory of measurability of set valued functions. The following theorem is the main result in the subject. In this theorem the third condition is what we will refer to as measurable.

**Theorem 29.1.2** The following are equivalent.

1. For all \( B \) a Borel set in \( X \), \( \Gamma^{-} (B) \in \mathcal{C} \).
2. For all \( F \) closed in \( X \), \( \Gamma^{-} (F) \in \mathcal{C} \).
3. For all \( U \) open in \( X \), \( \Gamma^{-} (U) \in \mathcal{C} \).
4. There exists a sequence, \( \{ \sigma_n \} \) of measurable functions satisfying \( \sigma_n (\omega) \in \Gamma (\omega) \) such that for all \( \omega \in \Omega \),

\[
\Gamma (\omega) = \{ \sigma_n (\omega) : n \in \mathbb{N} \}
\]

These functions are called measurable selections.

5. For all \( x \in X \), \( \omega \to \text{dist} (x, \Gamma (\omega)) \) is a measurable real valued function.
29.1. MULTIFUNCTIONS AND THEIR MEASURABILITY

6. \( G(\Gamma) \equiv \{ (\omega, x) : x \in \Gamma(\omega) \} \subseteq \mathcal{C} \times B(X) \).

**Proof:** It is obvious that 1.) \( \Rightarrow \) 2.). To see that 2.) \( \Rightarrow \) 3.) note that \( \Gamma^- (\bigcup_{n=1}^{\infty} F_i) = \bigcup_{n=1}^{\infty} \Gamma^- (F_i) \). Since any open set in \( X \) can be obtained as a countable union of closed sets, this implies 2.) \( \Rightarrow \) 3.).

Now we verify that 3.) \( \Rightarrow \) 4.). Let \( \{x_n\}_{n=1}^{\infty} \) be a countable dense subset of \( X \). For \( \omega \in \Omega \), let \( \psi_1(\omega) = x_n \) where \( n \) is the smallest integer such that \( \Gamma(\omega) \cap B(x_n, 1) \neq \emptyset \). Therefore, \( \psi_1(\omega) \) has countably many values, \( x_{n_1}, x_{n_2}, \ldots \) where \( n_1 < n_2 < \cdots \). Now

\[
\{ \omega : \psi_1 = x_n \} = \{ \omega : \Gamma(\omega) \cap B(x_n, 1) \neq \emptyset \} \cap \{ \omega : x \in \Gamma(\omega) \} \subseteq \mathcal{C}.
\]

Thus we see that \( \psi_1 \) is measurable and \( \text{dist}(\psi_1(\omega), \Gamma(\omega)) < 1 \). Let

\[
\Omega_n = \{ \omega \in \Omega : \psi_1(\omega) = x_n \}.
\]

Then \( \Omega_n \subseteq \mathcal{C} \) and \( \Omega_n \cap \Omega_m = \emptyset \) for \( n \neq m \). Let \( \bigcup_{n=1}^{\infty} \Omega_n = \Omega \). Let

\[
D_n = \{ x_k : x_k \in B(x_n, 1) \}.
\]

Now for each \( n \) and \( \omega \in \Omega_n \), let \( \psi_2(\omega) = x_k \) where \( k \) is the smallest index such that \( x_k \in D_n \) and \( B(x_k, 1) \cap \Gamma(\omega) \neq \emptyset \). Thus \( \text{dist}(\psi_2(\omega), \Gamma(\omega)) < \frac{1}{2} \) and

\[
d(\psi_2(\omega), \psi_1(\omega)) < 1.
\]

Continue this way obtaining \( \psi_k \) a measurable function such that

\[
\text{dist}(\psi_k(\omega), \Gamma(\omega)) < \frac{1}{2^{k-1}}, d(\psi_k(\omega), \psi_{k+1}(\omega)) < \frac{1}{2^{k-2}}.
\]

Then for each \( \omega \), \( \{ \psi_k(\omega) \} \) is a Cauchy sequence converging to a point, \( \sigma(\omega) \in \Gamma(\omega) \). This has shown that if \( \Gamma \) is measurable there exists a measurable selection, \( \sigma(\omega) \in \Gamma(\omega) \). Note that this had nothing to do with the measure. It remains to show there exists a sequence of these measurable selections, \( \sigma_n \) such that the conclusion of 4.) holds. To do this we define

\[
\Gamma_{ni}(\omega) \equiv \begin{cases} 
\Gamma(\omega) \cap B(x_n, 2^{-i}) & \text{if } \Gamma(\omega) \cap B(x_n, 2^{-i}) \neq \emptyset \\
\Gamma(\omega) & \text{otherwise.}
\end{cases}
\]

First we show that \( \Gamma_{ni} \) is measurable. Let \( U \) be open. Then

\[
\{ \omega : \Gamma_{ni}(\omega) \cap U \neq \emptyset \} = \{ \omega : \Gamma(\omega) \cap B(x_n, 2^{-i}) \cap U \neq \emptyset \} \cup
\]

\[
\left[ \{ \omega : \Gamma(\omega) \cap B(x_n, 2^{-i}) = \emptyset \} \cap \{ \omega : \Gamma(\omega) \cap U \neq \emptyset \} \right]
\]

\[
= \{ \omega : \Gamma(\omega) \cap B(x_n, 2^{-i}) \cap U \neq \emptyset \} \cup
\]

\[
[\{ \Omega \setminus \{ \omega : \Gamma(\omega) \cap B(x_n, 2^{-i}) \neq \emptyset \} \} \cap \{ \omega : \Gamma(\omega) \cap U \neq \emptyset \}],
\]

a measurable set. By what was just shown there exists \( \sigma_{ni}(\omega) \), a measurable function such that \( \sigma_{ni}(\omega) \in \Gamma_{ni}(\omega) \subseteq \Gamma(\omega) \) for all \( \omega \in \Omega \). For \( x \in \Gamma(\omega) \), then \( x \in B(x_n, 2^{-i}) \) whenever \( x_n \) is close enough to \( x \). Therefore, \( |\sigma_{ni}(\omega) - x| < 2^{-i} \). And it follows that condition 4.) holds. Note that this had nothing to do with the measure.

Now we verify that 4.) \( \Rightarrow \) 3.). Suppose there exist measurable selections \( \sigma_n(\omega) \in \Gamma(\omega) \) satisfying condition 4.). Let \( U \) be open. Then

\[
\{ \omega : \Gamma(\omega) \cap U \neq \emptyset \} = \bigcup_{n=1}^{\infty} \sigma_n^{-1}(U) \subseteq \mathcal{C}.
\]

Now we verify that 4.) \( \Rightarrow \) 5.). Let \( F(\omega) \equiv \text{dist}(x, \Gamma(\omega)) \). Then letting \( U \) be an open set in \( [0, \infty) \), \( F(\omega) \in U \) if and only if \( d(x, \sigma_n(\omega)) \in U \) for some \( \sigma_n(\omega) \). Let \( h_n(\omega) \equiv d(x, \sigma_n(\omega)) \). Then \( h_n \) is measurable and \( F^{-1}(U) = \bigcup_{n=1}^{\infty} h_n^{-1}(U) \subseteq \mathcal{C} \). This shows that for all \( \omega \), \( \omega \rightarrow \text{dist}(x, \Gamma(\omega)) \) is measurable and this proves 5.).

Now we verify that 5.) \( \Rightarrow \) 4.). We know \( \text{dist}(x, \Gamma(\omega)) \) is measurable and we show \( \{ \omega : \Gamma(\omega) \cap U \neq \emptyset \} \subseteq \mathcal{C} \) whenever \( U \) is open and then use 3.) \( \Rightarrow \) 4.). Since \( X \) is separable, there exists \( B(x_i, r_i) \) such that \( U = \bigcup_{i=1}^{\infty} B(x_i, r_i) \). Then

\[
\{ \omega : \Gamma(\omega) \cap U \neq \emptyset \} = \bigcup_{i=1}^{\infty} \{ \omega : \Gamma(\omega) \cap B(x_i, r_i) \neq \emptyset \}
\]

\[
= \bigcup_{i=1}^{\infty} \{ \omega : \text{dist}(x_i, \Gamma(\omega)) < r_i \} \subseteq \mathcal{C}.
\]

Therefore, 5.) \( \Rightarrow \) 4.) as claimed.
Now we must prove 5. $\Rightarrow$ 6.). We note that $\omega \to \text{dist}(x, \Gamma(\omega))$ is measurable and $x \to \text{dist}(x, \Gamma(\omega))$ is continuous. Also, the graph of $\Gamma, \mathcal{G}(\Gamma)$ is given by

$$\mathcal{G}(\Gamma) = \{ (\omega, x) : \text{dist}(x, \Gamma(\omega)) = 0 \}.$$  

We wish to show that $(\omega, x) \to \text{dist}(x, \Gamma(\omega))$ is product measurable because then $\mathcal{G}(\Gamma)$, being the inverse image of $\{0\}$ will be product measurable. Let $\{x_k\}$ be a countable dense set in $X$ and let

$$\phi_k(\omega, x) \equiv \text{dist}(x_n, \Gamma(\omega))$$

where $n$ is the first index such that $x \in B(x_n, 2^{-k})$. Then $\phi_k(\omega, x) \to \text{dist}(x, \Gamma(\omega))$ due to the continuity of $x \to \text{dist}(x, \Gamma(\omega))$ and so we must argue that $\phi_k$ is product measurable. On

$$E_n \equiv \Omega \times \left(B(x_n, 2^{-k}) \setminus \cup_{m<n} B(x_m, 2^{-k})\right),$$

$\phi_k(\omega, x) = \text{dist}(x_n, \Gamma(\omega))$. Thus, on this set, $\phi_k$ equals a measurable function of $\omega$ and does not depend on $x$ on $E_n$. It follows that there are measurable simple $\mathcal{C}$ measurable functions, $s_m(\omega)$ which increase pointwise to $\text{dist}(x_n, \Gamma(\omega))$ on $E_n$. Thus $s_m(\omega)x_{E_m}(x)$ increases to $\phi_k(\omega, x)$ on $E_n$ showing that $\phi_k x_{E_m}$ is product measurable with respect to $\mathcal{C} \times \sigma(\tau)$ since $E_n$ is a measurable rectangle with respect to $\mathcal{C}$ and $\sigma(\tau)$. Therefore, $\phi_k$ is product measurable and so $(\omega, x) \to \text{dist}(x, \Gamma(\omega))$ is also product measurable.

It remains to prove 6. $\Rightarrow$ 1.). This follows from Theorem 29.1.21.

$$\Gamma^-(B) \equiv \{ \omega : \Gamma(\omega) \cap B \neq \emptyset \} = \text{proj}_\Omega (\mathcal{G}(\Gamma) \cap (\Omega \cap B)).$$

But from Theorem 29.1.21, $\text{proj}_\Omega (\mathcal{G}(\Gamma) \cap (\Omega \cap B)) \in \tilde{\mathcal{C}} \subseteq \mathcal{C}_\mu = \mathcal{C}$. The last part results from $(\Omega, \mathcal{C}, \mu)$ being a complete measure space. Note that without this assumption we could not draw the conclusion desired. This required consideration of the measure.

For much more on multifunctions, you should see the book by Hu and Papageorgiou. The above proof follows the presentation in this book.

### 29.1.2 A Special Case Which Is Easier

The above is a pretty long and difficult argument to show that $\Gamma^-(U) \in \mathcal{C}$ for all $U$ open is equivalent to $\Gamma^-(F)$ for all $F$ closed. However, there is a special case for which this is much easier to show. Suppose $\Gamma(\omega)$ is not just closed but is also compact. Then as above, if $\Gamma^-(F) \in \mathcal{C}$ for all $F$ closed, then $\Gamma^-(U) = \cup_n \Gamma^-(F_n)$ where $F_n$ is an increasing sequence of closed sets whose union is $U$. This follows from the observation that

$$\Gamma(\omega) \cap U = \cup_n \Gamma(\omega) \cap F_n$$

and so to say the set on the left is nonempty is to say that at least one of the sets on the right is nonempty. Thus if $\Gamma^-(F) \in \mathcal{C}$ for all $F$ closed, then $\Gamma^-(U) \in \mathcal{C}$ for all $U$ open. This requires no special considerations.

Now suppose $\Gamma(\omega)$ is compact for every $\omega$ and that $\Gamma^-(U) \in \mathcal{C}$ for every $U$ open. Then let $F$ be a closed set and let $\{U_n\}$ be a decreasing sequence of open sets whose intersection equals $F$ such that also, for all $n, U_n \supseteq U_{n+1}$. Then

$$\Gamma(\omega) \cap F = \cap_n \Gamma(\omega) \cap U_n = \cap_n \Gamma(\omega) \cap \overline{U_n}.$$  

Now because of compactness, the set on the left is nonempty if and only if each set on the right is also nonempty. Thus $\Gamma^-(F) = \cap_n \Gamma^-(U_n) \in \mathcal{C}$. Thus in this special case, it is much easier to see that these two conditions for measurability are equivalent.
Chapter 30

Stochastic Inclusions

30.1 The General Context

The situation is as follows. There are spaces $V \subseteq W$ where $V, W$ are reflexive separable Banach spaces. It is assumed that $V$ is dense in $W$. Define the space for $p > 1$

$$V \equiv L^p ([0, T]; V)$$

where in each case, the $\sigma$ algebra of measurable sets will be $\mathcal{B} ([0, T])$ the Borel measurable sets. Thus, from the Riesz representation theorem,

$$V' = L^{p'} ([0, T]; V')$$

We also assume $(\Omega, \mathcal{F}, P)$ is a complete probability space. That is, if $P(E) = 0$ and $F \subseteq E$, then $F \in \mathcal{F}$. Also

$$V \subseteq W, \ W' \subseteq V'$$

$B(\omega)$ will be a linear operator, $B(\omega): W \to W'$ which satisfies

1. $\langle B(\omega)x, y \rangle = \langle B(\omega)y, x \rangle$
2. $\langle B(\omega)x, x \rangle \geq 0$ and equals 0 if and only if $x = 0$.
3. $\omega \to B(\omega)$ is a measurable $L(W, W')$ valued function.

In the above formulae, $\langle \cdot, \cdot \rangle$ denotes the duality pairing of the Banach space $W$, with its dual space. We will use this notation in the present paper, the exact specification of which Banach space being determined by the context in which this notation occurs.

For example, you could simply take $W = H = H'$ and $B$ the identity and consider a standard Gelfand triple where $H$ is a Hilbert space and $B$ equal to the identity. An interesting feature is the requirement that $B(\omega)$ be one to one. It would be interesting to include the case of degenerate $B$, but $B$ one to one includes the case of most interest just mentioned. Also a more general set of assumptions will allow the inclusion of this case of degenerate $B(\omega)$ also.

We assume always that the norm on the various reflexive Banach spaces is strictly convex.

30.2 Some Fundamental Theorems

The following fundamental result will be very useful. It says essentially that if $(Bu)' \in L^{p'} (0, T; V')$ and $u \in L^p (0, T; V)$ then the map $u \to Bu(t)$ is continuous as a map from

$$X \equiv \left\{ u \in L^p ([0, T]; V) : (Bu)' \in L^{p'} ([0, T]; V') \right\}$$

having norm equal to

$$\|u\|_X \equiv \|u\|_{L^p (0,T;V)} + \|(Bu)\|_{L^{p'} (0,T;V')}$$

to $W'$. There is also a convenient integration by parts formula, Theorem 21.4.3. For convenience, the dependence of $B$ on $\omega$ is often suppressed. This is not a problem because the entire approach will be to consider the situation for fixed $\omega$. 727
**Theorem 30.2.1** Let \( V \subseteq W, W' \subseteq V' \) be separable Banach spaces, and let \( Y \in L^{p'}(0,T;V') \) and

\[
Bu(t) = Bu_0 + \int_0^t Y(s) \, ds \quad \text{in} \ V', \ u_0 \in W, Bu(t) = B(u(t)) \quad \text{for a.e.} \ t
\]

(30.2.1)

Thus \( Y = (Bu)' \) as a weak derivative in the sense of \( V' \) valued distributions. It is known that \( u \in L^p(0,T,V) \) for \( p > 1 \). Then \( t \to Bu(t) \) is continuous into \( W' \) for \( t \) off a set of measure zero \( N \) and also there exists a continuous function \( t \to \langle Bu, u \rangle(t) \) such that for all \( t \notin N, \langle Bu, u \rangle(t) = \langle B(u(t)), u(t) \rangle, Bu(t) = B(u(t)) \), and for all \( t, \)

\[
\frac{1}{2} \langle Bu, u \rangle(t) = \frac{1}{2} \langle Bu_0, u_0 \rangle + \int_0^t \langle Y(s), u(s) \rangle \, ds
\]

Note that the formula (30.2.1) shows that \( Bu_0 = Bu(0) \). Also it shows that \( t \to \langle Bu, u \rangle(t) \) is continuous. To emphasize this a little more, \( Bu \) is the name of a function. \( Bu(t) = B(u(t)) \) for a.e. \( t \) and \( t \to Bu(t) \) is continuous into \( V' \) on \([0,T]\) because of the integral equation.

**Theorem 30.2.2** In the above corollary, the map \( u \to Bu(t) \) is continuous as a map from \( X \) to \( V' \). Also if \( Y \) denotes those \( f \in L^p([0,T];V) \) for which \( f' \in L^{p'}([0,T];V') \), so that \( f \) has a representative such that \( f(t) = f(0) + \int_0^t f'(s) \, ds \), then if \( \|f\|_Y \equiv \|f\|_{L^p([0,T];V)} + \|f'\|_{L^{p'}([0,T];V')} \) the map \( f \to f(t) \) is continuous.

**Proof:** First, why is \( u \to Bu(0) \) continuous? Say \( u, v \in X \) and say \( p \geq 2 \) first.

\[
Bu(t) - Bv(t) = Bu(0) - Bv(0) + \int_0^t (Bu)'(s) - (Bv)'(s) \, ds
\]

and so,

\[
\left( \int_0^T \|Bu(0) - Bv(0)\|_{V'}^{p'} \, dt \right)^{1/p'} \leq \left( \int_0^T \|Bu(t) - Bv(t)\|_{V'}^{p'} \, dt \right)^{1/p'}
\]

\[
+ \left( \int_0^T \left\| \int_0^t (Bu)'(s) - (Bv)'(s) \, ds \right\|_{V'}^p \, dt \right)^{1/p'}
\]

and so

\[
\|Bu(0) - Bv(0)\|_{V'}, T^{1/p'} \leq \left( \|B\|_{L^p([0,T];V)} + T^{1/p'} \left\| (Bu)' - (Bv)' \right\|_{L^{p'}([0,T];V')} \right)
\]

\[
\leq C (\|B\|, T) \|u - v\|_X
\]

Thus \( u \to Bu(0) \) is continuous into \( V' \). If \( p < 2 \), then you do something similar.

\[
\left( \int_0^T \|Bu(0) - Bv(0)\|_{V'}^p \, dt \right)^{1/p} \leq \left( \int_0^T \|Bu(t) - Bv(t)\|_{V'}^p \, dt \right)^{1/p}
\]

\[
+ \left( \int_0^T \left\| \int_0^t (Bu)'(s) - (Bv)'(s) \, ds \right\|_{V'}^p \, dt \right)^{1/p}
\]

\[
\|Bu(0) - Bv(0)\|_{V'}, T^{1/p} \leq \|B\|_{L^p([0,T];V)} + C (T) \left\| (Bu)' - (Bv)' \right\|_{L^{p'}([0,T];V')}
\]

\[
\leq C (\|B\|, T) \|u - v\|_X.
\]

However, one could just as easily have done this for an arbitrary \( s < T \) by repeating the argument for

\[
Bu(t) = Bu(s) + \int_s^t (Bu)'(r) \, dr
\]

Thus this mapping is certainly continuous into \( V' \). The last assertion is similar. ■

Also of use will be the following generalization of the Ascoli Arzela theorem. [78], Theorem 24.6.3.
30.2. SOME FUNDAMENTAL THEOREMS

Theorem 30.2.3 Let $q > 1$ and let $E \subseteq W \subseteq X$ where the injection map is continuous from $W$ to $X$ and compact from $E$ to $W$. Let $S$ be defined by

$$\left\{ u \text{ such that } \|u(t)\|_E \leq R \text{ for all } t \in [a,b], \text{ and } \|u(s) - u(t)\|_X \leq R|t - s|^{1/q} \right\}.$$  

Thus $S$ is bounded in $L^\infty (0,T;E)$ and in addition, the functions are uniformly Holder continuous into $X$. Then $S \subseteq C ([a,b];W)$ and if $\{ u_n \} \subseteq S$, there exists a subsequence, $\{ u_{n_k} \}$ which converges to a function $u \in C ([a,b];W)$ in the following way.

$$\lim_{k \to \infty} \|u_{n_k} - u\|_{\infty, W} = 0.$$  

Next is a major measurable selection theorem which forms an essential part of showing the existence of measurable solutions. See Theorem 22.2.3.

Theorem 30.2.4 Let $V$ be a reflexive separable Banach space with dual $V'$, and let $p, p'$ be such that $p > 1$ and $\frac{1}{p} + \frac{1}{p'} = 1$. Let the functions $t \to u_n (t, \omega)$, for $n \in \mathbb{N}$, be in $L^{p'} ([0,T] ; V')$ and $(t, \omega) \to u_n (t, \omega)$ be $\mathcal{B} ([0,T]) \times \mathcal{F} \equiv \mathcal{P}$ measurable into $V'$. Suppose there is a set of measure zero $N \subseteq \Omega$ such that if $\omega \notin N$, then

$$\sup_{t \in [0,T]} \|u_n (t, \omega)\|_V \leq C (\omega),$$  

for all $n$. Also, suppose for each $\omega \notin N$, each subsequence of $\{ u_n \}$ has a further subsequence that converges weakly in $L^{p'} ([0,T] ; V')$ to $v (\cdot, \omega) \in L^{p'} ([0,T] ; V')$ such that the function $t \to v (t, \omega)$ is weakly continuous into $V'$.

Then, there exists a product measurable function $u$ such that $t \to u (t, \omega)$ is weakly continuous into $V'$ and for each $\omega \notin N$, a subsequence $u_{n(\omega)}$ such that $u_{n(\omega)} (\cdot, \omega) \to u (\cdot, \omega)$ weakly in $L^{p'} ([0,T] ; V')$.

We prove the theorem in steps given below. Let $X = \prod_{k=1}^\infty C ([0,T])$ and note that when it is equipped with the product topology, then one can consider $X$ as a metric space using the metric

$$d (f, g) = \sum_{k=1}^\infty 2^{-k} \frac{\|f_k - g_k\|}{1 + \|f_k - g_k\|},$$  

where $f = (f_1, f_2, \ldots), g = (g_1, g_2, \ldots) \in X$, and the norm is the maximum norm in $C ([0,T])$. With this metric, $X$ is complete and separable.

Lemma 30.2.5 Let $\{ f_n \}$ be a sequence in $X$ and suppose that each one of the components $f_{nk}$ is bounded by $C = C (k)$ in $C^{0,1} ([0,T])$. Then, there exists a subsequence $\{ f_{n_j} \}$ that converges to some $f \in X$ as $n_j \to \infty$. Thus, $\{ f_n \}$ is pre-compact in $X$.

Proof: By the Ascoli–Arzelà theorem, there exists a subsequence $\{ f_{n_1} \}$ such that the sequence of the first components $f_{n_1}$ converges in $C ([0,T])$. Then, taking a subsequence, one can obtain $\{ n_2 \}$ a subsequence of $\{ n_1 \}$ such that both the first and second components of $f_{n_2}$ converge. Continuing in this way one obtains a sequence of subsequences, each a subsequence of the previous one such that $f_{n_j}$ has the first $j$ components converging to functions in $C ([0,T])$. Therefore, the diagonal subsequence has the property that it has every component converging to a function in $C ([0,T])$. The resulting function is $f \in \prod_k C ([0,T]).$ 

Now, for $m \in \mathbb{N}$ and $\phi \in V$, define $l_m (t) \equiv \max (0, t - (1/m))$ and $\psi_{m, \phi} : L^{p'} ([0,T] ; V') \to C ([0,T])$ by

$$\psi_{m, \phi} u (t) \equiv \int_0^T \langle m \phi \chi_{[l_m(t),t]} (s), u (s) \rangle_V ds = m \int_{l_m (t)}^t \langle \phi, u (s) \rangle_V ds.$$  

Here, $\chi_{[l_m(t),t]} (\cdot)$ is the indicator function of the interval $[l_m(t),t]$ and $\langle \cdot, \cdot \rangle_V = \langle \cdot, \cdot \rangle_V$ is the duality pairing between $V$ and $V'$.

Let $\mathcal{D} = \{ \phi_r \}_{r=1}^\infty$ denote a countable dense subset of $V$. Then the pairs $(\phi, m)$ for $\phi \in \mathcal{D}$ and $m \in \mathbb{N}$ form a countable set. Let $(m_k, \phi_{r_k})$ denote an enumeration of the pairs $(m, \phi) \in \mathbb{N} \times \mathcal{D}$. To simplify the notation, we set

$$f_k (u) (t) \equiv \psi_{m_k, \phi_{r_k}} (u) (t) = m_k \int_{l_{m_k} (t)}^t \langle \phi_{r_k}, u (s) \rangle_V ds.$$
For fixed $\omega \notin N$ and $k$, the functions $\{ t \to f_k (u_j (\cdot, \omega) (t)) \}_{j}$ are uniformly bounded and equicontinuous because they are in $C^{0,1} ([0, T])$. Indeed, we have for $\omega \notin N$,

$$| f_k (u_j (\cdot, \omega) (t)) | = \left| m_k \int_{l_{mk} (t)}^{t} \langle \phi_{rk}, u_j (s, \omega) \rangle_{V} \, ds \right| \leq C (\omega) \| \phi_{rk} \|_{V}$$

and for $t \leq t'$

$$| f_k (u_j (\cdot, \omega) (t)) - f_k (u_j (\cdot, \omega) (t')) | \leq m_k \left| \int_{l_{mk} (t)}^{t} \langle \phi_{rk}, u_j (s, \omega) \rangle_{V} \, ds - \int_{l_{mk} (t')}^{t'} \langle \phi_{rk}, u_j (s, \omega) \rangle_{V} \, ds \right| \leq 2m_k | t' - t | C (\omega) \| \phi_{rk} \|_{V}.$$ 

By Lemma 30.2.2, the set of functions $\{ X_{NC} (\omega) f (u_j (\cdot, \omega)) \}_{j=n}^{\infty}$ is pre-compact in $X = \prod_k C ([0, T])$. We now define a set valued map $\Gamma^{n} : \Omega \to X$ by

$$\Gamma^{n} (\omega) \equiv \bigcup_{j \geq n} \{ X_{NC} (\omega) f (u_j (\cdot, \omega)) \},$$

where the closure is taken in $X$. Then $\Gamma^{n} (\omega)$ is the closure of a pre-compact set in $X$ and so $\Gamma^{n} (\omega)$ is compact in $X$. From the definition, a function $f$ is in $\Gamma^{n} (\omega)$ if and only if $d (f, X_{NC} (\omega) f (w_l)) \to 0$ as $l \to \infty$, where each $w_l$ is one of the $u_j (\cdot, \omega)$ for $j \geq n$. In the topology on $X$, this happens iff for every $k$,

$$f_k (t) = \lim_{l \to \infty} m_k \int_{l_{mk} (t)}^{t} \langle \phi_{rk}, X_{NC} (\omega) w_l (s, \omega) \rangle_{V} \, ds,$$

where the limit is the uniform limit in $t$. 

**Lemma 30.2.6** The mapping $\omega \to \Gamma^{n} (\omega)$ is an $F$ measurable set-valued map with values in $X$. If $\sigma$ is a measurable selection, then for each $t$, $\omega \to \sigma (t, \omega)$ is $F$ measurable and $(t, \omega) \to \sigma (t, \omega)$ is $B ([0, T]) \times F$ measurable.

We note that if $\sigma$ is a measurable selection then $\sigma (\omega) \in \Gamma^{n} (\omega)$, so $\sigma = \sigma (\cdot, \omega)$ is a continuous function. To have $\sigma$ measurable would mean that $\sigma^{-1} (open) \in F$, where the open set is in $C ([0, T])$. 

**Proof:** Let $O$ be a basic open set in $X$. Then $O = \bigcap_{k=1}^{\infty} O_k$, where $O_k$ is a proper open set of $C ([0, T])$ only for $k \in \{ k_1, \ldots, k_r \}$. Thus there is a proper open set in these positions and in every other position the open set is the whole space $C ([0, T])$. We need to show that

$$\Gamma^{n}^{-1} (O) \equiv \{ \omega : \Gamma^{n} (\omega) \cap O \neq \emptyset \} \in F.$$ 

Now, $\Gamma^{n}^{-1} (O) = \bigcap_{k=1}^{r} \{ \omega : \Gamma^{n} (\omega)_{k_i} \cap O_{k_i} \neq \emptyset \}$, so we consider whether

$$\{ \omega : \Gamma^{n} (\omega)_{k_i} \cap O_{k_i} \neq \emptyset \} \in F. \quad (30.2.2)$$

From the definition of $\Gamma^{n} (\omega)$, this is equivalent to the condition that

$$f_{k_i} (X_{NC} (\omega) u_j (\cdot, \omega)) = (f (X_{NC} (\omega) u_j (\cdot, \omega)))_{k_i} \in O_{k_i},$$

for some $j \geq n$, and so the set in (30.2.2) is of the form

$$\bigcup_{j=n}^{\infty} \{ \omega : (f (X_{NC} (\omega) u_j (\cdot, \omega)))_{k_i} \in O_{k_i} \}.$$

Now $\omega \to (f (X_{NC} (\omega) u_j (\cdot, \omega)))_{k_i}$ is $F$ measurable into $C ([0, T])$ and so the above set is in $F$. To see this, let $g \in C ([0, T])$ and consider the inverse image of the ball with radius $r$ and center $g$,

$$B (g, r) = \{ \omega : \| (X_{NC} (\omega) f (u_j (\cdot, \omega)))_{k_i} - g \|_{C ([0, T])} < r \}.$$

By continuity considerations,

$$\| (X_{NC} (\omega) f (u_j (\cdot, \omega)))_{k_i} - g \|_{C ([0, T])} = \sup_{t \in [0, T]} \| (X_{NC} (\omega) f (u_j (t, \omega)))_{k_i} - g (t) \|,$$
which is the sup over countably many $\mathcal{F}$ measurable functions. Thus, it is $\mathcal{F}$ measurable. Since every open set is the countable union of such balls, it follows that the claim about $\mathcal{F}$ measurability is valid. Hence, $\Gamma^n(\omega)$ is $\mathcal{F}$ measurable whenever $O$ is a basic open set.

Now, $X$ is a separable metric space and so every open set is a countable union of these basic sets. Let $U \subseteq X$ be open with $U = \bigcup_{i=1}^{\infty} O_i$ where $O_i$ is a basic open set as above. Then,

$$\Gamma^n(U) = \bigcup_{i=1}^{\infty} \Gamma^n(O_i) \in \mathcal{F}.$$ 

The existence of a measurable selection follows from the standard theory of measurable multi-functions [3, 15] see [15] starting on Page 141 for all the necessary stuff on measurable multifunctions or Section 29.1. If $\sigma$ is one of these measurable selections, the evaluation at $t$ is $\mathcal{F}$ measurable. Thus, $\omega \to \sigma(t, \omega)$ is $\mathcal{F}$ measurable with values in $\mathbb{R}^\infty$. Also, $t \to \sigma(t, \omega)$ is continuous, and so it follows that in fact $\sigma$ is product measurable as claimed. 

**Definition 30.2.7** Let $\Gamma(\omega) \equiv \cap_{n=1}^{\infty} \Gamma^n(\omega)$.

**Lemma 30.2.8** $\Gamma$ is a nonempty $\mathcal{F}$ measurable set-valued function with values in compact subsets of $X$. There exists a measurable selection $\gamma$ such that $(t, \omega) \to \gamma(t, \omega)$ is $\mathcal{P}$ measurable. Also, for each $\omega$, there exists a subsequence,

$$u_{n(\omega)}(\cdot, \omega)$$

such that for each $k$,

$$\gamma_k(t, \omega) = \lim_{n(\omega) \to \infty} f(X)_{\gamma(t, \omega)\cdot} = \lim_{n(\omega) \to \infty} m_{k(\omega)} \int_{t_{m_{k(\omega)}}(t)}^{t_{m_{k(\omega)}}(t)} \langle \phi_{j_k}, X_{\gamma(t, \omega)\cdot} \rangle \, ds.$$ 

**Proof:** From the definition of $\Gamma(\omega) = \cap_{n=1}^{\infty} \Gamma^n(\omega)$ it follows that $\omega \to \Gamma(\omega)$ is a compact set-valued map in $X$ and is nonempty because each $\Gamma^n(\omega)$ is nonempty and compact, and the $\Gamma^n(\omega)$ are nested. We next show that $\omega \to \Gamma(\omega)$ is $\mathcal{F}$ measurable. Indeed, each $\Gamma^n$ is compact valued and $\mathcal{F}$ measurable so, if $F$ is closed,

$$\Gamma(\omega) \cap F = \cap_{n=1}^{\infty} \Gamma^n(\omega) \cap F,$$

and the left-hand side is not empty if and only if each $\Gamma^n(\omega) \cap F \neq \emptyset$. Thus, for $F$ closed,

$$\{\omega : \Gamma(\omega) \cap F \neq \emptyset\} = \bigcap_n \{\omega : \Gamma^n(\omega) \cap F \neq \emptyset\},$$

and so

$$\Gamma^{-}(F) = \bigcap_n \Gamma^{-}(F) \in \mathcal{F}.$$ 

The last claim follows from the theory of multi-functions, see, e.g., [3, 15] or Section 29.1. The fact that $\Gamma^n(\omega)$ is compact, $\Gamma^n$ is measurable and $\Gamma^{-}(U) \in \mathcal{F}$, for $U$ open, imply the strong measurability of $\Gamma^n$ see also Section 29.1, and also that $\Gamma^{-}(F) \in \mathcal{F}$. Thus, $\omega \to \Gamma(\omega)$ is a nonempty compact valued in $X$ and $\mathcal{F}$ measurable. We are using the theorem which says that when $\omega$ has compact values, then one can conclude that strong measurability and measurability coincide. This is why we can say that $\Gamma^{-}(F) \in \mathcal{F}$.

