Chapter 1

Set Theory and General Topology

Chapter 2

Compactness and Continuous Functions

Chapter 3

Banach Spaces

3.1 Baire category theorem

3.2 Uniform boundedness closed graph and open mapping theorems

3.3 Hahn Banach theorem

3.4 Exercises

1. first problem
2. second
3. third
4. fourth
5. fifth
6. sixth
7. seventh
8. number eight

$\sum_{k=-n}^{n} a_k e^{ikx}$. Show $S_n f(x) = \int_{-\pi}^{\pi} f(y) D_n(x-y) dy$ where

$$D_n(t) = \frac{\sin((n+\frac{1}{2})t)}{2\pi \sin(\frac{t}{2})}.$$

This is called the Dirichlet kernel.

9. ↑ Let $Y = \{f$ such that f is continuous, defined on $\mathbb{R}$, and 2π periodic$\}$. Define $||f||_Y = \sup\{|f(x)| : x \in [-\pi, \pi]\}$. Show that $(Y, ||\ ||_Y)$ is a Banach space. Let $x \in \mathbb{R}$ and define $L_n(f) = S_n f(x)$. Show $L_n \in Y'$ but $\lim_{n\to\infty} ||L_n|| = \infty$. **Hint:** Let $f(y)$ approximate $sign(D_n(x-y))$.

10. ↑ Show there exists a dense G_δ subset of Y such that for f in this set, $|S_n f(x)|$ is unbounded. Show there is a dense G_δ subset of Y having the property that $|S_n f(x)|$ is unbounded on a dense G_δ subset of $\mathbb{R}$. This shows Fourier series can fail to converge pointwise to continuous periodic functions in a fairly spectacular way.

11. Let X be a normed linear space and let M be a convex open set containing 0. Define
$$\rho(x) = \inf\{t > 0 : \frac{x}{t} \in M\}.$$
Show ρ is a gauge function defined on X. This particular example is called a Minkowski functional. Recall a set, M, is convex if $\lambda x + (1-\lambda)y \in M$ whenever $\lambda \in [0,1]$ and $x, y \in M$.

12. ↑ This problem explores the use of the Hahn Banach theorem in establishing separation theorems. Let M be an open convex set containing 0. Let $x \notin M$. Show there exists $x^* \in X'$ such that $\operatorname{Re} x^*(x) \geq 1 > \operatorname{Re} x^*(y)$ for all $y \in M$. **Hint:** If $y \in M, \rho(y) < 1$. Show this. If $x \notin M,\ \rho(x) \geq 1$. Try $f(\alpha x) = \alpha\rho(x)$ for $\alpha \in \mathbb{R}$. Then extend f to F, show F is continuous, then fix it so F is the real part of $x^* \in X'$.

13. A Banach space is said to be strictly convex if whenever $||x|| = ||y||$ and $x \neq y$, then
$$\left|\left|\frac{x+y}{2}\right|\right| < ||x||.$$
$F : X \to X'$ is said to be a duality map if it satisfies the following: a.) $||F(x)|| = ||x||$. b.) $F(x)(x) = ||x||^2$. Show that if X' is strictly convex, then such a duality map exists. **Hint:** Let $f(\alpha x) = \alpha||x||^2$ and use Hahn Banach theorem, then strict convexity.

Chapter 4

Hilbert Spaces

Chapter 5

Calculus in Banach Space

Chapter 6

Locally Convex Topological Vector Space

6.1 Separation theorems

6.2 The weak and weak$*$ topologies

6.3 Set-valued maps

vertex, $\mathbf{x}$, pick $A_\epsilon \mathbf{x} \in A\mathbf{x}$ and define A_ϵ on all of $\mathbb{C}^n$ by the following rule. If

$$\mathbf{x} \in [\mathbf{x}_0, \cdots, \mathbf{x}_{2n}],$$

so $\mathbf{x} = \sum_{i=0}^{2n} t_i \mathbf{x}_i$, then

$$A_\epsilon \mathbf{x} \equiv \sum_{k=0}^{2n} t_k A_\epsilon \mathbf{x}_k.$$

Thus A_ϵ is a continuous map defined on $\mathbb{C}^n$ thanks to the local finiteness of the collection of simplicies. Let P_K denote the projection on the convex set K. By the Brouwer fixed point theorem, there exists a fixed point, $\mathbf{x}_\epsilon \in K$ such that

$$P_K (\mathbf{y} - A_\epsilon \mathbf{x}_\epsilon + \mathbf{x}_\epsilon) = \mathbf{x}_\epsilon.$$

By Corollary 4.8 this requires

$$\operatorname{Re} (\mathbf{y} - A_\epsilon \mathbf{x}_\epsilon, \mathbf{z} - \mathbf{x}_\epsilon) \leq 0$$

for all $\mathbf{z} \in K$.

Suppose $\mathbf{x}_\epsilon \in [\mathbf{x}_0^\epsilon, \cdots, \mathbf{x}_{2n}^\epsilon]$ so $\mathbf{x}_\epsilon = \sum_{k=0}^{2n} t_k^\epsilon \mathbf{x}_k^\epsilon$. Then since $\mathbf{x}_\epsilon$ is contained in K, a compact set, and the diameter of each simplex is less than 1, it follows that $A_\epsilon \mathbf{x}_k^\epsilon$ is contained in $A(\overline{K + B(\mathbf{0}, 1)})$, which is contained in a compact set thanks to Lemma 6.29. Taking a subsequence, we may obtain from the Heine Borel theorem that for some sequence, $\epsilon \to 0$

$$t_k^\epsilon \to t_k, \mathbf{x}_\epsilon \to \mathbf{x}, A_\epsilon \mathbf{x}_k^\epsilon \to \mathbf{y}_k$$

for $k = 0, \cdots, 2n$. Since the diameter of the simplex containing $\mathbf{x}_\epsilon$ converges to 0, it follows

$$\mathbf{x}_k^\epsilon \to \mathbf{x},\ A_\epsilon \mathbf{x}_k^\epsilon \to \mathbf{y}_k.$$

Since the graph of A is closed and $A_\epsilon \mathbf{x}_k^\epsilon \in A\mathbf{x}_k^\epsilon$, this implies $\mathbf{y}_k \in A\mathbf{x}$. Since $A\mathbf{x}$ is convex,

$$\sum_{k=1}^{2n} t_k \mathbf{y}_k \in A\mathbf{x}.$$

Hence for all $\mathbf{z} \in K$,

$$\operatorname{Re}\left(\mathbf{y} - \sum_{k=1}^{2n} t_k \mathbf{y}_k, \mathbf{z} - \mathbf{x}\right) = \lim_{\epsilon \to 0} \operatorname{Re}\left(\mathbf{y} - \sum_{k=1}^{2n} t_k^\epsilon A_\epsilon \mathbf{x}_k^\epsilon, \mathbf{z} - \mathbf{x}_\epsilon\right)$$

$$= \lim_{\epsilon \to 0} \operatorname{Re} (\mathbf{y} - A_\epsilon \mathbf{x}_\epsilon, \mathbf{z} - \mathbf{x}_\epsilon) \leq 0.$$

Let $\mathbf{w} = \sum_{k=1}^{2n} t_k \mathbf{y}_k$. This proves the lemma.

Chapter 7

Measures and Measurable Functions

7.1 σ Algebras

7.2 Monotone classes and algebras

Corollary 7.2.8 *Let* $(Z_1, \mathcal{R}_1, \mathcal{E}_1)$ *and* $(Z_2, \mathcal{R}_2, \mathcal{E}_2)$ *be as just described in Lemma 7.6. Then* $(Z_1 \times Z_2, \mathcal{R}, \mathcal{E})$ *also satisfies the conditions of Lemma 7.6 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 $R_1^1 \times R_2^1$ and $R_1^2 \times R_2^2$ be two elements of $\mathcal{R}$.

$$R_1^1 \times R_2^1 \cap R_1^2 \times R_2^2 = R_1^1 \cap R_1^2 \times R_2^1 \cap R_2^2 \in \mathcal{R}$$

by assumption.

$$R_1^1 \times R_2^1 \setminus \left(R_1^2 \times R_2^2\right) =$$

$$R_1^1 \times \left(R_2^1 \setminus R_2^2\right) \cup \left(R_1^1 \setminus R_1^2\right) \times \left(R_2^1 \cap R_2^2\right)$$

$$= R_1^1 \times A_2 \cup A_1 \times R_2$$

where $A_2 \in \mathcal{E}_2$, $A_1 \in \mathcal{E}_1$, and $R_2 \in \mathcal{R}_2$. Since the two sets in the above expression on the right do not intersect, and each A_i is a finite union of disjoint elements of $\mathcal{R}_i$, it follows the above expression is in $\mathcal{E}$. This proves the corollary. The following example will be referred to frequently.

Example 7.2.9 *Consider for* $\mathcal{R}$*, sets of the form* $I = (a, b] \cap (-\infty, \infty)$ *where* $a \in [-\infty, \infty]$ *and* $b \in [-\infty, \infty]$*. Then, clearly,* $\emptyset, (-\infty, \infty) \in \mathcal{R}$ *and it is not hard to see that all conditions for Corollary 7.7 are satisfied. Applying this corollary repeatedly, we find that for*

$$\mathcal{R} \equiv \left\{ \prod_{i=1}^{n} I_i : I_i = (a_i, b_i] \cap (-\infty, \infty) \right\}$$

and $\mathcal{E}$ *is defined as finite disjoint unions of sets of* $\mathcal{R}$*,*

$$(\mathbb{R}^n, \mathcal{R}, \mathcal{E})$$

satisfies the conditions of Corollary 7.7 and in particular $\mathcal{E}$ *is an algebra of sets of* $\mathbb{R}^n$*. It is clear that the same would hold if* I *were of the form* $[a, b) \cap (-\infty, \infty)$*.*

Chapter 8

The Abstract Lebesgue Integral

Chapter 9

The Construction of Measures

9.1 Outer measures

But we know Formula (2) holds because A is measurable. Apply the Definition 9.1 to $S \cap T$ instead of S.

The next theorem is the main result on outer measures. It is a very general result which applies whenever one has an outer measure on the power set of any set. This theorem will be referred to as Caratheodory's procedure in the rest of the book.

Theorem 9.1.4 *The collection of μ measurable sets, $\mathcal{S}$, forms a σ algebra and*

$$\text{If } F_i \in \mathcal{S},\ F_i \cap F_j = \emptyset, \text{ then } \mu(\cup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} \mu(F_i). \tag{9.1.3}$$

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

$$\mu(F) = \lim_{n\to\infty} \mu(F_n). \tag{9.1.4}$$

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

$$\mu(F) = \lim_{n\to\infty} \mu(F_n). \tag{9.1.5}$$

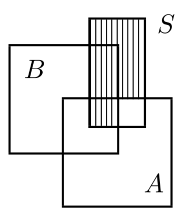
Also, $(\mathcal{S}, \mu)$ is complete. By this we mean that if $F \in \mathcal{S}$ and if $E \subseteq \Omega$ with $\mu(E \setminus F) + \mu(F \setminus E) = 0$, then $E \in \mathcal{S}$.

Proof: First note that $\emptyset$ and Ω are obviously in $\mathcal{S}$. Now suppose that $A, B \in \mathcal{S}$. We show $A \setminus B = A \cap B^C$ is in $\mathcal{S}$. Using the assumption that $B \in \mathcal{S}$ in the second equation below, in which $S \cap A$ plays the role of S in the definition for B being μ measurable,

$$\mu(S \cap (A \cap B^C)) + \mu(S \setminus (A \cap B^C)) = \mu(S \cap A \cap B^C) + \mu(S \cap (A^C \cup B))$$

$$= \mu(S \cap (A^C \cup B)) + \mu(S \cap A) - \mu(S \cap A \cap B). \tag{9.1.6}$$

The following picture of $S \cap (A^C \cup B)$ may be of use.



From the picture, and the measurablility of A, we see that Formula (6) is no larger than

$$\leq \mu(S \cap A \cap B) + \mu(S \setminus A) + \mu(S \cap A) - \mu(S \cap A \cap B) = \mu(S).$$

This has shown that if $A, B \in \mathcal{S}$, then $A \setminus B \in \mathcal{S}$. 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 $\cup_{i=1}^n A_i$. 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 $A_i \in \mathcal{S}$, $\mu(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$.

Now let $A = \cup_{i=1}^\infty A_i$ where $A_i \cap A_j = \emptyset$ for $i \neq j$.

$$\sum_{i=1}^\infty \mu(A_i) \geq \mu(A) \geq \mu(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i).$$

Since this holds for all n, we conclude, since μ is assumed to be an outer measure, that $\mu(A) = \sum_{i=1}^\infty \mu(A_i)$ which establishes Formula 9.1.3. Part 9.1.4 follows from part 9.1.3 just as in the proof of Theorem 7.12.

In order to establish 9.1.5, let the F_n be as given there. Then, since $(F_1 \setminus F_n)$ increases to $(F_1 \setminus F)$, we may use part 9.1.4 to conclude

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

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

$$\lim_{n\to\infty} (\mu(F_1) - \mu(F_n)) \geq \mu(F_1) - \mu(F)$$

which implies that, since $F \subseteq F_n$ for all n,

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

It remains to show $\mathcal{S}$ is closed under countable unions. We already know that if $A \in \mathcal{S}$, then $A^C \in \mathcal{S}$ and $\mathcal{S}$ is closed under finite unions. Let $A_i \in \mathcal{S}$, $A = \cup_{i=1}^\infty A_i$, $B_n = \cup_{i=1}^n A_i$. Then

$$\begin{aligned} \mu(S) &= \mu(S \cap B_n) + \mu(S \setminus B_n) \\ &= (\mu\lfloor S)(B_n) + (\mu\lfloor S)(B_n^C). \end{aligned} \tag{9.1.7}$$

By Lemma 9.3 we know B_n is $(\mu\lfloor S)$ measurable and so is B_n^C. We 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 we apply Parts 9.1.5 and 9.1.4 to Formula 9.1.7 and let $n \to \infty$. Thus

$$B_n \uparrow A, \;\; B_n^C \downarrow A^C$$

and this yields $\mu(S) = (\mu\lfloor S)(A) + (\mu\lfloor S)(A^C) = \mu(S \cap A) + \mu(S \setminus A)$.

Thus $A \in \mathcal{S}$ and this proves Parts 9.1.3, 9.1.4, and 9.1.5.

Chapter 10

Lebesgue Measure

10.1 Lebesgue Measure

10.2 Iterated integrals

10.3 Change of variables

10.4 Polar coordinates

Proof: Let E be an open set in X and let

$$\mathcal{S}_E \equiv \{F \text{ Borel in } Y \text{ such that } E \times F \text{ is Borel in } X \times Y\}.$$

Then $\mathcal{S}_E$ contains the open sets and is clearly closed with respect to countable unions. Let $F \in \mathcal{S}_E$. Then

$$E \times F^C \cup E \times F = E \times Y = \text{ a Borel set.}$$

Therefore, since $E \times F$ is Borel, it follows $E \times F^C$ is Borel. Therefore, $\mathcal{S}_E$ is a σ algebra. It follows $\mathcal{S}_E =$ Borel sets, and so, we have shown- open $\times$ Borel $=$ Borel. Now let F be a fixed Borel set in Y and define

$$\mathcal{S}_F \equiv \{E \text{ Borel in } X \text{ such that } E \times F \text{ is Borel in } X \times Y\}.$$

The same argument which was just used shows $\mathcal{S}_F$ is a σ algebra containing the open sets. Therefore, $\mathcal{S}_F =$ the Borel sets, and this proves the lemma since F was an arbitrary Borel set.

