Math 346 Lecture #2
6.2 The Fréchet Derivative in $\mathbb{R}^n$

Definition 6.2.1. Let U be an open subset of $\mathbb{R}^n$, $\mathrm{x} \in U$, and $f : U \to \mathbb{R}^m$. We say f is Fréchet differentiable at x if there is an $A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ such that

$$\lim_{h \to 0} \frac{\|f(\mathrm{x}+h) - f(\mathrm{x}) - Ah\|}{\|h\|} = 0.$$

[We did not specify which norms on $\mathbb{R}^n$ and $\mathbb{R}^m$ are being used here. Since all norms on $\mathbb{R}^p$ are topologically equivalent, the existence of the limit is independent of which norms on $\mathbb{R}^n$ and on $\mathbb{R}^m$ we use.]

When $f : U \to \mathbb{R}^m$ is Fréchet differentiable at $\mathrm{x} \in U$, we write $Df(\mathrm{x})$ for the linear transformation $A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ that appears in the limit.

The Fréchet derivative is sometimes called "the" derivative (we have not proven uniqueness of A but will shortly), or the total derivative to distinguish it from the directional (or Gâteaux) derivative.

We often refer to Fréchet differentiable simply as differentiable.

We say that f is differentiable on U if f is differentiable at each $\mathrm{x} \in U$ and write $Df : U \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ for the derivative.

Example (in lieu of 6.2.2). In standard coordinates, let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $f(x, y, z) = (x^2 + y, yz)$.

We show that

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 3 & 2 \end{bmatrix}$$

is the Fréchet derivative of f at $(1, 2, 3)$.

To this end we have for $h = (h_1, h_2, h_3)$ that

$$\begin{aligned} &f((1,2,3) + (h_1, h_2, h_3)) - f(1,2,3) \\ &\quad = ((1+h_1)^2 + (2+h_2), (2+h_2)(3+h_3)) - (3,6) \\ &\quad = (1 + 2h_1 + h_1^2 + 2 + h_2, 6 + 2h_3 + 3h_2 + h_2h_3) - (3,6) \\ &\quad = (2h_1 + h_2 + h_1^2, 2h_3 + 3h_2 + h_2h_3) \end{aligned}$$

and that

$$Ah = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 3 & 2 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = \begin{bmatrix} 2h_1 + h_2 \\ 3h_2 + 2h_3 \end{bmatrix}.$$

Thus

$$f((1,2,3) + (h_1, h_2, h_3)) - f(1,2,3) - Ah = (h_1^2, h_2h_3)$$

and so

$$\|f((1,2,3) + (h_1, h_2, h_3)) - f(1,2,3) - Ah\|_2 = \|(h_1^2, h_2h_3)\|_2 = \sqrt{h_1^4 + h_2^2h_3^2}.$$

Since

$$\|(h_1, h_2, h_3)\|_2 = \sqrt{h_1^2 + h_2^2 + h_3^2}$$

we have

$$\lim_{h \to 0} \frac{\sqrt{h_1^4 + h_2^2 h_3^2}}{\sqrt{h_1^2 + h_2^2 + h_3^2}} \leq \lim_{h \to 0} \frac{\sqrt{(h_1^2 + h_2^2)(h_1^2 + h_2^2 + h_3^2)}}{\sqrt{h_1^2 + h_2^2 + h_3^2}} = \lim_{h \to 0} \sqrt{h_1^2 + h_2^2} = 0$$

because

$$\begin{aligned} h_1^4 + h_2^2 h_3^2 &\leq h_1^2(h_1^2 + 2h_2^2 + h_3^2) + h_2^2 h_3^2 + h_2^2 h_2^2 \\ &= h_1^2(h_1^2 + h_2^2 + h_3^2) + h_1^2 h_2^2 + h_2^2 h_2^2 + h_3^2 h_2^2 \\ &= (h_1^2 + h_2^2)(h_1^2 + h_2^2 + h_3^2). \end{aligned}$$

Example (in lieu of 6.2.3). At an arbitrary point (x, y, z) the derivative of $f(x, y, z) = (x^2 + y, yz)$ is the matrix function

$$Df(x, y, z) = \begin{bmatrix} 2x & 1 & 0 \\ 0 & z & y \end{bmatrix}.$$

Nota Bene 6.2.4. Beware that $(x, y, z) \to Df(x, y, z)$ in the Example is not linear in (x, y, z), i.e., $Df(\alpha x, \alpha y, \alpha z)$ is not equal to $\alpha Df(x, y, z)$.