The standard theory [3, 15], Section 29.1, also guarantees the existence of an $\mathcal{F}$ measurable selection $\omega \to \gamma(\omega)$ with $\gamma(\omega) \in \Gamma(\omega)$, for each $\omega$, and also that $t \to \gamma_k(t, \omega)$ (the $k$th component of $\gamma$) is continuous. Next, we consider the product measurability of $\gamma_k$. We know that $\omega \to \gamma_k(\omega)$ is $\mathcal{F}$ measurable into $C([0,T])$ and since pointwise evaluation is continuous, $\omega \to \gamma_k(t, \omega)$ is $\mathcal{F}$ measurable. (This is nothing more than a case of the general result that a continuous function of a measurable function is measurable.) Then, since $t \to \gamma_k(t, \omega)$ is continuous, it follows that $\gamma_k$ is a $\mathcal{P}$ measurable real valued function and that $\gamma$ is a $\mathcal{P}$ measurable $\mathbb{R}^\infty$ valued function. Since $\gamma(\omega) \in \Gamma(\omega)$, it follows that for each $n, \gamma(\omega) \in \Gamma^n(\omega)$. Therefore, there exists $j_n \geq n$ such that for each $\omega$,

$$d\left( f(X\gamma(t, \omega)\cdot), \gamma(\omega) \right) < 2^{-n}.$$ 

Therefore, for a suitable subsequence $\{u_{n(\omega)}(\cdot, \omega)\}$, we have

$$\gamma(\omega) = \lim_{n(\omega) \to \infty} f(X\gamma(t, \omega)\cdot) = \lim_{n(\omega) \to \infty} f(X\gamma(t, \omega)\cdot)$$

for each $\omega$. In particular, for each $k$

$$\gamma_k(t, \omega) = \lim_{n(\omega) \to \infty} f(X\gamma(t, \omega)\cdot)$$
\[
= \lim_{n(\omega) \to \infty} m_k \int_{l_{m_k}(t)}^t \langle \phi_{r_k}, X_{N^C}(\omega) u_{n(\omega)}(s, \omega) \rangle_V \, ds, \tag{30.2.3}
\]
for each \( t \).

Note that it is not clear that \( (t, \omega) \to f(X_{N^C}(\omega) u_{n(\omega)}(t, \omega)) \) is \( \mathcal{P} \) measurable, although \( (t, \omega) \to \gamma(t, \omega) \) is \( \mathcal{P} \) measurable.

Now here is the proof of the theorem.

**Proof of Theorem 30.2.3** By assumption, there exists a further subsequence, still denoted by \( n(\omega) \), such that, in addition to 30.2.2, the weak limit

\[
\lim_{n(\omega) \to \infty} X_{N^C}(\omega) u_{n(\omega)}(\cdot, \omega) = u(\cdot, \omega)
\]
exists in \( L^p([0, T]; V') \) such that \( t \to u(t, \omega) \) is weakly continuous into \( V' \). Then, 30.2.3 also holds for this further subsequence and in addition,

\[
\begin{align*}
& m_k \int_{l_{m_k}(t)}^t \langle \phi_{r_k}, u(s, \omega) \rangle_V \, ds \\
= & \lim_{n(\omega) \to \infty} m_k \int_{l_{m_k}(t)}^t \langle \phi_{r_k}, X_{N^C}(\omega) u_{n(\omega)}(s, \omega) \rangle_V \, ds \\
= & \gamma_k(t, \omega).
\end{align*}
\]

Letting \( \phi \in \mathcal{D} \) be given, there exists a subsequence, denoted by \( k \), such that \( m_k \to \infty \) and \( \phi_{r_k} = \phi \). Recall \( (m_k, \phi_{r_k}) \) denoted an enumeration of the pairs \((m, \phi) \in \mathbb{N} \times \mathcal{D} \). Then, passing to the limit and using the assumed continuity of \( s \to u(s, \omega) \), the left-hand side of this equality converges to \( \langle \phi, u(s, \omega) \rangle_V \) and so the right-hand side, \( \gamma_k(t, \omega) \), must also converge and for each \( \omega \). Since the right-hand side is a product measurable function of \( (t, \omega) \), it follows that the pointwise limit is also product measurable. Hence, \( (t, \omega) \to \langle \phi, u(t, \omega) \rangle_V \) is product measurable for each \( \phi \in \mathcal{D} \). Since \( \mathcal{D} \) is a dense set, it follows that \( (t, \omega) \to \langle \phi, u(t, \omega) \rangle_V \) is \( \mathcal{P} \) measurable for all \( \phi \in V \) and so by the Pettis theorem, 30.2.3, \( (t, \omega) \to u(t, \omega) \) is \( \mathcal{P} \) measurable into \( V' \).

Actually, one can say more about the measurability of the approximating sequence and in fact, we can obtain one for which \( \omega \to u_{n(\omega)}(t, \omega) \) is also \( \mathcal{F} \) measurable.

**Lemma 30.2.9** Suppose that \( u_{n(\omega)} \to u \) weakly in \( L^p([0, T]; V') \), where \( u \) is product measurable, and \( \{u_{n(\omega)}\} \) is a subsequence of \( \{u_n\} \), such that there exists a set of measure zero \( N \subseteq \Omega \) and

\[
\sup_{t \in [0, T]} \|u_n(t, \omega)\|_V < C(\omega), \quad \text{for } \omega \notin N.
\]

Then, there exists a subsequence of \( \{u_n\} \), denoted as \( \{u_k(\omega)\} \), such that \( u_k(\omega) \to u \) weakly in \( L^p([0, T]; V') \), \( \omega \to k(\omega) \) is \( \mathcal{F} \) measurable, and \( \omega \to u_k(\omega)(t, \omega) \) is also \( \mathcal{F} \) measurable, for each \( \omega \notin N \).

**Proof:** Assume that \( f, g \in L^p([0, T]; V') \) and let \( \{\phi_k\} \) be a countable dense subset of \( L^p([0, T]; V) \). Then, a bounded set in \( L^p([0, T]; V') \) with the weak topology can be considered a complete metric space using the metric

\[
d(f, g) = \sum_{j=1}^{\infty} 2^{-j} \frac{|\langle \phi_j, f - g \rangle|}{1 + |\langle \phi_j, f - g \rangle|}.
\]

Now, let \( k(\omega) \) be the first index of \( \{u_n\} \) that is at least as large as \( k \) and such that

\[
d(X_{N^C}(\omega) u_{k(\omega)}, u) \leq 2^{-k}.
\]

Such an index exists because there exists a convergent sequence \( X_{N^C}(\omega) u_{n(\omega)} \) that converge weakly to \( u \). In fact,

\[
\{\omega : k(\omega) = l\} = \left\{ \omega : d(u_l, u) \leq 2^{-k}\right\} \cap \left\{ \omega : d(u_j, u) > 2^{-k}\right\}.
\]

Since \( u \) is product measurable and each \( u_0 \) is also product measurable, these are all measurable sets with respect to \( \mathcal{F} \) and so \( \omega \to k(\omega) \) is \( \mathcal{F} \) measurable. Now, we have that \( X_{N^C}(\omega) u_{k(\omega)} \to u \) weakly in \( L^p([0, T]; V') \), for each \( \omega \), and each function is \( \mathcal{F} \) measurable because

\[
u_{k(\omega)}(t, \omega) = \sum_{j=1}^{\infty} X_{k(\omega)=j} u_j(t, \omega),
\]
and every term in the sum is $F$ measurable. ■

**Theorem 30.2.10** Let $V$ be a reflexive separable Banach space with dual $V'$, and let $p, p'$ be such that $p > 1$ and $\frac{1}{p} + \frac{1}{p'} = 1$. Let the functions $t \to u_n(t, \omega)$, for $n \in \mathbb{N}$, be in $L^p([0, T]; V) = \mathcal{V}$ and $(t, \omega) \to u_n(t, \omega)$ be $\mathcal{B}([0, T]) \times \mathcal{F} \equiv \mathcal{F}$ measurable into $V$. Suppose there is a set of measure zero $N \subseteq \Omega$ such that if $\omega \notin N$, then

$$\|u_n(\cdot, \omega)\|_V \leq C(\omega),$$

for all $n$. (Thus, by weak compactness, for each $\omega$, each subsequence of $\{u_n\}$ has a further subsequence that converges weakly in $V$ to $v(\cdot, \omega) \in \mathcal{V}$. (v not known to be $\mathcal{P}$ measurable))

Then, there exists a product measurable function $u$ such that $t \to u(t, \omega)$ is in $\mathcal{V}$ and for each $\omega \notin N$, a subsequence $u_{n(\omega)}(\cdot, \omega) \to u(\cdot, \omega)$ weakly in $\mathcal{V}$.

We prove the theorem in steps given below. Let $X = \prod_{k=1}^{\infty} C([0, T])$ and note that when it is equipped with the product topology, then one can consider $X$ as a metric space using the metric

$$d(f, g) = \sum_{k=1}^{\infty} 2^{-k} \frac{\|f_k - g_k\|}{1 + \|f_k - g_k\|},$$

where $f = (f_1, f_2, \ldots)$, $g = (g_1, g_2, \ldots) \in X$, and the norm is the maximum norm in $C([0, T])$. With this metric, $X$ is complete and separable.

**Lemma 30.2.11** Let $\{f_n\}$ be a sequence in $X$ and suppose that each one of the components $f_{nk}$ is bounded by $C = C(k)$ in $C^{0, (1/p')}$([0, T]). Then, there exists a subsequence $\{f_{nj}\}$ that converges to some $f \in X$ as $n_j \to \infty$. Thus, $\{f_n\}$ is pre-compact in $X$.

**Proof:** This follows right away from Tychonoff’s theorem and the compactness of the embedding of the Holder space into $C([0, 1])$. ■

Now, for $m \in \mathbb{N}$ and $\phi \in V'$, define $l_m(t) \equiv \max (0, t - (1/m))$ and $\psi_{m, \phi} : V \to C([0, T])$ by

$$\psi_{m, \phi}(u)(t) = \int_0^T \langle m\phi X_{[l_m(t), t]}(s), u(s) \rangle_V ds = m \int_{l_m(t)}^t \langle \phi, u(s) \rangle_V ds.$$

Here, $X_{[l_m(t), t]}(\cdot)$ is the characteristic function of the interval $[l_m(t), t]$ and $\langle \cdot, \cdot \rangle_V = \langle \cdot, \cdot \rangle_{V'}$ is the duality pairing between $V$ and $V'$.

Let $D = \{\phi_r\}_{r=1}^{\infty}$ denote a countable subset of $V'$. Then the pairs $(\phi, m)$ for $\phi \in D$ and $m \in \mathbb{N}$ form a countable set. Let $(m_k, \phi_{rk})$ denote an enumeration of the pairs $(m, \phi) \in \mathbb{N} \times D$. To simplify the notation, we set

$$f_k(u)(t) \equiv \psi_{m_k, \phi_{rk}}(u)(t) = m_k \int_{l_{m_k}(t)}^t \langle \phi_{rk}, u(s) \rangle_V ds.$$

For fixed $\omega \notin N$ and $k$, the functions $\{t \to f_k(u_j(\cdot, \omega))(t)\}_{j}$ are uniformly bounded and equicontinuous because they are in $C^{0, 1/p'}([0, T])$. Indeed, we have for $\omega \notin N$,

$$|f_k(u_j(\cdot, \omega))(t)| = \left| m_k \int_{l_{m_k}(t)}^t \langle \phi_{rk}, u_j(s, \omega) \rangle_V ds \right| \leq m \left\| \phi_{rk} \right\|_{T^{1/p'}} \left\| \int_{l_{m_k}(t)}^t \left\| u_j(s, \omega) \right\|_{V'} ds \right\|_{1/p'} \leq C(\omega) m \left\| \phi_{rk} \right\|_{V'}, T^{1/p'}$$

and for $t \leq t'$

$$|f_k(u_j(\cdot, \omega))(t) - f_k(u_j(\cdot, \omega))(t')| \leq \left| m_k \int_{l_{m_k}(t)}^t \left\langle \phi_{rk}, u_j(s, \omega) \right\rangle_V ds - m_k \int_{l_{m_k}(t')}^t \left\langle \phi_{rk}, u_j(s, \omega) \right\rangle_V ds \right| \leq 2m_k |t' - t|^{1/p'} C(\omega) \left\| \phi_{rk} \right\|_{V'},$$
By Lemma 30.2.11, the set of functions \( \{ X_{NC}(\omega) f(u_j(\cdot, \omega)) \}_{j=1}^{\infty} \) is pre-compact in \( X = \prod_k C([0,T]) \). We now define a set valued map \( \Gamma^n : \Omega \to X \) by

\[
\Gamma^n(\omega) \equiv \bigcup_{j \geq n} \{ X_{NC}(\omega) f(u_j(\cdot, \omega)) \},
\]

where the closure is taken in \( X \). Then \( \Gamma^n(\omega) \) is the closure of a pre-compact set in \( X \) and so \( \Gamma^n(\omega) \) is compact in \( X \). From the definition, a function \( f \) is in \( \Gamma^n(\omega) \) if and only if \( d(f, X_{NC}(\omega) f(w_l)) \to 0 \) as \( l \to \infty \), where each \( w_l \) is one of the \( u_j(\cdot, \omega) \) for \( j \geq n \). In the topology on \( X \), this happens iff for every \( k \),

\[
f_k(t) = \lim_{l \to \infty} m_k \int_{l_m(t)}^{t} \langle \phi_{kn}, X_{NC}(\omega) w_l(s, \omega) \rangle \, ds,
\]

where the limit is the uniform limit in \( t \).

**Lemma 30.2.12** The mapping \( \omega \to \Gamma^n(\omega) \) is an \( F \) measurable set-valued map with values in \( X \). If \( \sigma \) is a measurable selection, then for each \( t, \omega \to \sigma(t, \omega) \) is \( F \) measurable and \( t, \omega \to \sigma(t, \omega) \) is \( B([0,T]) \times F \) measurable.

We note that if \( \sigma \) is a measurable selection then \( \sigma(\omega) \in \Gamma^n(\omega) \), so \( \sigma = \sigma(\cdot, \omega) \) is a continuous function. To have \( \sigma \) measurable would mean that \( \sigma^{-1}(\text{open}) \in F \), where the open set is in \( C([0,T]) \).

**Proof:** Let \( O \) be a basic open set in \( X \). Then \( O = \bigcap_{k=1}^{\infty} O_k \), where \( O_k \) is a proper open set of \( C([0,T]) \) only for \( k \in \{ k_1, \ldots, k_r \} \). Thus there is a proper open set in these positions and in every other position the open set is the whole space \( C([0,T]) \). We need to show that

\[
\Gamma^n(O) = \{ \omega : \Gamma^n(\omega) \cap O \neq \emptyset \} \in F.
\]

Now, \( \Gamma^n(O) = \bigcap_{k=1}^{\infty} \{ \omega : \Gamma^n(\omega)_{k_i} \cap O_{k_i} \neq \emptyset \} \), so we consider whether

\[
\{ \omega : \Gamma^n(\omega)_{k_i} \cap O_{k_i} \neq \emptyset \} \in F. \tag{30.2.4}
\]

From the definition of \( \Gamma^n(\omega) \), this is equivalent to the condition that

\[
f_{k_i}(X_{NC}(\omega) u_j(\cdot, \omega)) = (f(X_{NC}(\omega) u_j(\cdot, \omega)))_{k_i} \in O_{k_i}
\]

for some \( j \geq n \), and so the set in (30.2.4) is of the form

\[
\bigcup_{j=n}^{\infty} \{ \omega : \langle f(X_{NC}(\omega) u_j(\cdot, \omega)) \rangle_{k_i} \in O_{k_i} \}.
\]

Now \( \omega \to \langle f(X_{NC}(\omega) u_j(\cdot, \omega)) \rangle_{k_i} \) is \( F \) measurable into \( C([0,T]) \) and so the above set is in \( F \). To see this, let \( g \in C([0,T]) \) and consider the inverse image of the ball with radius \( r \) and center \( g \),

\[
B(g, r) = \{ \omega : \| (X_{NC}(\omega) f(u_j(\cdot, \omega)))_{k_i} - g \|_{C([0,T])} < r \}.
\]

By continuity considerations,

\[
\| (X_{NC}(\omega) f(u_j(\cdot, \omega)))_{k_i} - g \|_{C([0,T])} = \sup_{t \in [0,T]} \| (X_{NC}(\omega) f(u_j(t, \omega)))_{k_i} - g(t) \|,
\]

which is the sup over countably many \( F \) measurable functions. Thus, it is \( F \) measurable. Since every open set is the countable union of such balls, it follows that the claim about \( F \) measurability is valid. Hence, \( \Gamma^n(O) \) is \( F \) measurable whenever \( O \) is a basic open set.

Now, \( X \) is a separable metric space and so every open set is a countable union of these basic sets. Let \( U \subseteq X \) be open with \( U = \bigcup_{l=1}^{\infty} O_l \) where \( O_l \) is a basic open set as above. Then,

\[
\Gamma^n(U) = \bigcup_{l=1}^{\infty} \Gamma^n(O_l) \in F.
\]

The existence of a measurable selection follows from the standard theory of measurable multi-functions [8, 18] see [13] starting on Page 141 for all the necessary stuff on measurable multifunctions or Section 24.1. If \( \sigma \) is one of these measurable selections, the evaluation at \( t \) is \( F \) measurable. Thus, \( \omega \to \sigma(t, \omega) \) is \( F \) measurable with values in \( \mathbb{R}^{\infty} \). Also, \( t \to \sigma(t, \omega) \) is continuous, and so it follows that in fact \( \sigma \) is product measurable as claimed. \( \blacksquare \)
Definition 30.2.13 Let $\Gamma (\omega) = \cap_{n=1}^{\infty} \Gamma^n (\omega)$.

Lemma 30.2.14 $\Gamma$ is a nonempty $\mathcal{F}$ measurable set-valued function with values in compact subsets of $X$. There exists a measurable selection $\gamma$ such that $(t, \omega) \to \gamma (t, \omega)$ is $\mathcal{P}$ measurable. Also, for each $\omega$, there exists a subsequence, $u_{n(\omega)} (\cdot, \omega)$ such that for each $k$,

$$
\gamma_k (t, \omega) = \lim_{n(\omega) \to \infty} f \left( X_{NC} (\omega) u_{n(\omega)} (t, \omega) \right)_k
$$

$$
= \lim_{n(\omega) \to \infty} m_k \int_{l_m (t)}^{t} \langle \phi_{rk}, X_{NC} (\omega) u_{n(\omega)} (s, \omega) \rangle_V \, ds.
$$

Proof: From the definition of $\Gamma (\omega) = \cap_{n=1}^{\infty} \Gamma^n (\omega)$ it follows that $\omega \to \Gamma (\omega)$ is a compact set-valued map in $X$ and is nonempty because each $\Gamma^n (\omega)$ is nonempty and compact, and the $\Gamma^n (\omega)$ are nested. We next show that $\omega \to \Gamma (\omega)$ is $\mathcal{F}$ measurable. Indeed, each $\Gamma^n$ is compact valued and $\mathcal{F}$ measurable so, if $F$ is closed,

$$
\Gamma (\omega) \cap F = \cap_{n=1}^{\infty} \Gamma^n (\omega) \cap F,
$$

and the left-hand side is not empty if and only if each $\Gamma^n (\omega) \cap F \neq \emptyset$. Thus, for $F$ closed,

$$
\{ \omega : \Gamma (\omega) \cap F \neq \emptyset \} = \cap_n \{ \omega : \Gamma^n (\omega) \cap F \neq \emptyset \},
$$

and so

$$
\Gamma^- (F) = \cap_n \Gamma^- (F) \in \mathcal{F}.
$$

The last claim follows from the theory of multi-functions, see, e.g., Section 30.3. The fact that $\Gamma^n (\omega)$ is compact, $\Gamma^n$ is measurable and $\Gamma^n^- (U) \in \mathcal{F}$, for $U$ open, imply the strong measurability of $\Gamma^n$ see also Section 30.4 and also that $\Gamma^- (F) \in \mathcal{F}$. Thus, $\omega \to \Gamma (\omega)$ is nonempty compact valued in $X$ and $\mathcal{F}$ measurable. We are using the theorem which says that when $\Gamma$ has compact values, then one can conclude that strong measurability and measurability coincide. This is why we can say that $\Gamma^- (F) \in \mathcal{F}$.

The standard theory also guarantees the existence of an $\mathcal{F}$ measurable selection $\omega \to \gamma (\omega)$ with $\gamma (\omega) \in \Gamma (\omega)$, for each $\omega$, and also that $t \to \gamma_k (t, \omega)$ (the $k^{th}$ component of $\gamma$) is continuous. Next, we consider the product measurability of $\gamma_k$. We know that $\omega \to \gamma_k (\omega)$ is $\mathcal{F}$ measurable into $C([0, T])$ and since pointwise evaluation is continuous, $\omega \to \gamma_k (t, \omega)$ is $\mathcal{F}$ measurable. (This is nothing more than a case of the general result that a continuous function of a measurable function is measurable.) Then, since $t \to \gamma_k (t, \omega)$ is continuous, it follows that $\gamma_k$ is a $\mathcal{P}$ measurable real valued function and that $\gamma$ is a $\mathcal{P}$ measurable $\mathbb{R}^\infty$ valued function. Since $\gamma (\omega) \in \Gamma (\omega)$, it follows that for each $n, \gamma (\omega) \in \Gamma^n (\omega)$. Therefore, there exists $j_n \geq n$ such that for each $\omega$,

$$
d (f (X_{NC} (\omega) u_{j_n} (\cdot, \omega)), \gamma (\omega)) < 2^{-n}.
$$

Therefore, for a suitable subsequence $\{u_{n(\omega)} (\cdot, \omega)\}$, we have

$$
\gamma (\omega) = \lim_{n(\omega) \to \infty} f \left( X_{NC} (\omega) u_{n(\omega)} (\cdot, \omega) \right),
$$

for each $\omega$. In particular, for each $k$

$$
\gamma_k (t, \omega) = \lim_{n(\omega) \to \infty} f \left( X_{NC} (\omega) u_{n(\omega)} (t, \omega) \right)_k
$$

$$
= \lim_{n(\omega) \to \infty} m_k \int_{l_m (t)}^{t} \langle \phi_{rk}, X_{NC} (\omega) u_{n(\omega)} (s, \omega) \rangle_V \, ds, \tag{30.2.5}
$$

for each $t$.

Note that it is not clear that $(t, \omega) \to f \left( X_{NC} (\omega) u_{n(\omega)} (t, \omega) \right)$ is $\mathcal{P}$ measurable, although $(t, \omega) \to \gamma (t, \omega)$ is $\mathcal{P}$ measurable.

Now here is the proof of the theorem.

Proof of Theorem 30.2.14 By assumption, there exists a further subsequence, still denoted by $n(\omega)$, such that, the weak limit

$$
\lim_{n(\omega) \to \infty} X_{NC} (\omega) u_{n(\omega)} (\cdot, \omega) = v (\cdot, \omega)
$$
exists in \( V \). Then,

\[
m_k \int_{l_{m_k}(t)}^t \langle \phi_{r_k}, v (s, \omega) \rangle_V \, ds
\]

\[
= \lim_{m(\omega) \to \infty} m_k \int_{l_{m_k}(t)}^t \langle \phi_{r_k}, X_{\mathcal{N}C} (\omega) u_{n(\omega)} (s, \omega) \rangle_V \, ds
\]

\[
= \gamma_k (t, \omega), \text{ product measurable.}
\]

Letting \( \phi \in \mathcal{D} \) be given, there exists a subsequence, denoted by \( k \), such that \( m_k \to \infty \) and \( \phi_{r_k} = \phi \). Recall \((m, \phi)\) denoted an enumeration of the pairs \((m, \phi) \in \mathbb{N} \times \mathcal{D} \). For a given \( \phi \in \mathcal{D} \) denote this sequence by \( m_\phi \). Thus we have measurability of

\[
(t, \omega) \to m_\phi \int_{l_{m_\phi}(t)}^t \langle \phi, v (s, \omega) \rangle_V \, ds
\]

for each \( \phi \in \mathcal{D} \).

Now we will be a little more careful about the countable set \( \mathcal{D} \). Iterate the following. Let \( \phi_1 \neq 0 \). Let \( \mathcal{F} \) denote linearly independent subsets of \( V' \) which contain \( \phi_1 \) such that the elements are further apart than \( 1/5 \). Let \( \mathcal{C} \) denote a maximal chain. Thus \( \cup \mathcal{C} \) is also in \( \mathcal{F} \). If \( W := \text{span} \cup \mathcal{C} \) fails to be all of \( V' \), then there would exist \( \psi \notin W \) such that the distance of \( \psi \) to the closed subspace \( W \) is at least \( 1/5 \). Now \( \mathcal{C}, \cup \{ \mathcal{C} \cup \{ \psi \} \} \) would violate maximality of \( \mathcal{C} \). Hence \( W = V' \). Now it follows that \( \mathcal{C} \) must be countable since otherwise, \( V' \) would fail to be separable. Let \( M \) be the rational linear combinations of \( \mathcal{D} \). It must be dense in \( V' \). Note that linear combinations of the \( \phi_i \) are uniquely determined because none is a linear combination of the others. Now we define a linear mapping on \( M \) which makes sense for \((t, \omega)\) on a certain set.

**Definition 30.2.15** Let \( E \) be those points \((t, \omega)\) such that the following limit exists for each \( \phi \in \mathcal{D} \)

\[
\Lambda (t, \omega) \phi \equiv \lim_{m_\phi \to \infty} m_\phi \int_{l_{m_\phi}(t)}^t \langle \phi, v (s, \omega) \rangle_V \, ds
\]

Extend this mapping linearly. That is, for \( \psi \in M, \psi \equiv \sum a_i \phi_i \),

\[
\Lambda (t, \omega) \psi \equiv \sum a_i \Lambda (t, \omega) \phi_i = \sum a_i \left( \lim_{m_\phi \to \infty} m_\phi \int_{l_{m_\phi}(t)}^t \langle \phi_i, v (s, \omega) \rangle_V \, ds \right)
\]

Thus \((t, \omega) \to \Lambda (t, \omega) \psi \) is product measurable being the sum of limits of product measurable functions. Let \( G \) denote those \((t, \omega)\) in \( E \) such that there exists a constant \( C (t, \omega) \) such that for all \( \psi \in M \),

\[
| \Lambda (t, \omega) \psi | \leq C (t, \omega) || \psi ||
\]

**Lemma 30.2.16** \( G \) is product measurable.

**Proof:** This follows from the formula

\[
E \cap G^C = \cap_{n \in \mathbb{N}} \cup_{\psi \in M} \{ (t, \omega) : | \Lambda (t, \omega) \psi | > n || \psi || \}
\]

which is clearly product measurable because \((t, \omega) \to \Lambda (t, \omega) \psi \) is. Thus, since \( E \) is measurable, it follows that \( E \cap G = G \) is also. ■

For \((t, \omega) \in G, \Lambda (t, \omega) \) has a unique extension to all of \( V \), the dual space of \( V' \), still denoted as \( \Lambda (t, \omega) \). By the Riesz representation theorem, for \((t, \omega) \in G, \) there exists \( u (t, \omega) \in V \),

\[
\Lambda (t, \omega) \psi = \langle \psi, u(t, \omega) \rangle_{V', V}
\]

Thus \((t, \omega) \to \Lambda (t, \omega) u (t, \omega) \) is product measurable by the Pettis theorem. Let \( u = 0 \) off \( G \).

Fix \( \omega \). By the fundamental theorem of calculus,

\[
\lim_{m \to \infty} m \int_{l_m(t)}^t v (s, \omega) \, ds = v (t, \omega) \text{ in } V
\]
for a.e. \( t \) say for all \( t \notin N(\omega) \subseteq [0,T] \). Of course we do not know that \( \omega \to v(t,\omega) \) is measurable. However, the existence of this limit for \( t \notin N(\omega) \) implies that for every \( \phi \in V' \),
\[
\lim_{m \to \infty} \left| \int_{l_m(t)}^t \langle \phi, v(s,\omega) \rangle \, ds \right| \leq C(t,\omega) \|\phi\|
\]
for some \( C(t,\omega) \). Here \( m \) does not depend on \( \phi \). Thus, in particular, this holds for a subsequence and so for each \( t \notin N(\omega), (t,\omega) \in G \) because for each \( \phi \in \mathcal{D} \),
\[
\lim_{m_\phi \to \infty} m_\phi \int_{l_{m_\phi}(t)}^t \langle \phi, v(s,\omega) \rangle \, ds \text{ exists and satisfies the above inequality.}
\]
Hence, for all \( \psi \in M \),
\[
\Lambda(t,\omega) \psi = \langle \psi, u(t,\omega) \rangle_{V',V},
\]
where \( u \) is product measurable.

Therefore, for all \( \phi \in M \)
\[
\langle \phi, u(t,\omega) \rangle_{V',V} = \Lambda(t,\omega) \phi \equiv \lim_{m_\phi \to \infty} m_\phi \int_{l_{m_\phi}(t)}^t \langle \phi, v(s,\omega) \rangle \, ds = \langle \phi, v(t,\omega) \rangle_{V',V}
\]
and hence \( u(t,\omega) = v(t,\omega) \). Thus, for each \( \omega \), the product measurable function \( u \) satisfies \( u(t,\omega) = v(t,\omega) \) for a.e. \( t \). Hence \( u(\cdot,\omega) = v(\cdot,\omega) \) in \( V \).  

**Theorem 30.2.17** Let \( V \) be a reflexive separable Banach space with dual \( V' \), and let \( p, p' \) be such that \( p > 1 \) and \( \frac{1}{p} + \frac{1}{p'} = 1 \). Let the functions \( t \to u_n(t,\omega) \), for \( n \in \mathbb{N} \), be in \( L^p([0,T];V) = V \) and \( (t,\omega) \to u_n(t,\omega) \) be \( \mathcal{B}([0,T]) \times \mathcal{F} \equiv \mathcal{P} \) measurable into \( V \). Also let the functions \( t \to y_n(t,\omega) \) be in \( V' \) and \( (t,\omega) \to y_n(t,\omega) \) is \( \mathcal{P} \) measurable into \( V' \). Suppose there is a set of measure zero \( N \subseteq \Omega \) such that if \( \omega \notin N \), then
\[
\sup_{t \in [0,T]} \|y_n(t,\omega)\|_{V'} + \|u_n(\cdot,\omega)\|_V \leq C(\omega),
\]
for all \( n \). (Thus, by weak compactness, for each \( \omega \), each subsequence of \( \{u_n\} \) has a further subsequence that converges weakly in \( V \) to \( v(\cdot,\omega) \in V \). (\( v \) not known to be \( \mathcal{P} \) measurable).) Suppose that each subsequence of \( \{y_n(\cdot,\omega)\} \) has a subsequence which converges weakly in \( V' \) to \( z(\cdot,\omega) \in V' \) such that the function \( t \to z(t,\omega) \) is weakly continuous into \( V' \).

Then, there exist product measurable functions \( u, y \) such that \( t \to u(t,\omega) \) is in \( V \), \( t \to y(t,\omega) \) is weakly continuous into \( V' \) and for each \( \omega \notin N \), a subsequence of \( \mathbb{N} \) denoted by \( \{n(\omega)\} \) such that \( u_{n(\omega)}(\cdot,\omega) \to u(\cdot,\omega) \) weakly in \( V \) and \( y_{n(\omega)}(\cdot,\omega) \) weakly converges to \( y(\cdot,\omega) \) in \( V' \).

When can you obtain such an estimate in Theorem 30.2.17? There is a convenient result which follows.

**Proposition 30.2.18** Let \( \{u_n(\cdot,\omega)\}_{n=1}^\infty \) be a sequence of functions in \( L^p(\Omega, V) \) such that
\[
\sup_n \int_\Omega \|u_n\|_V^p \, dP = C < \infty. \tag{30.2.6}
\]
Then, there is a set \( N \) of measure zero and a subsequence \( \{u_{n,n}(\cdot,\omega)\}_{n=1}^\infty \) such that for all \( \omega \notin N \),
\[
\sup_n \|u_{n,n}(\cdot,\omega)\|_V \leq C(\omega) < \infty.
\]

**Proof:** First, we know that there is a set \( \hat{N} \) of measure zero such that for \( \omega \notin \hat{N}, \|u_n(\cdot,\omega)\|_V^p < \infty \) for all \( n \). Indeed, we just take the union of the exceptional sets, one for each \( n \), we have,
\[
P \left( \left\{ \omega : \liminf_{n \to \infty} \|u_n(\cdot,\omega)\|_V^p \geq M_1 \right\} \right)
\leq \frac{1}{M_1} \int_\Omega \liminf_{n \to \infty} \|u_n(\cdot,\omega)\|_V^p \, dP
\leq \frac{1}{M_1} \liminf_{n \to \infty} \int_\Omega \|u_n(\cdot,\omega)\|_V^p \, dP \leq \frac{C}{M_1}.
\]
Therefore, by choosing a sufficiently large \( M_1 \), we obtain that the set
\[
B_1 \equiv \left\{ \omega : \liminf_{n \to \infty} \| u_n (\cdot, \omega) \|^p_V \geq M_1 \right\}
\]
has measure less than 1/2. Letting \( G_1 \equiv \Omega \setminus B_1 \), it follows that \( P(G_1) > 1/2 \) and for \( \omega \in G_1 \),
\[
\liminf_{n \to \infty} \| u_n (\cdot, \omega) \|^p_V < M_1.
\]
It follows that there is a subsequence \( \{ u_{1,n}(\cdot, \omega) \}_{n=1}^{\infty} \) of \( \{ u_n(\cdot, \omega) \}_{n=1}^{\infty} \) such that for all \( n \) and \( \omega \in G_1 \),
\[
\| u_{1,n} (\cdot, \omega) \|^p_V < M_1.
\]
Next, by the same reasoning as above,
\[
P \left( \left\{ \omega \in B_1 : \liminf_{n \to \infty} \| u_{1,n} (\cdot, \omega) \|^p_V \geq M_2 \right\} \right) \leq \frac{C}{M_2},
\]
and so for a sufficiently large \( M_2 > M_1 \),
\[
B_2 \equiv \left\{ \omega \in B_1 : \liminf_{n \to \infty} \| u_{1,n} (\cdot, \omega) \|^p_V \geq M_2 \right\}
\]
has measure less than 1/4. Let \( G_2 \) be such that \( B_2 \cup G_2 = B_1 \). Then, for \( \omega \in G_2 \) it follows that \( \liminf_{n \to \infty} \| u_{1,n} (\cdot, \omega) \|^p_V < M_2 \). Thus there is a further subsequence \( \{ u_{2,n}(\cdot, \omega) \}_{n=1}^{\infty} \) of \( \{ u_{1,n}(\cdot, \omega) \}_{n=1}^{\infty} \) such that for \( \omega \in G_2 \),
\[
\| u_{2,n} (\cdot, \omega) \|^p_V < M_2.
\]
Continuing this way, there is a sequence of subsequences, the subsequence \( \{ u_{i,n} \}_{n=1}^{\infty} \) being a subsequence of the \( \{ u_{(i-1),n} \}_{n=1}^{\infty} \) such that for all \( n \),
\[
\| u_{i,n} (\cdot, \omega) \|^p_V < M_i \quad \text{if} \quad \omega \in \bigcup_{j=1}^{i} G_i,
\]
where \( \Omega \setminus \bigcup_{j=1}^{i} G_i \equiv B_i \) satisfying \( P(B_i) < 2^{-i} \), \( B_{i+1} \subseteq B_i \). Letting \( N \equiv \cap_i B_i \cup \hat{N} \), it follows that \( P(N) = 0 \). Now, we consider the diagonal sequence \( \{ u_{n,n}(\cdot, \omega) \}_{n=1}^{\infty} \). If \( \omega \not\in N \), then it is in some \( G_i \) and so for all \( n \) sufficiently large, say \( n \geq k \), \( \| u_{n,n} (\cdot, \omega) \|^p_V \leq M_i \). Therefore, for that \( \omega \), it follows that for all \( n \),
\[
\| u_{n,n} (\cdot, \omega) \|^p_V \leq M_i + \max \{ \| u_{m,m} (\cdot, \omega) \|^p_V, m < k \} \equiv C(\omega) < \infty. \]

The following theorem is also useful. It is really a generalization of the familiar Gram Schmidt process. It is Lemma 24.4.2.