Now we define the unit sphere in $\mathbb{R}^n$, S^{n-1}, by

$$S^{n-1} \equiv \{\mathbf{w} \in \mathbb{R}^n : |\mathbf{w}| = 1\}.$$

Then S^{n-1} is a compact metric space using the usual metric on $\mathbb{R}^n$. We define a map

$$\theta : S^{n-1} \times (0, \infty) \to \mathbb{R}^n \setminus \{\mathbf{0}\}$$

by

$$\theta(\mathbf{w}, \rho) \equiv \rho \mathbf{w}.$$

It is clear that θ is one to one and onto with a continuous inverse. Therefore, if $\mathcal{B}_1$ is the set of Borel sets in $S^{n-1} \times (0, \infty)$, and $\mathcal{B}$ are the Borel sets in $\mathbb{R}^n \setminus \{\mathbf{0}\}$, it follows

$$\mathcal{B} = \{\theta(F) : F \in \mathcal{B}_1\}. \tag{10.4.8}$$

Observe also that the Borel sets of S^{n-1} satisfy the conditions of Lemma 7.6 with Z defined as $S^{n-1}$and the same is true of the sets $(a, b] \cap (0, \infty)$ where $0 \leq a, b \leq \infty$ if Z is defined as $(0, \infty)$. By Corollary 7.7, finite disjoint unions of sets of the form

$$\{E \times I : E \text{ is Borel in } S^{n-1}$$

$$\text{and } I = (a, b] \cap (0, \infty) \text{ where } 0 \leq a, b \leq \infty\}$$

form an algebra of sets, $\mathcal{A}$. It is also clear that $\sigma(\mathcal{A})$ contains the open sets and so $\sigma(\mathcal{A}) = \mathcal{B}_1$ because every set in $\mathcal{A}$ is in $\mathcal{B}_1$ thanks to Lemma 10.18. Let $A_r \equiv S^{n-1} \times (0, r]$ and let

$$\mathcal{M} \equiv \left\{F \in \mathcal{B}_1 : \int_{\mathbb{R}^n} \mathcal{X}_{\theta(F \cap A_r)} dm_n\right.$$

Chapter 11

Product Measure

11.1 Product Measure

Let $K_x \subseteq (P \cap X_n)$ and $K_y \subseteq (Q \cap Y_n)$ be such that

$$\mu(K_x) + \varepsilon > \mu(P \cap X_n)$$

and

$$\lambda(K_y) + \varepsilon > \lambda(Q \cap Y_n).$$

By Theorem 1.36 $K_x \times K_y$ is compact and from the definition of product measure,

$$(\mu \times \lambda)(K_x \times K_y) = \mu(K_x)\lambda(K_y)$$

$$\geq \mu(P \cap X_n)\lambda(Q \cap Y_n) - \varepsilon(\lambda(Q \cap Y_n) + \mu(P \cap X_n)) + \varepsilon^2.$$

Since ε is arbitrary, this verifies that $(\mu \times \lambda)$ is inner regular on $S \cap R_n$ whenever S is an elementary set. Similarly, $(\mu \times \lambda)$ is outer regular on $S \cap R_n$ whenever S is an elementary set. Thus $\mathcal{G}_n$ contains the elementary sets.

Next we show that $\mathcal{G}_n$ is a monotone class. If $S_k \downarrow S$ and $S_k \in \mathcal{G}_n$, let K_k be a compact subset of $S_k \cap R_n$ with

$$(\mu \times \lambda)(K_k) + \varepsilon 2^{-k} > (\mu \times \lambda)(S_k \cap R_n).$$

Let $K = \cap_{k=1}^{\infty} K_k$. Then

$$S \cap R_n \setminus K \subseteq \cup_{k=1}^{\infty}(S_k \cap R_n \setminus K_k).$$

Therefore

$$\begin{aligned}(\mu \times \lambda)(S \cap R_n \setminus K) &\leq \sum_{k=1}^{\infty}(\mu \times \lambda)(S_k \cap R_n \setminus K_k) \\ &\leq \sum_{k=1}^{\infty}\varepsilon 2^{-k} = \varepsilon.\end{aligned}$$

Now let $V_k \supseteq S_k \cap R_n$, V_k is open and

$$(\mu \times \lambda)(S_k \cap R_n) + \varepsilon > (\mu \times \lambda)(V_k).$$

Let k be large enough that

$$(\mu \times \lambda)(S_k \cap R_n) - \varepsilon < (\mu \times \lambda)(S \cap R_n).$$

Then $(\mu \times \lambda)(S \cap R_n) + 2\varepsilon > (\mu \times \lambda)(V_k)$. This shows $\mathcal{G}_n$ is closed with respect to intersections of decreasing sequences of its elements. The consideration of increasing sequences is similar. By the monotone class theorem, $\mathcal{G}_n = \mathcal{S} \times \mathcal{F}$.

Now let $S \in \mathcal{S} \times \mathcal{F}$ and let $l < (\mu \times \lambda)(S)$. Then $l < (\mu \times \lambda)(S \cap R_n)$ for some n. It follows from the first part of this proof that there exists a

Chapter 12

The L^p spaces

Chapter 13

Representation Theorems

Chapter 14

Fundamental Theorem of Calculus

Chapter 15

General Radon Measures

Chapter 16

Fourier Transforms

Chapter 17

Probability

Chapter 18

Weak Derivatives

Chapter 19

Hausdorff Measures

$$l(\mathbf{x}_1) - g(\mathbf{x}_1) \geq m(A_{P_i\mathbf{x}_1}) \geq 2|y_1|, \tag{19.0.3}$$

$$l(\mathbf{x}_2) - g(\mathbf{x}_2) \geq m(A_{P_i\mathbf{x}_2}) \geq 2|y_2|. \tag{19.0.4}$$

Claim: $|y_1 - y_2| \leq |l(\mathbf{x}_1) - g(\mathbf{x}_2)|$ or $|y_1 - y_2| \leq |l(\mathbf{x}_2) - g(\mathbf{x}_1)|$.

Proof of Claim: If not,

$$\begin{aligned} 2|y_1 - y_2| &> |l(\mathbf{x}_1) - g(\mathbf{x}_2)| + |l(\mathbf{x}_2) - g(\mathbf{x}_1)| \\ &\geq |l(\mathbf{x}_1) - g(\mathbf{x}_1) + l(\mathbf{x}_2) - g(\mathbf{x}_2)| \\ &= l(\mathbf{x}_1) - g(\mathbf{x}_1) + l(\mathbf{x}_2) - g(\mathbf{x}_2). \\ &\geq 2\,|y_1| + 2\,|y_2| \end{aligned}$$

by 19.0.3 and 19.0.4 contradicting the triangle inequality.

Now suppose $|y_1 - y_2| \leq |l(\mathbf{x}_1) - g(\mathbf{x}_2)|$. From the claim,

$$\begin{aligned} |\mathbf{x}_1 - \mathbf{x}_2| &= (|P_i\mathbf{x}_1 - P_i\mathbf{x}_2|^2 + |y_1 - y_2|^2)^{1/2} \\ &\leq (|P_i\mathbf{x}_1 - P_i\mathbf{x}_2|^2 + |l(\mathbf{x}_1) - g(\mathbf{x}_2)|^2)^{1/2} \\ &\leq (|P_i\mathbf{x}_1 - P_i\mathbf{x}_2|^2 + (|z_1 - z_2| + 2\varepsilon)^2)^{1/2} \\ &\leq diam(A) + O(\sqrt{\varepsilon}) \end{aligned}$$

where z_1 and z_2 are such that $P_i\mathbf{x}_1 + z_1\mathbf{e}_i \in A$, $P_i\mathbf{x}_2 + z_2\mathbf{e}_i \in A$, and

$$|z_1 - l(\mathbf{x}_1)| < \varepsilon \text{ and } |z_2 - g(\mathbf{x}_2)| < \varepsilon.$$

If $|y_1 - y_2| \leq |l(\mathbf{x}_2) - g(\mathbf{x}_1)|$, then we use the same argument but let

$$|z_1 - g(\mathbf{x}_1)| < \varepsilon \text{ and } |z_2 - l(\mathbf{x}_2)| < \varepsilon,$$

Since $\mathbf{x}_1, \mathbf{x}_2$ are arbitrary elements of $S(A, \mathbf{e}_i)$ and ε is arbitrary, this proves Formula 2.

The next lemma says that if A is already symmetric with respect to the *jth* direction, then this symmetry is not destroyed by taking $S(A, \mathbf{e}_i)$.

Lemma 19.0.5 *Suppose A is a Borel set in $\mathbb{R}^n$ such that $P_j\mathbf{x} + \mathbf{e}_j x_j \in A$ if and only if $P_j\mathbf{x} + (-x_j)\mathbf{e}_j \in A$. Then if $i \neq j$, $P_j\mathbf{x} + \mathbf{e}_j x_j \in S(A, \mathbf{e}_i)$ if and only if $P_j\mathbf{x} + (-x_j)\mathbf{e}_j \in S(A, \mathbf{e}_i)$.*

Chapter 20

The Area Formula

20.1 The Area Formula

20.1.1 Preliminary Results

It was shown in Lemma **??** that

$$\mathcal{H}^n(FA) = \det(U)m_n(A)$$

where $F = RU$ with R preserving distances and U a symmetric matrix having all positive eigenvalues. The area formula gives a generalization of this simple relationship to the case where F is replaced by a nonlinear mapping, $\mathbf{h}$. It contains as a special case the earlier change of variables formula. There are two parts to this development. The first part is to generalize Lemma **??** to the case of nonlinear maps. When this is done, the area formula can be presented.

In the first part of this, $\mathbf{h}$ will be a Lipschitz function,

$$|\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{y})| \leq K|\mathbf{x} - \mathbf{y}|$$

defined on $\mathbb{R}^n$ which is one to one on G, a measurable subset of $\mathbb{R}^n$. This is no loss of generality because of Theorem **??**.

The following lemma states that Lipschitz maps take sets of measure zero to sets of measure zero.

Lemma 20.1.1 *If* $\mathbf{h}$ *is Lipschitz and* $m_n(T) = 0$*, then*

$$\mathcal{H}^n(\mathbf{h}(T)) = 0$$

Proof: Let $\varepsilon > 0$ be given. By outer regularity, there exists an open set V containing T such that $m_n(V) < \frac{\varepsilon}{k^n 6^n}$. For $\mathbf{x} \in T$ it follows there exists $r_{\mathbf{x}} < 1$ such that the ball centered at $\mathbf{x}$ with radius $5r_{\mathbf{x}}$ is contained in V and $5Kr_{\mathbf{x}} < \delta$ for $\delta > 0$. Here K is the Lipschitz constant of $\mathbf{h}$. Then by the Vitali covering theorem, there are disjoint balls $\{B(\mathbf{x}_i, r_i)\}$ such that the enlarged balls $\hat{B}_i$ having five times the radius cover T, each being contained in V. Then

$$\mathcal{H}_\delta^n(\mathbf{h}(T)) \leq \mathcal{H}_\delta^n\left(\mathbf{h}\left(\cup_{i=1}^\infty \hat{B}_i\right)\right) \leq \mathcal{H}_\delta^n\left(\cup_{i=1}^\infty \mathbf{h}\left(\hat{B}_i\right)\right)$$

$$\leq \sum_{i=1}^\infty \mathcal{H}_\delta^n\left(\mathbf{h}\left(\hat{B}_i\right)\right) \leq \sum_{i=1}^\infty \mathcal{H}_\delta^n\left(B\left(\mathbf{h}(\mathbf{x}_i), 5Kr_{\mathbf{x}_i}\right)\right)$$

$$\begin{aligned} &\leq \sum_{i=1}^\infty \alpha(n)\left(5Kr_{\mathbf{x}_i}\right)^n = (5K)^n \sum_{i=1}^\infty \alpha(n) r_{\mathbf{x}_i}^n \\ &= (5K)^n \sum_{i=1}^\infty m_n\left(B\left(\mathbf{x}_i, r_{\mathbf{x}_i}\right)\right) \\ &\leq (5K)^n m_n(V) \leq (5K)^n \frac{\varepsilon}{K^n 6^n} < \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, this shows $\mathcal{H}_\delta^n(\mathbf{h}(T)) = 0$. ■

Recall the following theorem on Vitali covers. It is Corollary **??** on Page **??**.

Corollary 20.1.2 *Let $F \subseteq \mathbb{R}^n$ be a set and let $\mathcal{F}$ be a collection of balls which come from some norm, open or not, but having bounded radii such that $\mathcal{F}$ covers F in the sense of Vitali. Then there exists a countable collection of balls from $\mathcal{F}$ having disjoint closures, denoted by $\{B_j\}_{j=1}^\infty$, such that $\bar{m}_n(F \setminus \cup_{j=1}^\infty B_j) = 0$.*

Lemma 20.1.3 *If S is a Lebesgue measurable set and $\mathbf{h}$ is Lipschitz then $\mathbf{h}(S)$ is $\mathcal{H}^n$ measurable. Also, if $\mathbf{h}$ is Lipschitz with constant K,*

$$\mathcal{H}^n(\mathbf{h}(S)) \leq K^n m_n(S)$$

Proof: Let $S_k = S \cap B(\mathbf{0}, k), k \in \mathbb{N}$. By inner regularity of Lebesgue measure, there exists a set F, which is the countable union of compact sets and a set T with $m_n(T) = 0$ such that

$$F \cup T = S_k.$$

Then $\mathbf{h}(F) \subseteq \mathbf{h}(S_k) \subseteq \mathbf{h}(F) \cup \mathbf{h}(T)$. By continuity of $\mathbf{h}$, $\mathbf{h}(F)$ is a countable union of compact sets and so it is Borel. By Lemma 20.1.1, $\mathcal{H}^n(\mathbf{h}(T)) = 0$ and so $\mathbf{h}(S_k)$ is $\mathcal{H}^n$ measurable because of completeness of Hausdorff measure, which comes from $\mathcal{H}^n$ being obtained from an outer measure. Now $\mathbf{h}(S) = \cup_{k=1}^\infty \mathbf{h}(S_k)$ and so it is also true that $\mathbf{h}(S)$ is $\mathcal{H}^n$ measurable.

Consider the estimate. Assume first that $S \subseteq B(\mathbf{0}, R)$. Let V be an open set containing S which is contained in $B(\mathbf{0}, R)$ such that $m_n(V \setminus S) < \varepsilon$. Then each point of S is contained in an open ball which is contained in V. Considering all such open balls having radius less than $\frac{\delta}{K}$, this forms a Vitali cover of S. Therefore, there are disjoint balls $\{B_i\}$ centered at points of S such that $m_n(S \setminus \cup_i B_i) = 0$. Lemma 20.1.1 says that $\mathcal{H}^n(\mathbf{h}(S \setminus \cup_i B_i)) = 0$. Thus $\mathbf{h}(B_i)$ is a set of radius no larger than δ. Also,

$$\mathbf{h}(S) = \mathbf{h}(S \setminus \cup_i B_i) + \mathbf{h}(\cup_i B_i)$$

Hence,

$$\begin{aligned}\mathcal{H}_\delta^n(\mathbf{h}(S)) &\leq \mathcal{H}_\delta^n(\mathbf{h}(S\setminus\cup_i B_i)) + \sum_i \beta(n)\, r(\mathbf{h}(B_i))^n \\ &\leq \mathcal{H}_\delta^n(\mathbf{h}(S\setminus\cup_i B_i)) + \sum_i \beta(n)\, K^n r(B_i)^n \\ &\leq \mathcal{H}_\delta^n(\mathbf{h}(S\setminus\cup_i B_i)) + K^n m_n(V)\end{aligned}$$

Then take the limit as $\delta \to 0$ to obain

$$\mathcal{H}^n(\mathbf{h}(S)) \leq K^n m_n(V) \leq K^n(m_n(S)+\varepsilon)$$

Since ε is arbitrary, this verifies the inequality. For the general case,

$$\begin{aligned}\mathcal{H}^n(\mathbf{h}(S)) &= \lim_{m\to\infty}\mathcal{H}^n(\mathbf{h}(S\cap B(\mathbf{0},m))) \\ &\leq \liminf_{m\to\infty} K^n m_n(S\cap B(\mathbf{0},m)) = K^n m_n(S) \ \blacksquare\end{aligned}$$

By Theorem **??** on Page **??**, when $D\mathbf{h}(\mathbf{x})$ exists,

$$D\mathbf{h}(\mathbf{x}) = R(\mathbf{x})\,U(\mathbf{x})$$

where $(U(\mathbf{x})\mathbf{u},\mathbf{v}) = (U(\mathbf{x})\mathbf{v},\mathbf{u}), (U(\mathbf{x})\mathbf{u},\mathbf{u}) \geq 0$ and $R^*R = I$ so R preserves lengths. This convention will be used in what follows.