In general we expect for a differentiable $f : U \to \mathbb{R}^m$ that $\mathrm{x} \to Df(\mathrm{x})$ is not linear in x.

For fixed $\mathrm{x} \in U$ the function $\mathrm{v} \to Df(\mathrm{x})\mathrm{v}$ is linear in v because $Df(\mathrm{x}) \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$.

Example 6.2.5. A linear transformation $L \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ is differentiable on $\mathbb{R}^n$ because for any $\mathrm{x} \in \mathbb{R}^n$ we have

$$\lim_{h \to 0} \frac{\|L(\mathrm{x} + h) - L(\mathrm{x}) - L(h)\|}{\|h\|} = \lim_{h \to 0} \frac{\|0\|}{\|h\|} = 0.$$

Thus we have $DL(\mathrm{x}) = L$ so that $DL(\mathrm{x})\mathrm{v} = L\mathrm{v}$ for every $\mathrm{x} \in \mathbb{R}^n$, i.e., the derivative is independent of x.

If A is the matrix representation of L in the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, i.e., $[L(\mathrm{x})] = A[\mathrm{x}]$, then $DL(\mathrm{x}) = A$ in the standard bases, i.e., $[DL(\mathrm{x})\mathrm{v}] = A[\mathrm{v}]$.

If the matrix A is the transpose of a $n \times 1$ real matrix a, i.e, $A = a^{\mathrm{T}}$, then the linear functional $L(\mathrm{x}) = \langle a, \mathrm{x} \rangle = a^{\mathrm{T}}\mathrm{x}$ from $\mathbb{R}^n$ to $\mathbb{R}$ can also be expressed as $L(\mathrm{x}) = \langle \mathrm{x}, a \rangle = \mathrm{x}^{\mathrm{T}} a$ because $\langle a, \mathrm{x} \rangle = \langle \mathrm{x}, a \rangle$, but the derivative of L is $DL(\mathrm{x}) = a^{\mathrm{T}}$, i.e., $[DL(\mathrm{x})\mathrm{v}] = a^{\mathrm{T}}[\mathrm{v}]$, not a because $a[\mathrm{v}]$ makes no sense.

Example 6.2.6. The **dual space** of $\mathbb{R}^m$, equipped with the standard inner product $\langle \mathrm{x}, \mathrm{y} \rangle = \mathrm{x}^T \mathrm{y}$, is the vector space $(\mathbb{R}^m)^* = \mathscr{L}(\mathbb{R}^m, \mathbb{R}) = \mathscr{B}(\mathbb{R}^m, \mathbb{R})$.

The vector space $(\mathbb{R}^m)^*$ is isomorphic to $\mathbb{R}^m$ by the Finite Dimensional Reisz Representation Theorem: for each linear function $L : \mathbb{R}^m \to \mathbb{R}$, there exists a unique $\mathrm{y} \in \mathbb{R}^m$ such that $L(\mathrm{x}) = \langle \mathrm{y}, \mathrm{x} \rangle = \mathrm{y}^T \mathrm{x}$, which gives an isomorphism $\mathrm{y} \to \langle \mathrm{y}, \mathrm{x} \rangle$ from $\mathbb{R}^m$ to $(\mathbb{R}^m)^*$.