**Theorem 30.2.19** Suppose \( V, W \) are separable Banach spaces, such that \( V \) is dense in \( W \) and \( B \in \mathcal{L}(W, W') \) satisfies
\[
\langle Bx, x \rangle \geq 0, \quad \langle Bx, y \rangle = \langle By, x \rangle, \quad B \neq 0.
\]
Then there exists a countable set \( \{ e_i \} \) of vectors in \( V \) such that
\[
\langle Be_i, e_j \rangle = \delta_{ij}
\]
and for each \( x \in W \),
\[
\langle Bx, x \rangle = \sum_{i=1}^{\infty} |\langle Bx, e_i \rangle|^2,
\]
and also
\[
Bx = \sum_{i=1}^{\infty} \langle Bx, e_i \rangle Be_i,
\]
the series converging in \( W' \). In case \( B = B(\omega) \) where \( \omega \to B(\omega) \) is measurable into \( \mathcal{L}(W, W') \), these vectors \( e_i \) will also depend on \( \omega \) and will be measurable functions of \( \omega \).
30.3 Preliminary Results

We use the following well known theorem \([10]\).

**Theorem 30.3.1** Let \(E \subseteq F \subseteq G\) where the injection map is continuous from \(F\) to \(G\) and compact from \(E\) to \(F\). Let \(p \geq 1\), let \(q > 1\), and define

\[
S \equiv \{ u \in L^p ([a,b], E) : \text{for some } C, \| u(t) - u(s) \|_G \leq C |t - s|^{1/q} \}
\]

and \(\| u \|_{L^p([a,b], E)} \leq R\).

Thus \(S\) is bounded in \(L^p ([a,b], E)\) and Holder continuous into \(G\). Then \(S\) is precompact in \(L^p ([a,b], F)\). This means that if \(\{ u_n \}_{n=1}^\infty \subset S\), it has a subsequence \(\{ u_{n_k} \} \) which converges in \(L^p ([a,b], F)\).

We recall the following theorem which is proved in [17] and earlier, Theorem 27.3.1 for what will suffice here.

**Theorem 30.3.2** If \(A\) and \(B\) are pseudo monotone and bounded then \(A + B\) is also pseudo monotone and bounded.

Also the following result, found in [10] is well known.

**Theorem 30.3.3** If a single valued map, \(A : X \to X'\) is monotone, hemi-continuous, and bounded, then \(A\) is pseudo monotone. Furthermore, the duality map, \(J^{-1} : X \to X'\) which satisfies \((J^{-1} f, f) = \| f \|_X^2, \| J^{-1} f \|_X = \| f \|_X\) is strictly monotone hemi-continuous and bounded. So is the duality map \(F : X \to X'\) which satisfies \(\| Ff \|_{X'} = \| f \|_{X}^{-1}, \langle Ff, f \rangle = \| f \|_{X}^{-1} \) for \(p > 1\).

The following fundamental result will be of use in what follows. There is somewhat more in this than will be needed. In this paper, \(B\) is a possibly degenerate operator satisfying only the following:

\[
B \in \mathcal{L}(W,W'), \langle Bu, u \rangle \geq 0, \langle Bu, v \rangle = \langle Bv, u \rangle \quad (30.3.7)
\]

where here \(V \subseteq W\) and \(V\) is dense in \(W\). In the case where \(B = B(\omega)\), we will assume for the sake of simplicity that

\[
B(\omega) = k(\omega) B, \quad k(\omega) \geq 0, k \text{ being } \mathcal{F} \text{ measurable}
\]

Allowing \(B\) to depend on \(\omega\) introduces some technical considerations so if there is no interest in this, simply assume \(B\) is independent of \(\omega\). This includes all cases of most interest.

**Lemma 30.3.4** Suppose \(V, W\) are separable Banach spaces such that \(V\) is dense in \(W\) and \(B \in \mathcal{L}(W,W')\) satisfies

\[
\langle Bx, x \rangle \geq 0, \langle Bx, y \rangle = \langle By, x \rangle, B \neq 0.
\]

Then there exists a countable set \(\{ e_i \}\) of vectors in \(V\) such that

\[
\langle Be_i, e_j \rangle = \delta_{ij}
\]

and for each \(x \in W\),

\[
\langle Bx, x \rangle = \sum_{i=1}^\infty | \langle Bx, e_i \rangle |^2
\]

and also

\[
Bx = \sum_{i=1}^\infty \langle Bx, e_i \rangle Be_i
\]

the series converging in \(W'\). If \(B = B(\omega)\) and \(B\) is \(\mathcal{F}\) measurable into \(\mathcal{L}(W,W')\) and if the \(e_i = e_i(\omega)\) as described above, then these \(e_i\) are measurable into \(V\). If \(t \to B(t, \omega)\) is \(C^1 ([0,T], \mathcal{L}(W,W'))\) and if for each \(w \in W\),

\[
\langle B'(t, \omega) w, w \rangle \leq k_{w,\omega}(t) \langle B(t, \omega) w, w \rangle
\]

Where \(k_{w,\omega} \in L^1 ([0,T])\), then the vectors \(e_i(t)\) can be chosen to also be right continuous functions of \(t\).

The following has to do with the values of \(Bu\) and gives an integration by parts formula.
Corollary 30.3.5 Let \( V \subseteq W, W' \subseteq V' \) be separable Banach spaces, and \( B \in \mathcal{L}(W, W') \) is nonnegative and self adjoint. Also suppose \( t \to B(u(t)) \) has a weak derivative \((Bu)' \in L^p((0, T, V'))\) for \( u \in L^p(0, T, V)\). Then there is a continuous function denoted as \( t \to Bu(t) \) which equals \( B(u(t)) \) a.e. \( t \). Say for \( t \notin N \). Suppose \( Bu(0) = Bu_0, u_0 \in W \). Then

\[
Bu(t) = Bu_0 + \int_0^t (Bu)'(s) \, ds \quad \text{in } V'
\]

(30.3.8)

Then \( t \to Bu(t) \) is in \( C(N^C, W') \) and also for such \( t \),

\[
\frac{1}{2} \langle Bu(t), u(t) \rangle = \frac{1}{2} \langle Bu_0, u_0 \rangle + \int_0^t \langle (Bu)'(s), u(s) \rangle \, ds
\]

There exists a continuous function \( t \to \langle Bu, u \rangle(t) \) which equals the right side of the above for all \( t \) and equals \( \langle B(u(t)), u(t) \rangle \) off \( N \). This also satisfies

\[
\sup_{t \in [0, T]} \langle Bu, u \rangle(t) \leq C \left( \|Bu\|_{L^p(0, T; V')}, \|u\|_{L^p(0, T, V)} \right)
\]

This also makes it easy to verify continuity of pointwise evaluation of \( Bu \). Let \( Lu = (Bu)' \).

\[
d \equiv X \equiv \left\{ u \in L^p(0, T, V) : Lu \equiv (Bu)' \in L^p((0, T, V')) \right\}
\]

(30.3.9)

\[
\|u\|_X \equiv \max \left( \|u\|_{L^p(0, T, V)}, \|Lu\|_{L^p((0, T, V'))} \right)
\]

(30.3.10)

Since \( L \) is closed, this \( X \) is a Banach space.

Then the following theorem is obtained.

Theorem 30.3.6 Say \((Bu)' \in L^p((0, T, V'))\) so

\[
Bu(t) = Bu(0) + \int_0^t (Bu)'(s) \, ds \quad \text{in } V'
\]

the map \( u \to Bu(t) \) is continuous as a map from \( X \) to \( V' \). Also, if \( Y \) denotes those \( f \in L^p([0, T], V) \) for which \( f' \in L^p([0, T], V) \), so that \( f \) has a representative such that \( f(t) = f(0) + \int_0^t f'(s) \, ds \), then if \( \|f\|_Y \equiv \|f\|_{L^p([0, T], V)} + \|f'\|_{L^p([0, T], V)} \), the map \( f \to f(t) \) is continuous.

Also one can obtain the following for \( p > 1 \).

Proposition 30.3.7 Let

\[
X = \left\{ u \in L^p(0, T, V) : Lu = (Bu)' \in L^p((0, T, V')) \right\}
\]

where \( V \) is a reflexive Banach space. Let a norm on \( X \) be given by

\[
\|u\|_X \equiv \max \left( \|u\|_Y, \|Lu\|_{Y'} \right)
\]

Then there is a continuous function \( t \to \langle Bu, v \rangle(t) \) such that \( \langle Bu, v \rangle(t) = \langle B(u(t)), v(t) \rangle \) a.e. \( t \) such that

\[
\sup_{t \in [0, T]} \|Bu, v\|(t) \leq C \|u\|_X \|v\|_X
\]

and if \( K : X \to X' \)

\[
\langle Ku, v \rangle \equiv \int_0^T \langle Lu, v \rangle \, ds + \langle Bu, v \rangle(0)
\]

Then \( K \) is continuous and linear and

\[
\langle Ku, u \rangle = \frac{1}{2} \langle Bu, u \rangle(T) + \langle Bu, u \rangle(0)
\]

If \( u \in X \) and \( Bu(0) = 0 \) then there exists a sequence \( \{u_n\} \) such that \( \|u_n - u\|_X \to 0 \) but \( u_n(t) = 0 \) for all \( t \) close to \( 0 \).
30.4 Measurable Approximate Solutions

The main result in this section is the following theorem. Its proof follows a method due to Brezis and Lions [11] adapted to the case considered here where the operator is set valued. In this theorem, we let $F : V \to V'$ be the duality map $\langle Fu, u \rangle = \|u\|^p, \|Fu\| = \|u\|^{p-1}$ for $p > 1$.

As above, $Lu = (Bu)$. In addition to this, define $\Lambda$ to be the restriction of $L$ to those $u \in X$ which have $Bu(0) = 0$. Thus

$$D(\Lambda) = \{u \in X : Bu(0) = 0\}$$

Then one can show that $\Lambda^*$ is monotone. It is not hard to see that this should be the case. Let $v \in D(\Lambda^*)$ and suppose it is smooth. Then

$$\int_0^T \langle \Lambda u, v \rangle dt = \langle Bu(T), v(T) \rangle - \int_0^T \langle Bu, v' \rangle dt$$

and so, if $\left| \int_0^T \langle \Lambda u, v \rangle dt \right| \leq C \|u\|_V$, then we need to have $v(T) = 0$ and $\Lambda^* v = -Bv'$. Now it is just a matter of doing the computations to verify that

$$\langle \Lambda^* v, v \rangle \geq 0.$$

**Lemma 30.4.1** Let $K$ and $L$ be as in Proposition 30.3.1 and let 30.5.24 - 30.5.27 hold. Then for each $f \in V'$ and $u_0 \in W$, there exists a unique $u \in X$ such that

$$\langle Ku, v \rangle + Fu = \langle f, v \rangle + (Bv(0), u_0)$$

(30.4.11)

for all $v \in X$. Also, the mapping which takes $(f, u_0)$ to this solution is demicontinuous in the sense that if $f_n \to f$ strongly in $V'$ and $u_{n_0} \to u_0$ in $W$, then $u_n \to u$ weakly in $V$.

**Proof:** Let $J^{-1}$ be the duality map mentioned above and define $H_\varepsilon : X \to X'$ by

$$\langle H_\varepsilon(u), v \rangle = \varepsilon \langle Lv, J^{-1}Lu \rangle + \langle Fu, v \rangle + \langle Ku, v \rangle$$

for all $v \in X$. Then $H_\varepsilon$ is pseudomonotone because it is monotone, bounded, and hemicontinuous. This follows from Theorem 30.3.2 and 30.3.3. It is also easy to see that $H_\varepsilon$ is coercive.

$$\frac{\langle H_\varepsilon(u), u \rangle}{\|u\|_X} = \varepsilon \|Lu\|^2_{V'} \|u\|^p_X + \|u\|^p_X + \frac{1}{2} \{\langle Bu(T), u(T) \rangle + \langle Bu, u(0) \rangle\} \frac{1}{\|u\|_X}$$

If not, then there is $\|u_n\|_X \to \infty$ but for some $M$,

$$\varepsilon \frac{\|Lu_n\|^2_{V'}}{\|u_n\|^2_X} + \frac{\|u_n\|^p_{V'}}{\|u_n\|^p_X} + \frac{1}{2} \{\langle Bu_n(T), u_n(T) \rangle + \langle Bu_n, u_n(0) \rangle\} \frac{1}{\|u_n\|_X} \leq M$$

Then one of $\|u_n\|_V$ or $\|Lu_n\|_{V'}$ is unbounded. Either way, a contradiction is obtained. Thus $H_\varepsilon$ is coercive bounded, and pseudomonotone. It follows that it maps onto $X'$.

There exists $u_\varepsilon \in X$ such that for all $v \in X$,

$$\varepsilon \langle Lv, J^{-1}Lu_\varepsilon \rangle + \langle Fu_\varepsilon, v \rangle + \langle Ku_\varepsilon, v \rangle = \langle f, v \rangle + \langle Bv(0), u_0 \rangle.$$  

(30.4.12)

In 30.4.12, let $v = u_\varepsilon$. Using the inequality,

$$\|\langle Bu(0), u_0 \rangle\| \leq \|Bu, v\|^{1/2}(0) \|Bu_0, u_0\|^{1/2}$$

$$\leq \frac{1}{2} \|Bu, v\| + \frac{1}{2} \|Bu_0, u_0\|,$$

it follows that

$$\langle Fu_\varepsilon, u_\varepsilon \rangle + \frac{1}{2} \{\langle Bu_\varepsilon, u_\varepsilon \rangle(T) + \langle Bu_\varepsilon, u_\varepsilon \rangle(0)\}$$

$$\leq \|f\|_{V'} \|u_\varepsilon\|_V + \frac{1}{2} \|Bu_\varepsilon, u_\varepsilon \rangle(0) + \frac{1}{2} \|Bu_0, u_0\|$$

Thus

$$\|u_\varepsilon\|_V^p + \frac{1}{2} \|Bu_\varepsilon, u_\varepsilon \rangle(T) \leq \frac{1}{2} \|Bu_0, u_0\| + \|f\|_{V'} \|u_\varepsilon\|_V,$$
which implies that there exists a constant $C$ independent of $\epsilon$ such that
\[
||u_\epsilon||_\mathcal{V} \leq C. \tag{30.4.13}
\]

Now let $v \in D(\Lambda)$. Thus $v \in X$ and $Bv(0) = 0$ so the last term of (30.4.12) equals 0. The term, $\langle Bu_\epsilon, v \rangle (0)$ found in the definition of $\langle Ku_\epsilon, v \rangle$ also equals 0. This follows from
\[
\langle Bu_\epsilon, v \rangle (0) = \lim_{n \to \infty} \langle Bu_\epsilon, v_n \rangle (0) = 0.
\]

where $v_n = 0$ near 0 and converges to $v$ in $X$ by Proposition 30.4.7. Therefore, for $v \in D(\Lambda)$, a dense subset of $\mathcal{V}$,
\[
\epsilon \langle Av, J^{-1} Lu_\epsilon \rangle + \langle Fu_\epsilon, v \rangle + \langle Lu_\epsilon, v \rangle = \langle f, v \rangle.
\]

It follows that $J^{-1} Lu_\epsilon \in D(\Lambda^*)$ and so for all $v \in D(\Lambda)$,
\[
\epsilon \langle \Lambda^* J^{-1} Lu_\epsilon, v \rangle + \langle Fu_\epsilon, v \rangle + \langle Lu_\epsilon, v \rangle = \langle f, v \rangle. \tag{30.4.14}
\]

Since $D(\Lambda)$ is dense in $\mathcal{V}$, this equation holds for all $v \in \mathcal{V}$ and so in particular, it holds for $v = J^{-1} Lu_\epsilon$. Therefore,
\[
-||Fu_\epsilon||_{\mathcal{V}} ||Lu_\epsilon||_{\mathcal{V}} + ||Lu_\epsilon||_{\mathcal{V}}^2 \leq ||f||_{\mathcal{V}} ||Lu_\epsilon||_{\mathcal{V}}. \tag{30.4.15}
\]

It follows from (30.4.15) that $||Lu_\epsilon||_{\mathcal{V}}$ is bounded independent of $\epsilon$. Therefore, there exists a sequence $\epsilon \to 0$ such that
\[
u_\epsilon \to u \text{ in } \mathcal{V}, \tag{30.4.16}
\]
\[
Ku_\epsilon \to Ku \text{ in } X', \tag{30.4.17}
\]
\[
Fu_\epsilon \to u^* \text{ in } \mathcal{V}', \tag{30.4.18}
\]
\[
Bu_\epsilon (0) \to Bu (0) \text{ in } W'. \tag{30.4.19}
\]

In (30.4.12) replace $v$ with $u_\epsilon - u$. Using $J^{-1}$ is monotone,
\[
\epsilon \langle Lu_\epsilon - Lu, J^{-1} Lu \rangle + \langle Fu_\epsilon + Ku_\epsilon, u_\epsilon - u \rangle
\]
\[
\leq \langle f, u_\epsilon - u \rangle + \langle B(u_\epsilon - u) (0), u_0 \rangle. \tag{30.4.20}
\]

Formula (30.4.12) applied to the last term of (30.4.20) implies
\[
\limsup_{\epsilon \to 0} \langle Fu_\epsilon + Ku_\epsilon, u_\epsilon - u \rangle \leq 0. \tag{30.4.21}
\]

By pseudomonotonicity,
\[
\liminf_{\epsilon \to 0} \langle Fu_\epsilon + Ku_\epsilon, u_\epsilon - u \rangle \geq \langle Fu + Ku, u - v \rangle = 0
\]
so \lim_{\epsilon \to 0} \langle Fu_\epsilon + Ku_\epsilon, u_\epsilon - u \rangle = 0 and so
\[
\langle u^* + Ku, u - v \rangle
\]
\[
\liminf_{\epsilon \to 0} (\langle Fu_\epsilon + Ku_\epsilon, u_\epsilon - u \rangle + \langle Fu_\epsilon + Ku_\epsilon, u - v \rangle) =
\]
\[
\liminf_{\epsilon \to 0} (\langle Fu_\epsilon + Ku_\epsilon, u_\epsilon - v \rangle \geq \langle Ku, u - v \rangle
\]
and so $u^* = Fu$ and from (30.4.12)
\[
\langle Ku, v \rangle + \langle Fu, v \rangle = \langle f, v \rangle + \langle Bu (0), u_0 \rangle. \tag{30.4.22}
\]

Thus for every $v \in X$,
\[
\int_0^T \langle (Bu)'(s), v \rangle ds + \langle Bu, v \rangle (0) + \int_0^T \langle Fu, v \rangle ds = \int_0^T \langle f, v \rangle ds + \langle Bu (0), u_0 \rangle
\]

So let $v$ be smooth and equal to 0 except for $t \in [0, \delta]$ and equals $v_0$ at 0. Then as $\delta \to 0$, the integrals become increasingly small and so
\[
\langle Bu (0), v_0 \rangle = \langle Bv_0, u_0 \rangle = \langle Bu_0, v_0 \rangle
\]
and since \( \nu_0 \) is arbitrary in \( V \), then it follows that \( B u (0) = B u_0 \). Thus this has provided a solution \( u \) to the system
\[
(Bu)' + Fu = f, \quad Bu (0) = B u_0, \ u \in X
\]

It remains to consider the assertion about continuity. First note that the solution to the above initial value problem is unique due to the strict monotonicity of \( F \). In fact, if there are two solutions, \( u, w \), then
\[
\frac{1}{2} \| Bu (t) - Bw (t) \|_V^2 + \int_0^t \langle F u - F w, u - w \rangle \, ds = 0
\]
and so, in particular, \( \langle F u - F w, u - w \rangle_{V', V} = 0 \) which implies \( u = w \) in \( V \).

Let \( u \) be the solution which goes with \( (f, u_0) \) and let \( u_n \) denote the solution which goes with \( (f_n, u_{0n}) \) where it is assumed that \( f_n \to f \) in \( V' \) and \( u_{0n} \to u_0 \) in \( W \). It is desired to show that \( u_n \to u \) weakly in \( V \). First note that the \( u_n \) are bounded in \( V \) because
\[
\frac{1}{2} \langle Bu_n, u_n \rangle (T) - \frac{1}{2} \langle Bu_0, u_0 \rangle + \int_0^T \| u_n \|_V^2 \, ds = \int_0^T \langle f_n, u_n \rangle \, ds \leq \| f_n \|_{V'} \| u_n \|_V
\]
and this clearly implies that \( \| u_n \|_V \) is indeed bounded. Thus if this sequence fails to converge weakly to \( u \), it must be the case that there is a subsequence, still denoted as \( u_n \) which converges weakly to \( w \neq u \) in \( V \). Then by the fact that \( F \) is bounded, there is an estimate of the form
\[
\| u_n \|_V + \| L u_n \|_{V'} \leq C
\]
Thus, a further subsequence satisfies
\[
\begin{align*}
&u_n \to w \text{ weakly in } V \\
&L u_n \to w \text{ weakly in } V' \\
&F u_n \to \xi \text{ weakly in } V'
\end{align*}
\]
then
\[
\begin{align*}
\int_0^T \langle (B (u_n - w))', u_n - w \rangle \, dt &= \frac{1}{2} \langle B (u_n - w), (u_n - w) \rangle (T) - \frac{1}{2} \langle B (u_n - w), (u_n - w) \rangle (0) \\
&\geq - \frac{1}{2} \langle B (u_n - w), (u_n - w) \rangle (0) = - \frac{1}{2} \langle B (u_{0n} - u_0), u_{0n} - u_0 \rangle
\end{align*}
\]
It follows
\[
\langle Lu, u \rangle_V + \langle Fu, u \rangle_{V'} - \frac{1}{2} \langle B (u_{0n} - u_0), u_{0n} - u_0 \rangle \leq \langle f_n, u_n - w \rangle
\]
and so \( \limsup_{n \to \infty} \langle Fu_n, u_n - w \rangle \leq 0 \). Then as before, \( \xi = Fw \) and one obtains
\[
(Bw)' + Fw = f, \quad Bw (0) = B u_0, \ w \in X
\]
contradicting uniqueness. Hence \( u_n \to u \) weakly as claimed.

Now suppose \((\Omega, \mathcal{F})\) is a measurable space and \( B = B (\omega) \) and is measurable into \( L (W, W') \) and \( f : [0, T] \times \Omega \to V' \) is product measurable, \( B ([0, T]) \times \mathcal{F} \text{ measurable where } B ([0, T]) \) denotes the Borel sets. Also, it is assumed that for each \( \omega, f (\cdot, \omega) \in V' \). The following lemma ties together these ideas.

**Lemma 30.4.2** Let \( f (\cdot, \omega) \in V' \). Then if \( \omega \to f (\cdot, \omega) \) is measurable into \( V' \), it follows that for each \( \omega \), there exists a representative \( \hat{f} (\cdot, \omega) \in V' \), \( \hat{f} (\cdot, \omega) = f (\cdot, \omega) \) in \( V' \) such that \( (t, \omega) \to \hat{f} (t, \omega) \) is product measurable. If \( f (\cdot, \omega) \in V' \) and \( (t, \omega) \to f (t, \omega) \) is product measurable, then \( \omega \to f (\cdot, \omega) \) is measurable into \( V' \). The same holds replacing \( V' \) with \( V \).

**Proof:** If a function \( f \) is measurable into \( V' \), then there exist simple functions \( f_n \)
\[
\lim_{n \to \infty} \| f_n (\omega) - f (\omega) \|_{V'} = 0, \quad \| f_n (\omega) \| \leq 2 \| f (\omega) \|_{V'} \equiv C (\omega)
\]
Now one of these simple functions is of the form
\[ \sum_{i=1}^{M} c_i \mathcal{X}_{E_i}(\omega) \]
where \( c_i \in \mathcal{V}' \). Therefore, there is no loss of generality in assuming that \( c_i(t) = \sum_{j=1}^{N} d_{ij} \mathcal{X}_{F_j}(t) \) where \( d_{ij} \in \mathcal{V}' \). Hence we can assume each \( f_n \) is product measurable into \( \mathcal{B}(\mathcal{V}') \times \mathcal{F} \). Then by Theorem \[ \text{Lemma 30.4.3} \], there exists \( \hat{f}(\cdot, \omega) \in \mathcal{V}' \) such that \( f \) is product measurable and a subsequence \( f_{n(\omega)} \) converging weakly in \( \mathcal{V}' \) to \( \hat{f}(\cdot, \omega) \) for each \( \omega \). Thus \( f_{n(\omega)}(\omega) \rightarrow f(\omega) \) strongly in \( \mathcal{V}' \) and \( f_{n(\omega)}(\omega) \rightarrow \hat{f}(\omega) \) weakly in \( \mathcal{V}' \). Therefore, \( \hat{f}(\omega) = f(\omega) \) in \( \mathcal{V}' \) and so it can be assumed that if \( f \) is measurable into \( \mathcal{V}' \) then for each \( \omega \), it has a representative \( \hat{f}(\omega) \) such that \( (t, \omega) \rightarrow \hat{f}(t, \omega) \) is product measurable.

If \( f \) is product measurable into \( \mathcal{V}' \) and each \( f(\cdot, \omega) \in \mathcal{V}' \), does it follow that \( f \) is measurable into \( \mathcal{V}' \)? By measurability, \( f(t, \omega) = \lim_{n \to \infty} \sum_{i=1}^{m_n} e_i^n \mathcal{X}_{E_i^n}(t, \omega) = \lim_{n \to \infty} f_n(t, \omega) \) where \( E_i^n \) is product measurable into \( \mathcal{V}' \). Then by product measurability, \( \omega \rightarrow f_n(\cdot, \omega) \) is measurable into \( \mathcal{V}' \) because if \( g \in \mathcal{V} \),
\[ \omega \rightarrow \langle f_n(\cdot, \omega), g \rangle \]
is of the form
\[ \omega \rightarrow \sum_{i=1}^{m_n} \int_{0}^{T} \langle e_i^n \mathcal{X}_{E_i^n}(t, \omega), g(t) \rangle \, dt \]
which is \( \omega \rightarrow \sum_{i=1}^{m_n} \int_{0}^{T} \langle e_i^n, g(t) \rangle \mathcal{X}_{E_i^n}(t, \omega) \, dt \)
and this is \( \mathcal{F} \) measurable since \( E_i^n \) is product measurable. Thus, it is measurable into \( \mathcal{V}' \) as desired. Obviously, the conclusion is the same for these two conditions if \( \mathcal{V}' \) is replaced with \( \mathcal{V} \).

Now consider the initial value problem
\[
(B(\omega) u(\cdot, \omega))' + F u(\cdot, \omega) = f(\cdot, \omega), \\
B(\omega) u(0, \omega) = B(\omega) u_0(\omega), \ u(\cdot, \omega) \in X
\]
(30.4.23)
where we also assume \( u_0 \) is \( \mathcal{F} \) measurable into \( W \). From Lemma \[ \text{Lemma 30.4.2} \], \( \omega \rightarrow (f(\cdot, \omega), u_0(\omega)) \) is measurable into \( \mathcal{V}' \times \mathcal{W} \). That is,
\[ (f, u_0)^{-1}(U) = \{ \omega : (f(\cdot, \omega), u_0(\omega)) \in U \} \in \mathcal{F} \]
for \( U \) an open set in \( \mathcal{V}' \times \mathcal{W} \). From Lemma \[ \text{Lemma 30.4.1} \], the map \( \Phi_{\omega} \) which takes \( (f, u_0) \) to the solution \( u \) is demicontinuous. We desire to argue that \( u \) is measurable into \( \mathcal{V} \). In doing so, it is easiest to assume that \( B \) does not depend on \( \omega \). However, the dependence on \( \omega \) can be included by using the approximation assumption for \( B(\omega) \) mentioned earlier.

Letting \( f_n(\cdot, \omega) \rightarrow f(\cdot, \omega) \) where \( f_n \) is a simple function and \( u_{n0}(\cdot, \omega) \rightarrow u_0(\omega) \) where \( u_{n0} \) is also a simple function, it follows that
\[ \Phi_{\omega}(f_n(\cdot, \omega), u_{n0}(\omega)) \rightarrow \Phi_{\omega}(f(\cdot, \omega), u_0(\omega)) = u \]
weakly.

**Lemma 30.4.3** Suppose \( f(\cdot, \omega) \in \mathcal{V}' \) for each \( \omega \) and that \( (t, \omega) \rightarrow f(t, \omega) \) is product measurable into \( \mathcal{V}' \). Also \( u_0 \) is \( \mathcal{F} \) measurable into \( \mathcal{W} \) and
\[ B(\omega) = k(\omega) B, \ k(\omega) \geq 0, \ k \text{ measurable} \]
Then for each \( \omega \in \Omega \), there exists a unique solution \( u(\cdot, \omega) \) in \( \mathcal{V} \) satisfying
\[
(B(\omega) u(\cdot, \omega))' + F u(\cdot, \omega) = f(\cdot, \omega), \\
B(\omega) u(0, \omega) = B(\omega) u_0(\omega), \ u(\cdot, \omega) \in X
\]
This solution has a representative which satisfies \( (t, \omega) \rightarrow u(t, \omega) \) is product measurable into \( \mathcal{V} \).

**Proof:** Let \( B_n(\omega) \equiv k_n(\omega) B \) where \( \{ k_n(\omega) \} \) is an increasing sequence of simple functions converging pointwise to \( k(\omega) \). Replace \( B(\omega) \) with \( B_n(\omega) \). Then define
\[ \langle K_n u, v \rangle \equiv \int_{0}^{T} \langle L_n u, v \rangle \, ds + \langle B u, v \rangle(0) \]
where \( L_n \) is defined as
\[ L_n u = (B_n(\omega) u)' \]
for $B_n$ having values in $\mathcal{L}(W,W')$ such that $B_n(\omega) \to B(\omega)$ and each of these is self-adjoint and nonnegative. Let $u_n$ be the solution to the above initial value problem

$$\langle K_n u_n, v \rangle + Fu_n = \langle f_n, v \rangle + \langle B(0), u_n \rangle$$

in which $u_{0n}$ and $f_n$ are simple functions converging to $u_0$ and $f$ in $W$ and $W'$ respectively for each $\omega$. Thus these have constant values in $W'$ or $W$ on finitely many measurable subsets of $\Omega$. Since $B_n$ is constant on measurable sets, it follows that $u_n(\cdot,\omega)$ is also a constant element of $V$ on each of finitely many measurable sets. Hence $u_n(\cdot,\omega)$ is measurable into $V$. Then fixing $\omega$, and so in fact, $u_n(\cdot,\omega)$.

Thus, by monotonicity,

$$\langle k(\omega) (B_n) \nu, u_n - u \rangle_{V', V} + \langle F u_n, u_n - u \rangle_{V', V} = \langle f_n, u_n - u \rangle_{V', V}$$

$$\geq \langle k(\omega) (B) \nu, u_n - u \rangle_{V', V} + \langle (k(\omega) - k(\omega)) (B_n) \nu, u_n - u \rangle_{V', V}$$

The last term in the above expression converges to 0 due to the convergence of $k(\omega)$ to $k(\omega)$. Thus

$$\langle (B(\omega) \nu, u_n - u \rangle_{V', V} + \langle F u_n, u_n - u \rangle_{V', V} \leq \langle f_n, u_n - u \rangle_{V', V}$$

and so

$$\lim_{n \to \infty} \sup_{V' \in V} \langle f u_n, u_n - u \rangle_{V', V} \leq 0$$

Then as before, one can conclude that $F u = \xi$. Then passing to the limit gives the desired solution to the equation, this for each $\omega$. However, by uniqueness, it follows that if $\bar{u}$ is the solution to the evolution equation of Lemma 30.4.2, then for each $\omega$, $u = \bar{u}$ in $V$. Also this $u$ just obtained is measurable into $V$ thanks to the Pettis theorem. Therefore, $\bar{u}$ can be modified on a set of measure zero for each fixed $\omega$ to equal $u$ a function measurable into $V$. Hence there exists a solution to the evolution equation of this lemma $u$ which is measurable into $V$. By the Lemma 30.4.3, it follows that there is a representative for $u$ which is product measurable into $V$. ■
30.5 The Main Result

The main result is an existence theorem for product measurable solutions to the system

\[
(B(\omega) u(\cdot, \omega))' + u^*(\cdot, \omega) = f(\cdot, \omega) \text{ in } V'
\]

\[
B(\omega) u(0, \omega) = B(\omega) u_0(\omega)
\]

where \(u^*(\cdot, \omega) \in A(\cdot, \omega), \omega\). It is Theorem 30.5.3 below. First are some assumptions.

