Math 346 Lecture #3
6.3 The General Fréchet Derivative

We now extend the notion of the Fréchet derivative to the Banach space setting.

Throughout let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces, and U an open set in X.

6.3.1 The Fréchet Derivative

Definition 6.3.1. A function $f : U \to Y$ is Fréchet differentiable at $\mathrm{x} \in U$ if there exists $A \in \mathscr{B}(X, Y)$ such that

$$\lim_{h \to 0} \frac{\|f(\mathrm{x} + h) - f(\mathrm{x}) - A(h)\|_Y}{\|h\|_X} = 0,$$

and we write $Df(\mathrm{x}) = A$. We say f is Fréchet differentiable on U if f is Fréchet differentiable at each $\mathrm{x} \in U$. We often refer to Fréchet differentiable simply as differentiable.

Note. When f is Fréchet differentiable at $\mathrm{x} \in U$, the derivative $Df(\mathrm{x})$ is unique. This follows from Proposition 6.2.10 (uniqueness of the derivative for finite dimensional X and Y) whose proof carries over without change to the general Banach space setting (see Remark 6.3.9).

Remark 6.3.2. The existence of the Fréchet derivative does not change when the norm on X is replace by a topologically equivalent one and/or the norm on Y is replaced by a topologically equivalent one.

Example 6.3.3. Any $L \in \mathscr{B}(X, Y)$ is Fréchet differentiable with $DL(\mathrm{x})(\mathrm{v}) = L(\mathrm{v})$ for all $\mathrm{x} \in X$ and all $\mathrm{v} \in X$.

The proof of this is exactly the same as that given in Example 6.2.5 for finite-dimensional Banach spaces.

Example (in lieu of 6.3.4). For $X = C([0, 1], \mathbb{R})$ with $\|\cdot\|_X = \|\cdot\|_\infty$, $Y = \mathbb{R}$ with $\|\cdot\|_Y = |\cdot|$, and a fixed polynomial $p(t) \in \mathbb{R}[t]$, the function $L : X \to Y$ defined by

$$L(f) = \int_0^1 p(t) f(t)\, dt$$

is a bounded linear functional, bounded because

$$|L(f)| \leq \int_0^1 |p(t) f(t)|\, dt \leq \int_0^1 \|p\|_\infty \|f\|_\infty = \|p\|_\infty \|f\|_\infty,$$

so that

$$\|L\|_{X,Y} = \sup_{\|f\|_\infty > 0} \frac{|L(f)|}{\|f\|_\infty} \leq \|p\|_\infty,$$

and so by Example 6.3.3, we have for each $f \in X$ that $DL(f)g = L(g)$ for all $g \in X$.

Note. When X is an infinite-dimensional Banach space, functions from an open subset of X to Y are often defined by Banach-valued integration.

Example (in lieu of 6.3.5). Again with $X = C([0,1],\mathbb{R})$ with $\|\cdot\|_X = \|\cdot\|_\infty$, $Y = \mathbb{R}$ with $\|\cdot\|_Y = |\cdot|$, and a fixed polynomial $p(t) \in \mathbb{R}[t]$, the function $Q : X \to \mathbb{R}$ defined by

$$Q(f) = \int_0^1 p(t)[f(t)]^3 \, dt$$

is not linear (unless $p(t) = 0$) but Fréchet differentiable on X.

Because $DQ(f)g = D_gQ(f)$ if Q is Fréchet differentiable at f, we guess the form of the Fréchet derivative $DQ(f)$ by computing the Gâteaux derivative $D_gQ(f)$.

To this end we have for fixed $f \in X$, fixed $g \in X$, and $r > 0$ that

$$\begin{aligned}
D_gQ(f) &= \lim_{r\to 0} \frac{Q(f+rg) - Q(f)}{r} \\
&= \lim_{r\to 0} \frac{1}{r}\left\{ \int_0^1 p(t)[f(t)+rg(t)]^3 \, dt - \int_0^1 p(t)[f(t)]^3 \, dt \right\} \\
&= \lim_{r\to 0} \frac{1}{r} \int_0^1 p(t)\left[3rf^2(t)g(t) + 3r^2 f(t)g^2(t) + r^3 g^3(t)\right] dt \\
&= \lim_{r\to 0} \int_0^1 p(t)\left[3f^2(t)g(t) + 3rf(t)g^2(t) + r^2 g^3(t)\right] dt \\
&= \int_0^1 3p(t)[f(t)]^2 g(t) \, dt,
\end{aligned}$$

provided we can take the limit inside the integral. (Can you explain why this is justified?)