Because $\langle \mathrm{y}, \mathrm{x} \rangle = \mathrm{y}^{\mathrm{T}}\mathrm{x}$ we typically write the elements x of $\mathbb{R}^m$ as column vectors, i.e., $m \times 1$ matrices, and the elements y^T of $(\mathbb{R}^m)^*$ as row vectors, i.e., $1 \times m$ matrices (where y is a column vector).

We do this to distinguish vectors in the isomorphic vector spaces $\mathbb{R}^m$ and $(\mathbb{R}^m)^*$.

For $A \in M_{n\times m}(\mathbb{R})$, the function $f(\mathrm{x}) = A\mathrm{x}$ from $\mathbb{R}^m$ to $\mathbb{R}^n$ is differentiable on $\mathbb{R}^m$ with $Df(\mathrm{x}) = A$.

The matrix A also defines a function $g(\mathrm{x}) = \mathrm{x}^T A$ from $\mathbb{R}^n$ to $(\mathbb{R}^m)^*$, i.e., $\mathrm{x}^T A$ is a row vector, i.e., an $1 \times m$ vector.

It is HW (Exercise 6.10) to show that g is differentiable on $\mathbb{R}^n$ with derivative Dg satisfying $Dg(\mathrm{x})\mathrm{v} = \mathrm{v}^{\mathrm{T}} A$ for $\mathrm{v} \in \mathbb{R}^n$.

If, instead, we write elements of $(\mathbb{R}^m)^*$ as column vectors, i.e., $m \times 1$ matrices, then the matrix $A \in M_{n\times m}(\mathbb{R})$ also defines a function $g : \mathbb{R}^n \to (\mathbb{R}^m)^*$ given by $g(\mathrm{x}) = A^{\mathrm{T}}\mathrm{x}$.

It is HW (Exercise 6.10) to show that this g is differentiable on $\mathbb{R}^n$ with derivative Dg satisfying $Dg(\mathrm{x})\mathrm{v} = A^{\mathrm{T}}\mathrm{v}$ for $\mathrm{v} \in \mathbb{R}^n$.

Remark 6.2.7. A linear transformation $L : \mathbb{R} \to \mathbb{R}^m$ is given in (standard) coordinates by scalar product of a column vector $[\ell_1, \ldots, \ell_m]^{\mathrm{T}}$.

When $m = 1$, the linear transformation is the scalar product of a 1×1 matrix.

For a differentiable function $f : (a, b) \to \mathbb{R}$ the derivative $Df(x) = f'(x)$ is an element of $\mathscr{B}(\mathbb{R})$ represented in (standard) coordinates by $[f'(x)]$.

Example 6.2.8. For a differentiable curve $\gamma : (a, b) \to \mathbb{R}^n$, the derivative or velocity $\gamma'(x)$ is precisely the total derivative $D\gamma(x)$.

This is because for each $x \in (a, b)$ we have

$$\lim_{h\to 0} \frac{\|\gamma(x + h) - \gamma(x) - \gamma'(x)h\|}{|h|} = \lim_{h\to 0} \left\| \frac{\gamma(x + h) - \gamma(x)}{h} - \gamma'(x) \right\| = 0.$$

Remark 6.2.9. The Fréchet or total derivative $Df(\mathrm{x})$, when it exists, defines the best linear approximation $L(h) = Df(\mathrm{x})h$ of $f(\mathrm{x} + h) - f(\mathrm{x})$ in the sense that for all $\epsilon > 0$ there is a $\delta > 0$ such that for all $0 < \|h\| < \delta$ there holds

$$|\ \|f(\mathrm{x} + h) - f(\mathrm{x})\| - \|L(h)\|\ | \leq \|f(\mathrm{x} + h) - f(\mathrm{x}) - L(h)\| < \epsilon\|h\|,$$

from which follows

$$\|L(h)\| - \epsilon\|h\| < \|f(\mathrm{x} + h) - f(\mathrm{x})\| < \|L(h)\| + \epsilon\|h\|$$

(compare with Remark 6.1.2).

We now prove that if a Fréchet derivative exists, it is unique.