Here \(I\) will denote a subinterval of \([0, T]\), of the form \(I = [0, \hat{T}], \hat{T} \leq T\), and \(\mathcal{V}_I \equiv L^p(I, V)\) with similar things defined analogously. We assume only that \(p > 1\).

**Definition 30.5.1** For \(X\) a reflexive Banach space, we say \(A : X \to \mathcal{P}(X')\) is pseudomonotone and bounded if the following hold.

1. The set \(Au\) is nonempty, closed and convex for all \(u \in X\). \(A\) takes bounded sets to bounded sets.

2. If \(u_i \to u\) weakly in \(X\) and \(u_i^* \in Au_i\) is such that

\[
\lim \sup_{i \to \infty} \langle u_i^*, u_i - u \rangle \leq 0,
\]

then, for each \(v \in X\), there exists \(u^*(v) \in Au\) such that

\[
\lim \inf_{i \to \infty} \langle u_i^*, u_i - v \rangle \geq \langle u^*(v), u - v \rangle.
\]

Now suppose the following for the operator \(A(\cdot, \omega) : \mathcal{V}_I \to \mathcal{V}_I'\) for each \(I\) a subinterval of \([0, T]\) and \(A(\cdot, \omega) : \mathcal{V}_I \to \mathcal{P}(\mathcal{V}_I')\) is bounded,

\[
1^* \in A(u, \omega)
\]

for each \(T\) in an increasing sequence converging to \(T\), then

\[
1^* \in A(u, \omega)
\]

Assume the specific estimate

\[
\sup \left\{ \| u^* \|_{\mathcal{V}_I} : u^* \in A(u, \omega) \right\} \leq a(\omega) + b(\omega) \| u \|_{\mathcal{V}_I}^{p-1}
\]

where \(a(\omega), b(\omega)\) are nonnegative. Also assume the coercivity condition:

\[
\lim_{\| u \|_{\mathcal{V}_I} \to \infty} \inf_{u \in \mathcal{V}_I} \left\{ 2 \langle u^*, u \rangle_{\mathcal{V}_I, \mathcal{V}_I} + \langle Bu, u \rangle(T) : u^* \in A(u, \omega) \right\} = \infty,
\]

or alternatively the following specific estimate valid for each \(t \leq T\) and for some \(\lambda(\omega) \geq 0\),

\[
\inf \left( \int_0^t \langle u^*, u \rangle + \lambda(\omega) \langle Bu, u \rangle dt : u^* \in A(u, \omega) \right) \geq \delta(\omega) \int_0^t \| u \|_{\mathcal{V}_I}^p ds - m(\omega)
\]

where \(m(\omega)\) is some nonnegative constant, \(\delta(\omega) > 0\). Note that the estimate is a coercivity condition on \(\lambda B + A\) rather than on \(A\) but is more specific than \(\| u \|_{\mathcal{V}_I}^p\).

Let \(U\) be a Banach space dense in \(V\) and that if \(u_i \to u\) in \(\mathcal{V}_I\) and \(u_i^* \in A(u_i)\) with \(u_i^* \to u^*\) in \(\mathcal{V}_I'\) and \((Bu_i)' \to (Bu)'\) in \(U'_I\), \(\to\) denoting weak convergence, then if

\[
\lim \sup_{i \to \infty} \langle u_i^*, u_i - u \rangle_{\mathcal{V}_I', \mathcal{V}_I} \leq 0
\]

it follows that for all \(v \in \mathcal{V}_I\), there exists \(u^*(v) \in Au\) such that

\[
\lim \inf_{i \to \infty} \langle u_i^*, u_i - v \rangle_{\mathcal{V}_I', \mathcal{V}_I} \geq \langle u^*(v), u - v \rangle_{\mathcal{V}_I', \mathcal{V}_I}
\]
where $r > \max(p,2)$, and we replace $p$ with $r$ and $I$ an arbitrary subinterval of the form $[0,\hat{T}], \hat{T} < T$, for $[0,T]$, and $U$ for $V$ where indicated. Here

$$U_{t I} \equiv L^r(I;U)$$

Note that we are not assuming $A$ is pseudomonotone, just that it satisfies a similar limit condition. Typically, this limit condition holds because of a use of the compact embedding of theorem 30.3.4 or similar result and it does not matter whether $U$ is a small subset of $V$ as long as it is dense in $V$.

Here is an alternate limit condition. Let $U$ be a Banach space dense in $V$ and that if $u_i \to u$ in $V_I$ and $u_i^* \in A(u_i)$ with $u_i^* \to u^*$ in $V_I'$ and $t \to Bu_i(t)$ is continuous and measurable and for each $\omega$,

$$\sup_i \sup_{t \neq s} \frac{\|Bu_i(t) - Bu_i(s)\|_{V'}}{|t - s|^\alpha} \leq C$$

then if

$$\lim_{i \to \infty} \sup_{t \neq s} \langle u_i^*, u_i - u \rangle_{V_I'} \leq 0$$

it follows that for all $v \in V_I$, there exists $u^*(v) \in Au$ such that

$$\liminf_{i \to \infty} \langle u_i^*, u_i - v \rangle_{V_I'} \geq \langle u^*(v), u - v \rangle_{V_I'}$$

This alternate condition is implied by (30.5.32), but the conditions under which either condition holds are likely to depend on some sort of compactness which will be usable for either limit condition. Technically if you assume this alternate condition, you are assuming more, but I don’t have any examples to show that it would be actually assuming more.

For $\omega \to u(\cdot,\omega)$ measurable into $V$,

$$\omega \to A(u(\cdot,\omega),\omega) \text{ has a measurable selection into } V'.$$

This last condition means there is a function $\omega \to u^*(\omega)$ which is measurable into $V'$ such that $u^*(\omega) \in A(u(\cdot,\omega),\omega)$. This is assured to take place if the following standard measurability condition is satisfied for all $O$ open in $V$:

$$\{\omega : A(u(\cdot,\omega),\omega) \cap O \neq \emptyset\} \in \mathcal{F}$$

See for example [8, 9]. Our assumption is implied by this one but they are not equivalent.

Note that this condition would hold if $u \to A(t,u,\omega)$ is bounded and pseudomonotone as a single valued map from $V$ to $V'$ and $(t,\omega) \to A(t,u,\omega)$ is product measurable into $V'$. One would use the demicontinuity of $u \to A(\cdot,u,\omega)$ which comes from the pseudo monotone and bounded assumption and consider a sequence of simple functions $u_n(t,\omega) \to u(t,\omega)$ in $V$ for $u$ measurable, each $u(\cdot,\omega)$ being in $V$. Then the measurability of $A(t,u_n,\omega)$ would attach to $A(t,u,\omega)$ in the limit. More generally, here is a useful lemma.

**Lemma 30.5.2** Suppose $\omega \to A(u(\omega),\omega)$ has a measurable selection in $V'$ for $u$ a given element of $V$ not dependent on $\omega$ and for each $\omega$, $A(u,\omega)$ is a closed bounded, convex set in $V'$. Also suppose $u \to A(u,\omega)$ is upper semicontinuous from the strong topology of $V$ to the weak topology of $V'$. That is, if $u_n \to u$ in $V$ strongly, then if $O$ is a weakly open set containing $A(u,\omega)$, it follows that $A(u_n,\omega) \in O$ for all $n$ large enough. Then whenever $u$ is measurable into $V$, it follows that there is a measurable selection for $\omega \to A(u(\omega),\omega)$ into $V'$.

**Proof:** Let $\omega \to u(\omega)$ be measurable into $V$ and let $u_n(\omega) \to u(\omega)$ in $V$ where $u_n$ is a simple function

$$u_n(\omega) = \sum_{k=1}^{m_n} c_k^n \chi_{E_k^n}(\omega), \text{ the } E_k^n \text{ disjoint, } \Omega = \bigcup_k E_k^n,$$

each $c_k^n$ being in $V$. Then by assumption, there is a measurable selection for $\omega \to A(c_k^n,\omega)$ denoted as $\omega \to y_k^n(\omega)$. Thus $\omega \to y_k^n(\omega)$ is measurable into $V'$ and $y_k^n(\omega) \in A(c_k^n,\omega)$ for all $\omega \in \Omega$. Then consider

$$y^n(\omega) = \sum_{k=1}^{m_n} y_k^n(\omega) \chi_{E_k^n}(\omega)$$

It is measurable and for $\omega \in E_k^n$ it equals $y_k^n(\omega) \in A(c_k^n,\omega) = A(u_n(\omega),\omega)$. Thus $y^n$ is a measurable selection of $\omega \to A(u_n(\omega),\omega)$. By the estimates, for each $\omega$ these $y^n(\omega)$ lie in a bounded subset of $V'$. The bound might depend
Thus there is a measurable into \( \mathcal{V}' \) function \( \omega \to y(\omega) \) and a subsequence for each \( \omega, y^{n(\omega)}(\omega) \) such that \( y^{n(\omega)}(\omega) \to y(\omega) \) weakly in \( \mathcal{V}' \). By the Pettis theorem, \( y \) is measurable into \( \mathcal{V}' \). Where is \( y(\omega) \) if \( y(\omega) \not\in A(u(\omega), \omega) \), then there would exist \( z(\omega) \in \mathcal{V} \) such that \( \langle y(\omega), z \rangle > r > \langle w, z \rangle \) for all \( w \in A(u(\omega), \omega) \). Let \( O = \{ w \in \mathcal{V} \text{ such that } r > \langle w, z \rangle \} \). Then \( O \) contains \( A(u(\omega), \omega) \) and is a weakly open set. It follows from the upper semicontinuity assumption that \( y^{n(\omega)}(\omega) \in O \) for all \( n(\omega) \) large enough. Thus \( r > \langle y^{n(\omega)}(\omega), z \rangle \). But by weak convergence,

\[
\langle y(\omega), z \rangle > r \geq \lim_{n(\omega) \to \infty} \langle y^{n(\omega)}(\omega), z \rangle = \langle y(\omega), z \rangle
\]

contradicting \( y(\omega) \notin A(u(\omega), \omega) \). Hence \( y(\omega) \in A(u(\omega), \omega) \) and \( \omega \to y(\omega) \) is a measurable selection. ■

Note that \( \mathcal{V} \) could be replaced with \( L^p(I, \mathcal{V}) \) where \( I \) is any interval and nothing changes.

If \( u \to A(u, \omega) \) is upper semicontinuous into the strong topology of \( \mathcal{V}' \) and \( \omega \to A(u, \omega) \) is measurable, then the conclusion of the above lemma holds also. This is because weakly open sets are also strongly open. Hence if the conclusion holds for all strongly open sets containing \( A(u, \omega) \), then it holds for all weakly open sets containing \( A(u, \omega) \).

The condition leading to \( 30.5.28 \) will typically be satisfied. For example, suppose \( u^* \in A(u, \omega) \) means that

\[
u^*(t) \in A(t, u(t), \int_0^t u(s) \, ds, \omega)
\]

for a.e. \( t \) where \( A \) has values in \( \mathcal{P}(\mathcal{V}') \). Then to say that \( u^* \mathcal{X}_{[0,T]} \in A(t, u\mathcal{X}_{[0,T]}, \omega) \) for each \( T \) in an increasing sequence converging to \( T \) would imply the above holding for a.e. \( t \). While the above is the typical situation one would expect to see, the following proposition is also interesting.

**Proposition 30.5.3** Suppose \( A(\cdot, \omega) : \mathcal{V} \to \mathcal{P}(\mathcal{V}') \) is upper semicontinuous from the strong topology of \( \mathcal{V} \) to the weak topology of \( \mathcal{V}' \) and has closed convex values. Then if

\[
u^* \mathcal{X}_{[0,T]} \in A(t, u\mathcal{X}_{[0,T]}, \omega)
\]

for each \( T \) in an increasing sequence converging to \( T \), then

\[
u^* \in A(u, \omega)
\]  

**Proof:** Let \( T_n \uparrow T \) such that \( u^* \mathcal{X}_{[0,T_n]} \in A(t, u\mathcal{X}_{[0,T_n]}, \omega) \). Then if \( u^* \notin A(u, \omega) \), there exists \( z \in \mathcal{V} \) such that

\[
\langle u^*, z \rangle > r > \langle w^*, z \rangle
\]

for all \( w^* \in A(u, \omega) \). Now \( u^* \mathcal{X}_{[0,T_n]} \to u^* \) in \( \mathcal{V}' \) and \( u\mathcal{X}_{[0,T_n]} \to u \) in \( \mathcal{V} \). Letting \( O \) be the weakly open set, \( \{ z^*: \langle z^*, z \rangle < r \} \), it follows that this \( O \) is a weakly open set which contains \( A(u, \omega) \). Hence, by upper semicontinuity, \( \langle u^* \mathcal{X}_{[0,T_n]}, z \rangle \to \langle u^*, z \rangle < r \) for all \( n \) large enough. Hence, passing to a limit, one obtains \( \langle u^*, z \rangle > r \geq \langle u^*, z \rangle \), a contradiction. Thus \( u^* \in A(u, \omega) \). ■

Let \( r > \max(p, 2) \). Let \( \mathcal{U} \) and \( \mathcal{U}_1 \) be defined by analogy with \( \mathcal{V} \) and \( \mathcal{V}_1 \) where \( \mathcal{U} = L^r([0, T], \mathcal{U}_1) \). Here \( \mathcal{U} \) is a Hilbert space which is dense in \( \mathcal{V} \) and embeds compactly into \( \mathcal{V} \). \( \| u \|_\mathcal{U} \geq \| u \|_{\mathcal{V}} \). Also let \( F : \mathcal{U} \to \mathcal{V}' \) be the duality map for \( r \). Thus

\[
\| F u \|_{\mathcal{V}'} = \| u \|_{\mathcal{V}}^{r-1}, \quad \langle F u, u \rangle = \| u \|_{\mathcal{V}}^r
\]

Also define the following notation for small positive \( h \).

\[
\tau_h g(t) = \begin{cases} g(t-h) & \text{if } t > h \\ 0 & \text{if } t \leq h \end{cases}
\]

Let \( \omega \to u_0(\omega) \) be \( \mathcal{F} \) measurable into \( \mathcal{W} \). Also let \( \omega \to f(\cdot, \omega) \) be \( \mathcal{F} \) measurable into \( \mathcal{V}' \), \( \omega \to B(\omega) \) measurable into \( \mathcal{L}(\mathcal{W}, \mathcal{W}') \). Now let \( u_h \) for \( h > 0 \) and small, be the unique solution to the initial value problem

\[
\begin{align*}
(B(\omega) u_h (\cdot, \omega))' + \varepsilon F u_h (\cdot, \omega) &= f(\cdot, \omega) - u_h^k (\cdot, \omega) \quad \text{in } \mathcal{U}', \\
B u_h (0, \omega) &= B u_0 (\omega)
\end{align*}
\]  

(30.5.39)
where \( u_h \in A(\tau_h u, \omega) \) is a \( F \) measurable selection into \( V' \). Since \( F \) is monotone bounded and hemicontinuous, there is no problem with it being pseudomonotone from \( X_{\tau} \) to \( X_{\tau}' \). Such a solution exists on \([0, h] \) by the above reasoning. Let this solution be denoted by \( u_1 \). Then use it to define a solution to the evolution equation on \([0, 2h] \) called \( u_2 \). By uniqueness, these coincide on \([0, h] \). Then use \( u_2 \) to extend to a solution on \([0, 3h] \) called \( u_3 \). Then \( u_3 = u_2 \) on \([0, 2h] \). Continue this way to obtain a solution valid on \([0, T] \). By Lemma 30.5.3, this solution may be assumed to be measurable into \( U' \). One gets this by using the lemma on a succession of intervals \([0, h], [0, 2h], \) and so forth.

Now acting on \( u_h \) and suppressing the dependence on \( \omega \) in most places, it follows from the assumed estimates that

\[
\frac{1}{2} \langle Bu_h, u_h \rangle (T) - \frac{1}{2} \langle Bu_0, u_0 \rangle + \varepsilon \int_0^T \| u_h \|_{V'}^p \, ds \leq \left( \frac{\int_0^T \| f \|_{V'}^p \, ds}{\int_0^T \| u_h \|_{V'}^p} \right)^{1/p} \left( \int_0^T (a + b \| \tau_h u_h \|_{V'}^{p-1}) \| u_h \|_{V'} \, ds \right)^{1/p}
\]

\[
+ \int_0^T \left( a + b \| \tau_h u_h \|_{V'}^{p-1} \right) \| u_h \|_{V'} \, ds \leq \| f \|_{V'}^p + \| u_h \|_{V'}^p + \| u_h \|_{V'}^p + a T^{1/p} + b \| u_h \|_{V'}^p
\]

which is of the form

\[
\leq C \left( \| f \|_{V'}, a(\omega) T \right) + (2 + b) \| u_h \|_{V'}^p
\]

Now here is where it is good that \( p < r \).

\[
\| u_h \|_{V'}^p \leq \int_0^T \frac{1}{\delta} \| u_h \|_{U'}^p \, ds \leq \left( \int_0^T \frac{1}{\delta^p} \| u_h \|_{U'}^p \, ds \right)^{p/r} \left( \int_0^T \frac{1}{\delta^p} \| u_h \|_{U'}^r \, ds \right)^{(r-p)/r}
\]

\[
\leq \frac{1}{\delta^{r(r-p)/r}} T^{(r-p)/r} \left( r - p \right) + \frac{C \delta^{r(p) \| u_h \|_{U}^p}}{r}
\]

Thus this has shown

\[
\frac{1}{2} \langle Bu_h, u_h \rangle (T) - \frac{1}{2} \langle Bu_0, u_0 \rangle + \varepsilon \| u_h \|_{U}^r \leq C \left( \| f \|_{V'}, a(\omega) + 1 \right) T, \varepsilon \right) + \langle Bu_0, u_0 \rangle
\]

Then for \( \delta \) small enough, depending on \( \varepsilon \),

\[
\frac{p}{r} \delta^{r/p} < \frac{\varepsilon}{2}
\]

And so the inequality ending at 30.5.40 yields

\[
\langle Bu_h, u_h \rangle (T) + \varepsilon \| u_h \|_{U}^r \leq C \left( \| f \|_{V'}, a(\omega) + 1 \right) T, \varepsilon \right) + \langle Bu_0, u_0 \rangle
\]

From 30.5.34 and the boundedness of the various operators, \( (B(\omega) u_h (\cdot, \omega))' \) is bounded in \( U' \). Thus, summarizing these estimates yields

\[
\| (B(\omega) u_h (\cdot, \omega))' \|_{U'} + \| u_h \|_{U} + \| u_h \|_{V'} \leq C
\]

(30.5.41)

where \( C \) does not depend on \( h \) although it does depend on \( \varepsilon \) and of course on \( \omega \). Then one can get a subsequence, still denoted with \( h \) such that as \( h \to 0 \),

\[
u_h \to u \text{ weakly in } U
\]

(30.5.42)

\[
\tau_h u_h \to u \text{ weakly in } U
\]

(30.5.43)

\[
(B(\omega) u_h (\cdot, \omega))' \to (B(\omega) u)' \text{ weakly in } U'
\]

(30.5.44)

\[
u_h \to u \text{ strongly in } V
\]

(30.5.45)

\[
u_h \to u^* \text{ weakly in } V'
\]

(30.5.46)
\[ \begin{align*}
F_{u_h} & \to \xi \in \mathcal{U}' \\
B u(0, \omega) &= B_{u_0}(\omega)
\end{align*} \]  

(30.5.47)  

(30.5.48)

The fourth of these comes from a use of Theorem 30.5.45. We need to argue that \( u^* \in A(u, \omega) \). From the equation and initial conditions of (30.5.46), it follows from monotonicity conditions and the observation that \( \mathcal{V}' \) is contained in \( \mathcal{U}' \) that

\[
( (B(\omega) u_h(\cdot, \omega))', u_h - u ) + \langle \varepsilon F u_h(\cdot, \omega), u_h - u \rangle + \langle u^*_h(\cdot, \omega), u_h - u \rangle = \langle f(\cdot, \omega), u_h - u \rangle
\]

and so

\[
( (B(\omega) u(\cdot, \omega))', u_h - u ) + \langle \varepsilon F u_h(\cdot, \omega), u_h - u \rangle_{\mathcal{U}', \mathcal{U}} + \langle u^*_h(\cdot, \omega), u_h - u \rangle_{\mathcal{V}', \mathcal{V}} \leq \langle f(\cdot, \omega), u_h - u \rangle
\]

by the strong convergence of (30.5.46); it follows that the third term converges to 0 as \( h \to 0 \). This is because the estimate (30.5.45) implies that the \( u^*_h \) are bounded, and then the strong convergence gives the desired result. Hence

\[
\lim_{h \to 0} \sup \langle \varepsilon F u_h(\cdot, \omega), u_h - u \rangle_{\mathcal{U}', \mathcal{U}} \leq 0
\]

and since \( F \) is monotone and hemicontinuous, it follows that in fact,

\[
\lim_{h \to 0} \langle \varepsilon F u_h(\cdot, \omega), u_h - u \rangle_{\mathcal{U}', \mathcal{U}} = 0
\]

so for \( v \in \mathcal{U} \) arbitrary,

\[
\langle \varepsilon \xi, u - v \rangle = \lim_{h \to 0} \left( \langle \varepsilon F u_h(\cdot, \omega), u - u_h \rangle_{\mathcal{U}', \mathcal{U}} + \langle \varepsilon F u_h(\cdot, \omega), u_h - v \rangle_{\mathcal{U}', \mathcal{U}} \right)
\]

\[
= \lim_{h \to 0} \langle \varepsilon F u_h(\cdot, \omega), u_h - v \rangle_{\mathcal{U}', \mathcal{U}} \geq \langle F u, u - v \rangle
\]

and so, since \( v \) is an arbitrary element of \( \mathcal{U} \), it follows that \( \xi = F(u) \).

Now consider the other term involving \( u^*_h \). Recall that \( u^*_h \in A(\tau_h u_h, \omega) \).

\[
\| \tau_h u_h - u_h \|_{\mathcal{V}} \leq \| \tau_h u_h - \tau_h u \|_{\mathcal{V}} + \| \tau_h u - u \|_{\mathcal{V}}
\]

\[
\leq \| u_h - u \|_{\mathcal{V}} + \| \tau_h u - u \|_{\mathcal{V}}
\]

and both of these on the right converge to 0 thanks to continuity of translation and (30.5.45). Therefore,

\[
\lim_{h \to 0} \langle u^*_h(\cdot, \omega), \tau_h u_h - u \rangle_{\mathcal{V}', \mathcal{V}} = 0.
\]

It follows that

\[
\langle u^*, u - v \rangle_{\mathcal{V}', \mathcal{V}} = \lim_{h \to 0} \langle u^*_h(\cdot, \omega), u - \tau_h u_h \rangle_{\mathcal{V}', \mathcal{V}} + \langle u^*_h(\cdot, \omega), \tau_h u_h - v \rangle_{\mathcal{V}, \mathcal{V}} \geq \lim_{h \to 0} \langle u^*_h(\cdot, \omega), \tau_h u_h - v \rangle_{\mathcal{V}', \mathcal{V}} + \langle u^*(v), u - v \rangle
\]

where \( u^*(v) \in A(u, \omega) \). Then it follows that \( u^* \in A(u, \omega) \) because if not, then by separation theorems, there would exist \( v \) such that

\[
\langle u^*, u - v \rangle_{\mathcal{V}', \mathcal{V}} < \langle u^*, u - v \rangle_{\mathcal{V}', \mathcal{V}}
\]

for all \( u^* \in A(u, \omega) \) which contradicts the above inequality. Thus, passing to the limit in (30.5.45),

\[
(B(\omega) u(\cdot, \omega))' + \varepsilon F u(\cdot, \omega) + u^* = f(\cdot, \omega) \text{ in } \mathcal{U}',
\]

\[
B u(0, \omega) = B_{u_0}(\omega)
\]

(30.5.49)

Here \( u^* \in A(u, \omega) \). Of course nothing is known about the measurability of \( u^*, u \). All that has been obtained in the above is a solution for each fixed \( \omega \). However, each of the functions \( u_h, u^*_h \) is measurable. Also we have the estimate (30.5.45). By Theorem 30.5.45, there are functions \( \hat{u}(\cdot, \omega), \hat{u}^*(\cdot, \omega) \) and a subsequence with subscript \( h(\omega) \) such that the following weak convergences in \( \mathcal{V} \) and \( \mathcal{V}' \) take place

\[
\begin{align*}
\vspace{1em}
\vspace{1em}
\vspace{1em}
\end{align*}
\]

such that the functions \( (t, \omega) \to \hat{u}(t, \omega), (t, \omega) \to \hat{u}^*(t, \omega) \) are product measurable into \( \mathcal{V} \) and \( \mathcal{V}' \) respectively. The above argument shows that for each \( \omega \), there is a further subsequence, still denoted with subscript \( h(\omega) \) such that

\[
\text{such that } u_{h(\omega)}(\cdot, \omega) \to u(\cdot, \omega) \text{ weakly in } \mathcal{V} \text{ and } u^*_{h(\omega)}(\cdot, \omega) \to u^*(\cdot, \omega) \text{ weakly in } \mathcal{V}' \text{ such that } (u(\cdot, \omega), u^*(\cdot, \omega)) \text{ is a solution to the evolution equation for each } \omega. \text{ By uniqueness of limits, } u(\cdot, \omega) = \hat{u}(\cdot, \omega), \text{ similar for } \hat{u}^*. \text{ Thus this solution which is defined for each } \omega \text{ has a representative for each } \omega \text{ such that the resulting functions of } t, \omega \text{ are product measurable into } \mathcal{V}, \mathcal{V}' \text{ respectively. This proves the following lemma.}
Lemma 30.5.4 For each \( \varepsilon > 0 \) there exists a solution to

\[
(B(\omega)u(\cdot,\omega))' + \varepsilon Fu(\cdot,\omega) + u^*(\cdot,\omega) = f(\cdot,\omega) \text{ in } U',
\]

(30.5.50)

\[
Bu(0,\omega) = Bu_0(\omega)
\]

(30.5.51)

this solution satisfies \((t,\omega) \to u(t,\omega)\) is product measurable into \(V\). Also \((t,\omega) \to u^*(t,\omega)\), and \((t,\omega) \to B(\omega)u(t,\omega)\) are product measurable into \(V'\) and \(W'\) respectively.

Next it is desired to remove the regularizing term \(\varepsilon u\). This will involve another use of Theorem 30.2.10. Denote by \(u_\varepsilon\) the solution to the above lemma. Then act on \(u_\varepsilon\) on both sides. This yields

\[
\frac{1}{2} \langle Bu_\varepsilon, u_\varepsilon \rangle (T) - \frac{1}{2} \langle Bu_0, u_0 \rangle + \varepsilon \int_0^T \langle Fu_\varepsilon, u_\varepsilon \rangle \, ds \leq \int_0^T \langle f, u_\varepsilon \rangle \, ds
\]

(30.5.52)

Then by the coercivity assumption,

\[
\lim_{||u||_V \to \infty} \inf_{u \in X,} \{ \epsilon \langle Fu_\varepsilon, u \rangle_{U'} + ||u_\varepsilon||_V : u^* \in A(u,\omega) \} = \infty
\]

it follows that

\[
\varepsilon \langle Fu_\varepsilon, u_\varepsilon \rangle_{U'} + ||u_\varepsilon||_V \leq C(u_0, f)
\]

(30.5.53)

where the constant on the right does not depend on \(\varepsilon\). Then

\[
\varepsilon Fu_\varepsilon \to 0 \text{ strongly in } U'
\]

this follows because from properties of the duality map,

\[
\langle \varepsilon Fu_\varepsilon, v \rangle \leq \varepsilon \langle Fu_\varepsilon, u_\varepsilon \rangle^{1/r'} \langle Fv, v \rangle^{1/r} = \varepsilon^{1/r'} \langle Fu_\varepsilon, u_\varepsilon \rangle^{1/r} \varepsilon^{1/r} ||v||_U \leq C\varepsilon^{1/r} ||v||_U
\]

Then since \(A\) is bounded, there is a constant \(C\) independent of \(\varepsilon\) such that

\[
||u_\varepsilon^*||_{V'} + ||(Bu_\varepsilon)'||_{U'} + ||u_\varepsilon||_V \leq C
\]

(30.5.54)

It follows there is a subsequence, still denoted with \(\varepsilon\) such that

\[
u_\varepsilon^* \to u^* \text{ weakly in } V',
\]

(30.5.55)

\[
(Bu_\varepsilon)' \to (Bu)' \text{ weakly in } U',
\]

(30.5.56)

also

\[
Bu_\varepsilon(t) - Bu_0 + \int_0^t u_\varepsilon^* ds + \varepsilon \int_0^t Fu_\varepsilon ds = \int_0^t f ds
\]

and so \(Bu_\varepsilon(t)\) converges to \(Bu(t)\) in \(U'\) weakly. This follows right away from the convergence of \((Bu_\varepsilon)'\) in the above. Also from the above equation,

\[
Bu(t) - Bu_0 + \int_0^t u^* ds = \int_0^t f ds
\]
Thus

\[ Bu(0) = Bu_0 \]
\[ (Bu)' + u^* = f \text{ in } U' \]

Since \( V' \subseteq U' \),

\[
\frac{1}{2} \langle Bu, u \rangle (t) - \frac{1}{2} \langle Bu_0, u_0 \rangle + \int_0^t \langle u^*, u \rangle_{V', V} \, ds = \int_0^t \langle f, u \rangle \, ds
\]

Also

\[
\frac{1}{2} \langle Bu_\varepsilon, u_\varepsilon \rangle (t) - \frac{1}{2} \langle Bu_0, u_0 \rangle + \int_0^t \langle u^*_\varepsilon, u_\varepsilon \rangle_{V', V} \, ds
\]

\[
+ \int_0^t \langle \varepsilon Fu_\varepsilon, u_\varepsilon \rangle \, ds = \int_0^t \langle f, u_\varepsilon \rangle \, ds \quad (30.5.58)
\]

Now let \( \{e_i\} \) be the vectors of Lemma 30.3.4 where these are in \( U \). Thus

\[
\langle Bu_\varepsilon, u_\varepsilon \rangle (T) = \sum_{i=1}^{\infty} \langle B(u_\varepsilon (T)), e_i \rangle^2
\]

Hence, by Fatou’s lemma,

\[
\liminf_{\varepsilon \to 0} \langle Bu_\varepsilon, u_\varepsilon \rangle (T) = \liminf_{\varepsilon \to 0} \sum_{i=1}^{\infty} \langle B(u_\varepsilon (T)), e_i \rangle^2
\]

\[
\geq \sum_{i=1}^{\infty} \liminf_{\varepsilon \to 0} \langle B(u_\varepsilon (T)), e_i \rangle^2
\]

\[
= \sum_{i=1}^{\infty} \liminf_{\varepsilon \to 0} \langle Bu_\varepsilon (T), e_i \rangle^2
\]

\[
= \sum_{i=1}^{\infty} \langle Bu(T), e_i \rangle^2
\]

\[
= \langle B(u(T)), u(T) \rangle = \langle Bu, u \rangle (T)
\]

From 30.5.58, letting \( t = T \),

\[
\limsup_{\varepsilon \to 0} \langle u^*_\varepsilon, u_\varepsilon \rangle_{V', V} \leq \limsup_{\varepsilon \to 0} \left( \langle f, u_\varepsilon \rangle + \frac{1}{2} \langle Bu_0, u_0 \rangle - \frac{1}{2} \langle Bu_\varepsilon, u_\varepsilon \rangle (T) \right)
\]

\[
\leq \langle f, u \rangle_{V', V} + \frac{1}{2} \langle Bu_0, u_0 \rangle - \frac{1}{2} \langle Bu, u \rangle (T) = \langle u^*, u \rangle_{V', V}
\]

It follows that

\[
\limsup_{\varepsilon \to 0} \langle u^*_\varepsilon, u_\varepsilon - u \rangle \leq \langle u^*, u \rangle_{V', V} - \langle u^*, u \rangle_{V', V} = 0
\]

and so

\[
\liminf_{\varepsilon \to 0} \langle u^*_\varepsilon, u_\varepsilon - v \rangle \geq \langle u^* (v), u - v \rangle
\]

for any \( v \in V \) where \( u^* (v) \in A(u, \omega) \). In particular for \( v = u \). Hence

\[
\liminf_{\varepsilon \to 0} \langle u^*_\varepsilon, u_\varepsilon - u \rangle \geq \langle u^* (v), u - u \rangle = 0 \geq \limsup_{\varepsilon \to 0} \langle u^*_\varepsilon, u_\varepsilon - u \rangle
\]

showing that \( \lim_{\varepsilon \to 0} \langle u^*_\varepsilon, u_\varepsilon - u \rangle = 0 \). Thus

\[
\langle u^*, u - v \rangle \geq \liminf_{\varepsilon \to 0} \langle u^*_\varepsilon, u - u_\varepsilon \rangle + \langle u^*_\varepsilon, u_\varepsilon - v \rangle
\]

\[
= \liminf_{\varepsilon \to 0} \langle u^*_\varepsilon, u_\varepsilon - v \rangle \geq \langle u^* (v), u - v \rangle
\]

This implies \( u^* \in A(u, \omega) \) because if not, then by separation theorems, there exists \( v \in V \) such that for all \( w^* \in A(u, \omega) \),

\[
\langle u^*, u - v \rangle < \langle w^*, u - v \rangle
\]
contrary to what was shown above. Thus this obtains

\[ Bu(t) - Bu_0 + \int_0^t u^*(s) \, ds = \int_0^t f(s) \, ds \]

where \( u^* \in A(u, \omega) \). In case \( Bu_\varepsilon(T) \neq B(u_\varepsilon(T)) \), you do the same argument for \( \hat{T} < T \) where \( Bu_\varepsilon(\hat{T}) = B(u_\varepsilon(\hat{T})) \) for all \( \varepsilon \) and for \( u \). Then the above argument shows that \( u^* X_{[0,T]} \in A(\mathcal{X}_{[0,T]} u, \omega) \). This being true for every such \( \hat{T} < T \) implies that it holds on \([0,T]\) and shows part of the following theorem which is the main result.