Lemma 20.1.4 *In this situation,* $|R^*\mathbf{u}| \leq |\mathbf{u}|$.

Proof: First note that

$$\begin{aligned}(\mathbf{u}-RR^*\mathbf{u}, RR^*\mathbf{u}) &= (\mathbf{u}, RR^*\mathbf{u}) - |RR^*\mathbf{u}|^2 \\ &= |R^*\mathbf{u}|^2 - |R^*\mathbf{u}|^2 = 0,\end{aligned}$$

and so

$$\begin{aligned}|\mathbf{u}|^2 &= |\mathbf{u}-RR^*\mathbf{u}+RR^*\mathbf{u}|^2 \\ &= |\mathbf{u}-RR^*\mathbf{u}|^2 + |RR^*\mathbf{u}|^2 \\ &= |\mathbf{u}-RR^*\mathbf{u}|^2 + |R^*\mathbf{u}|^2. \ \blacksquare\end{aligned}$$

As discussed earlier, there is a convenient estimate involving Lipschitz maps.

Lemma 20.1.5 *If* $|P\mathbf{x}-P\mathbf{y}| \leq L|\mathbf{x}-\mathbf{y}|$, *then for* E *a set in* $\mathbb{R}^n$,

$$\mathcal{H}^n(PE) \leq L^n \mathcal{H}^n(E).$$

Proof: Without loss of generality, assume $\mathcal{H}^n(E) < \infty$. Let $\delta > 0$ and let $\{C_i\}_{i=1}^\infty$ be a covering of E such that $r(C_i) \leq \delta$ for each i and

$$\sum_{i=1}^\infty \alpha(n)\, r(C_i)^n \leq \mathcal{H}_\delta^n(E) + \varepsilon.$$

Then $\{PC_i\}_{i=1}^{\infty}$ is a covering of PE such that $r(PC_i) \leq L\delta$. Therefore,

$$\begin{aligned}\mathcal{H}_{L\delta}^n(PE) &\leq \sum_{i=1}^{\infty} \alpha(n)\, r(PC_i)^n \\ &\leq L^n \sum_{i=1}^{\infty} \alpha(n)\, r(C_i)^n \leq L^n \mathcal{H}_{\delta}^n(E) + L^n \varepsilon \\ &\leq L^n \mathcal{H}^n(E) + L^n \varepsilon.\end{aligned}$$

Letting $\delta \to 0$,

$$\mathcal{H}^n(PE) \leq L^n \mathcal{H}^n(E) + L^n \varepsilon$$

and since $\varepsilon > 0$ is arbitrary, this proves the Lemma. ■

Then the following corollary follows from Lemma 20.1.4.

Corollary 20.1.6 *Let $T \subseteq \mathbb{R}^m$. Then*

$$\mathcal{H}^n(T) \geq \mathcal{H}^n(R^*T).$$

a decomposition

First is a simple lemma which is fairly interesting.

Lemma 20.1.7 *Let S, T be $n \times n$ matrices which are invertible. Then*

$$\mathbf{o}(T\mathbf{v}) = \mathbf{o}(S\mathbf{v}) = \mathbf{o}(\mathbf{v})$$

and if L is a continuous linear transformation such that for $a < b$,

$$\sup_{\mathbf{v}\neq\mathbf{0}} \frac{|L\mathbf{v}|}{|S\mathbf{v}|} < b, \ \inf_{\mathbf{v}\neq\mathbf{0}} \frac{|L\mathbf{v}|}{|S\mathbf{v}|} > a$$

If $\|S - T\|$ is small enough, it follows that the same inequalities hold with S replaced with T. Here $\|\cdot\|$ denotes the operator norm.

Proof: Consider the first claim. For

$$\begin{aligned}|\mathbf{v}| &= \left|T^{-1}T\mathbf{v}\right| \leq \left\|T^{-1}\right\| |T\mathbf{v}|, \\ |T\mathbf{v}| &\leq \|T\| |\mathbf{v}|\end{aligned}$$

and so

$$\frac{1}{\|T\|} |T\mathbf{v}| \leq |\mathbf{v}| \leq \left\|T^{-1}\right\| |T\mathbf{v}|$$

so $|\mathbf{v}| \to 0$ is the same as saying that $|T\mathbf{v}| \to 0$. Similar considerations apply to S. Thus the first claim is clearly true.

Consider the second claim. To say that $\|S - T\|$ is small is the same as saying that $\|S - T\|_F$ is small where this refers to the Frobinius norm in which the $n \times n$

matrix is regarded as an element of $\mathbb{R}^{n^2}$, with the Euclidean norm, since all norms are equivalent. Then

$$|S\mathbf{v}| = \left|ST^{-1}T\mathbf{v}\right| \leq \left\|ST^{-1}\right\| |T\mathbf{v}|$$

$$|T\mathbf{v}| = \left|TS^{-1}S\mathbf{v}\right| \leq \left\|TS^{-1}\right\| |S\mathbf{v}|$$

Hence

$$\frac{1}{\|TS^{-1}\|} |T\mathbf{v}| \leq |S\mathbf{v}| \leq \left\|ST^{-1}\right\| |T\mathbf{v}|$$

Thus the second result follows if it is the case that for $\|T - S\|$ sufficiently small, both $\left\|TS^{-1}\right\|, \left\|ST^{-1}\right\|$ close to 1. This is because there is $\hat{a} > a$ such that

$$\inf_{\mathbf{v}\neq\mathbf{0}} \frac{|L\mathbf{v}|}{|S\mathbf{v}|} > \hat{a} > a$$

$$\inf_{\mathbf{v}\neq\mathbf{0}} \frac{|L\mathbf{v}|}{|T\mathbf{v}|} \geq \inf_{\mathbf{v}\neq\mathbf{0}} \frac{|L\mathbf{v}|}{\|TS^{-1}\| |S\mathbf{v}|} \geq \frac{1}{\|TS^{-1}\|} \inf_{\mathbf{v}\neq\mathbf{0}} \frac{|L\mathbf{v}|}{|S\mathbf{v}|} > \frac{1}{\|TS^{-1}\|}\hat{a} > a$$

provided $\left\|TS^{-1}\right\|$ is close enough to 1. The other inequality is similar.

From the formula for the inverse in terms of the determinant, the entries of T^{-1} are close to the corresponding entries of S^{-1} provided $\|T - S\|$ is small enough. Thus by continuity,

$$\left\|ST^{-1} - I\right\|_F, \left\|TS^{-1} - I\right\|_F \text{ are small}$$

and consequently $\left\|ST^{-1}\right\|, \left\|TS^{-1}\right\|$ are both close to 1, the operator norm of I. This is because being close to I in the Frobenius norm is equivalent to being close to I in the operator norm. ■

The following is a simplified version of an argument in [**?**]. In what follows, it is assumed also that $\mathbf{h}$ is one to one. In particular, we assume the following:

$$\mathbf{h} \text{ is one to one on } G \text{ a measurable subset of } \mathbb{R}^n \tag{20.1.1}$$

$$D\mathbf{h}(\mathbf{x}) \text{ exists at } a.e.\mathbf{x} \in G \text{ say at all } \mathbf{x} \in A \subseteq G \tag{20.1.2}$$

By Rademacher's theorem, these conditions are satisfied if $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz and this situation is the one of main interest here. The conditions 20.1.1 - 20.1.2 could likely be obtained in other situations also.

For $\mathbf{x} \in A$, let $D\mathbf{h}(\mathbf{x}) \equiv R(\mathbf{x}) U(\mathbf{x})$ where $R(\mathbf{x})$ preserves lengths and $U(\mathbf{x}) \equiv \left(D\mathbf{h}(\mathbf{x})^* D\mathbf{h}(\mathbf{x})\right)^{1/2}$. Let A^+ denote those points of A for which $U(\mathbf{x})^{-1}$ exists. Thus this is a measurable subset of A.

Let B be a Lebesgue measurable subset of A^+ and let $\mathbf{b} \in B$. Let $\mathcal{S}$ be a countable dense subset of the space of symmetric invertible matrices and let $\mathcal{C}$ be a countable dense subset of B. Let $t > 1$ and let ε be so small that

$$\frac{1}{t} + \varepsilon < 1 < t - \varepsilon$$

Now since $U(\mathbf{b})$ is invertible, Lemma 20.1.7 implies $\mathbf{o}(\mathbf{a}-\mathbf{b}) = \mathbf{o}(U(\mathbf{b})(\mathbf{a}-\mathbf{b}))$ and so

$$|\mathbf{h}(\mathbf{a}) - \mathbf{h}(\mathbf{b}) - D\mathbf{h}(\mathbf{b})(\mathbf{a}-\mathbf{b})| < \varepsilon |U(\mathbf{b})(\mathbf{a}-\mathbf{b})| \tag{20.1.3}$$

provided that $\mathbf{a} \in B\left(\mathbf{b}, \frac{2}{i}\right)$ assuming i is sufficiently large. In addition to this,

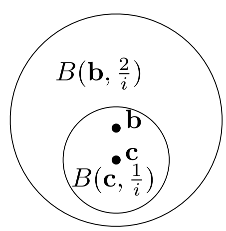
$$\begin{aligned}
\sup_{\mathbf{v}\neq\mathbf{0}} \frac{|D\mathbf{h}(\mathbf{b})\mathbf{v}|}{|U(\mathbf{b})\mathbf{v}|} &= \sup_{\mathbf{v}\neq\mathbf{0}} \frac{|R(\mathbf{b})U(\mathbf{b})\mathbf{v}|}{|U(\mathbf{b})\mathbf{v}|} = \sup_{\mathbf{v}\neq\mathbf{0}} \frac{|U(\mathbf{b})\mathbf{v}|}{|U(\mathbf{b})\mathbf{v}|} < (t-\varepsilon) \\
\inf_{\mathbf{v}\neq\mathbf{0}} \frac{|D\mathbf{h}(\mathbf{b})\mathbf{v}|}{|U(\mathbf{b})\mathbf{v}|} &= \inf_{\mathbf{v}\neq\mathbf{0}} \frac{|R(\mathbf{b})U(\mathbf{b})\mathbf{v}|}{|U(\mathbf{b})\mathbf{v}|} = \inf_{\mathbf{v}\neq\mathbf{0}} \frac{|U(\mathbf{b})\mathbf{v}|}{|U(\mathbf{b})\mathbf{v}|} > \frac{1}{t} + \varepsilon \qquad (20.1.4)
\end{aligned}$$

By Lemma 20.1.7, the inequalities 20.1.4, 20.1.3 continue to hold if $U(\mathbf{b})$ is replaced by another linear one to one and onto symmetric mapping T provided T is sufficiently close to $U(\mathbf{b})$. Let $T \in \mathcal{S}$ be such a linear transformation. Also let i be large enough that for all $\mathbf{a} \in B\left(\mathbf{b}, \frac{2}{i}\right)$,

$$|\mathbf{h}(\mathbf{a}) - \mathbf{h}(\mathbf{b}) - D\mathbf{h}(\mathbf{b})(\mathbf{a}-\mathbf{b})| < \varepsilon |T(\mathbf{a}-\mathbf{b})| \tag{20.1.5}$$

Now let $\mathbf{c} \in \mathcal{C}$ be close enough to $\mathbf{b}$ that $\mathbf{b} \in B\left(\mathbf{c}, \frac{1}{i}\right)$. Thus $\mathbf{b} \in E(T, \mathbf{c}, i)$ where for $i \in \mathbb{N}, \mathbf{c} \in \mathcal{C}$, $T \in \mathcal{S}$, $E(T, \mathbf{c}, i)$ consists of those $\mathbf{b} \in B\left(\mathbf{c}, \frac{1}{i}\right)$ such that for all $\mathbf{a} \in B\left(\mathbf{b}, \frac{2}{i}\right)$, 20.1.5 holds and also

$$\inf_{\mathbf{v}\neq\mathbf{0}} \frac{|D\mathbf{h}(\mathbf{b})\mathbf{v}|}{|T\mathbf{v}|} > \frac{1}{t} + \varepsilon, \ \sup_{\mathbf{v}\neq\mathbf{0}} \frac{|D\mathbf{h}(\mathbf{b})\mathbf{v}|}{|T\mathbf{v}|} < (t-\varepsilon) \tag{20.1.6}$$



Thus there are countably many of these sets $E(T, \mathbf{c}, i)$, some may be empty but as just shown, their union includes all of B. They are Borel measurable sets because $\mathbf{b} \to D\mathbf{h}(\mathbf{b})$ is Borel measurable.

For $\mathbf{a}, \mathbf{b} \in E(T, \mathbf{c}, i)$ it follows from 20.1.5, 20.1.6

$$|\mathbf{h}(\mathbf{a}) - \mathbf{h}(\mathbf{b})| < |D\mathbf{h}(\mathbf{b})(\mathbf{a}-\mathbf{b})| + \varepsilon |T(\mathbf{a}-\mathbf{b})| < |T(\mathbf{a}-\mathbf{b})|$$

and also

$$\left(\frac{1}{t} + \varepsilon\right) |T(\mathbf{a}-\mathbf{b})| < |D\mathbf{h}(\mathbf{b})(\mathbf{a}-\mathbf{b})| < |\mathbf{h}(\mathbf{a}) - \mathbf{h}(\mathbf{b})| + \varepsilon |T(\mathbf{a}-\mathbf{b})|$$

and so

$$t |T(\mathbf{a}-\mathbf{b})| > |\mathbf{h}(\mathbf{a}) - \mathbf{h}(\mathbf{b})| > \frac{1}{t} |T(\mathbf{a}-\mathbf{b})|$$

It follows from this that

$$\left|\mathbf{h}\left(T^{-1}(\mathbf{x})\right)-\mathbf{h}\left(T^{-1}(\mathbf{y})\right)\right| \leq t\,|\mathbf{x}-\mathbf{y}| \tag{20.1.7}$$

$$\frac{1}{t}\left|T\left(\mathbf{h}^{-1}(\mathbf{x})\right)-T\left(\mathbf{h}^{-1}(\mathbf{y})\right)\right| \leq |\mathbf{x}-\mathbf{y}| \tag{20.1.8}$$

Here the functions are defined on the appropriate sets, $T(E(T,\mathbf{c},i))$ in the first inequality and $\mathbf{h}(E(T,\mathbf{c},i))$ in the second.

Now let (E_k, T_k) result from a disjoint union of measurable subsets of the countably many $E(T,\mathbf{c},i)$ such that $B=\cup_k E_k$. Thus the above Lipschitz conditions hold for T_k in place of T. This proves most of the following lemma.