Proposition 6.2.10. Let U be open in $\mathbb{R}^n$. If $f : U \to \mathbb{R}^m$ is differentiable at $\mathrm{x} \in U$, then $Df(\mathrm{x})$ is unique.

Proof. Let $L_1, L_2 \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ satisfy

$$\lim_{h\to 0} \frac{\|f(\mathrm{x}+h) - f(\mathrm{x}) - L_i(h)\|}{\|h\|} = 0, \ \ i = 1, 2.$$

For $\mathrm{v} \neq 0$ and $t \neq 0$ we have

$$\frac{\|L_1(\mathrm{v}) - L_2(\mathrm{v})\|}{\|\mathrm{v}\|} = \frac{|t|}{|t|} \frac{\|L_1(\mathrm{v}) - L_2(\mathrm{v})\|}{\|\mathrm{v}\|} = \frac{\|L_1(t\mathrm{v}) - L_2(t\mathrm{v})\|}{\|t\mathrm{v}\|},$$

where

$$\begin{aligned}
&\frac{\|L_1(t\mathrm{v}) - L_2(t\mathrm{v})\|}{\|t\mathrm{v}\|} \\
&\quad = \frac{\|(f(\mathrm{x}+t\mathrm{v}) - f(\mathrm{x}) - L_2(t\mathrm{v})) - (f(\mathrm{x}+t\mathrm{v}) - f(\mathrm{x}) - L_1(t\mathrm{v}))\|}{\|t\mathrm{v}\|} \\
&\quad \leq \frac{\|f(\mathrm{x}+t\mathrm{v}) - f(\mathrm{x}) - L_2(t\mathrm{v})\|}{\|t\mathrm{v}\|} + \frac{\|f(\mathrm{x}+t\mathrm{v}) - f(\mathrm{x}) - L_1(t\mathrm{v})\|}{\|t\mathrm{v}\|}
\end{aligned}$$

These last two expressions go to 0 as $t \to 0$ by hypothesis.

Thus $L_1(\mathrm{v}) = L_2(\mathrm{v})$ for all $v \in \mathbb{R}^n$, which implies that $L_1 = L_2$. □

How could we more easily compute the Fréchet derivative when we know it exists?

Theorem 6.2.11 (the pointwise version). Let U be open in $\mathbb{R}^n$, and express $f : U \to \mathbb{R}^m$ in standard coordinates, i.e., $f = (f_1, \ldots, f_m)$, where $f_i : U \to \mathbb{R}$ for each $i = 1, \ldots, m$. If f is differentiable at $\mathrm{x} \in U$, then the partial derivatives $D_j f_i(\mathrm{x})$ exist for all $j = 1, \ldots, n$, and for all $i = 1, \ldots, m$, and the matrix representation of $Df(\mathrm{x})$ in the standard coordinates is the Jacobian matrix

$$J(\mathrm{x}) = \begin{bmatrix} D_1 f_1(\mathrm{x}) & D_2 f_1(\mathrm{x}) & \cdots & D_n f_1(\mathrm{x}) \\ D_1 f_2(\mathrm{x}) & D_2 f_2(\mathrm{x}) & \cdots & D_n f_2(\mathrm{x}) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 f_m(\mathrm{x}) & D_2 f_m(\mathrm{x}) & \cdots & D_n f_m(\mathrm{x}) \end{bmatrix}.$$

Proof. Let J_j be the j^{th} column of $J(\mathrm{x})$, i.e., $J_j = Df(\mathrm{x})e_j$, and let J_{ij} be the i^{th} entry of the j^{th} column of J.