**Theorem 30.5.5** Let the conditions on \( A \) hold \( \text{(30.5.27)} \) - \( \text{(30.5.36)} \). Also let \( B \) satisfy \( \text{(30.5.30)} \) and assume, if it depends on \( \omega \), it is of the form

\[ B(\omega) = k(\omega) B, \; k(\omega) \geq 0, \; k \text{ measurable} \]

Let \( u_0 \) be \( \mathcal{F} \) measurable into \( W \), and let \( f \) be product measurable into \( V' \). Then there exists a solution to the following evolution inclusion

\[
(B(\omega) u(\cdot, \omega))' + u^*(\cdot, \omega) = f(\cdot, \omega) \; \text{in} \; V', \\
B(\omega) u(0, \omega) = B(\omega) u_0(\omega)
\]

where \( u^*(\cdot, \omega) \in A(\mathcal{X}_{[0,T]} \omega) \). In addition to this, \( (t, \omega) \rightarrow u(t, \omega) \) is product measurable into \( V \) and \( (t, \omega) \rightarrow u^*(t, \omega) \) is product measurable into \( V' \).

**In place of the coercivity condition** \( \text{(30.5.30)} \) **assume the coercivity condition involving both** \( B \) **and** \( A \) **given in** \( \text{(30.5.27)} \). **Then**

\[
(B(\omega) u(\cdot, \omega))'(t) - B(\omega) u_0(\omega) + \int_0^t u^*(s, \omega) \, ds = \int_0^t f(s, \omega) \, ds \\
(Bu)' \in V'
\]

**Proof of Theorem 30.5.5**: First consider the claim about replacing the coercivity condition. Returning to \( \text{(30.5.27)} \) one obtains by integrating up to \( t \) and adding \( \lambda \int_0^t \langle Bu_\varepsilon, u_\varepsilon \rangle \, ds \) to both sides,

\[
\frac{1}{2} \langle Bu_\varepsilon, u_\varepsilon \rangle(t) - \frac{1}{2} \langle Bu_0, u_0 \rangle + \varepsilon \int_0^t \langle Fu_\varepsilon, u_\varepsilon \rangle \, ds \\
+ \int_0^t \langle u^*_\varepsilon, u_\varepsilon \rangle \, ds + \lambda \int_0^t \langle Bu_\varepsilon, u_\varepsilon \rangle \, ds = \int_0^t \langle f, u_\varepsilon \rangle \, ds + \lambda \int_0^t \langle Bu_\varepsilon, u_\varepsilon \rangle \, ds
\]

(30.5.62)

Then from the estimate \( \text{(30.5.31)} \),

\[
\frac{1}{2} \langle Bu_\varepsilon, u_\varepsilon \rangle(t) - \frac{1}{2} \langle Bu_0, u_0 \rangle + \varepsilon \int_0^t \langle Fu_\varepsilon, u_\varepsilon \rangle \, ds \\
+ \delta(\omega) \int_0^t \| u^*_\varepsilon \|_{V'} \, ds - m(\omega) = \int_0^t \langle f, u_\varepsilon \rangle \, ds + \lambda \int_0^t \langle Bu_\varepsilon, u_\varepsilon \rangle \, ds
\]

(30.5.63)

From this, it is a routine use of Gronwall’s inequality to obtain the estimate

\[
\varepsilon \langle Fu_\varepsilon, u_\varepsilon \rangle_{U', U} + \| u_\varepsilon \|_V \leq C(u_0, f, \lambda, \omega)
\]

(30.5.64)

Then the rest of the argument is the same. You obtain the following in \( U' \).

\[
B(\omega) u(t, \omega) - B(\omega) u_0(\omega) + \int_0^t u^*(s, \omega) \, ds = \int_0^t f(s, \omega) \, ds
\]
Since all terms but the first are in $V'$, the equation holds in $V'$. Also, the equation in (30.5.41) shows that $(B(\omega) u(\cdot, \omega))^\prime \in V'$.

It only remains to show that there is a product measurable solution. The above argument has shown that there exists a solution for each $\omega$. This is another application of Theorem 30.2.44. For the sequence defined in the convergences \[ \text{Theorem 30.2.44}, \] there is an estimate \[ \text{Theorem 30.2.44}. \] Therefore, the conditions of this theorem hold and there exists a subsequence denoted with $\varepsilon$ such that

\[ u_{\varepsilon(\omega)}(\cdot, \omega) \to \hat{u}(\cdot, \omega) \text{ weakly in } V, \]
\[ u^*_{\varepsilon(\omega)}(\cdot, \omega) \to \hat{u}^*(\cdot, \omega) \text{ weakly in } V' \]

where the $\hat{u}$ and $\hat{u}^*$ are product measurable. Now the above argument shows that for each $\omega$ there exists a further subsequence, still denoted with $\varepsilon$ such that convergence to a solution to the evolution inclusion is obtained $(u(\cdot, \omega), u^*(\cdot, \omega))$. Then by uniqueness of limits, $\hat{u}(\cdot, \omega) = u(\cdot, \omega)$ in $V$, similar for $u^*$ and $\hat{u}^*$. Hence there is a solution to the above evolution problem which satisfies the claimed product measurability. ■

One can give a very interesting generalization of the above theorem.

**Theorem 30.5.6** In the context of Theorem 30.5.4 let $q(t, \omega)$ be a product measurable function into $V$ such that $t \to q(t, \omega)$ is continuous, $q(0, \omega) = 0$.

Then, there exists a solution $u$ of the integral equation

\[ Bu(t, \omega) + \int_0^t z(s, \omega) \, ds = \int_0^t f(s, \omega) \, ds + Bu_0(\omega) + Bq(t, \omega), \]

where $(t, \omega) \to u(t, \omega)$ is measurable. Moreover, for each $\omega$, $Bu(t, \omega) = B(u(t, \omega))$ for a.e. $t$ and $z(\cdot, \omega) \in A(\omega, u(\cdot, \omega))$ for a.e. $t$. Also, for each $a \in [0, T]$,

\[ Bu(t, \omega) + \int_a^t z(s, \omega) \, ds = \int_a^t f(s, \omega) \, ds + Bu(a, \omega) + Bq(t, \omega) - Bq(a, \omega) \]

**Proof:** Define a stopping time

\[ \tau_r \equiv \inf \{ t : |q(t, \omega)| > r \} \]

Then this is the first hitting time of an open set by a continuous random variable and so it is a valid stopping time. Then for each $r$, let

\[ A_r(\omega, w) \equiv A(\omega, w + q^r(\cdot, \omega)) \]

where the notation means $q^r(t) \equiv q(t \wedge \tau_r)$. Then, since $q^r$ is uniformly bounded, all of the necessary estimates and measurability for the solution to the above corollary hold for $A_r$ replacing $A$. Therefore, there exists a solution $w_r$ to the inclusion

\[ (Bw_r)^\prime(\cdot, \omega) + A_r(w_r(\cdot, \omega), \omega) \ni f(\cdot, \omega), \quad Bw_r(0, \omega) = Bu_0(\omega) \]

Now for fixed $\omega, q^r(t, \omega)$ does not change for all $r$ large enough. This is because it is a continuous function of $t$ and so is bounded on the interval $[0, T]$. Thus, for $r$ large enough and fixed $\omega$, $q^r(t, \omega) = q(t, \omega)$. Thus, we obtain

\[ \langle Bw_r(t, \omega), w_r(t, \omega) \rangle + \int_0^t \| w_r(s, \omega) \|_V^p \, ds \leq C(\omega) \tag{30.5.65} \]

Now, as before one can pass to a limit involving a subsequence, as $r \to \infty$ and obtain a solution to the integral equation

\[ Bw(t, \omega) - Bu_0(\omega) + \int_0^t z(s, \omega) \, ds = \int_0^t f(s, \omega) \, ds \]

where $z(s, \omega) \in A(s, \omega, w(s, \omega), q(s, \omega))$ for a.e. $s$ and $z$ is product measurable. Then an application of Theorem 30.2.44 shows that there exists a solution $w$ to this integral equation for each $\omega$ which also has $(t, \omega) \to w(t, \omega)$ product measurable and $(t, \omega) \to z(t, \omega)$ product measurable. Now just let $u(t, \omega) = w(t, \omega) + q(t, \omega)$.

The last claim follows from letting $t = a$ in the top equation and then subtracting this from the top equation with $t > a$. ■
30.6 Variational Inequalities

We have some good theorems above in the context of $\mathcal{V}$ and $\mathcal{V}'$ and $B$ satisfies and assume, if it depends on $\omega$, it is of the form

$$B(\omega) = k(\omega)B, \ k(\omega) \geq 0, \ k \text{ measurable}$$

Now this will be used to consider variational inequalities.

Let $K$ be a closed convex subset of $\mathcal{V}$ containing 0. Let $P : \mathcal{V} \to \mathcal{V}'$ be an operator of penalization. Thus $P = 0$ on $K$ and is monotone and demicontinuous and nonzero off $K$.

$$Pu = F(u - \text{proj}_K(u))$$

where $F$ is the duality map such that $\langle Fu, u \rangle = \|u\|^2, \|Fu\| = \|u\|$. Then $A(\cdot, \omega) + nP$ satisfies the conditions for Theorem \ref{30.5.30} assuming $A(\cdot, \omega)$ satisfies the conditions of this theorem. Then by Theorem \ref{30.5.30} there exists a solution $u_n$ such that $(t, \omega) \to u_n(t, \omega), (t, \omega) \to u_n^*(t, \omega)$ are product measurable, and for each $\omega$,

$$\left(Bu_n\right)' + u_n^*(\cdot, \omega) + nP(u_n(\cdot, \omega)) = f(\cdot, \omega) \text{ in } \mathcal{V}'$$

$$Bu_n(0, \omega) = 0$$

(30.6.66)

Here $B$ is as described in that theorem. Using $0 \in K$ and monotonicity of $P$, the estimates for $A$ lead to an estimate of the form

$$\|u_n(\cdot, \omega)\|_{\mathcal{V}} + \|u_n^*(\cdot, \omega)\|_{\mathcal{V}'} \leq C(\omega)$$

Then there is a subsequence

$$u_n \to u \text{ weakly in } \mathcal{V}$$

$$u_n^* \to u^* \text{ weakly in } \mathcal{V}'$$

$$Pu_n \to \xi \text{ weakly in } \mathcal{V}'$$

Let $\Lambda$ denote those $v \in \mathcal{V}$ such that $(Bu)' \in \mathcal{V}'$ and $Bu(0) = 0$. Then for $v \in \Lambda$,

$$\langle (Bu_n)', u_n - v \rangle + \langle u_n^*(\cdot, \omega), u_n - v \rangle + n \langle P(u_n(\cdot, \omega)), u_n - v \rangle = \langle f(\cdot, \omega), u_n - v \rangle$$

Thus by monotonicity considerations,

$$\langle (Bu)' , u_n - v \rangle + \langle u_n^*(\cdot, \omega), u_n - v \rangle + n \langle P(u_n(\cdot, \omega)), u_n - v \rangle \leq \langle f(\cdot, \omega), u_n - v \rangle$$

(\*)

It follows that

$$\limsup_{n \to \infty} \langle P(u_n(\cdot, \omega)), u_n - v \rangle \leq 0$$

$$\limsup_{n \to \infty} \langle P(u_n(\cdot, \omega)), u_n - u \rangle \leq \langle -\xi, u - v \rangle$$

Now, since $\Lambda$ is dense, $v$ can be chosen as close as desired to $u$ and hence

$$\limsup_{n \to \infty} \langle P(u_n(\cdot, \omega)), u_n - u \rangle \leq 0$$

Since $P$ is monotone, in fact the limit exists in the above. Therefore, for any $v \in \Lambda$ and *,

$$\liminf_{n \to \infty} \langle (P(u_n(\cdot, \omega)), u_n - v) \rangle \geq \langle Pu, u - v \rangle$$

and so

$$\langle Pu, u - v \rangle \leq 0$$

for all $v \in \Lambda$. It follows that $Pu = 0$ and so $u \in K$.

Now for $v \in \Lambda \cap K$, monotonicity considerations imply

$$\langle (Bu)' , u_n - v \rangle + \langle u_n^*(\cdot, \omega), u_n - u \rangle + \langle u_n^*(\cdot, \omega), u - v \rangle \leq \langle f(\cdot, \omega), u_n - v \rangle$$

Then

$$\langle u_n^*(\cdot, \omega), u_n - u \rangle \leq \langle f(\cdot, \omega), u_n - v \rangle - \langle (Bu)', u_n - v \rangle - \langle u_n^*(\cdot, \omega), u - v \rangle$$

(30.6.67)
Then
\[ \limsup_{n \to \infty} \langle u_n^* (\cdot, \omega), u_n - u \rangle \leq \langle f (\cdot, \omega), u - v \rangle + \langle (Bv)', v - u \rangle + \langle u^* (\cdot, \omega), v - u \rangle \]

We assume the existence of a regularizing sequence. If \( u \in \mathcal{K} \) there exists \( u_i \to u \) weakly in \( \mathcal{V} \) such that
\[ \limsup_{i \to \infty} \langle (Bu_i)', u_i - u \rangle_{\mathcal{V}} \leq 0 \]

In the above inequality, let \( v = u_i \)
\[ \limsup_{n \to \infty} \langle u_n^* (\cdot, \omega), u_n - u \rangle \leq \langle f (\cdot, \omega), u - u_i \rangle + \langle (Bu_i)', u_i - u \rangle + \langle u^* (\cdot, \omega), u_i - u \rangle \]

Then take \( \limsup_{t \to 0} \) of both sides to obtain
\[ \limsup_{n \to \infty} \langle u_n^* (\cdot, \omega), u_n - u \rangle \leq 0. \]

Now assume the usual limit condition holds for \( A (\cdot, \omega) \). In practice, this typically means \( A (\cdot, \omega) \) will be single valued, monotone and hemicontinuous because there is no control on the time derivative. However, we will go ahead and assume just that the limit condition holds. This would also take place if \( A (\cdot, \omega) \) were not in \( \mathcal{V} \) valued, monotone and hemicontinuous because there is no control on the time derivative. However, we will go ahead and assume just that the limit condition holds. This would also take place if \( A (\cdot, \omega) \) were not in \( \mathcal{V} \) valued, monotone and hemicontinuous because there is no control on the time derivative. However, we will go ahead and assume just that the limit condition holds. This would also take place if \( A (\cdot, \omega) \) were not in \( \mathcal{V} \)

\[ \limsup_{n \to \infty} \langle u_n^* (\cdot, \omega), u_n - u \rangle \leq 0. \]

where \( u^* (v) \in A (u, \omega) \). In particular, this holds for \( v = u \) and so
\[ \liminf_{n \to \infty} \langle u_n^* (\cdot, \omega), u_n - u \rangle \geq 0 \geq \limsup_{n \to \infty} \langle u_n^* (\cdot, \omega), u_n - u \rangle \]
showing the the limit exists. Then
\[ \langle u^* (v), u - v \rangle \leq \liminf_{n \to \infty} \langle (u_n^* (\cdot, \omega), u_n - u) \rangle \]
\[ = \liminf_{n \to \infty} \langle (u_n^* (\cdot, \omega), u_n - u) + (u_n^* (\cdot, \omega), u - v) \rangle \]
\[ = \langle u^*, u - v \rangle \]

and since this is true for all \( v \in \mathcal{V} \) it follows that \( u^* \in A (u (\cdot, \omega), \omega) \) since otherwise, separation theorems would give a contradiction. If \( u^* \) were not in \( A (u (\cdot, \omega), \omega) \) there would exist \( v \) such that for all \( z^* \in A (u, \omega) \),
\[ \langle z^*, u - v \rangle > \langle u^*, u - v \rangle \]
contrary to the above. Therefore, in [30.6.17] we can take the limit of both sides and conclude that for every \( v \in \mathcal{K} \) such that \( (Bv)' \in \mathcal{V}' \), \( Bv (0) = 0 \),
\[ \langle (Bv)', u - v \rangle + \langle u^*, u - v \rangle \leq \langle f (\cdot, \omega), u - v \rangle \]
where \( u^* \in A (u, \omega) \)

This has proved the first part of the following theorem which gives measurable solutions to a variational inequality.

**Theorem 30.6.1** Suppose \( A (\cdot, \omega) \) is monotone hemicontinuous bounded and single valued and coercive as a map from \( \mathcal{V} \) to \( \mathcal{V}' \). Suppose also that for \( \omega \to u (\omega) \) measurable into \( \mathcal{V} \), it follows that \( \omega \to A (u (\omega), \omega) \) is measurable into \( \mathcal{V}' \). Let \( K \) be a closed convex subset of \( \mathcal{V} \) containing 0 and let \( B \in \mathcal{L} (\mathcal{W}, \mathcal{W}') \) be self adjoint and nonnegative as above. Let there be a regularizing sequence \( \{ u_i \} \) for each \( u \in K \) satisfying \( Bu_i (0) = 0 \), \( (Bu_i)' \in \mathcal{V}' \), \( u_i \in K \),
\[ \limsup_{i \to \infty} \langle (Bu_i)', u_i - u \rangle \leq 0 \]

Then for each \( \omega \), there exists a solution to
\[ \langle (Bv)', u - v \rangle + \langle A (u (\cdot, \omega), \omega), u (\cdot, \omega) - v \rangle \leq \langle f (\cdot, \omega), u - v \rangle \]
valid for all \( v \in \mathcal{K} \) such that \( (Bv)' \in \mathcal{V}' \), \( Bv (0) = 0 \), and \( (t, \omega) \to u (t, \omega) \), is \( \mathcal{B} ([0, T]) \times \mathcal{F} \) measurable.
Proof: This follows from Theorem 30.6.2. This is because there is an estimate of the right sort for the measurable functions \( u_n (\cdot, \omega) \) and \( u_n (\cdot, \omega) \) and the above argument which shows that a subsequence has a convergent subsequence which converges appropriately to a solution. ■

You can have \( K = K (\omega) \). There would be absolutely no change in the above theorem. You just need to have the operator of penalization satisfy \( \omega \to P (u (\omega), \omega) = F (u (\omega) - \text{proj}_{K(\omega)} u (\omega)) \) is measurable into \( V' \) provided \( \omega \to u (\omega) \) is measurable into \( V \). What are the conditions on the set valued \( \omega \to K (\omega) \) which will cause this to take place?

Lemma 30.6.2 Let \( \omega \to K (\omega) \) be measurable into \( V \). Then \( \omega \to \text{proj}_{K(\omega)} u (\omega) \) is also measurable into \( V \) if \( \omega \to u (\omega) \) is measurable.

Proof: It follows from Theorem 30.6.2 that there is a countable collection \( \{ w_n (\omega) \} \), \( \omega \to w_n (\omega) \) being measurable and \( w_n (\omega) \in K (\omega) \) for each \( \omega \) such that for each \( \omega \), \( K (\omega) = \bigcup_n w_n (\omega) \). Let

\[
d_n (\omega) \equiv \text{min} \{ \| u (\omega) - u_k (\omega) \| : k \leq n \}
\]

Let \( u_1 (\omega) \equiv w_1 (\omega) \). Let

\[
u_2 (\omega) = w_1 (\omega)
\]

on the set

\[
\{ \omega : \| u (\omega) - w_1 (\omega) \| < \{ \| u (\omega) - w_2 (\omega) \| \} \}
\]

and

\[
u_2 (\omega) \equiv w_2 (\omega) \text{ off the above set.}
\]

Thus \( \| u_2 (\omega) - u (\omega) \| = d_2 \). Let

\[
u_3 (\omega) = w_1 (\omega) \quad \text{on} \quad \left\{ \omega : \| u (\omega) - w_1 (\omega) \| < \| u (\omega) - w_j (\omega) \|, j = 2, 3 \right\} \equiv S_1
\]

\[
u_3 (\omega) = w_2 (\omega) \quad \text{on} \quad S_1 \cap \left\{ \omega : \| u (\omega) - w_1 (\omega) \| < \| u (\omega) - w_j (\omega) \|, j = 3 \right\}
\]

\[
u_3 (\omega) = w_3 (\omega) \quad \text{on the remainder of} \quad \Omega.
\]

Thus \( \| u_3 (\omega) - u (\omega) \| = d_3 \). Continue this way, obtaining \( u_n (\omega) \) such that \( \| u_n (\omega) - u (\omega) \| = d_n (\omega) \) and \( u_n (\omega) \in K (\omega) \) with \( u_n \) measurable. Thus, in effect one picks the closest of all the \( w_k (\omega) \) for \( k \leq n \) as the value of \( u_n (\omega) \) and \( u_n \) is measurable and by density in \( K (\omega) \) of \( \{ w_n (\omega) \} \) for each \( \omega \), \( \{ u_n (\omega) \} \) must be a minimizing sequence for

\[
\lambda (\omega) = \text{inf} \{ \| u (\omega) - z \| : z \in K (\omega) \}
\]

Then it follows that \( u_n (\omega) \to \text{proj}_{K(\omega)} u (\omega) \) weakly in \( V \). Here is why: Suppose it fails to converge to \( \text{proj}_{K(\omega)} u (\omega) \). Since it is minimizing, it is a bounded sequence. Thus there would be a subsequence, still denoted as \( u_n (\omega) \) which converges to some \( q (\omega) \neq \text{proj}_{K(\omega)} u (\omega) \). Then

\[
\lambda (\omega) = \lim_{n \to \infty} \| u (\omega) - u_n (\omega) \| \geq \| u (\omega) - q (\omega) \|
\]

because convex and lower semicontinuous is weakly lower semicontinuous. But this implies \( q (\omega) = \text{proj}_{K(\omega)} u (\omega) \) because the projection map is well defined thanks to strict convexity of the norm used. This is a contradiction. Hence \( \text{proj}_{K(\omega)} u (\omega) = \lim_{n \to \infty} u_n (\omega) \) and so is a measurable function. It follows that \( \omega \to P (u (\omega), \omega) \) is measurable into \( V \). ■

The following corollary is now immediate.

Corollary 30.6.3 Suppose \( A (\cdot, \omega) \) is monotone hemicontinuous bounded, single valued, and coercive as a map from \( V \) to \( V' \). Suppose also that for \( \omega \to u (\omega) \) measurable into \( V \), it follows that \( \omega \to A (u (\omega), \omega) \) is measurable into \( V' \). Let \( K (\omega) \) be a closed convex subset of \( V \) containing 0 and \( \omega \to K (\omega) \) is a set valued measurable multifunction. Let \( B \in \mathcal{L} (W, V') \) be self adjoint and nonnegative as above. Let there be a regularizing sequence \( \{ u_i \} \) for each \( u \in K \) satisfying \( Bu_i (0) = 0, (Bu_i)' \in V', u_i \in K \),

\[
\limsup_{i \to \infty} \langle (Bu_i)', u_i - u \rangle \leq 0
\]

Then for each \( \omega \), there exists a solution to

\[
\langle (Bv)', u - v \rangle + \langle A (u (\cdot, \omega), \omega), u (\cdot, \omega) - v \rangle \leq \langle f (\cdot, \omega), u (\cdot, \omega) - v \rangle
\]

valid for all \( v \in K (\omega) \) such that \( (Bv)' \in V', Bv (0) = 0 \), and \( (t, \omega) \to u (t, \omega) \) is \( \mathcal{B} ([0, T]) \times \mathcal{F} \) measurable.
Proof: The proof is identical to the above. One obtains a measurable solution to \( u_t = 0 \) in which \( P \) is replaced with \( P(\cdot, \omega) \). Then one proceeds in exactly the same steps as before and finally uses Theorem \( 30.5.36 \) to obtain the measurability of a solution to the variational inequality. 

What does it mean for \( u(\omega) \in K(\omega) \) for each \( \omega \)? It means that there is a sequence of the \( w_n \) \( \{w_n(\omega)\} \) such that each \( w_n \) is measurable into \( V \) implying that for each \( \omega \) there is a representative \( t \to w_n(t, \omega) \) such that the resulting \( (t, \omega) \to w_n(t, \omega) \) is product measurable and \( \|u(\cdot, \omega) - w_n(\cdot, \omega)\|_V \to 0 \). Thus there is no reason to think that \( (t, \omega) \to u(t, \omega) \) is product measurable. The message of the above corollary says that nevertheless, there is a measurable solution to the variational inequality.

### 30.7 Progressively Measurable Solutions

In the context of uniqueness of the evolution initial value problem for fixed \( \omega \), one can prove theorems about progressively measurable solutions fairly easily. First is a definition of the term progressively measurable.

**Definition 30.7.1** Let \( F_t \) be an increasing in \( t \sigma \) algebra of sets of \( \mathcal{F} \). Thus each \( F_t \) is a \( \sigma \) algebra and if \( s \leq t \), then \( F_s \subseteq F_t \). This set of \( \sigma \) algebras is called a filtration. A set \( S \subseteq [0, T] \times \Omega \) is called progressively measurable if for every \( t \in [0, T] \),

\[
S \cap [0, t] \times \Omega \in B([0, t]) \times F_t
\]

Denote by \( \mathcal{P} \) the progressively measurable sets. This is a \( \sigma \) algebra of subsets of \( [0, T] \times \Omega \). A function \( g \) is progressively measurable if \( X_{[0, t]}g \) is \( B([0, t]) \times F_t \) measurable for each \( t \).

Let \( A \) satisfy the bounded condition \( 30.5.27 \), the condition on subintervals \( 30.5.28 \), the specific boundedness estimate involving \( B \) and \( A \) in \( 30.6.66 \), and the limit condition \( 30.5.35 \). In place of the condition on the existence of a measurable selection \( 30.5.27 \), we will assume the following condition.

**Condition 30.7.2** For each \( t \leq T \), if \( \omega \to u(\cdot, \omega) \) is measurable into \( V_{[0, t]} \), then there exists a \( F_t \) measurable selection of \( A(u(\cdot, \omega), \omega) \) into \( V_{[0, t]}' \).

In this section, we assume that \( \omega \to B(\omega) \) is \( \mathcal{F}_0 \) measurable into \( \mathcal{L}(W, W') \).

The theorem to be shown is the following.

**Theorem 30.7.3** Assume the above conditions, the condition on subintervals, the specific boundedness estimate involving \( B \) and \( A \) in \( 30.5.27 \), and the Condition 30.7.2. Let \( u_0 \) be \( \mathcal{F}_0 \) measurable and \( \omega \to B(\omega) \) also \( \mathcal{F}_0 \) measurable and \( (t, \omega) \to X_{[0, t]}f(t, \omega) \) is \( B([0, t]) \times F_t \) measurable measurable into \( V' \) for each \( t \). Also assume that for each \( \omega \), there is at most one solution to the evolution equation

\[
(B(\omega) u(\cdot, \omega))(t) - B(\omega) u_0(\omega) + \int_0^t u^*(\cdot, \omega) ds = \int_0^t f(s, \omega) ds,
\]

\[
u^*(\cdot, \omega) \in A(u(\cdot, \omega), \omega)
\]

for \( t \in [0, \hat{T}] \) for each \( \hat{T} \leq T \). Then there exists a unique solution \( (u(\cdot, \omega), u^*(\cdot, \omega)) \) in \( V \times V' \) to the above integral equation for each \( \omega \). This solution satisfies \( (t, \omega) \to (u(t, \omega), u^*(t, \omega)) \) is progressively measurable into \( V \times V' \).

Proof: Let \( T \) denote subsets of \( [0, T] \) which contain \( T \) such that for \( S \in T \), there exists a solution \( u_S \) for each \( \omega \) to the above integral equation on \( [0, T] \) such that \( (t, \omega) \to X_{[0, s]}(t) u_S(t, \omega) \) is \( B([0, s]) \times F_s \) measurable for each \( s \in S \). Then \( \{T\} \in T \). If \( S, S' \) are in \( T \), then \( S \subseteq S' \) will mean that \( S \subseteq S' \) and also \( u_S(t, \omega) = u_{S'}(t, \omega) \) in \( V \) for all \( t \in S, \) similar for \( u_s^* \) and \( u_s^{**} \). Note that equality must hold in \( V \) by uniqueness. Now let \( C \) denote a maximal chain. Is \( \mathcal{C} \equiv S_\infty \) all of \( [0, T] \)? What is \( u_{S_\infty} \)? Define \( u_{S_\infty}(t, \omega) \) the common value of \( u_S(t, \omega) \) for all \( S \in C \), which contain \( t \in S_\infty \). If \( S \in S_\infty \), then it is in some \( S \in C \) and so the product measurability condition holds for this \( S \). Thus \( S_\infty \) is a maximal element of the partially ordered set. Is \( S_\infty \) all of \( [0, T] \)? Suppose \( \hat{s} \notin S_\infty, T > \hat{s} > 0 \).

From Theorem 30.5.36 there exists a solution to the integral equation on \( [0, \hat{s}] \) called \( u_1 \) such that \( (t, \omega) \to u_1(t, \omega) \) is \( B([0, \hat{s}]) \times F_1 \) measurable, similar for \( u_1^* \). By the same theorem, there is a solution on \( [0, T] \), \( u_2 \) which is \( B([0, T]) \times F_2 \) measurable. Now by uniqueness, \( u_2(\cdot, \omega) = u_1(\cdot, \omega) \) in \( V_{[0, \hat{s}]} \), similar for \( u_2^* \). Therefore, no harm is done in re-defining \( u_2 \) on \( [0, \hat{s}] \) so that \( u_2(\cdot, \omega) = u_1(\cdot, \omega) \) for all \( t \in [0, \hat{s}] \), similar for \( u^* \). Denote these functions as \( \hat{u}, \hat{u}^* \). By uniqueness, \( u_{S_\infty}(\cdot, \omega) = \hat{u}(\cdot, \omega) \) in \( L^p([0, \hat{s}], V) \). Thus no harm is done by re-defining \( \hat{u}(s, \omega) \) to equal \( u_{S_\infty}(s, \omega) \) for \( s < \hat{s} \) and \( \hat{u}(\hat{s}, \omega) \) at \( \hat{s} \). As to \( s > \hat{s} \) also re define \( \hat{u}(s, \omega) \equiv u_{S_\infty}(s, \omega) \) for such \( s \). By uniqueness, the two are equal in \( V_{[\hat{s}, T]} \) and no change occurs in the solution of the integral equation. Now \( S_\infty \) was not maximal after all. \( S_\infty \cup \{\hat{s}\} \) is larger. This contradiction shows that in fact, \( S_\infty = [0, T] \). ■
Theorem 30.7.4 Assume the above conditions, and the Condition 30.7.4. Let \( u_0 \) be \( \mathcal{F}_0 \) measurable and \( \omega \to B(\omega) \) also \( \mathcal{F}_0 \) measurable and \( (t, \omega) \to X_{[t,\hat{t}]}(t) f(t, \omega) \) is \( \mathcal{B}([0,\hat{t}]) \times \mathcal{F}_t \) product measurable into \( V' \) for each \( t \).

\[
B(\omega) = k(\omega) B, \ k(\omega) \geq 0, k \text{ measurable.}
\]

Also let \( t \to q(t, \omega) \) be continuous and \( q \) is progressively measurable into \( V \). Suppose there is at most one solution to

\[
Bu(t, \omega) + \int_0^t z(s, \omega) \, ds = \int_0^t f(s, \omega) \, ds + Bu_0(\omega) + Bq(t, \omega),
\]

for each \( \omega \). Then the solution \( u \) to the above integral equation is progressively measurable and so is \( z \). Moreover, for each \( \omega \), both \( Bu(t, \omega) = B(u(t, \omega)) \) a.e. \( t \) and \( z(\cdot, \omega) \in A(u(\cdot, \omega), \omega) \). Also, for each \( a \in [0, T] \),

\[
Bu(t, \omega) + \int_a^t z(s, \omega) \, ds = \int_a^t f(s, \omega) \, ds + Bu(a, \omega) + Bq(t, \omega) - Bq(a, \omega)
\]

**Proof:** By Theorem 30.7.4, there exists a solution to 30.7.68 which is \( \mathcal{B}([0, T]) \times \mathcal{F}_T \) measurable. Now, as in the proof of Theorem 30.7.4, one can define a new operator

\[
A_r(w, \omega) \equiv A(\omega, w + q^r(\cdot, \omega))
\]

where \( r \) is the stopping time defined there. Then, since \( q \) is progressively measurable, the progressively measurable condition is satisfied for this new operator. Hence by Theorem 30.7.4, there exists a unique solution \( w \) which is progressively measurable to the integral equation

\[
Bw_r(t, \omega) + \int_0^t z_r(s, \omega) \, ds = \int_0^t f(s, \omega) \, ds + Bu_0(\omega)
\]

where \( z_r(\cdot, \omega) \in A_r(w(\cdot, \omega), \omega) \). Then as in Theorem 30.7.4 you can let \( r \to \infty \) and eventually \( q^r(\cdot, \omega) = q(\cdot, \omega) \). Then, passing to a limit, it follows that for a given \( \omega \), there is a solution to

\[
Bw(t, \omega) + \int_0^t z(s, \omega) \, ds = \int_0^t f(s, \omega) \, ds + Bu_0(\omega)
\]

which is progressively measurable because \( w(\cdot, \omega) = \lim_{r \to \infty} w_r(\cdot, \omega) \) in \( V \) each \( w_r \) being progressively measurable. Thus for each \( \hat{T} < T, \omega \to w(\cdot, \omega) \) is measurable into \( V_{[0,\hat{T}]} \). Then by Lemma 30.7.4, \( w \) has a representative in \( V \) for each \( \omega \) such that the resulting function satisfies \( (t, \omega) \to X_{[0,\hat{T}]}(t) w(t, \omega) \) is \( \mathcal{B}([0, \hat{T}]) \times \mathcal{F}_{\hat{T}} \) measurable into \( V \). Thus one can assume that \( w \) is progressively measurable. Now as in Theorem 30.7.4, Define \( u = w + q \).

The last claim follows from letting \( t = a \) in the top equation and then subtracting this from the top equation with \( t > a \). \( \square \)

### 30.8 Adding Another Operator

Recall the following conditions for the various operators.