Thus the guess for the Fréchet derivative of Q at f is the linear transformation $B : X \to \mathbb{R}$ defined by

$$B(g) = \int_0^1 3p(t)[f(t)]^2 g(t) \, dt.$$

Notice that the guess for the Fréchet derivative of an function defined by an integral is another function defined by an integral. This is typical for functions defined by integration.

Now for each fixed f is our guess B for $DQ(f)$ a bounded linear transformation?

If we set $M = \|f\|_\infty$ and $K = \|p\|_\infty$, then for any $g \in X$ we have

$$\begin{aligned}
|B(g)| &= \left| \int_0^1 3p(t)[f(t)]^2 g(t) \, dt \right| \le \int_0^1 |3p(t)[f(t)]^2 g(t)| \, dt \\
&\le \int_0^1 3KM^2 |g(t)| \, dt \le 3KM^2 \int_0^1 \|g\|_\infty \, dt = 3KM^2 \|g\|_\infty,
\end{aligned}$$

so that

$$\|B\|_{X,Y} = \sup\left\{ \frac{|B(g)|}{\|g\|_\infty} : \text{ nonzero } g \in X \right\} \le 3KM^2.$$

Thus our guess B for $DQ(f)$ is a bounded linear transformation.

Now to show that our guess B is the Fréchet derivative of Q at f:

$$\begin{aligned}
&\lim_{h\to 0}\frac{|Q(f+h)-Q(f)-B(h)|}{\|h\|_\infty} \\
&\quad = \lim_{h\to 0}\frac{1}{\|h\|_\infty}\left|\int_0^1 p(t)[f(t)+h(t)]^3\ dt - \int_0^1 p(t)[f(t)]^3\ dt - \int_0^1 3p(t)[f(t)]^2h(t)\ dt\right| \\
&\quad = \lim_{h\to 0}\frac{1}{\|h\|_\infty}\left|\int_0^1 p(t)\big[3f(t)h^2(t)+h^3(t)\big]\ dt\right| \\
&\quad \le \lim_{h\to 0}\frac{1}{\|h\|_\infty}\int_0^1 |h^2(t)|\ \big|p(t)\big([f(t)]^2+h(t)\big)\big|\ dt \\
&\quad \le \lim_{h\to 0}\frac{\|h\|_\infty^2}{\|h\|_\infty}\int_0^1 \big|p(t)\big([f(t)]^2+h(t)\big)\big|\ dt = 0
\end{aligned}$$

because the integral is bounded for small h.

Therefore Q is Fréchet differentiable on X where for each $f \in X$ we have

$$DQ(f)(g) = \int_0^1 p(t)[f(t)]^2 g(t)\ dt,\ g \in X.$$

Definition 6.3.6. For $f : U \to Y$, if $Df : U \to \mathscr{B}(X, Y)$ given by $\mathrm{x} \to Df(\mathrm{x})$ for $\mathrm{x} \in U$, is continuous, then we say that f is continuously differentiable on U.

The set of continuously differentiable functions $f : U \to Y$ is denoted by $C^1(U, Y)$. We will show in the next section that $C^1(U, Y)$ is a vector space.

Note. The proof of Theorem 6.2.14 – for $f : U \to \mathbb{R}^m$, with U open in $\mathbb{R}^n$, if all the partial derivatives $D_i f_j(\mathrm{x})$ exist and are continuous for all $\mathrm{x} \in U$, then f is differentiable at each $\mathrm{x} \in U$ – actually shows that the function $\mathrm{x} \to Df(\mathrm{x})$ for $\mathrm{x} \in U$ is continuous because in the standard coordinates each entry of the Jacobian (a matrix representation of Df) is a continuous function on U. Thus the hypotheses of Theorem 6.2.14 imply that $f \in C^1(U, \mathbb{R}^m)$.

Example. This is from Math 634 – Theory of Ordinary Differential Equations – you are not responsible for knowing or reproducing this example.

But it is included to illustrate that the "same" function or operator may be continuously differentiable on one Banach space, i.e., $(C([0,1],\mathbb{R}), \|\cdot\|_\infty)$, but not differentiable on another Banach space, i.e., $(L^2([0,1],\mathbb{R}), \|\cdot\|_2)$.

For $X = C([0,1],\mathbb{R})$ equipped with the ∞-norm, we show that the operator $F : X \to X$ defined by

$$F(g)(t) = \sin(g(t))$$

is continuously differentiable on X.