Lemma 20.1.8 *There are disjoint measurable sets E_k whose union equals B and symmetric linear transformations T_k such that*

$$\left|\mathbf{h}\left(T_k^{-1}(\mathbf{x})\right)-\mathbf{h}\left(T_k^{-1}(\mathbf{y})\right)\right| \leq t\,|\mathbf{x}-\mathbf{y}| \tag{20.1.9}$$

$$\left|T_k\left(\mathbf{h}^{-1}(\mathbf{x})\right)-T_k\left(\mathbf{h}^{-1}(\mathbf{y})\right)\right| \leq t\,|\mathbf{x}-\mathbf{y}| \tag{20.1.10}$$

on $T_k(E_k)$ and $\mathbf{h}(E_k)$ respectively. Also, for $\mathbf{b}\in E_k$

$$\left(\frac{1}{t}+\varepsilon\right)|T_k\mathbf{v}| < |D\mathbf{h}(\mathbf{b})\,\mathbf{v}| = |U(\mathbf{b})\,\mathbf{v}| < (t-\varepsilon)\,|T_k\mathbf{v}| \tag{20.1.11}$$

One can also conclude that for $\mathbf{b}\in E_k$,

$$t^{-n}\,|\det(T_k)| \leq \det(U(\mathbf{b})) \leq t^n\,|\det(T_k)| \tag{20.1.12}$$

Proof: It only remains to verify the last claim. However, this follows right away from 20.1.11. This formula implies that

$$\frac{1}{t}\,|\mathbf{v}| < \left|U(\mathbf{b})\,T_k^{-1}\mathbf{v}\right| < t\,|\mathbf{v}|$$

and so

$$B\left(\mathbf{0},\frac{1}{t}\right) \subseteq U(\mathbf{b})\,T_k^{-1}B(\mathbf{0},1) \subseteq B(\mathbf{0},t)$$

This implies

$$\alpha(n)\,\frac{1}{t^n} \leq \det\left(U(\mathbf{b})\,T_k^{-1}\right)\alpha(n) \leq \alpha(n)\,t^n$$

and so

$$\frac{1}{t^n}\,|\det(T_k)| \leq \det(U(\mathbf{b})) \leq t^n\,|\det(T_k)| \;\blacksquare$$

This lemma, along with Lemma 20.1.3 about the relationship between Hausdorff measure and Lipschitz mappings, implies the following.

$$\mathcal{H}^n(\mathbf{h}(E_k)) = \mathcal{H}^n\left(\mathbf{h}\circ T_k^{-1}(T_k(E_k))\right) \leq t^n\mathcal{H}^n(T_k(E_k)) = t^n m_n(T_k(E_k))$$

$$m_n(T_k(E_k)) = m_n\left(\left(T_k \circ \mathbf{h}^{-1}(\mathbf{h}(E_k))\right)\right) \leq t^n \mathcal{H}^n(\mathbf{h}(E_k))$$

Summarizing,

$$t^n \mathcal{H}^n(\mathbf{h}(E_k)) \geq m_n(T_k(E_k)) \geq t^{-n} \mathcal{H}^n(\mathbf{h}(E_k))$$

Then the above inequality and 20.1.12 implies the following.

$$t^{-2n} \mathcal{H}^n(\mathbf{h}(E_k)) \leq t^{-n} m_n(T_k(E_k)) \leq t^{-n} |\det(T_k)| \, m_n(E_k)$$

$$\leq \int_{E_k} \det(U(\mathbf{x}))\, dm_n \leq t^n |\det(T_k)| \, m_n(E_k) = t^n m_n(T_k E_k) \leq t^{2n} \mathcal{H}^n(\mathbf{h}(E_k))$$

Summing over all E_k yields the following thanks to the assumption that $\mathbf{h}$ is one to one.

$$t^{-2n} \mathcal{H}^n(\mathbf{h}(B)) \leq \int_B \det(U(\mathbf{x}))\, dx \leq t^{2n} \mathcal{H}^n(\mathbf{h}(B))$$

Now B was completely arbitrary. Let it equal $B(\mathbf{x},r) \cap A^+$ where $\mathbf{x} \in A^+$. Then

$$\begin{aligned} t^{-2n} \mathcal{H}^n\left(\mathbf{h}\left(B(\mathbf{x},r) \cap A^+\right)\right) &\leq \int_{B(\mathbf{x},r)} \mathcal{X}_{A^+}(\mathbf{y}) \det(U(\mathbf{y}))\, dm_n(y) \\ &\leq t^{2n} \mathcal{H}^n\left(\mathbf{h}\left(B(\mathbf{x},r) \cap A^+\right)\right) \end{aligned}$$

Divide by $m_n(B(\mathbf{x},r))$ and use the fundamental theorem of calculus. This yields that for $\mathbf{x}$ off a set of m_n measure zero,

$$\begin{aligned} & t^{-2n} \limsup_{r \to 0+} \frac{\mathcal{H}^n(\mathbf{h}(B(\mathbf{x},r) \cap A^+))}{m_n(B(\mathbf{x},r))} \\ \leq\ & \mathcal{X}_{A^+}(\mathbf{x}) \det(U(\mathbf{x})) \\ \leq\ & t^{2n} \liminf_{r \to 0+} \frac{\mathcal{H}^n(\mathbf{h}(B(\mathbf{x},r) \cap A^+))}{m_n(B(\mathbf{x},r))} \end{aligned}$$

However, $t > 1$ was completely arbitrary and this shows that off a set of measure zero,

$$\lim_{r \to 0+} \frac{\mathcal{H}^n(\mathbf{h}(B(\mathbf{x},r) \cap A^+))}{m_n(B(\mathbf{x},r))} = \mathcal{X}_{A^+}(\mathbf{x}) \det(U(\mathbf{x})).$$

This has proved the following lemma.

Lemma 20.1.9 *There is a set of measure zero N such that for $\mathbf{x} \in A^+ \setminus N$,*

$$\lim_{r \to 0+} \frac{\mathcal{H}^n(\mathbf{h}(B(\mathbf{x},r) \cap A^+))}{m_n(B(\mathbf{x},r))} = \det(U(\mathbf{x}))$$

The next theorem removes the assumption that $U(\mathbf{x})^{-1}$ exists. From now on

$$J_*(\mathbf{x}) \equiv \det(U(\mathbf{x})).$$

Theorem 20.1.10 *Let $\mathbf{h} : U \to \mathbb{R}^m$ for $n \leq m, U$ an open set in $\mathbb{R}^n$, and suppose $\mathbf{h}$ is Lipschitz. Then for a.e. $\mathbf{x} \in A$, the set in G where $D\mathbf{h}(\mathbf{x})$ exists,*

$$J_*(\mathbf{x}) = \lim_{r \to 0} \frac{\mathcal{H}^n(\mathbf{h}(B(\mathbf{x}, r) \cap A))}{m_n(B(\mathbf{x},r))}, \tag{20.1.13}$$

where $J_*(\mathbf{x}) \equiv \det(U(\mathbf{x})) = \det\left(D\mathbf{h}(\mathbf{x})^* D\mathbf{h}(\mathbf{x})\right)^{1/2}$.

Proof: The above argument shows that the conclusion of the theorem holds when $J_*(\mathbf{x}) \neq 0$ at least with A replaced with A^+. I will apply this to a modified function in which the corresponding $U(\mathbf{x})$ always has an inverse. Let

$$\mathbf{k} : \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^n$$

be defined as

$$\mathbf{k}(\mathbf{x}) \equiv \begin{pmatrix} \mathbf{h}(\mathbf{x}) \\ \varepsilon \mathbf{x} \end{pmatrix}.$$

Then

$$D\mathbf{k}(\mathbf{x})^* D\mathbf{k}(\mathbf{x}) = D\mathbf{h}(\mathbf{x})^* D\mathbf{h}(\mathbf{x}) + \varepsilon^2 I_n$$

and so

$$\begin{aligned} J_*\mathbf{k}(\mathbf{x})^2 &\equiv \det\left(D\mathbf{h}(\mathbf{x})^* D\mathbf{h}(\mathbf{x}) + \varepsilon^2 I_n\right) \\ &= \det\left(Q^* D Q + \varepsilon^2 I_n\right) > 0 \end{aligned}$$

where D is a diagonal matrix having the nonnegative eigenvalues of $D\mathbf{h}(\mathbf{x})^* D\mathbf{h}(\mathbf{x})$ down the main diagonal, Q an orthogonal matrix.

Therefore, what was just shown applies to $\mathbf{k}$.Let

$$T \equiv \left\{(\mathbf{h}(\mathbf{w}), \mathbf{0})^T : \mathbf{w} \in B(\mathbf{x},r) \cap A\right\},$$

$$\begin{aligned} T_\varepsilon &\equiv \left\{(\mathbf{h}(\mathbf{w}), \varepsilon \mathbf{w})^T : \mathbf{w} \in B(\mathbf{x},r) \cap A\right\} \\ &\equiv \mathbf{k}(B(\mathbf{x},r) \cap A), \end{aligned}$$

then

$$T = PT_\varepsilon$$

where P is the projection map defined by

$$P\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \equiv \begin{pmatrix} \mathbf{x} \\ 0 \end{pmatrix}.$$

Since P decreases distances, it follows from Lemma 20.1.5

$$\begin{aligned} \mathcal{H}^n(\mathbf{h}(B(\mathbf{x},r) \cap A)) &= \mathcal{H}^n(T) = \mathcal{H}^n(PT_\varepsilon) \\ &\leq \mathcal{H}^n(T_\varepsilon) = \mathcal{H}^n(\mathbf{k}(B(\mathbf{x},r) \cap A)). \end{aligned}$$

Now from what was shown earlier,

$$\det \left(D\mathbf{k}(\mathbf{x})^* D\mathbf{k}(\mathbf{x})\right)^{1/2} = \lim_{r\to 0} \frac{\mathcal{H}^n\left(\mathbf{k}\left(B\left(\mathbf{x},r\right)\cap A\right)\right)}{m_n\left(B\left(\mathbf{x},r\right)\right)}$$

for a.e. $\mathbf{x} \in A$. This is because

$$\{\mathbf{x} : D\mathbf{k}(\mathbf{x}) \text{ exists}\} = A$$

and there is no exceptional set where $D\mathbf{k}(\mathbf{x})^* D\mathbf{k}(\mathbf{x})$ fails to have an inverse. Thus for a.e. $\mathbf{x}$,

$$\begin{aligned} \det \left(D\mathbf{k}(\mathbf{x})^* D\mathbf{k}(\mathbf{x})\right)^{1/2} &= \lim_{r\to 0} \frac{\mathcal{H}^n\left(\mathbf{k}\left(B\left(\mathbf{x},r\right)\cap A\right)\right)}{m_n\left(B\left(\mathbf{x},r\right)\right)} \\ &\geq \limsup_{r\to 0} \frac{\mathcal{H}^n\left(\mathbf{h}\left(B\left(\mathbf{x},r\right)\cap A\right)\right)}{m_n\left(B\left(\mathbf{x},r\right)\right)} \end{aligned}$$

Now let $\varepsilon_n \to 0+$ and pass to a limit. This yields that for $a.e.\mathbf{x}$,those not in the union of the exceptional sets corresponding to the exceptional sets which correspond to each ε_n in the sequence ε_n,

$$\det \left(D\mathbf{h}(\mathbf{x})^* D\mathbf{h}(\mathbf{x})\right)^{1/2} \geq \limsup_{r\to 0} \frac{\mathcal{H}^n\left(\mathbf{h}\left(B\left(\mathbf{x},r\right)\cap A\right)\right)}{m_n\left(B\left(\mathbf{x},r\right)\right)}$$

If $\mathbf{x} \in A^+$, then it was shown above, that

$$\det \left(D\mathbf{h}(\mathbf{x})^* D\mathbf{h}(\mathbf{x})\right)^{1/2} = \lim_{r\to 0} \frac{\mathcal{H}^n\left(\mathbf{h}\left(B\left(\mathbf{x},r\right)\cap A\right)\right)}{m_n\left(B\left(\mathbf{x},r\right)\right)}$$

and if $\mathbf{x} \notin A^+$, the above shows that

$$\det \left(D\mathbf{h}(\mathbf{x})^* D\mathbf{h}(\mathbf{x})\right)^{1/2} = 0 \geq \limsup_{r\to 0} \frac{\mathcal{H}^n\left(\mathbf{h}\left(B\left(\mathbf{x},r\right)\cap A\right)\right)}{m_n\left(B\left(\mathbf{x},r\right)\right)} \geq 0$$

and so this has shown that for $a.e.\mathbf{x}$,

$$\lim_{r\to 0} \frac{\mathcal{H}^n\left(\mathbf{h}\left(B\left(\mathbf{x},r\right)\cap A\right)\right)}{m_n\left(B\left(\mathbf{x},r\right)\right)} = J_*(\mathbf{x}) \ \blacksquare$$

Define the following set for future reference.

$$S \equiv \{\mathbf{x} \ \in A, D\mathbf{h}(\mathbf{x}) \text{ exists and } U(\mathbf{x})^{-1} \text{ does not exist}\} \tag{20.1.14}$$

20.1.2 The Area Formula

Since $\mathbf{h}$ is one to one on G, and Lipschitz on G, Lemma 20.1.3 implies one can define a measure ν, on the $\sigma-$ algebra of Lebesgue measurable subsets of G as follows.

$$\nu(E) \equiv \mathcal{H}^n\left(\mathbf{h}\left(E\cap A\right)\right).$$

Recall that A is the set in G on which $D\mathbf{h}(\mathbf{x})$ exists. This is all except a set of measure zero and so one could actually replace the right side with $\mathcal{H}^n(\mathbf{h}(E))$ because the new material has $\mathcal{H}^n$ measure zero. By Lemma 20.1.3, this is a measure and $\nu \ll m$. If $m_n(E) = 0$, then by Lemma 20.1.3, $\mathcal{H}^n(\mathbf{h}(E \cap A)) = 0$. In fact, by this lemma, $\mathcal{H}^n(\mathbf{h}(E \cap A)) \leq K^n m_n(E)$ so it is also clear that ν is σ finite.

Therefore by the corollary to the Radon Nikodym theorem, Corollary **??** on Page **??**, there exists $f \geq 0$, $f(\mathbf{x}) = 0$ if $\mathbf{x} \notin A$, and

$$\nu(E) = \int_E f dm = \int_{A\cap E} f dm_n.$$

In fact,

$$f \in L^1_{loc}(\mathbb{R}^n)$$

Indeed, for a ball B,

$$\infty > K^n m_n(B) \geq \mathcal{H}^n(\mathbf{h}(B \cap A)) \equiv \nu(B) = \int_B f dm_n$$

What is f? I will show that $f(\mathbf{x}) = J_*(\mathbf{x}) = \det(U(\mathbf{x}))$ a.e. Here $U(\mathbf{x}) \equiv \left(D\mathbf{h}(\mathbf{x})^* D\mathbf{h}(\mathbf{x})\right)^{1/2}$. Theorem 20.1.10 and the fundamental theorem of calculus implies that for a.e. $\mathbf{x} \in G$,

$$\begin{aligned} f(\mathbf{x}) &= \lim_{r\to 0} \frac{1}{m_n(B(\mathbf{x},r))} \int_{B(\mathbf{x},r)} f(\mathbf{y})\, dm \\ &= \lim_{r\to 0} \frac{\mathcal{H}^n(\mathbf{h}(B(\mathbf{x},r) \cap A))}{m_n(B(\mathbf{x},r))} = J_*(\mathbf{x}). \end{aligned}$$

Therefore, $f(\mathbf{x}) = J_*(\mathbf{x})$ a.e.

Note that one could define a measure ν_R as $\mathcal{H}^n(\mathbf{h}(E \cap A \cap B(\mathbf{0}, R)))$ and it would still be the case that $\nu_R \ll m_n$ and the Radon Nikodym derivative would still be $J_*(\mathbf{x})$ for $\mathbf{x} \in A \cap B(\mathbf{0}, R)$. You would just apply the above Theorem 20.1.10 for $G = A \cap B(\mathbf{0}, R)$ and the new A would be $A \cap B(\mathbf{0}, R)$.