**Bounded and coercive conditions**

\[
A(\cdot, \omega), A(\cdot, \omega) : \mathcal{V}_I \to \mathcal{V}_I' \text{ for each } I \text{ a subinterval of } [0, T] \quad I = [0, \hat{T}], \hat{T} \leq T
\]

\[
A(\cdot, \omega) : \mathcal{V}_I \to \mathcal{P}(\mathcal{V}_I') \text{ is bounded,}
\]

If, for \( u \in \mathcal{V} \),

\[
u^* \mathcal{X}_{[0, \hat{T}]} \in A\left(u \mathcal{X}_{[0, \hat{T}]}, \omega\right)
\]

for each \( \hat{T} \) in an increasing sequence converging to \( T \), then

\[
u^* \in A(u, \omega)
\]
Assume the specific estimate
\[
\sup \{ \| u^* \|_{V_t^t} : u^* \in A(u, \omega) \} \leq a(\omega) + b(\omega) \| u \|_{V_t^t}^{p-1}
\] (30.8.71)
where \( a(\omega), b(\omega) \) are nonnegative. Also assume the following coercivity estimate valid for each \( t \leq T \) and for some \( \lambda(\omega) \geq 0 \),
\[
\inf \left( \int_0^t \langle u^*, u \rangle + \lambda(\omega) \langle Bu, u \rangle \, dt : u^* \in A(u, \omega) \right) \geq \delta(\omega) \int_0^t \| u \|_{V_t}^p \, ds - m(\omega)
\] (30.8.72)
where \( m(\omega) \) is some nonnegative constant, \( \delta(\omega) > 0 \).

**Limit condition**

Let \( U \) be a Banach space dense in \( V \) and that if \( u_i \rightharpoonup u \) in \( V_I \) and \( u_i^* \in A(u_i) \) with \( u_i^* \rightharpoonup u^* \) in \( V_I' \) and \( (Bu_i)' \rightharpoonup (Bu)' \) in \( U_I' \), \( \rightharpoonup \) denoting weak convergence, then if
\[
\limsup_{i \to \infty} \langle u_i^*, u_i - u \rangle_{V_I', V_I} \leq 0
\]
it follows that for all \( v \in V_I \), there exists \( u^*(v) \in Au \) such that
\[
\liminf_{i \to \infty} \langle u_i^*, u_i - v \rangle_{V_I', V_I} \geq \langle u^*(v), u - v \rangle_{V_I', V_I}
\] (30.8.73)
where \( r > \max(p, 2) \), and we replace \( p \) with \( r \) and \( I \) an arbitrary subinterval of the form \([0, \hat{T}], \hat{T} < T\), for \([0, T]\), and \( U \) for \( V \) where indicated. Here
\[
U_{tI} \equiv L^r(I; U)
\]
Typically, \( U \) is compactly embedded in \( V \).

**Measurability condition**

For \( \omega \to u(\cdot, \omega) \) measurable into \( V \),
\[
\omega \to A(u(\cdot, \omega), \omega) \text{ has a measurable selection into } V'.
\] (30.8.74)
This last condition means there is a function \( \omega \to u^*(\omega) \) which is measurable into \( V' \) such that \( u^*(\omega) \in A(u(\cdot, \omega), \omega) \).

As for the operator \( B \) it is either independent of \( \omega \) and is a nonnegative self-adjoint operator mapping \( W \) to \( W' \) or else it is of the form \( k(\omega)B \) where \( k \geq 0 \) and is measurable.

We will assume here that \( p > 1 \). Then the following main result was obtained. It is Theorem 30.8.1.

**Theorem 30.8.1** If 30.8.72 and \( B \) as described above, let \( q(t, \omega) \) be a product measurable function into \( V \) such that \( t \to q(t, \omega) \) is continuous, \( q(0, \omega) = 0 \).

Then, there exists a solution \( u \) of the integral equation
\[
Bu(t, \omega) + \int_0^t z(s, \omega) \, ds = \int_0^t f(s, \omega) \, ds + Bu_0(\omega) + Bq(t, \omega),
\]
where \( t, \omega \to u(t, \omega) \) is measurable. Moreover, for each \( \omega \), \( Bu(t, \omega) = B(u(t, \omega)) \) for a.e. \( t \) and \( z(\cdot, \omega) \in A(\omega, u(\cdot, \omega)) \) for a.e. \( t \). Also, for each \( a \in [0, T] \),
\[
Bu(t, \omega) + \int_a^t z(s, \omega) \, ds = \int_a^t f(s, \omega) \, ds + Bu(a, \omega) + Bq(t, \omega) - Bq(a, \omega)
\]
The idea here is to show that everything works as well if a suitable unbounded maximal monotone operator is added in. The result is interesting but not as interesting as it might be. This is because the maximal monotone operator must be quasi bounded. Still it is interesting to note that the above holds for some unbounded operators. This has been pointed out in the case where there are no stochastic effects in a recent paper [11]. This generalizes this result by considering the measurability of solutions and allowing for possibly degenerate leading operator \( B \).

To begin with assume \( q = 0 \).

Now let \( G : D(G) \subseteq V \to P(V') \) be maximal monotone. Also assume that \( 0 \in D(G) \). Then you have
\[
\langle u^*, u \rangle \geq \langle g^*, u \rangle \text{ if } u^* \in Gu
\]
for every \( g^* \in G(0) \). Hence
\[
\langle u^*, u \rangle \geq - |G(0)| \|u\|_V \quad \text{if} \quad u^* \in Gu
\]
(30.8.75)
where \( |G(0)| = \inf \{ |y^*|_V : y^* \in G(0) \} \).

There is a standard way of approximating \( G \) with bounded demicontinuous operators which is reviewed next. It is all in Barbu [2]. See Section 761. Since \( G \) is maximal monotone,
\[
0 \in F(x_\mu - x) + \mu^{p-1}G(x_\mu)
\]
where \( F \) is a duality map for \( p \), the one used in the above theorem. Barbu uses only \( p = 2 \) but it works just as well for arbitrary \( p > 1 \) with the minor modifications used here. To see this, you consider \( \hat{G}(y) \equiv G(x + y) \). Then \( \hat{G} \) is also maximal monotone and so there exists a solution to
\[
0 \in F(\hat{x}) + \mu^{p-1}\hat{G}(\hat{x}) = F(\hat{x}) + \mu^{p-1}G(x + \hat{x})
\]
Now let \( x_\mu = x + \hat{x} \) so \( \hat{x} = x_\mu - x \). Hence
\[
0 \in F(x_\mu - x) + \mu^{p-1}Gx_\mu
\]
The symbol \( \limsup_{n,n \to \infty} a_{mn} \) means \( \limsup_{n \to \infty} \left( \sup_{n \geq N, m \geq N} a_{mn} \right) \).

**Lemma 30.8.2** Suppose \( \limsup_{n,n \to \infty} a_{mn} \leq 0 \). Then \( \limsup_{m \to \infty} (\limsup_{n \to \infty} a_{mn}) \leq 0 \).

**Proof:** Suppose the first inequality. Then for \( \varepsilon > 0 \), there exists \( N \) such that if \( n, m \) are both as large as \( N \), then \( a_{mn} \leq \varepsilon \). Thus \( \sup_{n \geq N} a_{mn} \leq \varepsilon \) provided \( m \geq N \) also. Hence for such \( m \),
\[
\lim_{n \to \infty} \left( \sup_{n \geq N} a_{mn} \right) \leq \varepsilon
\]
for each \( m \geq N \). It follows \( \limsup_{m \to \infty} (\limsup_{n \to \infty} a_{mn}) \leq \varepsilon \). Since \( \varepsilon \) is arbitrary, this proves the lemma. ■

Then here is a simple observation.

**Lemma 30.8.3** Let \( G : D(G) \subseteq X \to \mathcal{P}(X') \) where \( X \) is a Banach space be maximal monotone and let \( v_n \in Gu_n \) and
\[
u_n \to u, \quad v_n \to v \text{ weakly}.
\]
Also suppose that
\[
\limsup_{m,n \to \infty} \langle v_n - v_m, u_n - u_m \rangle \leq 0
\]
or
\[
\limsup_{n \to \infty} \langle v_n - v, u_n - u \rangle \leq 0
\]
Then \([u,v] \in \mathcal{G}(G)\) and \( \langle v_n, u_n \rangle \to \langle v, u \rangle \).

**Proof:** By monotonicity,
\[
0 \geq \limsup_{m,n \to \infty} \langle v_n - v_m, u_n - u_m \rangle
\]
\[
\geq \liminf_{m,n \to \infty} \langle v_n - v_m, u_n - u_m \rangle \geq 0
\]
and so
\[
\lim_{m,n \to \infty} \langle v_n - v_m, u_n - u_m \rangle = 0
\]
Suppose then that \( \langle v_n, u_n \rangle \) fails to converge to \( \langle v, u \rangle \). Then there is a subsequence, still denoted with subscript \( n \) such that \( \langle v_n, u_n \rangle \to \mu \neq \langle v, u \rangle \). Let \( \varepsilon > 0 \). Then there exists \( M \) such that if \( n, m > M \), then
\[
|\langle v_n, u_n \rangle - \mu| < \varepsilon, |\langle v_n - v_m, u_n - u_m \rangle| < \varepsilon
\]
Then if \( m, n > M \),
\[
|\langle v_n - v_m, u_n - u_m \rangle| = |\langle v_n, u_n \rangle + \langle v_m, u_m \rangle - \langle v_n, u_m \rangle - \langle v_m, u_n \rangle| < \varepsilon
\]
Hence it is also true that
\[ |\langle v_n, u_n \rangle + \langle v_m, u_m \rangle - \langle v_n, u_m \rangle - \langle v_m, u_n \rangle| \leq |2\mu - (\langle v_n, u_m \rangle + \langle v_m, u_n \rangle)| < 3\varepsilon \]
Now take a limit first with respect to \( n \) and then with respect to \( m \) to obtain
\[ |2\mu - (\langle v, u \rangle + \langle v, u \rangle)| < 3\varepsilon \]
Since \( \varepsilon \) is arbitrary, \( \mu = \langle v, u \rangle \) after all. Hence the claim that \( \langle v_n, u_m \rangle \to \langle v, u \rangle \) is verified. Next suppose \( [x, y] \in G(G) \) and consider
\[
\langle v - y, u - x \rangle = \langle v, u \rangle - \langle v, x \rangle - \langle y, u \rangle + \langle y, x \rangle
\]
\[
= \lim_{n \to \infty} (\langle v_n, u_n \rangle - \langle v_n, x \rangle - \langle y, u_n \rangle + \langle y, x \rangle)
\]
\[
\geq \lim \sup_{n \to \infty} (\langle v_n, u_n \rangle - \langle v_n, x \rangle - \langle y, u_n \rangle + \langle y, x \rangle)
\]
and since \( [x, y] \) is arbitrary, it follows that \( v \in G(u) \).

Next suppose \( \lim \sup_{n \to \infty} \langle v_n - v, u_n - u \rangle \leq 0 \). It is not known that \( [u, v] \in G(G) \).
\[
\lim \sup_{n \to \infty} (\langle v_n, u_n \rangle - \langle v_n, u_n \rangle + \langle v, u \rangle) \leq 0
\]
\[
\lim \sup_{n \to \infty} (\langle v_n, u_n \rangle - \langle v, u \rangle) \leq 0
\]
Thus \( \lim \sup_{n \to \infty} (v_n, u_n) \leq (v, u) \). Now let \( [x, y] \in G(G) \)
\[
\langle v - y, u - x \rangle = \langle v, u \rangle - \langle v, x \rangle - \langle y, u \rangle + \langle y, x \rangle
\]
\[
\geq \lim \sup_{n \to \infty} (\langle v_n, u_n \rangle - \langle v_n, x \rangle - \langle y, u_n \rangle + \langle y, x \rangle)
\]
\[
\geq \lim \inf_{n \to \infty} (\langle v_n - y, u_n - x \rangle) \geq 0
\]
Hence \( [u, v] \in G(G) \). Now
\[
\lim \sup_{n \to \infty} (v_n - v, u_n - u) \leq 0 \leq \lim \inf_{n \to \infty} (v_n - v, u_n - u)
\]
the second coming from monotonicity and the fact that \( v \in G_u \). Therefore,
\[
\lim_{n \to \infty} (v_n - v, u_n - u) = 0
\]
which shows that \( \lim_{n \to \infty} (v_n, u_n) = (v, u) \). [\( \Box \)]

Similar reasoning implies

**Lemma 30.8.4** Suppose \( A \) is a set valued operator, \( A : X \to P (X) \) and \( u_n^* \in Au_n \). Suppose also that \( u_n \to u \) weakly and \( u_n^* \to u^* \) weakly. Suppose also that
\[
\lim \sup_{m, n \to \infty} \langle u_n^* - u_m^*, u_n - u_m \rangle \leq 0
\]
Then one can conclude that
\[
\lim \sup_{n \to \infty} \langle u_n^*, u_n - u \rangle \leq 0
\]

**Proof:** It is assumed that
\[
\lim \sup_{m, n \to \infty} (\langle u_n^*, u_m \rangle + \langle u_m^*, u_m \rangle - \langle u_n^*, u_n \rangle + \langle u_m^*, u_n \rangle) \leq 0
\]
Then is it the case that \( \lim \sup_{n \to \infty} \langle u_n^*, u_n \rangle \leq \langle u^*, u \rangle \)? Let \( \mu \) equal \( \lim \sup_{n \to \infty} \langle u_n^*, u_n \rangle \). Then in the above, it implies
\[
(2\mu - (\langle u_n^*, u_m \rangle + \langle u_m^*, u_n \rangle)) < \varepsilon
\]
whenever \( m, n \) large enough. Thus taking \( \lim \sup_{n \to \infty} \lim \sup_{m \to \infty} \) of the above, you get
\[
(2\mu - (\langle u^*, u \rangle + \langle u^*, u \rangle)) < \varepsilon
\]
Thus you at least need \( \mu \leq \langle u^*, u \rangle \). That is, \( \lim \sup_{n \to \infty} \langle u_n^*, u_n \rangle \leq \langle u^*, u \rangle \). Hence
\[
\lim \sup_{n \to \infty} \langle u_n^*, u_n - u \rangle = \lim \sup_{n \to \infty} \langle u_n^*, u_n \rangle - \langle u^*, u \rangle \leq \langle u^*, u \rangle - \langle u^*, u \rangle = 0 \]
**Definition 30.8.5** Let \( x_\mu \) just defined be denoted by \( J_\mu x \) and define also

\[
G_\mu (x) \equiv -\mu^{-(p-1)} F(x_\mu - x) .
\]

This \( x_\mu \) is defined as follows.

\[
0 \in F(x_\mu - x) + \mu^{p-1} Gx_\mu
\]

Later, we will write \( J_\mu u \) for \( u_\mu \). Thus

\[
0 = F(J_\mu u - u) + \mu^{p-1} z_\mu, \quad z_\mu \in G(J_\mu u)
\]

Also from this definition,

\[
G_\mu (u) = -\mu^{-(p-1)} F(J_\mu u - u) = z_\mu \in G(J_\mu u)
\]

Then there are some things which can be said about these operators.

**Theorem 30.8.6** The following hold. Here \( V \) is a reflexive Banach space with strictly convex norm. \( G : D(G) \rightarrow \mathcal{P}(V') \) is maximal monotone. Then

1. \( J_\mu \) and \( G_\mu \) are bounded single valued operators defined on \( V \). Bounded means they take bounded sets to bounded sets. Also \( G_\mu \) is a monotone operator.

2. \( G_\mu, J_\mu \) are demicontinuous. That is, strongly convergent sequences are mapped to weakly convergent sequences.

3. For every \( x \in D(G), \|G_\mu (x)\| \leq |Gx| \equiv \inf \{\|y^*\| : y^* \in Gx\} \). For every \( x \in \text{conv}(D(G)) \), it follows that 

\[
\lim_{\mu \rightarrow 0} J_\mu (x) = x .
\]

The new symbol means the closure of the convex hull. It is the closure of the set of all convex combinations of points of \( D(G) \).

Then \( A(\cdot, \omega) + G_\mu \) will be bounded and have the same limit properties as \( A(\cdot, \omega) \). As to measurability, \( G \) and hence \( G_\mu \) do not depend on \( \omega \) and so the measurability condition will hold.

What about the estimates? We need to consider the estimates. Recall what these were:

\[
\sup \{\|u^*\|_{V'} : u^* \in A(u, \omega)\} \leq a(\omega) + b(\omega)\|u\|_{V'}^{p-1}
\]

where \( a(\omega), b(\omega) \) are nonnegative. Also assume the following coercivity estimate valid for each \( t \leq T \) and for some \( \lambda(\omega) \geq 0 \),

\[
\inf \left( \int_0^t \langle u^*, u \rangle + \lambda(\omega) \langle Bu, u \rangle \, dt : u^* \in A(u, \omega) \right) \geq \delta(\omega) \int_0^t \|u\|_{V'}^p \, ds - m(\omega)
\]

where \( m(\omega) \) is some nonnegative constant, \( \delta(\omega) > 0 \).

The coercivity is not too bad. This is because \( G_\mu \) is monotone and \( 0 \in D(G) \). Therefore,

\[
\langle G_\mu u, u \rangle = \langle G_\mu u - G_\mu (0), u \rangle \geq 0
\]

so

\[
\langle G_\mu u, u \rangle \geq -|G(0)||u|| \geq \frac{\delta(\omega)}{2} \|u\|^p - \tilde{m}(\omega)
\]

so the coercivity condition \((30.8.77)\) will end up holding for \( A + G_\mu \). However, more needs to be considered for the growth condition.

From the definition of \( u_\mu \), there exists \( z_\mu \in Gu_\mu \)

\[
0 = F(u_\mu - u) + \mu^{p-1} z_\mu
\]

Then from the choice of \( F \), it is also the duality map from \( V \) to \( V' \) corresponding to \( p > 2 \).
growth condition remains valid for $G$.

Thus

$$0 \geq ||u_\mu|| - ||u||^p - \mu |G(0)||u_\mu||$$

This requires that there is some constant $C$ such that $||u_\mu|| \leq C ||u|| + C$. The details follow.

Let $a = ||u_\mu||, b = ||u||$. Then they are both positive and

$$0 \geq |a - b|^p - b^p - \alpha a$$

where $\alpha = p\mu |G(0)|$. Want to say $a \leq Cb + C$ for some $C$. This is the conclusion of the following lemma.

Lemma 30.8.7 Suppose $0 \geq |a - b|^p - b^p - \alpha a$ for $a, b \geq 0$ and $\alpha > 0$. Then there exists a constant $C$ such that

$$a \leq Cb + C$$

Proof: If $b \geq a$, then there is nothing to show. Therefore, it suffices to show that the desired inequality holds for $a > b$. Thus from now on, $a > b$.

$$0 \geq (a - b)^p - b^p - \alpha a$$

Suppose $a > nb + n$. Let $x = b/a$. Then for $x \in [0, 1]$,

$$0 \geq (1 - x)^p - x^p - \alpha \frac{1}{a^{p-1}}$$

Now for all $n$ large enough, the right side is a decreasing function of $x$ which is positive at $x = 0$ and negative at $x = 1$. Thus $x$ corresponds to the place where this function is negative. Taking a limit as $n \to \infty$, it follows that we must have

$$x \geq \delta, \delta \in (0, 1)$$

It is where $(1 - x)^p - x^p = 0$. Thus $x = \frac{b}{a} \geq \delta$. Then, since $a > nb + n$,

$$\frac{1}{\delta}b \geq a > nb + n$$

Now this is a contradiction when $n$ is taken increasingly large. Hence, for large enough $n, a \leq nb + n$. ■

It follows that $||u_\mu|| \leq C ||u|| + C$ for some $C$. Hence,

$$||G_{\mu,u}|| \leq \frac{1}{\mu^{p-1}} ||u_\mu - u||^{p-1} \leq \frac{1}{\mu^{p-1}} (||u_\mu|| + ||u||)^{p-1}$$

$$\leq \frac{2p-2}{\mu^{p-1}} (||u_\mu||^{p-1} + ||u||^{p-1})$$

$$\leq \frac{2p-2}{\mu^{p-1}} ((C ||u|| + C)^{p-1} + ||u||^{p-1})$$

$$\leq \frac{2p-2}{\mu^{p-1}} \left( 2^{p-2} \left( C^{p-1} ||u||^{p-1} + C^{p-1} \right) + ||u||^{p-1} \right)$$

$$\leq C_\mu ||u||^{p-1} + C_\mu$$

This is the case that $p \geq 2$. The case that $p > 1$ but $p < 2$ is easier. In this case,

$$\frac{1}{\mu^{p-1}} (||u_\mu|| + ||u||)^{p-1} \leq \frac{1}{\mu^{p-1}} \left( ||u_\mu||^{p-1} + ||u||^{p-1} \right)$$

A similar inequality holds. Thus the necessary growth condition is obtained for $G_{\mu}$ and consequently, the necessary growth condition remains valid for $G_{\mu} + A$. It was noted earlier that the coercivity estimate continues to hold.
30.8. ADDING ANOTHER OPERATOR

It follows that there exists a solution to the integral equation

\[ Bu(t,\omega) + \int_0^t z(s,\omega) \, ds + \int_0^t G_\mu(u(s,\omega)) \, ds = \int_0^t f(s,\omega) \, ds + Bu_0(\omega) \]

where \( z(\cdot,\omega) \in A(u(\cdot,\omega),\omega) \) which has the measurability described above. That is, both \( u \) and \( z \) are product measurable. Then acting on \( u \chi_{[0,t]} \) and using the estimates valid for \( \lambda \) large enough, one can get an estimate of the form

\[
\frac{1}{2} \langle Bu, u \rangle(t) - \frac{1}{2} \langle Bu_0, u \rangle + \int_0^t \| u(s) \|_V^p \, ds + \int_0^t \langle G_\mu u, u \rangle \, ds \leq \lambda \int_0^t \langle Bu, u \rangle \, ds + C(f) \tag{30.8.78}
\]

Now \( G_\mu \) is monotone and so,

\[
\langle G_\mu u, u \rangle = \langle G_\mu u - G_\mu 0, u \rangle + \langle G_\mu (0), u \rangle \geq \langle G_\mu (0), u \rangle \geq - |G(0)||u|
\]

It follows easily from standard manipulations and (30.8.78) that \( \| u \|_V \) is bounded independent of \( \mu \). That is, there is a constant \( C \) independent of \( \mu \) such that

\[
\| u \|_V \leq C \tag{30.8.79}
\]

The details follow. The above inequality (30.8.78) implies that by acting on \( u \chi_{[0,t]} \),

\[
\frac{1}{2} \langle Bu, u \rangle(t) - \frac{1}{2} \langle Bu_0, u \rangle + \int_0^t \| u(s) \|_V^p \, ds - \int_0^t |G(0)||u|_V \, ds \leq \lambda \int_0^t \langle Bu, u \rangle \, ds + C(f)
\]

Then by Gronwall’s inequality and adjusting constants,

\[
\langle Bu, u \rangle(t) + \int_0^t \| u(s) \|_V^p \, ds \leq C(u_0, f, \lambda) + C(\lambda) \int_0^t |G(0)||u|_V \, ds \tag{30.8.80}
\]

so it is clear that there is an inequality of the form

\[
\sup_{t \in [0,T]} \langle Bu, u \rangle(t) + \int_0^T \| u(s) \|_V^p \, ds \leq C(u_0, f, \lambda)
\]

Then returning to (30.8.78) all terms are bounded except \( \int_0^T \langle G_\mu u, u \rangle \, dt \), so this term must also be bounded for \( t = T \) also. Thus

\[
\left| \int_0^T \langle G_\mu u, u \rangle \, dt \right| \leq C
\]

where \( C \) is independent of \( \mu \). We denote by \( u_\mu \) the solution to the above equation.

Here is the definition of quasi-bounded.

**Definition 30.8.8** A set valued operator \( G \) is quasi-bounded if whenever \( x \in D(G) \) and \( x^* \in Gx \) are such that

\[ |\langle x^*, x \rangle|, \| x \| \leq M, \]

it follows that \( \| x^* \| \leq K_M \). Bounded would mean that if \( \| x \| \leq M \), then \( \| x^* \| \leq K_M \). Here you only know this if there is another condition.

**Assumption 30.8.9** \( G : D(G) \to \mathcal{P}(V') \) is quasi-bounded and maximal monotone.

By Proposition 27.6.18 an example of a quasi-bounded operator is a maximal monotone operator \( G \) for which \( 0 \in \text{int}(D(G)) \). Now \( G_\mu u_\mu \in GJ_\mu u_\mu \) as noted above. Therefore, there exists \( g_\mu \in G(J_\mu u_\mu) \) such that

\[
C \geq \langle G_\mu u_\mu, u_\mu \rangle_{V',V} = \langle g_\mu, u_\mu \rangle_{V',V} = \langle g_\mu, J_\mu u_\mu \rangle_{V',V} + \langle g_\mu, u_\mu - J_\mu u_\mu \rangle_{V',V} \tag{30.8.81}
\]

\[
\geq - |G(0)||J_\mu u_\mu|_V + \left( - \frac{1}{\mu^{p-1}} F(J_\mu u_\mu - u_\mu), u_\mu - J_\mu u_\mu \right)_{V',V}
\]

\[
= - |G(0)||J_\mu u_\mu|_V + \frac{1}{\mu^{p-1}} \| J_\mu u_\mu - u_\mu \|_V^p \tag{30.8.82}
\]
Thus the fact that \( \|u_\mu\| \) is bounded independent of \( \mu \) implies that \( \|J_\mu u_\mu\| \) is also bounded and that in fact \( \|u_\mu - J_\mu u_\mu\|_V \to 0 \) as \( \mu \to 0 \). This follows from consideration of the last line of the above formula. Note also that
\[
\langle g_\mu, u_\mu - J_\mu u_\mu \rangle_{V', V} = \frac{1}{\mu^{p-1}} \|J_\mu u_\mu - u_\mu\|_V^p \text{ is bounded.} \tag{30.8.83}
\]

Then from \( \|u_\mu\| \) it follows that \( \langle g_\mu, J_\mu u_\mu \rangle_{V', V} \) is bounded. By the assumption that \( G \) is quasi-bounded, \( g_\mu \) must also be bounded.

Then we have shown
\[
Bu_\mu(t, \omega) + \int_0^t z_\mu(s, \omega) \, ds + \int_0^t g_\mu(s, \omega) \, ds = \int_0^t f(s, \omega) \, ds + Bu_0(\omega) \tag{30.8.84}
\]
where
\[
\|g_\mu\|_{V'} + \|z_\mu\|_{V'} + \sup_{t \in [0, T]} \langle Bu_\mu, u_\mu \rangle(t) + \|J_\mu u_\mu\|_V + \|u_\mu\|_V + \|\langle Bu_\mu \rangle\|_V' \leq C \tag{30.8.85}
\]
The last term in the sum being bounded follows from the integral equation and the fundamental theorem of calculus along with the boundedness \( f, g_\mu, z_\mu \). In addition to this, the estimate \( 30.8.82 \) implies
\[
\lim_{\mu \to 0} \|J_\mu u_\mu - u_\mu\|_V = 0. \tag{30.8.86}
\]

There is a subsequence, \( \mu \to 0 \) still denoted as \( \mu \) such that
\[
g_\mu \to g \text{ weakly in } V' \tag{30.8.87}
\]
\[
z_\mu \to z \text{ weakly in } V' \tag{30.8.88}
\]
\[
u_\mu \to u \text{ weakly in } V \tag{30.8.89}
\]
\[
J_\mu u_\mu \to u \text{ weakly in } V \tag{30.8.90}
\]
\[
(Bu_\mu)' \to (Bu)' \text{ weakly in } V' \tag{30.8.91}
\]
\[
Bu_\mu(t) \to Bu(t) \text{ weakly in } V' \tag{30.8.92}
\]
Now consider two of these for \( \mu \) and \( \nu \). Subtract and act on \( u_\mu - u_\nu \). Then one obtains
\[
\langle Bu_\mu - Bu_\nu, u_\mu - u_\nu \rangle(t) + \int_0^t \langle z_\mu - z_\nu, u_\mu - u_\nu \rangle + \int_0^t \langle g_\mu - g_\nu, u_\mu - u_\nu \rangle = 0 \tag{30.8.93}
\]
Consider that last term for \( t = T \). It equals
\[
\int_0^T \langle G_\mu u_\mu - G_\nu u_\nu, u_\mu - u_\nu \rangle + \int_0^T \langle G_\mu u_\mu - G_\nu u_\nu, J_\mu u_\mu - J_\nu u_\nu \rangle + \int_0^T \langle g_\mu - g_\nu, u_\mu - J_\mu u_\mu - (u_\nu - J_\nu u_\nu) \rangle + \int_0^T \langle g_\mu - g_\nu, J_\mu u_\mu - J_\nu u_\nu \rangle = 0 \tag{30.8.94}
\]
where
\[
|\varepsilon(\mu, \nu)| \leq \left( \int_0^T (\|g_\mu\| + \|g_\nu\|)^p \right)^{1/p} \left( \int_0^T (\|u_\mu - J_\mu u_\mu\| + \|u_\nu - J_\nu u_\nu\|)^p \right)^{1/p} \leq 2C \left( \|u_\mu - J_\mu u_\mu\|_V + \|u_\nu - J_\nu u_\nu\|_V \right)
\]
Adjusting constants and using \( 30.8.83 \),
\[
\leq C \left( \mu^{1-(1/p)} + \nu^{1-(1/p)} \right)
\]
Thus
\[ \int_0^T \langle G\mu u_\mu - G\nu u_\nu, u_\mu - u_\nu \rangle = \int_0^T \langle g_\mu - g_\nu, J_\mu u_\mu - J_\nu u_\nu \rangle + \varepsilon(\mu, \nu) \]
where \( \lim_{\mu, \nu \to 0} \varepsilon(\mu, \nu) = 0 \). It follows from (30.8.10)

\[ \limsup_{\mu, \nu \to 0} \left( \int_0^T \langle z_\mu - z_\nu, u_\mu - u_\nu \rangle ds + \varepsilon(\mu, \nu) \right) \]
\[ = \limsup_{\mu, \nu \to 0} \left( \int_0^T \langle z_\mu - z_\nu, u_\mu - u_\nu \rangle ds \right) \leq 0 \]

From Lemma (30.8.11),
\[ \limsup_{\mu \to 0} \langle z_\mu, u_\mu - u \rangle_{\nu', V} \leq 0 \]
By the limit condition for \( A(\cdot, \nu) \), for each \( v \in V \), there exists \( z(v) \in Au \) such that
\[ \liminf_{\mu \to 0} \langle z_\mu, u_\mu - v \rangle = \liminf_{\mu \to 0} (\langle z_\mu, u_\mu - u \rangle + \langle z_\mu, u - v \rangle) \]
\[ = \langle z, u - v \rangle \geq \langle z(v), u - v \rangle \]
Since \( A(u, \nu) \) is convex and closed, separation theorems imply that \( z \in Au \). Return to the equation solved.
\[ \langle Bu_\mu \rangle' + z_\mu + g_\mu = f \]
Then act on \( u_\mu - u \) and use monotonicity arguments to write
\[ \langle (Bu)' - u, u_\mu - u \rangle_{\nu', V} + \langle z_\mu, u_\mu - u \rangle_{\nu', V} + \langle g_\mu, u_\mu - u \rangle_{\nu', V} \leq \langle f, u_\mu - u \rangle_{\nu', V} \] (30.8.94)
Then it was shown above that
\[ 0 \geq \limsup_{\mu \to 0} \langle z_\mu, u_\mu - u \rangle_{\nu', V} \geq \liminf_{\mu \to 0} \langle z_\mu, u_\mu - u \rangle_{\nu', V} \geq \langle z(u), u - u \rangle_{\nu', V} = 0 \]
and so, from (30.8.11),
\[ \lim_{\mu \to 0} \langle g_\mu, u_\mu - u \rangle_{\nu', V} = \lim_{\mu \to 0} \langle g_\mu, J_\mu u_\mu - u \rangle_{\nu', V} = 0 \]
and so
\[ \lim_{\mu \to 0} \langle g_\mu, J_\mu u_\mu \rangle_{\nu', V} = \langle g, u \rangle_{\nu', V} \]
Now let \([a, b] \in G(G)\). Then
\[ \langle b - g, a - u \rangle = \lim_{\mu \to 0} \langle b - g_\mu, a - J_\mu u_\mu \rangle \geq 0 \]
because \( g_\mu \in G(J_\mu u_\mu) \). Since \( G \) is maximal monotone, it follows that \([u, g] \in G(G)\).

This has shown that for each \( \omega \) fixed, and every sequence of solutions to the integral equation \( \{u_\mu\} \), each function \( \{Bu_\mu\} \) being product measurable by Theorem (30.8.11), there exists a subsequence which converges to a solution \( u \) to the integral equation. In particular, \( t \to Bu(t) \) is weakly continuous into \( V' \). Then by the fundamental measurable selection theorem, Theorem (30.8.11), there exists a product measurable function \( \bar{u}(t, \omega) \) with values in \( V \) weakly continuous in \( t \) and a sequence depending on \( \omega \), \( \{u_\mu(\omega)\} \) such that for each \( \omega, \lim_{\mu(\omega) \to 0} u_\mu(\omega) (\cdot, \omega) = \bar{u}(\cdot, \omega) \) weakly in \( V \). However, from the above argument, for each \( \omega \), there is a further subsequence, still denoted with subscript \( \mu(\omega) \) such that in \( V' \),
\[ \lim_{\mu(\omega) \to 0} u_\mu(\omega) (\cdot, \omega) = u(\cdot, \omega) \]
where \( u \) is a solution to the integral equation. Since \( u(\cdot, \omega) = \bar{u}(\cdot, \omega) \) in \( V \) it follows that these must be equal a.e. and hence \( t, \omega \to u(t, \omega) \) is product measurable. This proves the following theorem.

**Theorem 30.8.10** Suppose \((30.8.7)\) and \( B \) as described above and \( u_0 \) is \( F \) measurable. Also let \( G : D(G) \subseteq V \to \mathcal{P}(V') \) be maximal monotone and quasi-bounded.

Then, there exists a solution \( u \) of the integral equation
\[ Bu(t, \omega) + \int_0^t z(s, \omega) ds + \int_0^t g(s, \omega) ds = \int_0^t f(s, \omega) ds + Bu_0(\omega) \]
where \((t, \omega) \to u(t, \omega) \) is product measurable \((t, \omega) \to z(t, \omega) \) also. Moreover, for each \( \omega \), \( Bu(t, \omega) = B(u(t, \omega)) \) a.e. \( t \) and \( z(\cdot, \omega) \in A(u(\cdot, \omega), \omega) \), and \( g(\cdot, \omega) \in G(u(\cdot, \omega)) \) for each \( \omega \).
Note that in the case of most interest where you have a Gelfand triple and $B$ is the identity, the fundamental theorem of calculus implies easily that $\omega \to z(s, \omega) + g(s, \omega)$ is measurable for a.e. $s$. One can also generalize to the following in which a measurable $q(t, \omega)$ is added.