For fixed $g \in X$, the linear operator $B \in \mathscr{L}(X)$ defined by

$$B(h)(t) = (\cos g(t))h(t)$$

is bounded because

$$|(\cos g(t))h(t)| \leq |h(t)| \leq \|h\|_\infty.$$

The operator F is differentiable with $(DF(g)h)(t) = (\cos g(t))h(t)$ for all $h \in X$ because

$$\begin{aligned}
&|F(g+h)(t) - F(g)(t) - (\cos g(t))h(t)| \\
&\quad = |\sin(g(t)+h(t)) - \sin g(t) - (\cos g(t))h(t)| \\
&\quad = |\sin g(t)\cos h(t) + \cos g(t)\sin h(t) - \sin g(t) - (\cos g(t))h(t)| \\
&\quad = |(-1+\cos h(t))\sin g(t) + (-h(t)+\sin h(t))\cos g(t)| \\
&\quad \leq |(-1+\cos h(t))\sin g(t)| + |(-h(t)+\sin h(t))\cos g(t)| \\
&\quad = |\sin g(t)|\,|-1+\cos h(t)| + |\cos g(t)|\,|-h(t)+\sin h(t)| \\
&\quad \leq |-1+\cos h(t)| + |-h(t)+\sin h(t)| \\
&\quad = \frac{|h(t_1)|^2}{2} + \frac{|h(t_2)|^3}{6} \\
&\quad \leq \frac{\|h\|_\infty^2}{2} + \frac{\|h\|_\infty^3}{6},
\end{aligned}$$

where we have used the existence of $t_1, t_2 \in [0,1]$ for which

$$|-1+\cos h(t)| = \frac{|h(t_1)|^2}{2}, \quad |-h(t)+\sin h(t)| = \frac{|h(t_2)|^3}{6},$$

that follow from Lagrange's Remainder Theorem applied to cos and sin respectively.

To show that $g \to DF(g)$ is continuous we start with

$$\begin{aligned}
\|DF(g_1) - DF(g_2)\|_\infty &= \sup_{\|h\|_\infty=1} \|DF(g_1)h - DF(g_2)h\|_\infty \\
&= \sup_{\|h\|_\infty=1} \sup_{t\in[0,1]} |(\cos g_1(t))h(t) - (\cos g_2(t))h(t)| \\
&= \sup_{\|h\|_\infty=1} \sup_{t\in[0,1]} |h(t)|\,|\cos g_1(t) - \cos g_2(t)|.
\end{aligned}$$

By the Mean Value Theorem applied to cos whose derivative is bounded in absolute value by 1, we have

$$|\cos g_1(t) - \cos g_2(t)| \leq |g_1(t) - g_2(t)|.$$

Thus

$$\begin{aligned}
\|DF(g_1) - DF(g_2)\|_\infty &\leq \sup_{\|h\|_\infty=1} \sup_{t\in[0,1]} |h(t)|\,|g_1(t) - g_2(t)| \\
&\leq \|g_1 - g_2\|_\infty
\end{aligned}$$

which says that DF is Lipschitz continuous on X, and hence DF continuous on X.

If we take the "same" operator F on the domain $X = L^2([0,1],\mathbb{R})$ with its 2-norm, then

$F(g)(t) = \sin g(t)$ is Lipschitz continuous because

$$\begin{aligned} \|F(g_1) - F(g_2)\|_2^2 &= \int_0^1 |\sin g_1(t) - \sin g_2(t)|^2 \, dt \\ &\leq \int_0^1 |g_1(t) - g_2(t)|^2 \, dt \\ &= \|g_1 - g_2\|_2^2 \end{aligned}$$

(where we have used the Mean Value Theorem to get $|\sin g_1(t) - \sin g_2(t)| \leq |g_1(t) - g_2(t)|$), but this F is not differentiable at $g = 0$ (to show this requires a long tedious argument that depends on the properties of Lebesgue integration).

We will now show that differentiability at a point implies continuity at that point. To this end we need to articular a notation that differentiability implies. This is where we will see the need for the Fréchet derivative to be a *bounded* linear transformation.