Now let F be a Borel set in $\mathbb{R}^m$. Recall this implies F is $\mathcal{H}^n$ measurable. Then

$$\begin{aligned} \int_{\mathbf{h}(A)} \mathcal{X}_F(\mathbf{y})\, d\mathcal{H}^n &= \int \mathcal{X}_{F\cap\mathbf{h}(A)}(\mathbf{y})\, d\mathcal{H}^n \\ &= \mathcal{H}^n\left(\mathbf{h}\left(\mathbf{h}^{-1}(F) \cap A\right)\right) \\ &= \nu\left(\mathbf{h}^{-1}(F)\right) = \int \mathcal{X}_{A\cap\mathbf{h}^{-1}(F)}(\mathbf{x})\, J_*(\mathbf{x})\, dm \\ &= \int_A \mathcal{X}_F(\mathbf{h}(\mathbf{x}))\, J_*(\mathbf{x})\, dm. \end{aligned} \tag{20.1.15}$$

Similarly, if

$$A_R \equiv A \cap B(\mathbf{0}, R),$$

$$\int \mathcal{X}_{F\cap\mathbf{h}(A_R)}(\mathbf{y})\, d\mathcal{H}^n = \int_{\mathbf{h}(A_R)} \mathcal{X}_F(\mathbf{y})\, d\mathcal{H}^n = \int_{A_R} \mathcal{X}_F(\mathbf{h}(\mathbf{x}))\, J_*(\mathbf{x})\, dm.$$

Note there are no measurability questions in the above formula because $\mathbf{h}^{-1}(F)$ is a Borel set due to the continuity of $\mathbf{h}$. The Borel measurability of $J_*(\mathbf{x})$ also follows from the observation that $\mathbf{h}$ is continuous and therefore, the partial derivatives are Borel measurable, being the limit of continuous functions. Then $J_*(\mathbf{x})$ is just a continuous function of these partial derivatives. However, things are not so clear if F is only assumed $\mathcal{H}^n$ measurable. Is there a similar formula for F only $\mathcal{H}^n$ measurable?

First consider the case where E is only $\mathcal{H}^n$ measurable but

$$\mathcal{H}^n(E\cap\mathbf{h}(A))=0.$$

By Theorem **??** on Page **??**, there exists a Borel set $F\supseteq E\cap\mathbf{h}(A)$ such that

$$\mathcal{H}^n(F)=\mathcal{H}^n(E\cap\mathbf{h}(A))=0.$$

Then from 20.1.15,

$$\mathcal{X}_{A\cap\mathbf{h}^{-1}(F)}(\mathbf{x})\,J_*(\mathbf{x})=0\text{ a.e.}$$

But

$$0\le\mathcal{X}_{A\cap\mathbf{h}^{-1}(E)}(\mathbf{x})\,J_*(\mathbf{x})\le\mathcal{X}_{A\cap\mathbf{h}^{-1}(F)}(\mathbf{x})\,J_*(\mathbf{x}) \tag{20.1.16}$$

which shows the two functions in 20.1.16 are equal a.e. Therefore by completeness of Lebesgue measure, $\mathcal{X}_{A\cap\mathbf{h}^{-1}(E)}(\mathbf{x})\,J_*(\mathbf{x})$ is Lebesgue measurable and so from 20.1.15,

$$0=\int\mathcal{X}_{E\cap\mathbf{h}(A)}(\mathbf{y})\,d\mathcal{H}^n=\int\mathcal{X}_{F\cap\mathbf{h}(A)}(\mathbf{y})\,d\mathcal{H}^n$$

$$=\int\mathcal{X}_{A\cap\mathbf{h}^{-1}(F)}(\mathbf{x})\,J_*(\mathbf{x})\,dm_n=\int\mathcal{X}_{A\cap\mathbf{h}^{-1}(E)}(\mathbf{x})\,J_*(\mathbf{x})\,dm_n, \tag{20.1.17}$$

which shows 20.1.15 holds in this case where E is $\mathcal{H}^n$ measurable and

$$\mathcal{H}^n(E\cap\mathbf{h}(A))=0.$$

Now let $A_R\equiv A\cap B(\mathbf{0},R)$ for large R and let E be $\mathcal{H}^n$ measurable. By Theorem **??**, there exists $F\supseteq E\cap\mathbf{h}(A_R)$ such that F is Borel and

$$\mathcal{H}^n(F\setminus(E\cap\mathbf{h}(A_R)))=0. \tag{20.1.18}$$

Then

$$(E\cap\mathbf{h}(A_R))\cup(F\setminus(E\cap\mathbf{h}(A_R))\cap\mathbf{h}(A_R))=F\cap\mathbf{h}(A_R)$$

and so

$$\mathcal{X}_{A_R\cap\mathbf{h}^{-1}(F)}J_*=\mathcal{X}_{A_R\cap\mathbf{h}^{-1}(E)}J_*+\mathcal{X}_{A_R\cap\mathbf{h}^{-1}(F\setminus(E\cap\mathbf{h}(A_R)))}J_*$$

where from 20.1.18 and 20.1.17, the second function on the right of the equal sign is Lebesgue measurable and equals zero a.e. Therefore, the first function on the right of the equal sign is also Lebesgue measurable and equals the function on the left a.e. Thus,

$$\int\mathcal{X}_{E\cap\mathbf{h}(A_R)}(\mathbf{y})\,d\mathcal{H}^n=\int\mathcal{X}_{F\cap\mathbf{h}(A_R)}(\mathbf{y})\,d\mathcal{H}^n=\int_{A_R}\mathcal{X}_F(\mathbf{h}(\mathbf{x}))\,J_*(\mathbf{x})\,dm$$

$$= \int \mathcal{X}_{A_R \cap \mathbf{h}^{-1}(F)} (\mathbf{x}) J_* (\mathbf{x}) \, dm_n = \int \mathcal{X}_{A_R \cap \mathbf{h}^{-1}(E)} (\mathbf{x}) J_* (\mathbf{x}) \, dm_n. \quad (20.1.19)$$

Since this holds for any R, it holds for 20.1.19 with A replacing A_R and the function

$$\mathbf{x} \to \mathcal{X}_{A \cap \mathbf{h}^{-1}(E)} (\mathbf{x}) J_* (\mathbf{x})$$

is Lebesgue measurable. Writing this in a more familiar form yields

$$\int_{\mathbf{h}(A)} \mathcal{X}_E (\mathbf{y}) \, d\mathcal{H}^n = \int_A \mathcal{X}_E (\mathbf{h}(\mathbf{x})) J_* (\mathbf{x}) \, dm_n. \quad (20.1.20)$$

From this, it follows that if s is a nonnegative, $\mathcal{H}^n$ measurable simple function, 20.1.20 continues to be valid with s in place of $\mathcal{X}_E$. Then approximating an arbitrary nonnegative $\mathcal{H}^n$ measurable function g by an increasing sequence of simple functions, it follows that 20.1.20 holds with g in place of $\mathcal{X}_E$ and there are no measurability problems because $\mathbf{x} \to g(\mathbf{h}(\mathbf{x})) J_* (\mathbf{x})$ is Lebesgue measurable. This proves the following theorem which is the area formula.

Theorem 20.1.11 *Let $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz continuous. Also let $\mathbf{h}$ be one to one on a measurable set $G \subseteq \mathbb{R}^n$ and let $m \geq n$. Let $A \subseteq G$ be the set of $\mathbf{x} \in G$ on which $D\mathbf{h}(\mathbf{x})$ exists which is all but a set of measure zero, and let $g : \mathbf{h}(A) \to [0, \infty]$ be $\mathcal{H}^n$ measurable. Then*

$$\mathbf{x} \to (g \circ \mathbf{h})(\mathbf{x}) J_* (\mathbf{x})$$

is Lebesgue measurable and

$$\int_{\mathbf{h}(A)} g(\mathbf{y}) \, d\mathcal{H}^n = \int_A g(\mathbf{h}(\mathbf{x})) J_* (\mathbf{x}) \, dm_n$$

where $J_* (\mathbf{x}) = \det (U(\mathbf{x})) = \det \left(D\mathbf{h}(\mathbf{x})^* D\mathbf{h}(\mathbf{x}) \right)^{1/2}$.

Since $\mathcal{H}^n = m_n$ on $\mathbb{R}^n$, this is just a generalization of the usual change of variables formula except that here, one does not even need to know that $\mathbf{h}$ is C^1 so this is much better and in addition it is not limited to $\mathbf{h}$ having values in $\mathbb{R}^n$. Also note that you could replace A with G since they differ by a set of measure zero.

It is easy to generalize the above theorem to the case where $\mathbf{h}$ is only locally Lipschitz. The definition follows.

Definition 20.1.12 *Let $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$. This function is said to be locally Lipschitz if for every $\mathbf{x} \in \mathbb{R}^n$, there exists a ball $B_{\mathbf{x}}$ containing $\mathbf{x}$ and a constant $K_{\mathbf{x}}$ such that for all $\mathbf{y}, \mathbf{z} \in B_{\mathbf{x}}$,*

$$|\mathbf{h}(\mathbf{z}) - \mathbf{h}(\mathbf{y})| \leq K_{\mathbf{x}} |\mathbf{z} - \mathbf{y}|$$

The proof uses a little generalization of Lemma 20.1.1.

Lemma 20.1.13 *If $\mathbf{h}$ is locally Lipschitz and $m_n(T) = 0$, then*

$$\mathcal{H}^n (\mathbf{h}(T)) = 0$$

Proof: Let

$$T_k \equiv \{\mathbf{x} \in T : \mathbf{h} \text{ has Lipschitz constant } k \text{ near } \mathbf{x}\}.$$

Thus $T = \cup_k T_k$. I will show $\mathbf{h}(T_k)$ has $\mathcal{H}^n$ measure zero and then it will follow that

$$\mathbf{h}(T) = \cup_{k=1}^{\infty} \mathbf{h}(T_k), \text{ the } \mathbf{h}(T_k) \text{ increasing in } k,$$

must also have measure zero.

Let $\varepsilon > 0$ be given. By outer regularity, there exists an open set V containing T_k such that $m_n(V) < \frac{\varepsilon}{k^n 6^n}$. For $\mathbf{x} \in T_k$ it follows there exists $r_{\mathbf{x}} < 1$ such that the ball centered at $\mathbf{x}$ with radius $5r_{\mathbf{x}}$ is contained in V and in this ball, $\mathbf{h}$ has Lipschitz constant k and $5kr_{\mathbf{x}} < \delta$ for $\delta > 0$. Then by the Vitali covering theorem, there are disjoint balls $\{B(\mathbf{x}_i, r_i)\}$ such that the enlarged balls $\hat{B}_i$ having five times the radius cover T_k, each being contained in V. Then

$$\mathcal{H}_\delta^n(\mathbf{h}(T_k)) \leq \mathcal{H}_\delta^n\left(\mathbf{h}\left(\cup_{i=1}^{\infty}\hat{B}_i\right)\right) \leq \mathcal{H}_\delta^n\left(\cup_{i=1}^{\infty}\mathbf{h}\left(\hat{B}_i\right)\right)$$

$$\leq \sum_{i=1}^{\infty} \mathcal{H}_\delta^n\left(\mathbf{h}\left(\hat{B}_i\right)\right) \leq \sum_{i=1}^{\infty} \mathcal{H}_\delta^n\left(B\left(\mathbf{h}(\mathbf{x}_i), 5kr_{\mathbf{x}_i}\right)\right)$$

$$\begin{aligned} &\leq \sum_{i=1}^{\infty} \alpha(n)(5kr_{\mathbf{x}_i})^n = (5k)^n \sum_{i=1}^{\infty} \alpha(n) r_{\mathbf{x}_i}^n \\ &= (5k)^n \sum_{i=1}^{\infty} m_n(B(\mathbf{x}_i, r_{\mathbf{x}_i})) \\ &\leq (5k)^n m_n(V) \leq (5k)^n \frac{\varepsilon}{k^n 6^n} < \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, this shows $\mathcal{H}_\delta^n(\mathbf{h}(T_k)) = 0$. Since δ is arbitrary, this implies $\mathcal{H}^n(\mathbf{h}(T_k)) = 0$. Now

$$\mathcal{H}^n(\mathbf{h}(T)) = \lim_{k\to\infty} \mathcal{H}^n(\mathbf{h}(T_k)) = 0. \blacksquare$$

Then an easy generalization of the above formula is the following.

Theorem 20.1.14 *Let $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz. Also suppose that $\mathbf{h}$ is one to one on G, a measurable subset of $\mathbb{R}^n$. Then let $g : \mathbf{h}(G) \to [0, \infty]$ be $\mathcal{H}^n$ measurable. It follows that*

$$\mathbf{x} \to (g \circ \mathbf{h})(\mathbf{x}) J_*(\mathbf{x})$$

is Lebesgue measurable and

$$\int_{\mathbf{h}(G)} g(\mathbf{y})\, d\mathcal{H}^n = \int_G g(\mathbf{h}(\mathbf{x})) J_*(\mathbf{x})\, dm_n$$

where $J_*(\mathbf{x}) = \det(U(\mathbf{x})) = \det\left(D\mathbf{h}(\mathbf{x})^* D\mathbf{h}(\mathbf{x})\right)^{1/2}$.

Proof: Let $\mathcal{C}$ consist of balls of radius less than 1 covering G such that for $B \in \mathcal{C}$, $\mathbf{h}$ is Lipschitz continuous on B. By the Vitali covering theorem, there exists a sequence of these balls $\{B_i\}$ such that they are disjoint and $G \setminus \cup_i B_i$ has measure zero. Then, using Theorem 20.1.14 and the above Lemma 20.1.13,

$$\begin{aligned}\int_{\mathbf{h}(G)} g(\mathbf{y})\, d\mathcal{H}^n &= \int_{\mathbf{h}(\cup_i B_i)} g(\mathbf{y})\, d\mathcal{H}^n = \sum_i \int_{\mathbf{h}(B_i)} g(\mathbf{y})\, d\mathcal{H}^n \\ &= \sum_i \int_{B_i} g(\mathbf{h}(\mathbf{x}))\, J_*(\mathbf{x})\, dm_n \\ &= \int_{\cup_i B_i} g(\mathbf{h}(\mathbf{x}))\, J_*(\mathbf{x})\, dm_n = \int_G g(\mathbf{h}(\mathbf{x}))\, J_*(\mathbf{x})\, dm_n \blacksquare\end{aligned}$$

20.1.3 Mappings That Are Not One To One

Let $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz. We drop the requirement that $\mathbf{h}$ be one to one. This follows [**?**]. See also [**?**] which is where I read it originally. This reference considers this case originally instead of splitting it into a case where $\mathbf{h}$ is one to one and a case where it is not. Again, let A be the set on which $D\mathbf{h}(\mathbf{x})$ exists.

Thus $m_n(N) = 0$ where N is the set where $D\mathbf{h}(\mathbf{x})$ does not exist and $\mathcal{H}^n(\mathbf{h}(S)) = 0$ and so by Lemma 20.1.3

$$\mathcal{H}^n(\mathbf{h}(S \cup N)) \leq \mathcal{H}^n(\mathbf{h}(S)) + \mathcal{H}^n(\mathbf{h}(N)) = 0. \tag{20.1.21}$$

Let $B \equiv \mathbb{R}^n \setminus (S \cup N)$.

A similar lemma to the following was proved in the section on the change of variables formula for a C^1 map. There the proof was based on the inverse function theorem. However, this is no longer possible so a slightly more technical argument is required.

Let S be given by

$$S \equiv \{\mathbf{x} \ \in A, \text{ such that } U(\mathbf{x})^{-1} \text{ does not exist}\}$$

Recall that $D\mathbf{h}(\mathbf{x}) = R(\mathbf{x})\, U(\mathbf{x})$ where R preserves distances.