For $h = re_j$ we have

$$\begin{aligned}
0 &= \lim_{h\to 0} \frac{\|f(\mathrm{x}+h) - f(\mathrm{x}) - Df(\mathrm{x})h\|}{\|h\|} \\
&= \lim_{r\to 0} \frac{\|f(\mathrm{x}+re_j) - f(\mathrm{x}) - rDf(\mathrm{x})e_j\|}{|r| \, \|e_j\|} \\
&= \lim_{r\to 0} \frac{\|f(x_1, \ldots, x_j + r, \ldots, x_n) - f(x_1, \ldots, x_n) - rJ_j\|}{|r|}
\end{aligned}$$

This implies the each component of the vector function in the numerator goes to 0 as $r \to 0$, i.e.,

$$\lim_{r\to 0} \frac{|f_i(x_1,\ldots,x_j+r,\ldots,x_n) - f_i(x_1,\ldots,x_n) - rJ_{ij}|}{|r|} = 0.$$

This shows that the partial derivative $D_j f_i(\mathrm{x})$ exists and equals J_{ij}, and that the entries of $J(\mathrm{x})$ are precisely the partial derivatives $D_j f_i(\mathrm{x})$. $\square$

Note. The statement of Theorem 6.2.11 in the book assumes that f is Fréchet differentiable on U and concludes that the partial derivatives $D_i f_j(\mathrm{x})$ exist for all $\mathrm{x} \in U$. But the proof of Theorem 6.2.11 in the book shows that Fréchet differentiability at a single point implies the existence of the partial derivatives at that point.

Remark 6.2.12. We often call $Df(\mathrm{x})$ the Jacobian matrix even though we have expressed the Jacobian as the matrix representation in the standard coordinates.

Example (in lieu of 6.2.13). The function $f(x,y,z) = (x^2+y, yz)$ is differentiable on $\mathbb{R}^3$, so by Theorem 6.2.11, the Jacobian of f is computed to be

$$Df(x,y,z) = \begin{bmatrix} 2x & 1 & 0 \\ 0 & z & y \end{bmatrix}$$

by computing the partial derivatives.

How do we determine if a function is differentiable on an open set?

Theorem 6.2.14. For U open in $\mathbb{R}^n$ and $f : U \to \mathbb{R}^m$ given by components $f = (f_1,\ldots,f_m)$ in standard coordinates, if the partial derivatives $D_i f_j(\mathrm{x})$ exist and are continuous on U for all $i = 1,\ldots,n$ and all $j = 1,\ldots,m$, then f is differentiable on U.

Proof. Supposing all of the partial derivatives exist we can form the matrix

$$J(\mathrm{x}) = \begin{bmatrix} D_1 f_1(\mathrm{x}) & D_2 f_1(\mathrm{x}) & \cdots & D_n f_1(\mathrm{x}) \\ D_1 f_2(\mathrm{x}) & D_2 f_2(\mathrm{x}) & \cdots & D_n f_2(\mathrm{x}) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 f_m(\mathrm{x}) & D_2 f_m(\mathrm{x}) & \cdots & D_n f_m(\mathrm{x}) \end{bmatrix}.$$

We will show for each $\mathrm{x} \in U$ that $J(\mathrm{x}) \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ satisfies

$$\lim_{h\to 0} \frac{\|f(\mathrm{x}+h) - f(\mathrm{x}) - J(\mathrm{x})h\|}{\|h\|} = 0.$$

Since all norms on $\mathbb{R}^p$ are topologically equivalent, we will use the ∞-norm for both the numerator and the denominator.

Choose $\delta > 0$ small enough so that $B(\mathrm{x},\delta) \subset U$ (we are using that U is open here).

For $\mathrm{y} \in B(\mathrm{x},\delta)$ with $\mathrm{y} \neq \mathrm{x}$, we have

$$\begin{aligned} f(\mathrm{y}) - f(\mathrm{x}) = f(y_1,\ldots,y_n) &- f(x_1,y_2,\ldots,y_n) + f(x_1,y_2,\ldots,y_n) \\ &- f(x_1,x_2,y_3,\ldots,y_n) + f(x_1,x_2,y_3,\ldots,y_n) \\ &+ \cdots + f(x_1,\ldots,x_{n-1},y_n) - f(x_1,\ldots,x_n). \end{aligned}$$

Passing to components, we have for $i = 1, \ldots, m$ and $j = 1, \ldots, n$ the differences

$$f_i(x_1, \ldots, x_{j-1}, y_j, y_{j+1}, \ldots, y_n) - f_i(x_1, \ldots, x_{j-1}, x_j, y_{j+1}, \ldots, y_n)$$

whose sum over $j = 1, \ldots, n$ adds to $f_i(\mathrm{u}) - f_i(\mathrm{x})$ for each $i = 1, \ldots, m$.