**Corollary 30.8.11** Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ and $B$ as described above and $u_0$ is $\mathcal{F}$ measurable. Also let $G : D(G) \subseteq V \to \mathcal{P}(V')$ be maximal monotone and quasi-bounded. Let $(t, \omega) \to q(t, \omega)$ be product measurable into $V$ and let $t \to q(t, \omega)$ be continuous, $q(0, \omega) = 0$. Then, there exists a solution $u$ of the integral equation

$$Bu(t, \omega) + \int_0^t z(s, \omega) ds + \int_0^t g(s, \omega) ds = \int_0^t f(s, \omega) ds + Bu_0(\omega) + Bq(t, \omega),$$

where $(t, \omega) \to u(t, \omega), (t, \omega) \to z(t, \omega), g(t, \omega)$ are product measurable. Moreover, for each $\omega$, $Bu(t, \omega) = B(u(t, \omega)) a.e. t$ and $z(\cdot, \omega) \in A(u(\cdot, \omega), \omega)$ for a.e. $t$, and $g(\cdot, \omega) \in G(u(\cdot, \omega))$ for each $\omega$.

**Proof:** Define a stopping time

$$\tau_n(\omega) = \inf \{ t : q(t, \omega) > n \}$$

Then let $\tilde{A}(\cdot, \omega) \equiv A(q^{\tau_n}(\cdot, \omega) + w, \omega)$. Then $\tilde{A}$ satisfies the same properties as $A$ and so there exists a solution to the integral equation

$$Bw_n(t, \omega) + \int_0^t z_n(s, \omega) ds + \int_0^t g_n(s, \omega) ds = \int_0^t f(s, \omega) ds + Bu_0(\omega)$$

where $w_n, z_n, g_n$ are product measurable, $z_n(\cdot, \omega) \in \tilde{A}(w_n(\cdot, \omega), \omega)$ a.e. $t$. By continuity of $t \to q(t, \omega), \tau_n = \infty$ for all $n$ sufficiently large and so $q(t, \omega) = q^{\tau_n}(t, \omega)$. As before, for each $\omega$, one obtains the convergences $\mathcal{U}_n \to \mathcal{U}$ as $n \to \infty$. As before, for each $\omega, z(\cdot, \omega) \in \tilde{A}(w(\cdot, \omega), \omega) a.e. t \to w(t, \omega)$ is the function to which $w_n(\cdot, \omega)$ converges weakly. Note that the estimates allowing this to happen are dependent on $\omega$. However, one can apply Theorem 30.8.11 as before and obtain a solution to

$$Bu(t, \omega) + \int_0^t z(s, \omega) ds + \int_0^t g(s, \omega) ds = \int_0^t f(s, \omega) ds + Bu_0(\omega)$$

such that $w, z, g$ are product measurable into $V$ or $V'$ and $z(\cdot, \omega) \in \tilde{A}(w(\cdot, \omega), \omega)$. Now let $u(t, \omega) = w(t, \omega) + q(t, \omega)$ to obtain the existence of the desired solution in the corollary. \(\blacksquare\)

### 30.9 Including Stochastic Integrals

You can include stochastic integrals in the above formulation. In this section and from now on, we will assume that $W$ is a Hilbert space because the stochastic integrals featured here will have values in $W$ and the version of the stochastic integral to be considered here will be the Ito integral. Here is a brief review of this integral.

Let $U$ be a separable real Hilbert space and let $Q : U \to U$ be self adjoint and nonnegative. Also $H$ will be a separable real Hilbert space. $L_2(Q^{1/2}U, H)$ will denote the Hilbert Schmidt operators which map $Q^{1/2}U$ to $H$. Here $Q^{1/2}U$ is the Hilbert space which has an inner product given by

$$(y, z) \equiv \left( Q^{-1/2}y, Q^{-1/2}z \right)$$

where $Q^{-1/2}y$ denotes $x$ such that $Q^{1/2}x = y$ and out of all such $x$, this is the one which has the smallest norm. It is like the Moore Penrose inverse in linear algebra. Then one can define a stochastic integral

$$\int_0^t \Phi dW$$

where $\Phi \in L^2([0, T] \times \Omega; L_2(Q^{1/2}U, H))$ where here $\Phi$ is progressively measurable with respect to the filtration $\mathcal{F}_t$. This filtration will be

$$\mathcal{F}_t = \cap_{p>1} \sigma \left( W(r) - W(s) : 0 \leq s \leq r \leq p \right)$$

The horizontal line indicates completion. The symbol

$$\sigma \left( W(r) - W(s) : 0 \leq s \leq r \leq p \right)$$

indicates the smallest $\sigma$ algebra for which all those increments are measurable. Here $W(t)$ is a Wiener process which has values in $U_1$, some other Hilbert space, maybe $H$. There is a Hilbert Schmidt operator $J \in L_2(Q^{1/2}U, U_1)$ such that $W(t) = \sum_{i=1}^{\infty} \psi_i(t) Je_i$ where here the $\psi_i$ are independent real Wiener processes. You could take $U, U_1$ to both be $H$. This is following [72]. Then the Stochastic integral has the following properties.
1. \( \int_0^t \Phi dW \) is a martingale with respect to \( \mathcal{F}_t \) with values in \( H \), equal to 0 when \( t = 0 \).

2. One has the Ito isometry

\[
E \left( \left\| \int_0^t \Phi dW \right\|^2_H \right) = \int_0^t \| \Phi \|^2_{L^2} ds
\]

3. One can localize as follows. For \( \tau \) a stopping time,

\[
\int_0^{t \land \tau} \Phi dW = \int_0^t \mathcal{A}_{[0, \tau]} \Phi dW
\]

4. One can also generalize to the case where \( \Phi \) is only progressively measurable and instead of being in \( L^2 ([0, T] \times \Omega; \mathcal{L}_2 (Q^{1/2} U, H)) \), you have only that

\[
P \left( \int_0^T \| \Phi (t) \|^2_{L^2} dt < \infty \right) = 1
\]

This is done by using an appropriate sequence of stopping times called a localizing sequence. More generally a local martingale is a stochastic process \( M(t) \) adapted to the filtration for which there is a localizing sequence of stopping times \( \{\tau_n\} \) such that \( \lim_{n \to \infty} \tau_n = \infty \) and \( M^{\tau_n} \) is a martingale. Local martingales will occur in the estimates which are encountered in what follows.

5. Denoting by \( M(t) \) the stochastic integral, \( M(t) = \int_0^t \Phi dW \), the quadratic variation is given by

\[
[M](t) = \int_0^t \| \Phi \|^2_{L^2} ds
\]

6. We will also need a part of the Burkholder Davis Gundy inequality \[53\], Theorem \[14.4.4\] which in terms of this stochastic integral is of the form

\[
\int_{\Omega} M^* dP \leq CE \left( \left( \int_0^T \| \Phi \|^2_{L^2} ds \right)^{1/2} \right)^2, \quad C \text{ some constant}
\]

where \( M(t) \) is the above stochastic integral and

\[
M^* \equiv \sup \{ \| M(t) \|_H : t \in [0, T] \}
\]

Now let \( \Phi \in L^2 ([0, T] \times \Omega; \mathcal{L}_2 (Q^{1/2} U, W)) \). Let an orthonormal basis for \( Q^{1/2} U \) be \( \{g_i\} \) and an orthonormal basis for \( W \) be \( \{f_i\} \). Then \( \{f_i \otimes g_j\} \) is an orthonormal basis for \( \mathcal{L}_2 (Q^{1/2} U, W) \). Hence,

\[
\Phi = \sum_i \sum_j \Phi_{ij} f_i \otimes g_j
\]

where \( f_i \otimes g_j (y) \equiv (g_j, y)_{Q^{1/2} U} f_i \). Let \( E \) be a separable real Hilbert space which is dense in \( V \). Then without loss of generality, one can assume that the orthonormal basis for \( W \) are all vectors in \( E \). Thus the orthogonal projection of \( \Phi \) onto the closed subspace span \( \{f_i \otimes g_i, i, j \leq n\} \) given by

\[
\Phi_n = \sum_{i=1}^n \sum_{j=1}^n \Phi_{ij} f_i \otimes g_j
\]

Then \( \Phi_n \in L^2 ([0, T] \times \Omega; \mathcal{L}_2 (Q^{1/2} U, E)) \) and also

\[
\lim_{n \to \infty} \| \Phi_n - \Phi \|_{L^2([0, T] \times \Omega; \mathcal{L}_2 (Q^{1/2} U, W))} = 0
\]

and \( \int_0^t \Phi_n dW \) is continuous and progressively measurable into \( E \) hence into \( V \). The following corollary will be useful.
Corollary 30.9.1 Let $\| \Phi_n (t, \omega) \|_{L_2 (Q^{1/2} U, W)} \leq \| \Phi (t, \omega) \|_{L_2 (Q^{1/2} U, W)}$ where $\| \Phi_n (t, \omega) \| \uparrow \| \Phi (t, \omega) \|$

$\Phi \in L^2 \left( \Omega; L^\infty \left( [0, T], L_2 \left( Q^{1/2} U, W \right) \right) \right) \cap L^2 \left( [0, T] \times \Omega, L_2 \left( Q^{1/2} U, W \right) \right)$

Then off a set of measure zero, the stochastic integrals $\int_0^t \Phi_n dW$ satisfy

$$\sup_n \sup_{t \neq s} \frac{\left\| \int_s^t \Phi_n dW \right\|}{|t - s|^\gamma} < C (\omega), \gamma < 1/2$$

**Proof:** Let $\alpha$ be large. Then by the Burkholder Davis Gundy inequality,

$$\int_\Omega \left( \left\| \int_s^t \Phi_n dW \right\| \right)^\alpha dP \leq C \int_\Omega \left( \int_s^t \| \Phi_n \|^2 d\tau \right)^{\alpha/2} dP$$

$$\leq C \int_\Omega \| \Phi_n \|_{L^\infty ([0, T], L_2 (Q^{1/2} U, W))} |t - s|^{\alpha/2}$$

$$\leq C |t - s|^{\alpha/2}$$

Then by the Kolmogorov Čentsov theorem,

$$E \left( \sup_{0 \leq s < t \leq T} \frac{\left\| \int_s^t \Phi_n dW \right\|}{(t - s)^\gamma} \right) < C < \infty$$

where $\gamma < \beta/\alpha$ where, $\beta + 1 = \alpha/2$. Thus

$$\sup_n E \left( \sup_{0 \leq s < t \leq T} \frac{\left\| \int_s^t \Phi_n dW \right\|}{(t - s)^\gamma} \right) < C$$

By Proposition 30.2.18 there is an enlarged set of measure zero and a subsequence, still denoted with $n$ such that off the set of measure zero,

$$\sup_n \left( \sup_{0 \leq s < t \leq T} \frac{\left\| \int_s^t \Phi_n dW \right\|}{(t - s)^\gamma} \right) \leq C (\omega) \quad \blacksquare$$

Recall the following conditions for the various operators.

**Bounded and coercive conditions**

$A (\cdot, \omega) : A (\cdot, \omega) : V_I \to V'_I$ for each $I$ a subinterval of $[0, T]$ $I = [0, \tilde{T}], \tilde{T} \leq T$

$$A (\cdot, \omega) : V_I \to P (V'_I) \text{ is bounded,} \quad (30.9.95)$$

If, for $u \in V,$

$$u^* X_{[0, \tilde{T}]} \in A \left( u X_{[0, \tilde{T}], \omega} \right)$$

for each $\tilde{T}$ in an increasing sequence converging to $T$, then

$$u^* \in A (u, \omega) \quad (30.9.96)$$

Assume the specific estimate

$$\sup \left\{ \left\| u^* \right\|_{V'_I} : u^* \in A (u, \omega) \right\} \leq a (\omega) + b (\omega) \| u \|_{V_I}^{p - 1}$$

(30.9.97)
where \( a(\omega), b(\omega) \) are nonnegative. Also assume the following coercivity estimate valid for each \( t \leq T \) and for some \( \lambda(\omega) \geq 0 \),
\[
\inf \left( \int_0^t (u^*, u) + \lambda(\omega) (Bu, u) \, dt : u^* \in A(u, \omega) \right) \geq \delta(\omega) \int_0^t \|u\|_{V'}^2 \, ds - m(\omega) \tag{30.9.98}
\]
where \( m(\omega) \) is some nonnegative constant, \( \delta(\omega) > 0 \).

**Monotonicity**

It will also be assumed that \( \lambda(\omega) B + A \) is monotone in the sense that
\[
\int_0^t \langle \lambda(\omega) Bu + u^* - \lambda(\omega) Bv + v^*, u - v \rangle \, ds \geq 0
\]
for a suitable choice of \( \lambda(\omega) \) whenever \( u^* \in A(u, \omega), v^* \in A(v, \omega) \).

**Limit condition**

Let \( U \) be a Banach space dense in \( V \) and that if \( u_i \rightharpoonup u \) in \( V_I \) and \( u_i^* \in A(u_i) \) with \( u_i^* \rightharpoonup u^* \) in \( V_I' \) and \( t \to Bu_i(t) \) is continuous and
\[
\sup_i \sup_{t \neq s} \frac{\|Bu_i(t) - Bu_i(s)\|_{U'}}{|t - s|} \leq C \tag{30.9.99}
\]
then if
\[
\limsup_{i \to \infty} \langle u_i^*, u_i - u \rangle_{V_I', V_I} < 0 \tag{30.9.100}
\]
it follows that for all \( v \in V_I \), there exists \( u^*(v) \in Au \) such that
\[
\liminf_{i \to \infty} \langle u_i^*, u_i - v \rangle_{V_I', V_I} \geq \langle u^*(v), u - v \rangle_{V_I', V_I} \tag{30.9.101}
\]
As to \( B(\omega) \), it is \( k(\omega) B \) where \( B \in \mathcal{L}(W, W') \) and is self adjoint and nonnegative where \( k \) is \( \mathcal{F}_0 \) measurable.

**Progressively measurable condition**

**Condition 30.9.2** For each \( t \leq T \), if \( \omega \to u(\cdot, \omega) \) is measurable into \( V_{[0, t]} \), then there exists a \( \mathcal{F}_t \) measurable selection of \( A(u(\cdot, \omega), \omega) \) into \( V_{[0, t]}' \).

Then there is a theorem. It was Theorem \ref{thm:progressively-measurable} which gave existence and uniqueness of progressively measurable solutions \( u \) to the integral equation.

**Theorem 30.9.3** Assume the above conditions, \ref{cond:30.7.4} - \ref{cond:30.9.2} along with the progressive measurability condition \ref{cond:30.9.2}. Let \( u_0 \) be \( \mathcal{F}_0 \) measurable and \( \omega \to B(\omega) \) also \( \mathcal{F}_0 \) measurable and \( (t, \omega) \to \mathcal{X}_{[0,t]}(t) f(t, \omega) \) is \( \mathcal{B}([0,t]) \times \mathcal{F}_t \) product measurable into \( V' \) for each \( t \).

\[
B(\omega) = k(\omega) B, \quad k(\omega) \geq 0, k \text{ measurable.}
\]

Also let \( t \to q(t, \omega) \) be continuous and \( q \) is progressively measurable into \( V \). Suppose there is at most one solution to
\[
Bu(t, \omega) + \int_0^t z(s, \omega) \, ds = \int_0^t f(s, \omega) \, ds + Bu_0(\omega) + Bq(t, \omega), \tag{30.9.102}
\]
for each \( \omega \). Then the solution to the above integral equation \( u \) is progressively measurable. Moreover, for each \( \omega \), both \( Bu(t, \omega) = B(u(t, \omega)) \) for a.e. \( t \) and \( z(\cdot, \omega) \in A(u(t, \omega), \omega) \). Also, for each \( a \in [0, T] \),
\[
Bu(t, \omega) + \int_a^t z(s, \omega) \, ds = \int_a^t f(s, \omega) \, ds + Bu(a, \omega) + Bq(t, \omega) - Bq(a, \omega)
\]

Letting \( q(t) = \int_0^t \Phi_u \, dW \) defined above with the filtration also being the one obtained from the Wiener process, this implies the following theorem. The \( \sigma \) algebra of progressively measurable sets will be denoted by \( \mathcal{P} \).
**Theorem 30.9.4** Assume the above conditions, along with the progressive measurability condition. Also assume there is at most one solution to where
\[ q(t, \cdot) \equiv \int_0^t \Phi_n dW \]
Then there exists a \( P \) measurable \( u_n \) such that also \( z_n \) is progressively measurable
\[ Bu_n(t, \omega) - Bu_0(\omega) + \int_0^t z_n(s, \omega) \, ds = \int_0^t f(s, \omega) \, ds + B \int_0^t \Phi_n dW \]
where for each \( \omega \), \( z_n(\cdot, \omega) \in A(u_n(\cdot, \omega), \omega) \). The function \( Bu_n(t, \omega) = B(u_n(t, \omega)) \) for a.e. \( t \).

This gives an existence theorem for the inclusion of a stochastic integral. However, it is desired to get a similar result for \( \Phi \) rather than \( \Phi_n \). Next is the Ito formula which is usable because of the progressive measurability of \( u_n, z_n \). This formula applies to the following situation.

**Situation 30.9.5** Let \( X \) have values in \( V \) and satisfy the following
\[ BX(t) = BX_0 + \int_0^t Y(s) \, ds + B \int_0^t Z(s) \, dW(s), \quad (30.9.103) \]
\( X_0 \in L^2(\Omega; W) \) and is \( \mathcal{F}_0 \) measurable, where \( Z \) is \( \mathcal{L}_2(Q^{1/2}U, W) \) progressively measurable and
\[ \|Z\|_{\mathcal{L}_2([0,T] \times \Omega; \mathcal{L}_2(Q^{1/2}U,W))} < \infty. \]
This is what is needed to define the stochastic integral in the above formula.

Assume \( X, Y \) satisfy
\[ BX, Y \in K' \equiv L^p([0, T] \times \Omega; V'), \]
the \( \sigma \) algebra of measurable sets defining \( K' \) will be the progressively measurable sets. Here \( 1/p' + 1/p = 1, \ p > 1 \).

Also the sense in which the equation holds is as follows. For a.e. \( \omega \), the equation holds in \( V' \) for all \( t \in [0, T] \). Thus we are considering a particular representative \( X \) of \( K \) for which this happens. Also it is only assumed that \( BX(t) = B(X(t)) \) for a.e. \( t \). Thus \( BX \) is the name of a function having values in \( V' \) for which \( BX(t) = B(X(t)) \) for a.e. \( t \), all \( t \notin N_\omega \) a set of measure zero. Assume that \( X \) is progressively measurable also and \( X \in L^p([0, T] \times \Omega; V) \).

Then in the above situation, we obtain the following integration by parts formula which is called the Ito formula. This particular version is presented in Theorem and is a generalization of work of Krylov. A proof of the special case of a Gelfand triple in which \( B = I \) is in [17].

**Theorem 30.9.6** In Situation \( 30.9.4 \), for \( \omega \) off a set of measure zero, for every \( t \in N_\omega^C \), the measure of \( N_\omega \) equalling 0,
\[ \langle BX(t), X(t) \rangle = \langle BX_0, X_0 \rangle + \int_0^t 2 \langle Y(s), X(s) \rangle \, ds + \int_0^t \langle BZ, Z \rangle_{\mathcal{L}_2} \, ds + 2M(t) \quad (30.9.104) \]
where \( M(t) \) is a stochastic integral and a local martingale equal to 0 when \( t = 0 \). Also, there exists a unique continuous, progressively measurable function denoted as \( \langle BX, X \rangle \) such that it equals \( \langle BX(t), X(t) \rangle \) for a.e. \( t \) and \( \langle BX, X \rangle(t) \) equals the right side of the above for all \( t \). In addition to this,
\[ E(\langle BX(t), X(t) \rangle) = \]
\[ E(\langle BX_0, X_0 \rangle) + E\left( \int_0^t 2 \langle Y(s), X(s) \rangle + \langle BZ, Z \rangle_{\mathcal{L}_2} \, ds \right) \quad (30.9.105) \]
Also the quadratic variation of \( M(t) \) in \( 30.9.102 \) is dominated by
\[ C \int_0^t \|Z\|^2_{\mathcal{L}_2} \|BX\|^2_{W'} \, ds \quad (30.9.106) \]
for a suitable constant \( C \). Also \( t \to BX(t) \) is continuous with values in \( W' \) for \( t \in N_\omega^C \). In fact, this martingale can be written as
\[ \int_0^t (Z \circ J^{-1})^* BX \circ JdW \]
That ugly integral displayed above can be written in the form
\[ \int_0^t \langle BX, dN \rangle \]
where \( N(t) = \int_0^t Z(s) \, dW \).

Now we consider the meaning of the symbol \( \langle BZ, Z \rangle_{\mathcal{L}_2} \). You begin with a complete orthonormal set \( \{g_k\} \) in \( Q^{1/2}U \). Then to say that \( Z \) has values in \( \mathcal{L}_2(Q^{1/2}U; W) \) is to say that \( \sum_j \sum_i (Z(g_i), e_j)^2 = \sum_i \|Z(g_i)\|^2_W < \infty \) where \( \{e_j\} \) is an orthonormal basis in \( W \). You can let it be the one used earlier where each is actually in \( V \) or even in \( E \). Then the symbol means
\[ (R^{-1}BZ, Z)_{\mathcal{L}_2} \]
where \( R \) is the Riesz map from the Hilbert space \( W \) to its dual space. Thus it equals
\[ \sum_i (R^{-1}BZ(g_i), Z(g_i))_W = \sum_i \langle BZ(g_i), Z(g_i) \rangle \]
so it is seen to be nonnegative.

Now apply this Ito formula to Theorem [SN2] in which we make the assumptions there on \( \|u_0\| \in L^2(\Omega) \) and that \( f \in L^p([0,T] \times \Omega; V') \) where the \( \sigma \) algebra is \( \mathcal{P} \) the progressively measurable \( \sigma \) algebra, and
\[ \Phi \in L^2\left(\Omega, L^2\left([0,T], \mathcal{L}_2\left(Q^{1/2}U, W\right)\right)\right) \]
which implies the same is true of \( \Phi_n \). This yields, from the assumed estimates, an expression of the form where \( \delta > 0 \) is a suitable constant.
\[
\frac{1}{2} \langle Bu_n, u_n \rangle(t) + \frac{1}{2} \langle Bu_0, u_0 \rangle + \delta \int_0^t \|u_n(s)\|_V^p \, ds \\
\leq \lambda \int_0^t \langle Bu_n, u_n \rangle(s) \, ds + \int_0^t \langle f, u_n \rangle_{V', V} \, ds + \int_0^t c(s, \omega) \, ds \\
+ \int_0^t \langle B\Phi_n, \Phi_n \rangle_{\mathcal{L}_2} \, ds + M_n(t)
\]  
(30.9.107)

where \( c \in L^1([0,T] \times \Omega) \). Then taking expectations or using that part of the Ito formula,
\[
\frac{1}{2} E\left(\langle Bu_n, u_n \rangle(t)\right) + \lambda E\left(\int_0^T \|u_n(s)\|_V^p \, ds\right) \\
\leq \lambda E\left(\int_0^T \langle Bu_n, u_n \rangle(s) \, ds\right) + \int_0^T E\left(\langle f, u_n \rangle_{V', V}\right) \, ds + C(\Phi, u_0)
\]
Then by Gronwall’s inequality and some simple manipulations,
\[
E\left(\langle Bu_n, u_n \rangle(t)\right) + E\left(\int_0^T \|u_n(s)\|_V^p \, ds\right) \leq C(T, f, u_0, \Phi)
\]
By Proposition [HMR1], there is a subsequence, still denoted with \( n \) and a set of measure zero \( N \) such that for \( \omega \notin N \),
\[
\int_0^T \|u_n(s, \omega)\|_V^p \, ds \leq C(\omega) < \infty
\]
This is a nice result but one is needed on \( \sup_t \langle Bu_n, u_n \rangle(t) \) also. This is because you end up having to estimate \( M^*(t) \).

Then using obvious estimates and Gronwall’s inequality in [HMR1], this yields an inequality of the form
\[
\langle Bu_n, u_n \rangle(t) - \langle Bu_0, u_0 \rangle + \int_0^t \|u_n(s)\|_V^p \, ds \leq C(f, \lambda, c) + \|B\| \int_0^t \|\Phi_n\|_{\mathcal{L}_2}^2 \, ds + M_n^*(t)
\]
where the random variable \( C(f, \lambda, c) \) is nonnegative and is integrable. Now \( t \to M_n^*(t) \) is increasing as is the integral on the right. Hence it follows that, modifying the constants,
\[
\sup_{s \in [0,t]} \langle Bu_n, u_n \rangle(s) + \int_0^t \|u_n(s)\|_V^p \, ds
\]
\[ \leq C(f, \lambda, c, u_0) + 2\|B\| \int_0^t \|\Phi_n\|_{\mathcal{L}_2}^2 \, ds + 2M_n^* (t) \quad (30.9.108) \]

Next take the expectation of both sides and use the Burkholder Davis Gundy inequality along with the description of the quadratic variation of the martingale \( M_n(t) \). This yields

\[
E \left( \sup_{s \in [0,t]} \langle Bu_n, u_n \rangle (s) \right) + E \left( \int_0^t \|u_n (s)\|_V^p \, ds \right) \\
\leq C + 2\|B\| \left( \int_0^t \|\Phi_n\|_{\mathcal{L}_2}^2 \, ds \right) \\
+ C \int_\Omega \left( \int_0^t \|Bu_n\|_W^2 \|\Phi_n\|_{\mathcal{L}_2}^2 \, ds \right)^{1/2} \, dP
\]

Now \( \|Bu\| = \sup_{\|v\| \leq 1} \langle Bu, v \rangle \leq \langle Bu, w \rangle^{1/2} \). Also \( \int_0^t \|\Phi_n\|_{\mathcal{L}_2}^2 \, ds \leq \int_0^T \|\Phi\|_{\mathcal{L}_2}^2 \, ds \) and so the above inequality implies

\[
E \left( \sup_{s \in [0,t]} \langle Bu_n, u_n \rangle (s) \right) + E \left( \int_0^t \|u_n (s)\|_V^p \, ds \right) \\
\leq C(f, \lambda, c, \Phi) + C \int_\Omega \sup_{s \in [0,t]} \langle Bu_n, u_n \rangle^{1/2} (s) \left( \int_0^t \|\Phi\|_{\mathcal{L}_2}^2 \right)^{1/2} \, dP
\]

Then adjusting the constants yields

\[
\frac{1}{2} E \left( \sup_{s \in [0,T]} \langle Bu_n, u_n \rangle (s) \right) + E \left( \int_0^T \|u_n (s)\|_V^p \, ds \right) \\
\leq C + C \int_0^T \|\Phi\|_{\mathcal{L}_2}^2 \, dt dP = C 
\]

If needed, you could use a stopping time to be sure that \( E \left( \sup_{s \in [0,T]} \langle Bu_n, u_n \rangle (s) \right) < \infty \) and then let it converge to \( \infty \).

From the integral equation,

\[
Bu_n(t) - Bu_m(t) + \int_0^t z_n - z_m \, ds = B \int_0^t (\Phi_n - \Phi_m) \, dW
\]

Then using the monotonicity assumption and the Ito formula,

\[
\frac{1}{2} \langle Bu_n - Bu_m, u_n - u_m \rangle (t) \leq \lambda \int_0^t \langle Bu_n - Bu_m, u_n - u_m \rangle \, ds
\]

\[
+ \int_0^t \langle B(\Phi_n - \Phi_m), \Phi_n - \Phi_m \rangle \, d + \int_0^t \langle (\Phi_n - \Phi_m) \circ J^{-1} \rangle B(\Phi_n - \Phi_m) \circ J \, dW
\]

and so, from Gronwall’s inequality, there is a constant \( C \) which is independent of \( m, n \) such that

\[
\langle Bu_n - Bu_m, u_n - u_m \rangle (t) \leq CM_{nm}(t) \leq CM_{nm}^* (t) + C \int_0^t \|\Phi_n - \Phi_m\|_{\mathcal{L}_2}^2 \, ds
\]

where \( M_{nm} \) refers to that local martingale on the right. Thus also

\[
\sup_{t \in [0,T]} \langle Bu_n - Bu_m, u_n - u_m \rangle (t) \leq CM_{nm}(t) \leq CM_{nm}^* (t) + C \int_0^T \|\Phi_n - \Phi_m\|_{\mathcal{L}_2}^2 \, ds
\]

Taking the expectation and using the Burkholder Davis Gundy inequality again in a similar manner to the above, we obtain

\[
E \left( \sup_{t \in [0,T]} \langle Bu_n - Bu_m, u_n - u_m \rangle (t) \right) \leq C \int_\Omega \int_0^T \|\Phi_n - \Phi_m\|_{\mathcal{L}_2}^2 \, dt dP
\]
Now the right side converges to 0 as $m, n \to \infty$ and so there is a subsequence, denoted with the index $k$ such that whenever $m > k$,

$$E \left( \sup_{t \in [0,T]} \langle Bu_k - Bu_m, u_k - u_m \rangle (t) \right) \leq \frac{1}{2^k}$$

Note how this implies

$$\int_{\Omega} \int_0^T \langle Bu_k - Bu_m, u_k - u_m \rangle \, dt \, dP \leq \frac{T}{2^k} \quad (30.9.111)$$

Then consider the martingales $M_k(t)$ considered earlier. One of these is of the form

$$M_k = \int_0^t (\Phi_k \circ J^{-1})^* Bu_k \circ J dW$$

Then by the Burkholder Davis Gundy inequality and modifying constants as appropriate,

$$E \left( (M_k - M_{k+1})^* \right)$$

$$\leq C \int_{\Omega} \left( \int_0^T \| (\Phi_k \circ J^{-1})^* Bu_k - (\Phi_{k+1} \circ J^{-1})^* Bu_{k+1} \|^2 \, dt \right)^{1/2} \, dP$$

$$\leq C \int_{\Omega} \left( \int_0^T \| \Phi_k - \Phi_{k+1} \|^2 \langle Bu_k, u_k \rangle \, dt \right)^{1/2} \, dP$$

$$+ C \int_{\Omega} \left( \int_0^T \| \Phi_{k+1} \|^2 \langle Bu_k - Bu_{k+1}, u_k - u_{k+1} \rangle \, dt \right)^{1/2} \, dP$$

$$\leq C \int_{\Omega} \sup_t \langle Bu_k, u_k \rangle^{1/2} \left( \int_0^T \| \Phi_k - \Phi_{k+1} \|^2 \, dt \right)^{1/2} \, dP$$

$$+ C \int_{\Omega} \sup_t \langle Bu_k - Bu_{k+1}, u_k - u_{k+1} \rangle^{1/2} \left( \int_0^T \| \Phi_{k+1} \|^2 \, dt \right)^{1/2} \, dP$$

$$\leq C \left( \int_{\Omega} \sup_t \langle Bu_k, u_k \rangle \, dP \right)^{1/2} \left( \int_{\Omega} \int_0^T \| \Phi_k - \Phi_{k+1} \|^2 \, dt \, dP \right)^{1/2}$$

$$+ C \left( \int_{\Omega} \sup_t \langle Bu_k - Bu_{k+1}, u_k - u_{k+1} \rangle \, dP \right)^{1/2} \left( \int_{\Omega} \int_0^T \| \Phi_{k+1} \|^2 \, dt \, dP \right)^{1/2}$$

From the above inequality, and after adjusting the constants, the above is no larger than an expression of the form $C \left( \frac{1}{2} \right)^{k/2}$ which is a summable sequence. Then

$$\sum_k \int_{\Omega} \sup_{t \in [0,T]} |M_k(t) - M_{k+1}(t)| \, dP < \infty$$

Thus $\{M_k\}$ is a Cauchy sequence in $M_T^1$ and so there is a continuous martingale $M$ such that

$$\lim_{k \to \infty} E \left( \sup_t |M_k(t) - M(t)| \right) = 0$$

Taking a further subsequence if needed, one can also have

$$P \left( \sup_t |M_k(t) - M(t)| > \frac{1}{k} \right) \leq \frac{1}{2^k}$$
and so by the Borel Cantelli lemma, there is a set of measure zero such that off this set, \( \sup_k |M_k(t) - M(t)| \) converges to 0. Hence for such \( \omega \), \( M_k^\omega(T) \) is bounded independent of \( k \) Thus for \( \omega \) off a set of measure zero, \( \text{LEMMA} \) implies that for such \( \omega \),

\[
\sup_{s \in [0,T]} \langle Bu_r, u_r \rangle(s) + \int_0^T \|u_r(s)\|_V^p \, ds \leq C(\omega)
\]

where \( C(\omega) \) does not depend on the index \( r \), this for the subsequence just described which will be the sequence of interest in what follows. Using the boundedness assumption for \( A \), one also obtains an estimate of the form

\[
\sup_{s \in [0,T]} \langle Bu_r, u_r \rangle(s) + \int_0^T \|u_r(s)\|_V^p \, ds + \int_0^T \|z_r\|_{V'}^p \leq C(\omega) \quad (30.9.112)
\]

The idea here is to take weak limits converging to a function \( u \) and then identify \( z(\cdot, \omega) \) as being in \( A(u, \omega) \) but this will involve a difficulty. It will require a use of the above Ito formula and this will need \( u \) to be progressively measurable. By uniqueness, it would seem that this could be concluded by arguing that one does not need to take a subsequence due to uniqueness but the problem is that we won’t know the limit of the sequence is a solution unless we use the Ito formula. This is why we make the extra assumption that for \( z(\cdot, \omega) \in A(u, \omega) \) and for all \( \lambda \) large enough,

\[
\langle \lambda Bu_1 + z_1 - (\lambda Bu_2 + z_2), u_1 - u_2 \rangle \geq \delta \|u_1 - u_2\|_V^p, \quad \alpha \geq 1 \quad (30.9.113)
\]

where here \( \hat{V} \) will be a Banach space such that \( V \) is dense in \( \hat{V} \) and the embedding is continuous. As mentioned, this is not surprising in the case of most interest where there is a Gelfand triple and \( B = I \). Then using the integral equation for \( r = p, q, p < q \) along with the conclusion of the Ito formula above,

\[
E(\langle B(u_n - u_m), u_n - u_m \rangle(t)) + E\left( \int_0^t \|u_n - u_m\|_V^p \, ds \right) \\
\leq E\left( \int_0^t \|B\| \|\Phi_n - \Phi_m\|_{\mathcal{L}_2}^2 \, ds \right) \equiv e(m,n)
\]

Hence, the right side converges to 0 as \( m, n \to \infty \) from the dominated convergence theorem. In particular,

\[
E\left( \int_0^T \|u_n - u_m\|_V^p \, ds \right) \leq E\left( \int_0^T \|B\| \|\Phi_n - \Phi_m\|_{\mathcal{L}_2}^2 \, ds \right) \equiv e(m,n) \quad (30.9.114)
\]

Then also

\[
P\left( \int_0^T \|u_n - u_m\|_V^p \, ds > \lambda \right) \leq \frac{e(m,n)}{\lambda}
\]

and so there exists a subsequence, denoted by \( r \) such that

\[
P\left( \int_0^T \|u_r - u_{r+1}\|_V^p \, ds \leq 2^{-r} \right) < 2^{-r}
\]

Thus, by the Borel Cantelli lemma, there is a further enlarged set of measure zero, still denoted as \( N \) such that for \( \omega \notin N \)

\[
\int_0^T \|u_r - u_{r+1}\|_V^p \, ds \leq 2^{-r}
\]

for all \( r \) large enough. Hence, by the usual proof of completeness, for these \( \omega \),

\[
\{u_r(\cdot, \omega)\}
\]

is Cauchy in \( L^\alpha \left( [0,T], \hat{V} \right) \) and also \( u_r(t, \omega) \) converges to some \( u(t, \omega) \) pointwise in \( \hat{V} \) for a.e. \( t \). In addition, from \( \text{LEMMA} \) these functions are a Cauchy sequence in \( L^\alpha \left( [0,T] \times \Omega; \hat{V} \right) \) with respect to the \( \sigma \) algebra of progressively measurable sets. Thus from Lemma \( \text{(30.5.1)} \), it can be assumed that for \( \omega \) off the set of measure zero, \( (t, \omega) \to (u(t, \omega)) \) is progressively measurable. From now on, this will be the sequence or a further subsequence. For \( \omega \notin N \), a set of measure zero and \( \text{LEMMA} \), there is a further subsequence for which the following convergences occur as \( r \to \infty \).

\[
u_r \to u \text{ weakly in } V \quad (30.9.115)
\]
Then, by the Burkholder Davis Gundy inequality,

\[ B(u_r) \to B(u) \text{ weakly in } V' \]  \hspace{1cm} (30.9.116)

\[ z_r \to z \text{ weakly in } V' \]  \hspace{1cm} (30.9.117)

\[ \left( B \left( u_r - \int_0^{t} \Phi_r dW \right) \right) \to \left( B \left( u - \int_0^{t} \Phi dW \right) \right) \text{ weakly in } V' \]  \hspace{1cm} (30.9.118)

\[ \int_0^{t} \Phi_r dW \to \int_0^{t} \Phi dW \text{ uniformly in } C([0,T];W) \]  \hspace{1cm} (30.9.119)

\[ B(u_r)(t) \to B(u)(t) \text{ weakly in } V' \]  \hspace{1cm} (30.9.120)

\[ B(u)(0) = B(u_0), \]  \hspace{1cm} (30.9.121)

\[ B(u)(t) = B(u(t)) \text{ a.e. } t \]  \hspace{1cm} (30.9.122)

In addition to this, we can choose the subsequence such that

\[ \sup_{r \neq s} \sup_{|t - s| < \gamma} \left| \int_t^s \Phi_r dW \right| < C(\omega) < \infty \]  \hspace{1cm} (30.9.123)

This is thanks to Corollary 30.9.1. The boundedness of the operator \( A \), in particular the given estimates, imply that \( z_r \) is bounded in \( L^{p'}([0,T] \times \Omega, V') \). Thus a subsequence can be obtained which yields weak convergence of \( z_r \) in \( L^{p'}([0,T] \times \Omega, V') \) and then Lemma 30.9.2 implies that a subsequence can be obtained which yields weak convergence of \( z_r \) off a set of measure zero, \( z \) is progressively measurable.