Lemma 20.1.15 *For S defined above, $\mathcal{H}^n(\mathbf{h}(S)) = 0$.*

Proof: From Theorem 20.1.10, for *a.e.* $\mathbf{x} \in S$ and r is small enough,

$$\frac{\mathcal{H}^n(\mathbf{h}(B(\mathbf{x},r) \cap A))}{m_n(B(\mathbf{x},r))} = \frac{\mathcal{H}^n(\mathbf{h}(B(\mathbf{x},r)))}{m_n(B(\mathbf{x},r))} < \varepsilon.$$

Therefore, whenever $\mathbf{x} \in S$ and r small enough,

$$\mathcal{H}^n(\mathbf{h}(B(\mathbf{x},r))) \leq \varepsilon\alpha(n)\, r^n. \tag{20.1.22}$$

Let $S_k = S \cap B(\mathbf{0},k)$ and for each $\mathbf{x} \in S_k$, let $r_{\mathbf{x}}$ be such that 20.1.22 holds with r replaced by $5r_{\mathbf{x}}$ and

$$B(\mathbf{x},r_{\mathbf{x}}) \subseteq B(\mathbf{0},k).$$

By the Vitali covering theorem, there is a disjoint subsequence of these balls, $\{B(\mathbf{x}_i, r_i)\}$, with the property that $\{B(\mathbf{x}_i, 5r_i)\} \equiv \left\{\widehat{B}_i\right\}$ covers S_k. Then by the way these balls were defined, with 20.1.22 holding for $r = 5r_i$,

$$\mathcal{H}^n(\mathbf{h}(S_k)) \leq \sum_{i=1}^{\infty} \mathcal{H}^n\left(\mathbf{h}\left(\widehat{B}_i\right)\right) \leq 5^n \varepsilon \sum_{i=1}^{\infty} \alpha(n) r_i^n$$

$$= 5^n \varepsilon \sum_{i=1}^{\infty} m_n(B(\mathbf{x}_i, r_i)) \leq 5^n \varepsilon m_n(B(\mathbf{0}, k)).$$

Since ε is arbitrary, this shows $\mathcal{H}^n(\mathbf{h}(S_k)) = 0$. Now letting $k \to \infty$, this shows $\mathcal{H}^n(\mathbf{h}(S)) = 0$. ■

Lemma 20.1.16 *There exists a sequence of disjoint measurable sets, $\{F_i\}$, such that*

$$\cup_{i=1}^{\infty} F_i = B$$

and $\mathbf{h}$ is one to one on F_i.

Proof: $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ is a finite dimensional normed linear space. Let $\mathcal{I}$ be the elements of $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ which are invertible and let $\mathcal{F}$ be a countable dense subset of $\mathcal{I}$. Also let C be a countable dense subset of $B \equiv \mathbb{R}^n \setminus (S \cup N)$. For $\mathbf{c} \in C$ and $T \in \mathcal{F}$,

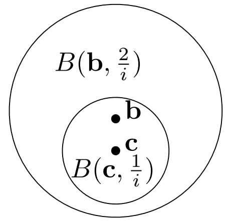
$$E(\mathbf{c}, T, i) \equiv \left\{\mathbf{b} \in B\left(\mathbf{c}, i^{-1}\right) \cap B \text{ such that } (a.) \text{ and } (b.) \text{ hold}\right\}$$

where the conditions *(a.)* and *(b.)* are as follows.

$$\frac{1}{1+\varepsilon} |T\mathbf{v}| \leq |U(\mathbf{b})\mathbf{v}| \text{ for all } \mathbf{v} \tag{a.}$$

$$|\mathbf{h}(\mathbf{a}) - \mathbf{h}(\mathbf{b}) - D\mathbf{h}(\mathbf{b})(\mathbf{a} - \mathbf{b})| \leq \varepsilon |T(\mathbf{a} - \mathbf{b})| \tag{b.}$$

for all $\mathbf{a} \in B\left(\mathbf{b}, 2i^{-1}\right)$. Here $0 < \varepsilon < 1/2$.



Obviously, there are countably many $E(\mathbf{c}, T, i)$. Now suppose $\mathbf{a}, \mathbf{b} \in E(\mathbf{c}, T, i)$ and $\mathbf{h}(\mathbf{a}) = \mathbf{h}(\mathbf{b})$. Then

$$|\mathbf{a} - \mathbf{b}| \leq |\mathbf{a} - \mathbf{c}| + |\mathbf{c} - \mathbf{b}| < \frac{2}{i}.$$

Therefore, from $(a.)$ and $(b.)$,

$$\begin{aligned}\frac{1}{1+\varepsilon}\left|T\left(\mathbf{a}-\mathbf{b}\right)\right| &\leq \left|U\left(\mathbf{b}\right)\left(\mathbf{a}-\mathbf{b}\right)\right| = \left|D\mathbf{h}\left(\mathbf{b}\right)\left(\mathbf{a}-\mathbf{b}\right)\right| \\ &= \left|\mathbf{h}\left(\mathbf{a}\right)-\mathbf{h}\left(\mathbf{b}\right)-D\mathbf{h}\left(\mathbf{b}\right)\left(\mathbf{a}-\mathbf{b}\right)\right| \leq \varepsilon\left|T\left(\mathbf{a}-\mathbf{b}\right)\right|.\end{aligned}$$

Since T is one to one, this shows that $\mathbf{a}=\mathbf{b}$. Thus $\mathbf{h}$ is one to one on $E\left(\mathbf{c},T,i\right)$.

Now let $\mathbf{b}\in B$. Choose $T\in\mathcal{F}$ such that

$$\left\|U\left(\mathbf{b}\right)-T\right\| < \varepsilon\left\|U\left(\mathbf{b}\right)^{-1}\right\|^{-1}.$$

Then for all $\mathbf{v}\in\mathbb{R}^n$,

$$\left|T\mathbf{v}-U\left(\mathbf{b}\right)\mathbf{v}\right| \leq \varepsilon\left\|U\left(\mathbf{b}\right)^{-1}\right\|^{-1}\left|\mathbf{v}\right| \leq \varepsilon\left|U\left(\mathbf{b}\right)\mathbf{v}\right|$$

and so

$$\left|T\mathbf{v}\right| \leq \left(1+\varepsilon\right)\left|U\left(\mathbf{b}\right)\mathbf{v}\right|$$

which yields $(a.)$. Now choose i large enough that for $\left|\mathbf{a}-\mathbf{b}\right|<2i^{-1}$,

$$\begin{aligned}\left|\mathbf{h}\left(\mathbf{a}\right)-\mathbf{h}\left(\mathbf{b}\right)-D\mathbf{h}\left(\mathbf{b}\right)\left(\mathbf{a}-\mathbf{b}\right)\right| &< \frac{\varepsilon}{\left|\left|T^{-1}\right|\right|}\left|\mathbf{a}-\mathbf{b}\right| \\ &\leq \varepsilon\left|T\left(\mathbf{a}-\mathbf{b}\right)\right|\end{aligned}$$

and pick $\mathbf{c}\in C\cap B\left(\mathbf{b},i^{-1}\right)$. Then $\mathbf{b}\in E\left(\mathbf{c},T,i\right)$ and this shows that B equals the union of these sets.

Let $\{E_i\}$ be an enumeration of these sets and define $F_1\equiv E_1$, and if $F_1,\cdots,F_n$ have been chosen, $F_{n+1}\equiv E_{n+1}\setminus\cup_{i=1}^n F_i$. Then $\{F_i\}$ satisfies the conditions of the lemma and this proves the lemma. ■

The following corollary will not be needed right away but it is of interest.

Corollary 20.1.17 *For each E_i in Lemma 20.1.16, $\mathbf{h}^{-1}$ is Lipschitz on $\mathbf{h}\left(E_i\right)$.*

Proof: Pick $\mathbf{a},\mathbf{b}\in E_i$. Then by condition $a.$ and $b.$,

$$\left|\mathbf{h}\left(\mathbf{a}\right)-\mathbf{h}\left(\mathbf{b}\right)\right| \geq \left|D\mathbf{h}\left(\mathbf{b}\right)\left(\mathbf{a}-\mathbf{b}\right)\right|-\varepsilon\left|T\left(\mathbf{a}-\mathbf{b}\right)\right|$$

$$\geq \left(\frac{1}{1+\varepsilon}-\varepsilon\right)\left|T\left(\mathbf{a}-\mathbf{b}\right)\right| \geq r\left|\mathbf{a}-\mathbf{b}\right|$$

for some $r>0$ by the equivalence of all norms on a finite dimensional space. Therefore,

$$\left|\mathbf{h}^{-1}\left(\mathbf{h}\left(\mathbf{a}\right)\right)-\mathbf{h}^{-1}\left(\mathbf{h}\left(\mathbf{b}\right)\right)\right| \leq \frac{1}{r}\left|\mathbf{h}\left(\mathbf{a}\right)-\mathbf{h}\left(\mathbf{b}\right)\right|$$

and this proves the corollary. ■

Now let $g:\mathbf{h}\left(\mathbb{R}^n\right)\to[0,\infty]$ be $\mathcal{H}^n$ measurable. By Theorem 20.1.11,

$$\int_{\mathbf{h}(A)}\mathcal{X}_{\mathbf{h}(F_i)}\left(\mathbf{y}\right)g\left(\mathbf{y}\right)d\mathcal{H}^n = \int_{F_i}g\left(\mathbf{h}\left(\mathbf{x}\right)\right)J_*\left(\mathbf{x}\right)dm. \tag{20.1.23}$$

Now define

$$\mathfrak{n}(\mathbf{y}) = \sum_{i=1}^{\infty} \mathcal{X}_{\mathbf{h}(F_i)}(\mathbf{y}).$$

By Lemma 20.1.3, $\mathbf{h}(F_i)$ is $\mathcal{H}^n$ measurable and so $\mathfrak{n}$ is a $\mathcal{H}^n$ measurable function. For each $\mathbf{y} \in B$, $\mathfrak{n}(\mathbf{y})$ gives the number of elements in $\mathbf{h}^{-1}(\mathbf{y}) \cap B$. From 20.1.23,

$$\int_{\mathbf{h}(\mathbb{R}^n)} \mathfrak{n}(\mathbf{y})\, g(\mathbf{y})\, d\mathcal{H}^n = \int_B g(\mathbf{h}(\mathbf{x}))\, J_*(\mathbf{x})\, dm. \tag{20.1.24}$$

Now define

$$\#(\mathbf{y}) \equiv \text{ number of elements in } \mathbf{h}^{-1}(\mathbf{y}).$$

Theorem 20.1.18 *Let $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz. Then the function $\mathbf{y} \to \#(\mathbf{y})$ is $\mathcal{H}^n$ measurable and if*

$$g : \mathbf{h}(\mathbb{R}^n) \to [0, \infty]$$

is $\mathcal{H}^n$ measurable, then

$$\int_{\mathbf{h}(\mathbb{R}^n)} g(\mathbf{y})\, \#(\mathbf{y})\, d\mathcal{H}^n = \int_{\mathbb{R}^n} g(\mathbf{h}(\mathbf{x}))\, J_*(\mathbf{x})\, dm.$$

Proof: If $\mathbf{y} \notin \mathbf{h}(S \cup N)$, then $\mathfrak{n}(\mathbf{y}) = \#(\mathbf{y})$. By 20.1.21

$$\mathcal{H}^n(\mathbf{h}(S \cup N)) = 0$$

and so $\mathfrak{n}(\mathbf{y}) = \#(\mathbf{y})$ a.e. Since $\mathcal{H}^n$ is a complete measure, $\#(\cdot)$ is $\mathcal{H}^n$ measurable. Letting

$$G \equiv \mathbf{h}(\mathbb{R}^n) \setminus \mathbf{h}(S \cup N),$$

20.1.24 implies

$$\begin{aligned} \int_{\mathbf{h}(\mathbb{R}^n)} g(\mathbf{y})\, \#(\mathbf{y})\, d\mathcal{H}^n &= \int_G g(\mathbf{y})\, \mathfrak{n}(\mathbf{y})\, d\mathcal{H}^n \\ &= \int_B g(\mathbf{h}(\mathbf{x}))\, J_*(\mathbf{x})\, dm \\ &= \int_{\mathbb{R}^n} g(\mathbf{h}(\mathbf{x}))\, J_*(\mathbf{x})\, dm. \blacksquare \end{aligned}$$

As in Theorem 20.1.14, there is an easy generalization based on the Vitali covering theorem to the case where $\mathbf{h}$ is only locally Lipschitz on $\mathbb{R}^n$.

Theorem 20.1.19 *Let $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz. Then the function $\mathbf{y} \to \#(\mathbf{y})$ is $\mathcal{H}^n$ measurable and if*

$$g : \mathbf{h}(\mathbb{R}^n) \to [0, \infty]$$

is $\mathcal{H}^n$ measurable, then

$$\int_{\mathbf{h}(\mathbb{R}^n)} g(\mathbf{y})\, \#(\mathbf{y})\, d\mathcal{H}^n = \int_{\mathbb{R}^n} g(\mathbf{h}(\mathbf{x}))\, J_*(\mathbf{x})\, dm.$$

Chapter 21

The Coarea Formula

Chapter 22

Fourier Analysis in $\mathbb{R}^n$

Chapter 23

Integration for Vector Valued Functions

Chapter 24

Convex Functions

24.1 Continuity properties of convex functions

24.2 Separation properties

24.3 Conjugate functions

24.4 Subgradients

for all $z \in X$. Therefore,

$$\phi^* (y^*) \leq y^* (x) - \phi (x) \leq \phi^* (y^*).$$

Hence

$$y^* (x) = \phi^* (y^*) + \phi (x). \tag{24.4.7}$$

Now if $z^* \in X'$ is arbitrary, Formula 7 shows

$$(z^* - y^*) (x) = z^* (x) - \phi^* (y^*) - \phi (x) \leq \phi^* (z^*) - \phi^* (y^*)$$

and this shows $x \in \delta\phi^* (y^*)$.

Now suppose $x \in \delta\phi^* (y^*)$. Then for $z^* \in X'$,

$$(z^* - y^*) (x) \leq \phi^* (z^*) - \phi^* (y^*)$$

and so, taking sup over all z^*, and using Theorem 24.17,

$$\phi^{**} (x) = \phi (x) \leq y^* (x) - \phi^* (y^*) \leq \phi^{**} (x) .$$

Thus

$$y^* (x) = \phi^* (y^*) + \phi^{**} (x) = \phi^* (y^*) + \phi (x) \geq y^* (z) - \phi (z) + \phi (x)$$

for all $z \in X$ and this implies for all $z \in X$,

$$\phi (z) - \phi (x) \geq y^* (z - x)$$

so $y^* \in \delta\phi (x)$ and this proves the theorem.

Definition 24.4.22 *If X is a Banach space, $u \in H^1 (0, T; X)$ if there exists $g \in L^2 (0, T; X)$ such that*

$$u (t) = u (0) + \int_0^t g (s) \, ds$$

and we define $u' (\cdot) \equiv g (\cdot)$.

The next Lemma is quite interesting for its own sake but it is also used in the next theorem. We leave its proof as an interesting exercise for the reader.

Lemma 24.4.23 *Suppose $g \in L^2 (0, T; X)$. Then*

$$\int_{(\cdot)}^{(\cdot)+h} g (s) \, ds \mathcal{X}_{[0,T-h]} (\cdot) \to g$$

in $L^2 (0, T; X)$.

The following theorem is a form of the chain rule in which the derivative is replaced by the subgradient.

Theorem 24.4.24 *Suppose* $u \in H^1(0,T;X), z \in L^2(0,T;X')$, *and* $z(t) \in \delta\phi(u(t))$ *a.e* $t \in [0,T]$. *Then the function,* $t \to \phi(u(t))$ *is in* $L^1(0,T)$ *and its weak derivative equals* $z(u')$.