Define functions g_{ij} defined by

$$z \to f_i(x_1, \ldots, x_{j-1}, z, y_{j+1}, \ldots, y_n).$$

For $j = 1$, each function $g_{i1}(z) = f_i(z, y_2, \ldots, y_n)$, $i = 1, \ldots, m$, is differentiable on the open interval with endpoints x_1 and y_1, and continuous on the closed interval with endpoints x_1 and y_1.

By the Mean Value Theorem, there exists ξ_{i1} in the closed interval with endpoints x_1 and y_1 such that

$$g_{i1}(y_1) - g_{i1}(x_1) = D_1 f_i(\xi_{i1}, y_2, \ldots, y_n)(y_1 - x_1).$$

Continuing to apply the Mean Value Theorem to g_{ij} for $j > 1$ there exist ξ_{ij} in the closed interval with endpoints x_j and y_j such that

$$\begin{aligned}
&f_i(x_1, \ldots, x_{j-1}, y_j, y_{j+1}, \ldots, y_n) - f_i(x_1, \ldots, x_{j-1}, x_j, y_{j+1}, \ldots, y_n) \\
&\quad = g_{ij}(y_j) - g_{ij}(x_j) \\
&\quad = D_j f_i(\xi_{ij})(y_j - x_j).
\end{aligned}$$

Thus the sum of these differences for each $i = 1, \ldots, m$ equals

$$\begin{aligned}
f_i(\mathrm{y}) - f_i(\mathrm{x}) = {} & D_1 f_i(\xi_{i1}, y_2, \ldots, y_n)(y_1 - x_1) + D_2 f_i(x_1, \xi_{2i}, y_3, \ldots, y_n)(y_2 - x_2) \\
& + \cdots + D_n f_i(x_1, \ldots, x_{n-1}, \xi_{ni})(y_n - x_n).
\end{aligned}$$

Now the i^{th} entry of $J(\mathrm{x})(\mathrm{y} - \mathrm{x})$ is

$$[J(\mathrm{x})(\mathrm{y} - \mathrm{x})]_i = \sum_{j=1}^{n} D_j f_i(x_1, \ldots, x_n)(y_j - x_j).$$

Thus

$$\begin{aligned}
&|f_i(\mathrm{y}) - f_i(\mathrm{x}) - [J(\mathrm{x})(\mathrm{y} - \mathrm{x})]_i| \\
&\quad = |(D_1 f_i(\xi_{i1}, y_2, \ldots, y_n) - D_1 f_i(x_1, \ldots, x_n))(y_1 - x_1) \\
&\qquad + (D_2 f_i(x_1, \xi_{i2}, y_3, \ldots, y_n) - D_2 f_i(x_1, \ldots, x_n))(y_2 - x_2) \\
&\qquad + \cdots \\
&\qquad + (D_n f_i(x_1, \ldots, x_{n-1}, \xi_{in}) - D_n f_i(x_1, \ldots, x_n))(y_n - x_n)| \\
&\quad \leq |D_1 f_i(\xi_{i1}, y_2, \ldots, y_n) - D_1 f_i(x_1, \ldots, x_n)|\, |y_1 - x_1| \\
&\qquad + |D_2 f_i(x_1, \xi_{i2}, y_3, \ldots, y_n) - D_2 f_i(x_1, \ldots, x_n)|\, |y_2 - x_2| \\
&\qquad + \cdots \\
&\qquad + |D_n f_i(x_1, \ldots, x_{n-1}, \xi_{in}) - D_n f_i(x_1, \ldots, x_n)|\, |y_n - x_n|
\end{aligned}$$

Since $|y_j - x_j| \leq \|\mathrm{y} - \mathrm{x}\|_\infty$, we can replace each $|y_j - x_j|$ with $\|\mathrm{y} - \mathrm{x}\|_\infty$.