The claim 30.9.120 and 30.9.121 follow from the continuity of the evaluation map defined on \( X \), Theorem 30.9.24. The claim 30.9.122 follows from 30.9.121 and the convergence 30.9.117. To see this, let \( \psi \in C_c^\infty(0,T) \).

\[
\int_0^T B(u)(t) \psi(t) \, dt = \lim_{r \to \infty} \int_0^T B(u_r)(t) \psi(t) \, dt = \lim_{r \to \infty} \int_0^T B(u(t)) \psi(t) \, dt = \int_0^T B(u(t)) \psi(t) \, dt
\]

Since this is true for all such \( \psi \), it follows that \( B(u)(t) = B(u(t)) \) for a.e. \( t \). Passing to a limit in the integral equation yields the following for \( \omega \) off a set of measure zero,

\[ B(u)(t,\omega) - B(u_0)(\omega) + \int_0^t z(s,\omega) \, ds = \int_0^t f(s,\omega) \, ds + B \int_0^t \Phi_n dW \]

In the following claim, assume \( \Phi \in L^2 \left( \Omega, L^\infty \left( [0,T], L^2 \left( Q^{1/2}V, W \right) \right) \right) \)

Claim: \( \lim_{r \to \infty} \int_0^T \left( \Phi_r \circ J^{-1} \right)^* B(u_r) \circ JdW = \int_0^T \left( \Phi \circ J^{-1} \right)^* B(u) \circ JdW \) off a set of measure zero.

Proof of claim:

\[
E \left( \left| \int_0^T \left( \Phi_r \circ J^{-1} \right)^* B(u_r) \circ JdW - \int_0^T \left( \Phi \circ J^{-1} \right)^* B(u) \circ JdW \right| \right)
\leq E \left( \left| \int_0^T \left( \Phi_r \circ J^{-1} \right)^* B(u_r) \circ JdW - \int_0^T \left( \Phi \circ J^{-1} \right)^* B(u) \circ JdW \right| \right)
+ E \left( \left| \int_0^T \left( \Phi \circ J^{-1} \right)^* B(u_r) \circ JdW - \int_0^T \left( \Phi \circ J^{-1} \right)^* B(u) \circ JdW \right| \right)
\]

Then, by the Burkholder Davis Gundy inequality,

\[
\leq \int_\Omega \left( \int_0^T \left\| \Phi_r - \Phi \right\|^2 \left| B(u_r, t) \right| \right)^{1/2} \, dP
+ \int_\Omega \left( \int_0^T \left\| \Phi \right\|^2 \left| B(u_r - u, t) \right| \right)^{1/2} \, dP
\]
Also, this shows that we are taking a countable subset there is a set of measure zero such that \((t, \omega) \mapsto B_u(t, \omega)\) is progressively measurable in the above convergences. From the above considerations using the space \(\hat{\mathcal{V}}\), it follows that \(\mathsf{J} \in B_{u} \mathcal{V}\). Thus we take \(\mathsf{J} \) to be this common function. Hence there is a set of measure zero such that for \(\omega \notin N\), \(B_u(t, \omega)\) is progressively measurable in the above convergences. Also, this shows that we are taking \(u \in L^p([0, T] \times \Omega; V)\). From the measurability of \(u_r\), \(u\), we can obtain a dense countable subset \(\{ t_k \}\) and an enlarged set of measure zero \(N\) such that for \(\omega \notin N\), \(B_u(t_k, \omega) = B(u(t_k, \omega))\) and \(B_u(t, \omega) = B(u(t, \omega))\) for all \(t_k\) and \(r\). This uses the same argument as in Lemma 30.9.111.

It remains to verify that \(z(\cdot, \omega) \in A(u(\cdot, \omega), \omega)\). It follows from the above considerations that the Ito formula above can be used at will. Assume that for a given \(\omega \notin N\), \(B_u(T, \omega) = B(u(T, \omega))\), similar for \(B_u\). If not, just do the following argument for all \(T'\) close to \(T\), letting \(T'\) be in the dense subset just described. Then from the integral
Then there exists a $P$ tion

Theorem 30.9.7

This has proved the following Theorem. A follows from separation theorems and the fact that $Bu$ norm and so the same is true of the

satisfy a Holder condition into $V$. The reason the limit condition applies is the estimate

Now from the limit condition, for any $v \in V$, there exists a $z \in A (u \cdot, \omega, \omega)$ such that

Thus also

A similar formula to (30.9.124) holds for $u$. Thus

It follows from (30.9.125) and the other convergences that

Hence

Now from the limit condition, for any $v \in V$, there exists a $z \in A (u \cdot, \omega, \omega)$ such that

The reason the limit condition applies is the estimate (30.9.125) and the convergence (30.9.126) which shows that

satisfy a Holder condition into $V'$. Then the estimate (30.9.127) implies that the $B \int_0^{(t)} \Phi_r dW$ are bounded in a Holder norm and so the same is true of the $Bu_r$. Thus the situation of the limit condition (30.9.119) is obtained. Then it follows from separation theorems and the fact that $A (u \cdot, \omega, \omega)$ is closed and convex that $z (\cdot, \omega) \in A (u \cdot, \omega, \omega)$. This has proved the following Theorem.

**Theorem 30.9.7** Assume the above conditions, (30.9.114) - (30.9.119) along with the progressive measurability condition (30.9.123). Also assume there is at most one solution to (30.9.125) where

$q (t, \cdot) \equiv \int_0^t \Phi_n dW$

Then there exists a $P$ measurable $u_n$ such that also $z_n$ is progressively measurable

$$Bu (t, \omega) - Bu_0 (\omega) + \int_0^t z (s, \omega) ds = \int_0^t f (s, \omega) ds + B \int_0^t \Phi dW$$

where for each $\omega$, $z (\cdot, \omega) \in A (u \cdot, \omega, \omega)$. The function $Bu (t, \omega) = B (\Phi (t, \omega))$ for a.e. $t$. Here

$\Phi \in L^2 ([0, T] \times \Omega), L_2 (Q^{1/2} U, W) \cap L^2 (\Omega, L^\infty ([0, T], L_2 (Q^{1/2} U, W)))$
This has an easy generalization to the case where \( \Phi(t, \omega) \) is replaced with \( \sigma(t, \omega, u) \) for \( u \) the solution to the integral equation. In this generalization, we assume that

\[
\langle Bu, u \rangle = \|u\|_W^2
\]

For example, it could be \( W = W' \) and \( B = I \) which is the usual case of a Gelfand triple. We will strengthen the monotonicity assumption to for \( z_i \in A(u_i, \omega), \)

\[
(\lambda Bu_1 + z_1 - (\lambda Bu_2 + z_2), u_1 - u_2) \geq \delta \|u_1 - u_2\|_W^2
\]

(30.9.126)

The case of most interest is the usual one where \( V \subseteq W = W' \subseteq V' \), the case of a Gelfand triple in which \( B \) is the identity. As to \( \sigma \), the assumption is made that

\[
\|\sigma(t, \omega, u)\|_W \leq C + C \|u\|_W
\]

\[
\|\sigma(t, \omega, u) - \sigma(t, \omega, u_2)\|_{\mathcal{L}_2(Q^{1/2}U, W)} \leq K \|u_1 - u_2\|_W
\]

Of course it is also assumed that whenever \( u \) has values in \( W \) and is progressively measurable, \( (t, \omega) \to \sigma(t, \omega, u(t, \omega)) \) is also progressively measurable into \( \mathcal{L}_2(Q^{1/2}U, W) \).

Letting \( w_i \in L^2([0, T] \times \Omega, W) \cap L^2(\Omega, L^\infty([0, T], W)) \) each \( w_i \) being progressively measurable, the above assumptions and Theorem 30.9.127, there exists a solution \( u_i \) to the integral equation

\[
Bu_i(t, \omega) - Bu_0(\omega) + \int_0^t z_i(s, \omega) \, ds = \int_0^t f(s, \omega) \, ds + B \int_0^t \sigma(w_i) \, dW
\]

here we write \( \sigma(w_i) \) for short instead of \( \sigma(t, \omega, w_i) \). Then from the estimates,

\[
\langle Bu, u \rangle(t) - \langle Bu_0, u_0 \rangle + \delta \int_0^t \|u\|_V^p \, ds = 2 \int_0^t \langle f, u \rangle \, ds + C(b_3, b_4, b_5)
\]

\[
+ \lambda \int_0^t \langle Bu, u \rangle \, ds + \int_0^t \langle B\sigma(w), \sigma(u) \rangle_{\mathcal{L}_2} \, ds + 2M^*(t)
\]

\[
\leq 2 \int_0^t \langle f, u \rangle \, ds + C(b_3, b_4, b_5) + \lambda \int_0^t \langle Bu, u \rangle \, ds + \int_0^t \left( C + C \|w\|_W^2 \right) \, ds + 2M^*(t)
\]

where \( M^*(t) = \sup_{s \in [0, t]} |M(s)| \) and the quadratic variation of \( M \) is no larger than

\[
\int_0^t \|\sigma(w)\|^2 \langle B(u), u \rangle \, ds
\]

Then using Gronwall’s inequality, one obtains an inequality of the form

\[
\sup_{s \in [0, T]} \langle Bu, u \rangle(s) \leq C + C \left( M^*(t) + \int_0^t \|w\|_W^2 \, ds \right)
\]

where \( C = (u_0, f, \delta, \lambda, b_3, b_4, b_5, T) \) and is integrable. Then take expectation. By Burkholder Davis Gundy inequality and adjusting constants as needed,

\[
E \left( \sup_{s \in [0, T]} \langle Bu, u \rangle(s) \right)
\]

\[
\leq C + C \int_\Omega \int_0^T \|w\|_W^2 \, ds dP + C \int_\Omega \left( \int_0^T \|\sigma(w)\|^2 \langle B(u), u \rangle \, ds \right)^{1/2} dP
\]

\[
\leq C + C \int_\Omega \int_0^T \|w\|_W^2 \, ds dP + C \int_\Omega \sup_{s \in [0, T]} \langle Bu, u \rangle^{1/2} (s) \left( \int_0^T \|\sigma(w)\|^2 \, ds \right)^{1/2} dP
\]

\[
\leq C + C \int_\Omega \int_0^T \|w\|_W^2 \, ds dP + \frac{1}{2} E \left( \sup_{s \in [0, T]} \langle Bu, u \rangle(s) \right) + C \int_\Omega \int_0^T \left( C + C \|w\|_W^2 \right)
\]
Thus
\[ E \left( \langle Bu, u \rangle(t) \right) \leq E \left( \sup_{s \in [0,T]} \langle Bu, u \rangle(s) \right) \leq C + C \int_0^T \|w\|^2_W ds dP \]
and so
\[ \|u\|^2_{L^2(\Omega, \mathbb{L}^\infty([0,T], W))} \leq C + C \int_0^T \|w\|^2_W ds dP \]
which implies \( u \in L^2(\Omega, \mathbb{L}^\infty([0,T], W)) \) and is progressively measurable.

Using the monotonicity assumption, there is a suitable \( \lambda \) such that
\[
\frac{1}{2} \langle B(u_1 - u_2), u_1 - u_2 \rangle(t) + r \int_0^t \|u_1 - u_2\|^2_W ds - \lambda \int_0^t \langle B(u_1 - u_2), u_1 - u_2 \rangle ds - \int_0^t \langle B\sigma(u_1) - B\sigma(u_2), \sigma(u_1) - \sigma(u_2) \rangle_{\mathbb{L}_2^2} ds \leq M^*(t)
\]
where the right side is of the form \( \sup_{s \in [0,t]} |M(s)| \) where \( M(t) \) is a local martingale having quadratic variation dominated by
\[
C \int_0^t \|\sigma(w_1) - \sigma(w_2)\|^2 \langle B(u_1 - u_2), u_1 - u_2 \rangle ds
\]
Then by assumption and using Gronwall’s inequality, there is a constant \( C = C(\lambda, K, T) \) such that
\[
\langle B(u_1 - u_2), u_1 - u_2 \rangle(t) \leq CM^*(t)
\]
Then also, since \( M^* \) is increasing,
\[
\sup_{s \in [0,t]} \langle B(u_1 - u_2), u_1 - u_2 \rangle(s) \leq CM^*(t)
\]
Taking expectations and from the Burkholder Davis Gundy inequality,
\[
E \left( \sup_{s \in [0,t]} \langle B(u_1 - u_2), u_1 - u_2 \rangle(s) \right) \leq C \int \left( \int_0^t \|\sigma(w_1) - \sigma(w_2)\|^2 \langle B(u_1 - u_2), u_1 - u_2 \rangle \right)^{1/2} dP
\]
\[
\leq C \int \sup_{s \in [0,t]} \langle B(u_1 - u_2), u_1 - u_2 \rangle^{1/2}(s) \left( \int_0^t \|\sigma(w_1) - \sigma(w_2)\|^2 \right)^{1/2} dP
\]
Then it follows after adjusting constants that there exists an inequality of the form
\[
E \left( \sup_{s \in [0,t]} \langle B(u_1 - u_2), u_1 - u_2 \rangle(s) \right) \leq CE \left( \int_0^t \|\sigma(w_1) - \sigma(w_2)\|^2_{\mathbb{L}_2^2} ds \right)
\]
Hence
\[
E \left( \sup_{s \in [0,t]} \langle B(u_1 - u_2), u_1 - u_2 \rangle(t) \right) \leq CK^2E \left( \int_0^t \|w_1 - w_2\|^2_W ds \right)
\]
Thus, for each \( t \leq T \)
\[
\int \langle B(u_1 - u_2), u_1 - u_2 \rangle(t) dP \leq CK^2E \left( \int_0^t \|w_1 - w_2\|^2_W ds \right)
\]
once one can consider the map \( \psi(w) \equiv u \) as described above. The function \( t \to \langle B(\psi^n w_1 - \psi^n w_2), \psi^n w_1 - \psi^n w_2 \rangle(t) \) is continuous. So there exists \( t(\omega) \) where the maximum occurs. Then let \( t \) be this \( t(\omega) \). Then the above inequality implies
\[
E \left( \|\psi^n w_1 - \psi^n w_2\|^2_{L^\infty([0,T], W)} \right)
\]
\begin{align*}
\mathbb{E}(B(\psi^n w_1 - \psi^n w_2), \psi^n w_1 - \psi^n w_2)(t)) & \leq CK^2 \mathbb{E}\left(\int_0^t \|\psi^{n-1} w_1 - \psi^{n-1} w_2\|_W^2 dt_1 \right) \\
& = CK^2 \mathbb{E}\left(\int_0^t \left\langle B(\psi^{n-1} w_1 - \psi^{n-1} w_2), \psi^{n-1} w_1 - \psi^{n-1} w_2 (t_1) \right\rangle dt_1 \right) \\
& \leq (CK^2)^2 \mathbb{E}\left(\int_0^t \int_0^{t_1} \left\langle B(\psi^{n-2} w_1 - \psi^{n-2} w_2), \psi^{n-2} w_1 - \psi^{n-2} w_2 (t_2) \right\rangle dt_2 dt_1 \right) \\
& \cdots \leq (CK^2)^n \mathbb{E}\left(\int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} \left\langle B(w_1 - w_2), w_1 - w_2 (t_n) \right\rangle dt_n \cdots dt_2 dt_1 \right) \\
& \leq (CK^2)^n \mathbb{E}\left(\sup_t \left\langle B(w_1 - w_2), w_1 - w_2 (t) \right\rangle \frac{T^n}{n!} \leq \frac{1}{2} \|w_1 - w_2\|_{L^2(\Omega,L^\infty([0,T],W))}^2 \right)
\end{align*}

provided \(n\) is sufficiently large. It follows that

\[\|\psi^n w_1 - \psi^n w_2\|_{L^2(\Omega,L^\infty([0,T],W))}^2 \leq \frac{1}{2} \|w_1 - w_2\|_{L^2(\Omega,L^\infty([0,T],W))}^2\]

for all \(n\) sufficiently large. Hence, if one begins with \(w \in L^2(\Omega,L^\infty([0,T],W)) \cap L^2([0,T] \times \Omega,W)\), the sequence of iterates \(\{\psi^n w\}_{n=1}^\infty\) must converge to some fixed point \(u\) in \(L^2(\Omega,L^\infty([0,T],W))\). This \(u\) is automatically in \(L^2([0,T] \times \Omega,W)\) and is progressively measurable since each of the iterates is progressively measurable.
Bibliography


[38] Figueiredo I. and Trabucho L., A class of contact and friction dynamic problems in thermoelasticity and in thermoviscoelasticity, (preprint).


Index

\( G_\delta \), 52
\( L^1_{\text{loc}} \), 110
\( L^p (\Omega; X) \), 196
\( \pi \) systems, 4

a density result, 180
absolutely continuous, 111
adapted, 224, 262
Alexander subbasis theorem, 33
algebra of sets, 30
analytic sets, 718
application of Bair theorem, 638
Banach Alaoglu theorem, 65
Banach Steinhaus theorem, 71
Bochner integrable, 133
Bochner's theorem, 222
Borel Cantelli lemma, 153
bounded continuous linear functions, 54
box topology, 34
Brouwer fixed point theorem, 85
Browder lemma, 621
Burkholder Davis Gundy inequality, 309
Caratheodory extension theorem, 32
Caratheodory's procedure, 29
Cauchy Schwarz inequality, 83
central limit theorem, 209
closed graph theorem, 58
compact, 12
compact injection map, 499
covariation, 51
dominated convergence theorem, 120
dual space, 52
duality map
  strictly monotone, 108, 109
duality mapping, 112
duality maps
  application of duality, 112
  properties, 112
Eberlein Smulian theorem, 68
Egoroff theorem, 139
elementary function, 51
elementary function integral, 776
elementary functions, 776
equis integrable, 246
epsilon net, 12
ergodic, 26
  individual ergodic theorem, 26
Erling's lemma, 499
events, 157

filtration, 206
filtration
  normal, 206
  first hitting time, 276
fixed point theorem
  Kakutani, 620
Fourier transform \( L^1 \), 45
functions of Wiener processes, 458
general Radon measures, 110
  general Radon measures, 110
  fundamental theorem of calculus, 110
  integral, 51
  properties, 613
  application of duality, 112
  dual space, 52
gauge function, 51
Gaussian measure, 620, 652
Gelfand triple, 177
generalized gradient
  pseudomonotone, 226
generalized gradients, 226
good lambda inequality, 13

Hahn Banach theorem, 51
Hardy Littlewood maximal function, 117
Hausdorff maximal principle, 51
Heine Borel theorem, 51
hemicontinuous, 607
Doob Dynkin lemma, 110
Doob estimate, 122
Doob's submartingale estimate, 122
Duality map
  strictly monotone, 108, 109
duality maps
  application of duality, 112
  properties, 112
application of Bair theorem, 638
Bannach Alaoglu theorem, 65
Bannach Steinhaus theorem, 71
Bochner integrable, 133
Bochner's theorem, 222
Borel Cantelli lemma, 153
bounded continuous linear functions, 54
box topology, 34
Brouwer fixed point theorem, 85
Browder lemma, 621
Burkholder Davis Gundy inequality, 309
Caratheodory extension theorem, 32
Caratheodory's procedure, 29
Cauchy Schwarz inequality, 83
central limit theorem, 209
closed graph theorem, 58
compact, 12
compact injection map, 499
covariation, 51
dominated convergence theorem, 120
dual space, 52
duality map
  strictly monotone, 108, 109
duality mapping, 112
duality maps
  application of duality, 112
  properties, 112
Eberlein Smulian theorem, 68
Egoroff theorem, 139
elementary function, 51
elementary function integral, 776
elementary functions, 776
equis integrable, 246
epsilon net, 12
ergodic, 26
  individual ergodic theorem, 26
Erling's lemma, 499
events, 157

filtration, 206
filtration
  normal, 206
  first hitting time, 276
fixed point theorem
  Kakutani, 620
Fourier transform \( L^1 \), 45
functions of Wiener processes, 458
general Radon measures, 110
  general Radon measures, 110
  fundamental theorem of calculus, 110
  integral, 51
  properties, 613
  application of duality, 112
  dual space, 52
gauge function, 51
Gaussian measure, 620, 652
Gelfand triple, 177
generalized gradient
  pseudomonotone, 226
generalized gradients, 226
good lambda inequality, 13
Hahn Banach theorem, 51
Hardy Littlewood maximal function, 117
Hausdorff maximal principle, 51
Heine Borel theorem, 51
hemicontinuous, 607
787
<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hilbert Schmidt operators</td>
<td>724</td>
</tr>
<tr>
<td>Hilbert Schmidt theorem</td>
<td>724</td>
</tr>
<tr>
<td>Hilbert space</td>
<td>724</td>
</tr>
<tr>
<td>independent events</td>
<td>724</td>
</tr>
<tr>
<td>independent random vectors</td>
<td>724</td>
</tr>
<tr>
<td>independent sigma algebras</td>
<td>724</td>
</tr>
<tr>
<td>inner product space</td>
<td>724</td>
</tr>
<tr>
<td>inner regular</td>
<td>724</td>
</tr>
<tr>
<td>isometric</td>
<td>724</td>
</tr>
<tr>
<td>Ito isometry</td>
<td>724</td>
</tr>
<tr>
<td>Ito representation theorem</td>
<td>724</td>
</tr>
<tr>
<td>James map</td>
<td>724</td>
</tr>
<tr>
<td>Jensen's inequality</td>
<td>724</td>
</tr>
<tr>
<td>Kakutani fixed point theorem</td>
<td>724</td>
</tr>
<tr>
<td>Kolmogorov Centsov theorem</td>
<td>724</td>
</tr>
<tr>
<td>Kolmogorov extension theorem</td>
<td>724</td>
</tr>
<tr>
<td>Kolmogorov zero one law</td>
<td>724</td>
</tr>
<tr>
<td>Kolmogorov's inequality</td>
<td>724</td>
</tr>
<tr>
<td>Lebesgue set</td>
<td>724</td>
</tr>
<tr>
<td>Lebesgue decomposition</td>
<td>724</td>
</tr>
<tr>
<td>Lebesgue point</td>
<td>724</td>
</tr>
<tr>
<td>limits</td>
<td>724</td>
</tr>
<tr>
<td>independent random variables</td>
<td>724</td>
</tr>
<tr>
<td>local martingale</td>
<td>724</td>
</tr>
<tr>
<td>local martingale stochastic integral</td>
<td>724</td>
</tr>
<tr>
<td>local submartingale</td>
<td>724</td>
</tr>
<tr>
<td>localization</td>
<td>724</td>
</tr>
<tr>
<td>elementary functions</td>
<td>724</td>
</tr>
<tr>
<td>general case</td>
<td>724</td>
</tr>
<tr>
<td>stochastically square integrable functions</td>
<td>724</td>
</tr>
<tr>
<td>localizing sequence</td>
<td>724</td>
</tr>
<tr>
<td>locally bounded</td>
<td>724</td>
</tr>
<tr>
<td>locally convex topological vector space</td>
<td>724</td>
</tr>
<tr>
<td>lower semicontinuous</td>
<td>724</td>
</tr>
<tr>
<td>martingale</td>
<td>724</td>
</tr>
<tr>
<td>maximal monotone approximation</td>
<td>724</td>
</tr>
<tr>
<td>onto</td>
<td>724</td>
</tr>
<tr>
<td>pseudomonotone</td>
<td>724</td>
</tr>
<tr>
<td>maximal monotone</td>
<td>724</td>
</tr>
<tr>
<td>Banach space</td>
<td>724</td>
</tr>
<tr>
<td>equivalent conditions</td>
<td>724</td>
</tr>
<tr>
<td>interior of domain</td>
<td>724</td>
</tr>
<tr>
<td>mean ergodic theorem</td>
<td>724</td>
</tr>
<tr>
<td>measurable</td>
<td>724</td>
</tr>
<tr>
<td>multifunctions</td>
<td>724</td>
</tr>
<tr>
<td>measurable representative</td>
<td>724</td>
</tr>
<tr>
<td>measurable selection</td>
<td>724</td>
</tr>
<tr>
<td>measurable sets</td>
<td>724</td>
</tr>
<tr>
<td>measure space</td>
<td>724</td>
</tr>
<tr>
<td>isomorphic</td>
<td>724</td>
</tr>
<tr>
<td>separable</td>
<td>724</td>
</tr>
<tr>
<td>min max theorem</td>
<td>724</td>
</tr>
<tr>
<td>Minkowski functional</td>
<td>724</td>
</tr>
<tr>
<td>modification</td>
<td>724</td>
</tr>
<tr>
<td>monotone</td>
<td>724</td>
</tr>
<tr>
<td>monotone and hemicontinuous coercive</td>
<td>724</td>
</tr>
<tr>
<td>monotone graph extended</td>
<td>724</td>
</tr>
<tr>
<td>monotone operators</td>
<td>724</td>
</tr>
<tr>
<td>Moreau's theorem</td>
<td>724</td>
</tr>
<tr>
<td>multi-index</td>
<td>724</td>
</tr>
<tr>
<td>multifunctions</td>
<td>724</td>
</tr>
<tr>
<td>normal</td>
<td>724</td>
</tr>
<tr>
<td>normal filtration</td>
<td>724</td>
</tr>
<tr>
<td>nuclear operator</td>
<td>724</td>
</tr>
<tr>
<td>open mapping theorem</td>
<td>724</td>
</tr>
<tr>
<td>operator norm</td>
<td>724</td>
</tr>
<tr>
<td>optional sampling theorem</td>
<td>724</td>
</tr>
<tr>
<td>outer measure</td>
<td>724</td>
</tr>
<tr>
<td>outer regular</td>
<td>724</td>
</tr>
<tr>
<td>pap97</td>
<td>724</td>
</tr>
<tr>
<td>partial order</td>
<td>724</td>
</tr>
<tr>
<td>penalization operators</td>
<td>724</td>
</tr>
<tr>
<td>pi systems</td>
<td>724</td>
</tr>
<tr>
<td>pivot space</td>
<td>724</td>
</tr>
<tr>
<td>Plancherel theorem</td>
<td>724</td>
</tr>
<tr>
<td>pointwise limits of measurable functions</td>
<td>724</td>
</tr>
<tr>
<td>Polish space</td>
<td>724</td>
</tr>
<tr>
<td>polish space</td>
<td>724</td>
</tr>
<tr>
<td>polynomial</td>
<td>724</td>
</tr>
<tr>
<td>positive</td>
<td>724</td>
</tr>
<tr>
<td>positive definite functions</td>
<td>724</td>
</tr>
<tr>
<td>precompact</td>
<td>724</td>
</tr>
<tr>
<td>predictable</td>
<td>724</td>
</tr>
<tr>
<td>probability space</td>
<td>724</td>
</tr>
<tr>
<td>product topology</td>
<td>724</td>
</tr>
<tr>
<td>progressively measurable</td>
<td>724</td>
</tr>
<tr>
<td>progressively measurable composition</td>
<td>724</td>
</tr>
<tr>
<td>integral of</td>
<td>724</td>
</tr>
<tr>
<td>progressively measurable version</td>
<td>724</td>
</tr>
<tr>
<td>projection in Hilbert space</td>
<td>724</td>
</tr>
<tr>
<td>Prokhorov's theorem</td>
<td>724</td>
</tr>
<tr>
<td>pseudomonotone</td>
<td>724</td>
</tr>
<tr>
<td>bounded</td>
<td>724</td>
</tr>
<tr>
<td>pseudomonotone</td>
<td>724</td>
</tr>
<tr>
<td>generalized perturbation</td>
<td>724</td>
</tr>
<tr>
<td>modified bounded</td>
<td>724</td>
</tr>
<tr>
<td>perturbation</td>
<td>724</td>
</tr>
<tr>
<td>sum</td>
<td>724</td>
</tr>
<tr>
<td>pseudomonotone</td>
<td>724</td>
</tr>
<tr>
<td>onto</td>
<td>724</td>
</tr>
</tbody>
</table>
INDEX

pseudomonotone operator, 622, 623, 640
set valued, 622, 623

Q Wiener process, 371
quasi-bounded, 622

Rademacher’s theorem, 114
Radon Nikodym property, 144
Radon Nikodym Theorem
  finite measures, 103
random variable, 154
random vector, 154
  independent, 154
reflexive Banach Space, 60
representation of martingales, 456
Riesz map, 87
Riesz representation theorem
  Hilbert space, 86
Riesz representation theorem $L^p$
  $\sigma$ finite case, 150

self adjoint, 96
seminorms, 96
separability of $C(H)$, 170
sequential weak* compactness, 170
set valued functions
  measurability, 505, 724
set valued map
  locally bounded, 639
simple function, 127
Skorokhod’s theorem, 170
step functions
  approximation result, 376
stochastic integral, 373
stochastic integral, 393
  definition, 383
  linear transformation, 397
  main result, 384
  quadratic variation, 395
  stochastically square integrable, 392
stochastic process, 456
stochastically square integrable, 392
strong law of large numbers, 169, 250
strong topology, 169
stopped martingale, 226, 241
stopped process, 228
stopped submartingale, 284
stopping time, 229
strictly convex, 610
strong law of large numbers, 169, 243
strongly measurable, 127
subbasis, 33
submartingale, 226
submartingale convergence theorem, 228
sum of maximal monotone and pseudomonotone, 640
sums
  independent random variables, 164, 167, 243
supermartingale, 226

Suslin space, 722

tail event, 169
tight, 170, 206
topological vector space dual, 171
total variation, 111
 totally bounded set, 12
trace, 97
trace class operator, 94
trajectories, 253

type M, 609
demicontinuous, 310

uniform boundedness theorem, 50
upcrossing, 447, 467
upper semicontinuity
  set valued map, 714
  upper semicontinuous composition, 714

variational inequality, 57
vector measures, 142
Vector valued distributions, 467
version, 247
Vitali
  convergence theorem, 140

weak convergence of measures, 178, 247
weak topology, 178, 247
weak* measurable, 141
weak* topology, 57
weakly measurable, 127
Wiener process, 339, 346, 364
Wiener process in Hilbert Space, 347, 366

Yankov von Neumann Aumann, 722, 723