Proof: Modify u on a set of measure zero such that $\delta\phi(u(t)) \neq \emptyset$ for all t. Next modify z on a set of measure zero such that for $\widetilde{u}$ and $\widetilde{z}$ the modified functions, $\widetilde{z}(t) \in \delta\phi(\widetilde{u}(t))$ for all t. First we show $t \to \phi(\widetilde{u}(t))$ is in $L^1(0,T)$. Pick $t_0 \in [0,T]$ and let $\widetilde{z}(t_0) \in \delta\phi(\widetilde{u}(t_0))$. Then for $t \in [0,T]$,

$$\widetilde{z}(t_0)(\widetilde{u}(t) - \widetilde{u}(t_0)) + \phi(\widetilde{u}(t_0)) \leq \phi(\widetilde{u}(t)) \leq \widetilde{z}(t)(\widetilde{u}(t) - \widetilde{u}(t_0)) + \phi(\widetilde{u}(t_0))$$

since $\widetilde{z}(t)(\widetilde{u}(t_0) - \widetilde{u}(t)) \leq \phi(\widetilde{u}(t_0)) - \phi(\widetilde{u}(t))$. This inequality shows $t \to \phi(\widetilde{u}(t))$ is in $L^1(0,T)$ since $\widetilde{z} \in L^2(0,T;X')$ and $\widetilde{u} \in L^2(0,T;X)$. Also, for $t \in [0,T-h]$,

$$\mathcal{X}_{[0,T-h]}(t)\,\widetilde{z}(t)\left(\frac{\widetilde{u}(t+h) - \widetilde{u}(t)}{h}\right) \leq \mathcal{X}_{[0,T-h]}(t)\,\frac{\phi(\widetilde{u}(t+h)) - \phi(\widetilde{u}(t))}{h}$$

$$\leq \mathcal{X}_{[0,T-h]}(t)\,\widetilde{z}(t+h)\left(\frac{\widetilde{u}(t+h) - \widetilde{u}(t)}{h}\right)$$

Now $\mathcal{X}_{[0,T-h]}(\cdot)\,\widetilde{z}(\cdot + h) \to z(\cdot)$ in $L^2(0,T;X')$ by continuity of translation. Also,

$$\mathcal{X}_{[0,T-h]}(\cdot)\,\frac{\widetilde{u}(\cdot + h) - \widetilde{u}(\cdot)}{h} = \mathcal{X}_{[0,T-h]}(\cdot)\,\frac{u(\cdot + h) - u(\cdot)}{h}$$

$$= \mathcal{X}_{[0,T-h]}(\cdot)\,\frac{1}{h}\int_{(\cdot)}^{(\cdot)+h} u'(s)\,ds$$

in $L^2(0,T;X)$ and so by Lemma 24.4.23,

$$\mathcal{X}_{[0,T-h]}(\cdot)\,\frac{\phi(\widetilde{u}(\cdot + h)) - \phi(\widetilde{u}(\cdot))}{h} \to z(u')$$

in $L^1(0,T)$.

It follows from the definition of weak derivatives that in the sense of weak derivatives,

$$\frac{d}{dt}(\phi(u(\cdot))) = z(u') \in L^1(0,T).$$

Note that by Theorem 18.2, this implies that for a.e. $t \in [0,T]$, $\phi(u(t))$ is equal to a continuous function, $\phi \circ u$, and that

$$(\phi \circ u)(t) - (\phi \circ u)(0) = \int_0^t z(u')\,ds.$$

There are other rules of calculus which have a generalization to subgradients. The following theorem is on such a generalization. It generalizes the theorem which states that the derivative of a sum equals the sum of the derivatives.

Theorem 24.4.25 *Let ϕ_1 and ϕ_2 be convex, l.s.c. and proper. Then*

$$\delta(\lambda\phi_i)(x) = \lambda\delta\phi_i(x), \ \delta(\phi_1+\phi_2)(x) \supseteq \delta\phi_1(x) + \delta\phi_2(x) \tag{24.4.8}$$

if $\lambda > 0$. If there exists $\overline{x} \in dom(\phi_1) \cap dom(\phi_2)$ and ϕ_1 is continuous at $\overline{x}$ then for all $x \in X$,

$$\delta(\phi_1+\phi_2)(x) = \delta\phi_1(x) + \delta\phi_2(x). \tag{24.4.9}$$

Proof: Formula 24.4.8 is obvious so we only need to show Formula 24.4.9. Suppose $\overline{x}$ is as described. It is clear Formula 24.4.9 holds whenever $x \notin dom(\phi_1) \cap dom(\phi_2)$ since then both sides equal $\emptyset$. Therefore, we will assume $x \in dom(\phi_1) \cap dom(\phi_2)$ in what follows. Let $x^* \in \delta(\phi_1+\phi_2)(x)$. We need to show x^* is the sum of an element of $\delta\phi_1(x)$ and $\delta\phi_2(x)$. Define

$$C_1 \equiv \{(y,a) \in X \times \mathbb{R} : \phi_1(y) - x^*(y-x) - \phi_1(x) \leq a\},$$

$$C_2 \equiv \{(y,a) \in X \times \mathbb{R} : a \leq \phi_2(x) - \phi_2(y)\}.$$

Both C_1 and C_2 are convex and nonempty. In addition to this,

$$(\overline{x}, \phi_1(\overline{x}) - x^*(\overline{x}-x) - \phi_1(x) + 1) \in int(C_1)$$

due to the assumed continuity of ϕ_1 at $\overline{x}$. If $(y,a) \in int(C_1)$ then

$$\phi_1(y) - x^*(y-x) - \phi_1(x) \leq a - \epsilon$$

whenever ϵ is small enough. Therefore, if (y,a) is also in C_2, the assumption that $x^* \in \delta(\phi_1+\phi_2)(x)$ implies

$$a - \epsilon \geq \phi_1(y) - x^*(y-x) - \phi_1(x) \geq \phi_2(x) - \phi_2(y) \geq a,$$

a contradiction. Therefore $int(C_1) \cap C_2 = \emptyset$ and so by Corollary 6.11 and Lemma 24.10, there exists $(w^*, \beta) \in X' \times \mathbb{R}$ with

$$(w^*, \beta) \neq (0,0), \tag{24.4.10}$$

and

$$w^*(y) + \beta a \geq w^*(y_1) + \beta a_1, \tag{24.4.11}$$

whenever $(y,a) \in C_1$ and $(y_1, a_1) \in C_2$.

Claim: $\beta > 0$.

Proof of claim: If $\beta < 0$ let

$$a = \phi_1(\overline{x}) - x^*(\overline{x}-x) - \phi_1(x) + 1,$$

$$a_1 = \phi_2(x) - \phi_2(\overline{x}), \text{ and } y = y_1 = \overline{x}.$$

Then

$$\beta(\phi_1(\overline{x}) - x^*(\overline{x}-x) - \phi_1(x) + 1) \geq \beta(\phi_2(x) - \phi_2(\overline{x})).$$

Dividing by β yields

$$\phi_1(\overline{x}) - x^*(\overline{x} - x) - \phi_1(x) + 1 \leq \phi_2(x) - \phi_2(\overline{x})$$

and so

$$\begin{aligned}\phi_1(\overline{x}) + \phi_2(\overline{x}) - (\phi_1(x) + \phi_2(x)) + 1 &\leq x^*(\overline{x} - x) \\ &\leq \phi_1(\overline{x}) + \phi_2(\overline{x}) - (\phi_1(x) + \phi_2(x)),\end{aligned}$$

a contradiction. Therefore, $\beta \geq 0$.

Now suppose $\beta = 0$. Letting

$$a = \phi_1(\overline{x}) - x^*(\overline{x} - x) - \phi_1(x) + 1,$$

$$(\overline{x}, a) \in int(C_1),$$

and so there exists an open set U containing 0 and $\eta > 0$ such that

$$\overline{x} + U \times (a - \eta, a + \eta) \subseteq C_1.$$

Therefore, Formula 24.4.11 applied to $(\overline{x} + z, a) \in C_1$ and $(\overline{x}, \phi_2(x) - \phi_2(\overline{x})) \in C_2$ for $z \in U$ yields

$$w^*(\overline{x} + z) \geq w^*(\overline{x})$$

for all $z \in U$. Hence $w^*(z) = 0$ on U which implies $w^* = 0$, contradicting Formula 24.4.10. This proves the claim.

Now with the claim, it follows $\beta > 0$ and so, letting $z^* = w^*/\beta$, Formula 24.4.11 implies

$$z^*(y) + a \geq z^*(y_1) + a_1$$

whenever $(y, a) \in C_1$ and $(y_1, a_1) \in C_2$. In particular,

$$\begin{aligned}(y, \phi_1(y) - x^*(y - x) - \phi_1(x)) \in C_1 \text{ and} \\ (y_1, \phi_2(x) - \phi_2(y_1)) \in C_2.\end{aligned} \tag{24.4.12}$$

So letting $y = x$,

$$z^*(x) + (\phi_1(x) - x^*(x - x) - \phi_1(x)) \geq z^*(y_1) + \phi_2(x) - \phi_2(y_1).$$

Therefore,

$$z^*(y_1 - x) \leq \phi_2(y_1) - \phi_2(x)$$

for all y_1 and so $z^* \in \delta\phi_2(x)$. Now let $y_1 = x$ in Formula 24.4.12. Then

$$z^*(y) + \phi_1(y) - x^*(y - x) - \phi_1(x) \geq z^*(x)$$

and so $x^* - z^* \in \delta\phi_1(x)$ so $x^* = z^* + (x^* - z^*) \in \delta\phi_2(x) + \delta\phi_1(x)$ and this proves the theorem.

24.5 Hilbert space

24.6 Exercises

1. For A a maximal monotone operator defined on a Hilbert space H, let

$$G(A) \equiv \{[x,y] : x \in D(A) \text{ and } y \in Ax\}.$$

Show that for $\lambda > 0$, λA is also maximal monotone. We define $J_\lambda(A)$, written as J_λ for short, by

$$J_\lambda(A)(x) \equiv (I + \lambda A)^{-1} x.$$

Show

$$|J_\lambda x - J_\lambda y| \leq |x - y|.$$

Hint: For $r \in (-1,1)$ and $f \in H$, show there exists a solution, u, to the equation,

$$(1+r)u + Au \ni (1+r)f,$$

as follows. Let

$$J_1 = (I + A)^{-1}$$

and show J_1 is Lipschitz continuous with Lipschitz constant 1. This equation has a solution if and only if

$$u = J_1((1+r)f - ru) = Tu.$$

Show T is a contraction map.

2. ↑ Define for A maximal monotone,

$$A_\lambda x \equiv \frac{1}{\lambda}x - \frac{1}{\lambda}J_\lambda x.$$

Show A_λ is Lipschitz continuous with Lipschitz constant no more than $\frac{2}{\lambda}$. Also verify that

$$A_\lambda x \in AJ_\lambda x,$$

and

$$|A_\lambda x| \leq |y|$$

for all $y \in Ax$ if $x \in D(A)$. This operator, A_λ, is called the Yosida approximation to A.

3. ↑ Suppose

$$(y_1 - y, x_1 - x) \geq 0$$

for all $[x,y] \in G(A)$ where A is maximal monotone. Show that this implies $x_1 \in D(A)$ and $y_1 \in Ax_1$. **Hint:** Try to show

$$J_\lambda(x_1 + \lambda y_1) = x_1$$

because then it will follow $x_1 \in D(A)$ and $y_1 \in Ax_1$. To verify this, use the assumption and Problem 2 to conclude

$$0 \leq (y_1 - A_\lambda (x_1 + \lambda y_1), x_1 - J_\lambda (x_1 + \lambda y_1)).$$

Then simplify to find

$$0 \leq -\frac{1}{\lambda} (x_1 - J_\lambda (x_1 + \lambda y_1), x_1 - J_\lambda (x_1 + \lambda y_1)).$$

The problem shows the graphs of these operators are maximal with respect to also being monotone and this is the reason for the name, maximal monotone.

4. $\uparrow$ Suppose $[x_k, y_k] \in G(A)$ and

$$x_k \to x, y_k \rightharpoonup y$$

where the half arrow denotes weak convergence. Show that then $[x, y] \in G(A)$.

5. $\uparrow$ Let A be maximal monotone and let B be Lipschitz and monotone. Then $A + B$ is maximal monotone. **Hint:** First suppose B has Lipschitz constant less than one. Then consider

$$Tx \equiv (I + A)^{-1} (y - Bx).$$

Show T is a contraction map and consequently has a fixed point x which satisfies

$$y \in x + Ax + Bx.$$

Next let $A + B$ play the role of A to conclude that $A + B + B$ is maximal monotone. Continuing in this way, show that any Lipschitz constant is all right.

6. $\uparrow$ Let A and B be maximal monotone, let

$$y \in x_\lambda + B_\lambda x_\lambda + Ax_\lambda,$$

and suppose $B_\lambda x_\lambda$ is bounded independent of λ. Show there exists

$$x_1 \in D(A) \cap D(B)$$

such that

$$y \in x_1 + Bx_1 + Ax_1.$$

Hint: $y - x_\lambda - B_\lambda x_\lambda \in Ax_\lambda$ and so

$$|x_\lambda - x_\mu|^2 \leq (B_\mu x_\mu - B_\lambda x_\lambda, x_\lambda - x_\mu)$$

$$= -(B_\lambda x_\lambda - B_\mu x_\mu, x_\lambda - x_\mu)$$

$$= -(B_\lambda x_\lambda - B_\mu x_\mu, J_\lambda (B) x_\lambda - J_\mu (B) x_\mu) -$$

$$(B_\lambda x_\lambda - B_\mu x_\mu, \lambda B_\lambda x_\lambda - \mu B_\mu x_\mu)$$
$$\leq |(B_\lambda x_\lambda - B_\mu x_\mu, \lambda B_\lambda x_\lambda - \mu B_\mu x_\mu)|.$$

Conclude $\{x_\lambda\}$ is Cauchy as $\lambda \to 0$. Select a subsequence

$$x_\lambda \to x_1, B_\lambda x_\lambda \rightharpoonup z_1, \text{ and } y - x_\lambda - B_\lambda x_\lambda \rightharpoonup z_2.$$

Use Problem 4 and the observation that $J_\lambda(B) x_\lambda - x_\lambda \to 0$ to conclude

$$z_1 \in Bx_1, z_2 \in Ax_1,$$

and

$$y = x + z_1 + z_2.$$

7. $\uparrow$ Let A be maximal monotone and let $B = \partial\phi$ where ϕ is proper, lower semicontinuous, and convex. Suppose

$$\phi(J_\lambda(A)x) \leq \phi(x) + C\lambda$$

and there exists $\xi \in D(A) \cap D(\phi)$. Then $A + \partial\phi$ is maximal monotone. **Hint:** Let $y \in H$ be arbitrary and let x_λ be given by

$$y \in x_\lambda + \partial\phi(x_\lambda) + A_\lambda x_\lambda$$

and show $A_\lambda x_\lambda$ is bounded. Using Problem 6 it will follow $A + \partial\phi$ is maximal monotone. To do this, note

$$(y - x_\lambda - A_\lambda x_\lambda, J_\lambda(x_\lambda) - x_\lambda) \leq C\lambda$$

because $y - x_\lambda - A_\lambda x_\lambda \in \partial\phi(x_\lambda)$. Thus,

$$-(y - x_\lambda - A_\lambda x_\lambda, A_\lambda x_\lambda) \leq C. \tag{24.6.7}$$

Also since $A_\lambda \xi$ is bounded independent of λ, (Problem 2), and A_λ is monotone,

$$\phi(\xi) - \phi(x_\lambda) \geq (y - x_\lambda - A_\lambda x_\lambda, \xi - x_\lambda)$$
$$\geq (y - x_\lambda, \xi - x_\lambda) - (A_\lambda x_\lambda, \xi - x_\lambda)$$
$$\geq (y - x_\lambda, \xi - x_\lambda) - (A_\lambda \xi, \xi - x_\lambda) \geq |x_\lambda|^2 - C|x_\lambda|$$

for some C independent of λ. Hence $C \geq \phi(x_\lambda) + |x_\lambda|^2$. By Theorem **??** we can find $|x_\lambda|$ is bounded and then Formula 24.6.7 shows $A_\lambda x_\lambda$ is bounded.

8. Let ϕ be a proper convex function defined on a normed linear space. Show ϕ is lower semicontinuous if and only if whenever $u_n \to u$, $\phi(u) \leq \liminf_{n\to\infty} \phi(u_n)$.

9. Let $L : D(L) \subseteq L^2(\Omega) \to L^2(\Omega; \mathbb{R}^n)$ be given by $Lu \equiv \nabla u$ where $D(L)$ is defined to be the space of functions in $L^2(\Omega)$ whose weak

The Hausdorff Maximal theorem

The purpose of this appendix is to prove the equivalence between the axiom of choice, the Hausdorff maximal theorem, and the well-ordering principle. The Hausdorff maximal theorem and the well-ordering principle are very useful but a little hard to believe; so, it may be surprising that they are equivalent to the axiom of choice. First we give a proof that the axiom of choice implies the Hausdorff maximal theorem, a remarkable theorem about partially ordered sets.