By the assumed continuity of the partial derivatives, we can for each $\epsilon > 0$ choose δ small enough so that

$$\begin{aligned}
|D_1 f_i(\xi_{i1}, y_2, \ldots, y_n) - D_1 f_i(x_1, \ldots, x_n)| &< \frac{\epsilon}{n} \\
|D_2 f_i(x_1, \xi_{i2}, y_3, \ldots, y_n) - D_2 f_i(x_1, \ldots, x_n)| &< \frac{\epsilon}{n} \\
&\vdots \\
|D_n f_i(x_1, \ldots, x_{n-1}, \xi_{in}) - D_n f_i(x_1, \ldots, x_n)| &\leq \frac{\epsilon}{n}.
\end{aligned}$$

Thus

$$|f_i(\mathrm{y}) - f_i(\mathrm{x}) - [J(\mathrm{x})(\mathrm{y} - \mathrm{x})]_i| < \epsilon \|\mathrm{y} - \mathrm{x}\|_\infty.$$

This implies that for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\|f(\mathrm{y}) - f(\mathrm{x}) - J(\mathrm{x}(\mathrm{y} - \mathrm{x})\|_\infty < \epsilon \|\mathrm{y} - \mathrm{x}\|_\infty$$

holds whenever $\mathrm{y} \in B(\mathrm{x}, \delta)$.

Therefore f is differentiable at x and $Df(\mathrm{x}) = J(\mathrm{x})$ by uniqueness. □

Fréchet differentiability implies Gâteaux differentiability, and permits the computation of directional derivatives with relative ease (unlike the earlier computational tour du force).

Theorem 6.2.15. For U open in $\mathbb{R}^n$, if $f : U \to \mathbb{R}^m$ differentiable at $\mathrm{x} \in U$, then the directional or Gâteaux derivative along any $v \in \mathbb{R}^n$ exists and satisfies

$$D_\mathrm{v} f(\mathrm{x}) = Df(\mathrm{x})\mathrm{v}.$$

Consequentially, the directional derivative of f at x is linear in v when f is differentiable at x.

Proof. For $\mathrm{v} = 0$, there is nothing to show, because the directional derivative exists and is 0.

For $\mathrm{v} \neq 0$, consider the function

$$\alpha(t) = \left\| \frac{f(\mathrm{x} + t\mathrm{v}) - f(\mathrm{x})}{t} - Df(\mathrm{x})\mathrm{v} \right\|.$$

We will show that $\lim_{t \to 0} \alpha(t) = 0$.

For $\epsilon > 0$ there exists by the differentiability of f at x a $\delta > 0$ such that for all $0 < \|h\| < \delta$ there holds

$$\|f(\mathrm{x} + h) - f(\mathrm{x}) - Df(\mathrm{x})h\| < \frac{\epsilon \|h\|}{\|\mathrm{v}\|}.$$

This implies for nonzero $h = t\mathrm{v}$ satisfying $\|t\mathrm{v}\| < \delta$, i.e., $|t| < \delta / \|\mathrm{v}\|$, that

$$\|f(\mathrm{x} + t\mathrm{v}) - f(\mathrm{x}) - tDf(\mathrm{x})\mathrm{v}\| < \epsilon |t|.$$

Dividing through by $|t|$ and bringing $|t|$ inside the norm on the left gives

$$\alpha(t) = \left\| \frac{f(\mathrm{x} + t\mathrm{v}) - f(\mathrm{x})}{t} - Df(\mathrm{x})\mathrm{v} \right\| < \epsilon.$$

Therefore $\lim_{t \to 0} \alpha(t) = 0$. □