We say that a nonempty set is partially ordered if there exists a partial order, $\prec$, satisfying

$$x \prec x$$

and

$$\text{if } x \prec y \text{ and } y \prec z \text{ then } x \prec z.$$

An example of a partially ordered set is the set of all subsets of a given set and $\prec \equiv \subseteq$. Note that we can not conclude that any two elements in a partially ordered set are related. In other words, just because $x,\ y$ are in the partially ordered set, it does not follow that either $x \prec y$ or $y \prec x$. We call a subset of a partially ordered set, $\mathcal{C}$, a chain if $x,\ y \in \mathcal{C}$ implies that either $x \prec y$ or $y \prec x$. Sometimes this is called a totally ordered set. We say $\mathcal{C}$ is a maximal chain if whenever $\widetilde{\mathcal{C}}$ is a chain containing $\mathcal{C}$, it follows the two chains are equal. In other words $\mathcal{C}$ is a maximal chain if there is no strictly larger chain.

Lemma 24.6.1 *Let $\mathcal{F}$ be a nonempty partially ordered set with partial order $\prec$. Then assuming the axiom of choice, there exists a maximal chain in $\mathcal{F}$.*

Proof: Let $\mathcal{X}$ be the set of all chains from $\mathcal{F}$. For $\mathcal{C} \in \mathcal{X}$, let

$$S_{\mathcal{C}} = \{x \in \mathcal{F} \text{ such that } \mathcal{C} \cup \{x\} \text{ is a chain strictly larger than } \mathcal{C}\}.$$

If $S_{\mathcal{C}} = \emptyset$ for any $\mathcal{C}$, then $\mathcal{C}$ is maximal and we are done. Thus, assume $S_{\mathcal{C}} \neq \emptyset$ for all $\mathcal{C} \in \mathcal{X}$. Let $f(\mathcal{C}) \in S_{\mathcal{C}}$. (This is where the axiom of choice is being used.) Let

$$g(\mathcal{C}) = \mathcal{C} \cup \{f(\mathcal{C})\}.$$

Thus $g(\mathcal{C}) \supsetneq \mathcal{C}$ and $g(\mathcal{C}) \setminus \mathcal{C} = \{f(\mathcal{C})\} = \{\text{a single element of } \mathcal{F}\}$. We call a subset $\mathcal{T}$ of $\mathcal{X}$ a tower if

$$\emptyset \in \mathcal{T},$$

$$\mathcal{C} \in \mathcal{T} \text{ implies } g(\mathcal{C}) \in \mathcal{T},$$

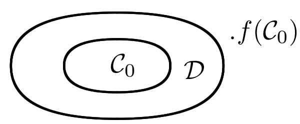
and if $\mathcal{S} \subseteq \mathcal{T}$ is totally ordered with respect to set inclusion, then

$$\cup \mathcal{S} \in \mathcal{T}.$$

Note that $\mathcal{X}$ is a tower. Let $\mathcal{T}_0$ be the intersection of all towers. Thus, $\mathcal{T}_0$ is a tower, the smallest tower. We wish to show that any two sets in $\mathcal{T}_0$ are comparable in the sense of set inclusion so that $\mathcal{T}_0$ is actually a chain. This will proceed in two steps. First let $\mathcal{C}_0$ be a set of $\mathcal{T}_0$ which is comparable to every set of $\mathcal{T}_0$. Such sets exist, $\emptyset$ being an example. Let

$$\mathcal{B} \equiv \{\mathcal{D} \in \mathcal{T}_0 : \mathcal{D} \supsetneq \mathcal{C}_0 \text{ and } f(\mathcal{C}_0) \notin \mathcal{D}\}.$$

The picture represents sets of $\mathcal{B}$.



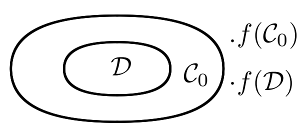
We wish to show $\mathcal{B} = \emptyset$. This will be accomplished by showing $\widetilde{\mathcal{T}_0} \equiv \mathcal{T}_0 \setminus \mathcal{B}$ is a tower. Since $\mathcal{T}_0$ is the smallest tower, this will require that $\widetilde{\mathcal{T}_0} = \mathcal{T}_0$ and so $\mathcal{B} = \emptyset$. Note that for $\mathcal{D} \in \widetilde{\mathcal{T}_0}$, to say $\mathcal{D} \in \widetilde{\mathcal{T}_0}$ is the same as saying $\mathcal{D} \notin \mathcal{B}$.

Claim: $\widetilde{\mathcal{T}_0}$ is a tower and so $\mathcal{B} = \emptyset$.

Proof of the claim: It is clear that $\emptyset \in \widetilde{\mathcal{T}_0}$. Suppose $\mathcal{D} \in \widetilde{\mathcal{T}_0}$. We need to verify $g(\mathcal{D}) \in \widetilde{\mathcal{T}_0}$.

Case 1: $f(\mathcal{D}) \in \mathcal{C}_0$. If $\mathcal{D} \subseteq \mathcal{C}_0$, then if $g(\mathcal{D}) \nsubseteq \mathcal{C}_0$, it would follow $g(\mathcal{D}) \supsetneq \mathcal{C}_0$ due to the assumption that $\mathcal{C}_0$ is comparable. However, both $\{f(\mathcal{D})\}$ and $\mathcal{D}$ are contained in $\mathcal{C}_0$ and so $g(\mathcal{D}) \subseteq \mathcal{C}_0$ which implies $g(\mathcal{D}) \notin \mathcal{B}$. On the other hand, if $\mathcal{D} \supsetneq \mathcal{C}_0$, then since $\mathcal{D} \in \widetilde{\mathcal{T}_0}$, we know $f(\mathcal{C}_0) \in \mathcal{D}$ and so $g(\mathcal{D})$ also contains $f(\mathcal{C}_0)$ implying $g(\mathcal{D}) \notin \mathcal{B}$.

Case 2: $f(\mathcal{D}) \notin \mathcal{C}_0$. If $\mathcal{D} \subsetneq \mathcal{C}_0$ then we can't have $f(\mathcal{D}) \notin \mathcal{C}_0$ because if this were so, $g(\mathcal{D}\ )$ would not compare to $\mathcal{C}_0$.



Hence if $f(\mathcal{D}) \notin \mathcal{C}_0$, then $\mathcal{D} \supseteq \mathcal{C}_0$. If $\mathcal{D} = \mathcal{C}_0$, then $g(\mathcal{D}) = g(\mathcal{C}_0)$ so $g(\mathcal{D}) \notin \mathcal{B}$. Therefore, assume $\mathcal{D} \supsetneq \mathcal{C}_0$. Then, since $\mathcal{D}$ is in $\widetilde{\mathcal{T}_0}$, $f(\mathcal{C}_0) \in \mathcal{D}$ and so $f(\mathcal{C}_0) \in g(\mathcal{D})$. Therefore, $g(\mathcal{D}) \in \widetilde{\mathcal{T}_0}$.

Now suppose $\mathcal{S}$ is a totally ordered subset of $\widetilde{\mathcal{T}_0}$. Then if every element of $\mathcal{S}$ is contained in $\mathcal{C}_0$, so is $\cup\mathcal{S}$ and so $\cup\mathcal{S} \in \widetilde{\mathcal{T}_0}$. If, on the other hand, some chain from $\mathcal{S}$, $\mathcal{C}$, contains $\mathcal{C}_0$ properly, then since $\mathcal{C} \notin \mathcal{B}$, $f(\mathcal{C}_0) \in \mathcal{C} \subseteq \cup\mathcal{S}$ showing that $\cup\mathcal{S} \notin \mathcal{B}$ also. This has proved $\widetilde{\mathcal{T}_0}$ is a tower and since $\mathcal{T}_0$ is the smallest tower, it follows $\widetilde{\mathcal{T}_0} = \mathcal{T}_0$.

Now we define $\mathcal{T}_1$ to be the set of all chains from $\mathcal{T}_0$ which are comparable to every chain from $\mathcal{T}_0$.

Claim: $\mathcal{T}_1$ is a tower.

Proof of the claim: It is clear that $\emptyset \in \mathcal{T}_1$. Suppose $\mathcal{C}_0 \in \mathcal{T}_1$. We need to verify that $g(\mathcal{C}_0) \in \mathcal{T}_1$. Let $\mathcal{D} \in \mathcal{T}_0$ and consider $g(\mathcal{C}_0) \equiv \mathcal{C}_0 \cup \{f(\mathcal{C}_0)\}$. If $\mathcal{D} \subseteq \mathcal{C}_0$, then $\mathcal{D} \subseteq g(\mathcal{C}_0)$. If $\mathcal{D} \supsetneq \mathcal{C}_0$, then $\mathcal{D} \supseteq g(\mathcal{C}_0)$ by what was just shown $\left(\widetilde{\mathcal{T}_0} = \mathcal{T}_0\right)$.

Hence $g(\mathcal{C}_0)$ is comparable to every set of $\mathcal{T}_1$. Now suppose $\mathcal{S}$ is a chain of elements of $\mathcal{T}_1$ and let $\mathcal{D}$ be an element of $\mathcal{T}_0$. If every element of $\mathcal{S}$ is contained in $\mathcal{D}$, then $\cup\mathcal{S}$ is also contained in $\mathcal{D}$. On the other hand, if some set, $\mathcal{C}$, from $\mathcal{S}$ contains $\mathcal{D}$ properly, then $\cup\mathcal{S}$ also contains $\mathcal{D}$. Thus $\cup\mathcal{S} \in \mathcal{T}_1$.

This shows $\mathcal{T}_1$ is a tower and proves therefore, that $\mathcal{T}_0 = \mathcal{T}_1$. Thus every set of $\mathcal{T}_0$ compares with every other set of $\mathcal{T}_0$ showing $\mathcal{T}_0$ is a chain in addition to being a tower.

Now $\cup\mathcal{T}_0, g(\cup\mathcal{T}_0) \in \mathcal{T}_0$. Hence, because $g(\cup\mathcal{T}_0)$ is a chain in $\mathcal{T}_0$, and $\mathcal{T}_0$ is a chain, it follows $g(\cup\mathcal{T}_0) \subseteq \cup\mathcal{T}_0$. Thus

$$\cup\mathcal{T}_0 \supseteq g(\cup\mathcal{T}_0) \supsetneq \cup\mathcal{T}_0,$$

a contradiction. Hence there must exist a maximal chain after all. This proves the lemma.

If X is a nonempty set, we say $\leq$ is an order on X if

$$x \leq x,$$

and if $x,\ y \in X$, then

$$\text{either } x \leq y \text{ or } y \leq x$$

and

$$\text{if } x \leq y \text{ and } y \leq z \text{ then } x \leq z.$$

We say that $\leq$ is a well order and say that $(X, \leq)$ is a well-ordered set if every nonempty subset of X has a smallest element. More precisely, if $S \neq \emptyset$ and $S \subseteq X$ then there exists an $x \in S$ such that $x \leq y$ for all $y \in S$. A familiar example of a well-ordered set is the natural numbers.

Lemma 24.6.2 *The Hausdorff maximal principle implies every nonempty set can be well-ordered.*

Proof: Let X be a nonempty set and let $a \in X$. Then $\{a\}$ is a well-ordered subset of X. Let

$$\mathcal{F} = \{S \subseteq X : \text{ there exists a well order for } S\}.$$

Thus $\mathcal{F} \neq \emptyset$. We will say that for $S_1,\ S_2 \in \mathcal{F}$, $S_1 \prec S_2$ if $S_1 \subseteq S_2$ and there exists a well order for S_2, $\leq_2$ such that

$$(S_2, \leq_2) \text{ is well-ordered}$$

and if

$$y \in S_2 \setminus S_1 \text{ then } x \leq_2 y \text{ for all } x \in S_1,$$

and if $\leq_1$is the well order of S_1 then the two orders are consistent on S_1. Then we observe that $\prec$ is a partial order on $\mathcal{F}$. By the Hausdorff maximal principle, we let $\mathcal{C}$ be a maximal chain in $\mathcal{F}$ and let

$$X_\infty = \cup\mathcal{C}.$$

We also define an order, $\leq$, on X_∞ as follows. If $x,\ y$ are elements of X_∞, we pick $S \in \mathcal{C}$ such that $x,\ y$ are both in S. Then if $\leq_S$ is the order on S, we let $x \leq y$ if and only if $x \leq_S y$. This definition is well defined because of the definition of the order, $\prec$. Now let U be any nonempty subset of X_∞. Then $S \cap U \neq \emptyset$ for some $S \in \mathcal{C}$. Because of the definition of $\leq$, if $y \in S_2 \setminus S_1,\ S_i \in \mathcal{C}$, then $x \leq y$ for all $x \in S_1$. Thus, if $y \in X_\infty \setminus S$ then $x \leq y$ for all $x \in S$ and so the smallest element of $S \cap U$ exists and is the smallest element in U. Therefore X_∞ is well-ordered. Now suppose there exists $z \in X \setminus X_\infty$. Define the following order, $\leq_1$, on $X_\infty \cup \{z\}$.

$$x \leq_1 y \text{ if and only if } x \leq y \ \text{ whenever } x, y \in X_\infty$$

$$x \leq_1 z \text{ whenever } x \in X_\infty.$$

Then let

$$\widetilde{\mathcal{C}} = \{S \in \mathcal{C} \text{ or } X_\infty \cup \{z\}\}.$$

Then $\widetilde{\mathcal{C}}$ is a strictly larger chain than $\mathcal{C}$ contradicting maximality of $\mathcal{C}$. Thus $X \setminus X_\infty = \emptyset$ and this shows X is well-ordered by $\leq$. This proves the lemma.

With these two lemmas we can now state the main result.

Theorem 24.6.3 *The following are equivalent.*

The axiom of choice

The Hausdorff maximal principle

The well-ordering principle.

Proof: It only remains to prove that the well-ordering principle implies the axiom of choice. Let I be a nonempty set and let X_i be a nonempty set for each $i \in I$. Let $X = \cup\{X_i : i \in I\}$ and well order X. Let $f(i)$ be the smallest element of X_i. Then

$$f \in \prod_{i \in I} X_i.$$

24.7 Exercises

1. Zorn's lemma states that in a nonempty partially ordered set, if every chain has an upper bound, there exists a maximal element, x in the partially ordered set. When we say x is maximal, we mean that if $x \prec y$, it follows $y = x$. Show Zorn's lemma is equivalent to the Hausdorff maximal theorem.

2. Let X be a vector space. We say $Y \subseteq X$ is a Hamel basis if every element of X can be written in a unique way as a finite linear combination of elements in Y. Show that every vector space has a Hamel basis and that if $Y,\ Y_1$ are two Hamel bases of X, then there exists a one to one and onto map from Y to Y_1.

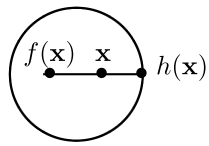
3. ↑ Using the Baire category theorem of Chapter 3 show that any Hamel basis of a Banach space is either finite or uncountable.

Stone's Theorem and Partitions of Unity

Taylor Series and Analytic Functions

The Brouwer fixed point theorem

Proof: Suppose there is no fixed point for f. Then define $h(\mathbf{x})$ as shown in the following picture.



Then $h : B \to \partial B$ is a continuous retraction contradicting Lemma 24.63. This proves the theorem.

Corollary 24.7.65 *Let K be any compact convex subset of $\mathbb{R}^n$ and let $f : K \to K$ be continuous. Then f has a fixed point.*

Proof: Let $K \subseteq B(0, r)$ and define $g : \overline{B(0, r)} \to \overline{B(0, r)}$ by

$$g(\mathbf{x}) \equiv f \circ proj_K(\mathbf{x}).$$

Then g is continuous so $g(\mathbf{x}) = \mathbf{x}$ for some $\mathbf{x} \in B(0, r)$. Thus

$$f(proj_K(\mathbf{x})) = \mathbf{x}.$$

Since $f(K) \subseteq K$, it follows that $\mathbf{x} \in K$ and so $proj_K(\mathbf{x}) = \mathbf{x}$. Hence $f(\mathbf{x}) = \mathbf{x}$. This proves the corollary.