Definition. A function $f : U \to Y$ is locally Lipschitz at a point $\mathrm{x}_0 \in U$ if there exists $\delta > 0$ and $L > 0$ such that $B(\mathrm{x}_0, \delta) \subset U$ and for all $\mathrm{x} \in B(\mathrm{x}_0, \delta)$ there holds

$$\|f(\mathrm{x}) - f(\mathrm{x}_0)\|_Y \leq L\|\mathrm{x} - \mathrm{x}_0\|_X.$$

Remark. Locally Lipschitz at a point is not the same as locally Lipschitz on U where the latter means for every $\mathrm{x}_0 \in U$ there exists $\delta > 0$ and $L > 0$ such that $B(\mathrm{x}_0, \delta) \subset U$ and for all $\mathrm{x}, \mathrm{z} \in B(\mathrm{x}_0, \delta)$ there holds

$$\|f(\mathrm{x}) - f(\mathrm{z})\|_Y \leq L\|\mathrm{x} - \mathrm{z}\|_X.$$

Neither of these are the same as Lipschitz on U which means there exists $L > 0$ such that for all $\mathrm{x}, \mathrm{z} \in U$ there holds

$$\|f(\mathrm{x}) - f(\mathrm{z})\|_Y \leq L\|\mathrm{x} - \mathrm{z}\|_X.$$

Lipschitz on U implies locally Lipschitz on U with the same constant L, and locally Lipschitz on U implies locally Lipschitz at every point of U with the same constant L.

Proposition 6.3.7. If $f : U \to Y$ is Fréchet differentiable at $\mathrm{x}_0 \in U$ then f is locally Lipschitz at x_0. (This is not the same as in the book; more about this after the proof.)

Proof. By the assumed differentiability of f at x_0, we have for $\epsilon = 1$ the existence of $\delta > 0$ such that for all $\mathrm{x} \in U$ satisfying $\|\mathrm{x} - \mathrm{x}_0\|_X < \delta$ there holds

$$\frac{\|f(\mathrm{x}) - f(\mathrm{x}) - Df(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0)\|_Y}{\|\mathrm{x} - \mathrm{x}_0\|_X} < 1.$$

This can be rewritten as

$$\|f(\mathrm{x}) - f(\mathrm{x}) - Df(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0)\|_Y < \|\mathrm{x} - \mathrm{x}_0\|_X.$$

Now by the triangle inequality we have

$$\begin{aligned} \|f(\mathrm{x}) - f(\mathrm{x}_0)\|_Y &= \|f(\mathrm{x}) - f(\mathrm{x}_0) - Df(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0) + Df(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0)\|_Y \\ &\leq \|f(\mathrm{x}) - f(\mathrm{x}_0) - Df(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0)\|_Y + \|Df(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0)\|_Y \\ &\leq \|\mathrm{x} - \mathrm{x}_0\|_X + \|Df(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0)\|_Y. \end{aligned}$$

Since $Df(\mathrm{x}_0)$ is a bounded linear transformation, we have (see Remark 3.5.12) that

$$\|Df(\mathrm{x}_0)(\mathrm{x}-\mathrm{x}_0)\|_Y \leq \|Df(\mathrm{x}_0)\|_{X,Y}\|\mathrm{x}-\mathrm{x}_0\|_X.$$

Thus we obtain

$$\|f(\mathrm{x})-f(\mathrm{x}_0)\|_Y \leq \big(1+\|Df(\mathrm{x}_0)\|_{X,Y}\big)\|\mathrm{x}-\mathrm{x}_0\|_X.$$

If $B(\mathrm{x}_0,\delta)$ is not a subset of U, then by the openness of U there is a small enough value of δ for which it is.

Taking $L = 1+\|Df(\mathrm{x}_0)\|_{X,Y}$ now gives that f is locally Lipschitz at x_0. □

Note. The statement of Proposition 6.3.7 given in the book – if f is differentiable on U, then f is locally Lipschitz at every point of U – follows from what is proved above.

Corollary 6.3.8. If $f: U \to Y$ is differentiable at $\mathrm{x}_0 \in U$, then f is continuous at x_0.

This is not the same as in the book, but the statement of Corollary in 6.3.8 in the book follows from this. It is HW (Exercise 6.14) to prove the version of Corollary 6.3.8 in the book. (Remember that differentiability and continuity are point properties.)

Note. It is the boundedness of the Fréchet derivative that makes differentiability of a function at a point imply continuity of the function at that point.

Proposition 6.3.10. If $f: U \to Y$ is differentiable at $\mathrm{x} \in U$ and f is locally Lipschitz at x with constant L, then $\|Df(\mathrm{x})\|_{X,Y} \leq L$.

Proof. By the assumed differentiability of f at x, for each $\epsilon > 0$ there exists $\delta > 0$ such that for all $0 < \|h\| < \delta$ there holds

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

By the assumed local Lipschitz at x with constant $L > 0$ there is $\nu > 0$ such that for all $\mathrm{x}+h \in B(\mathrm{x},\nu) \subset U$ there holds

$$\|f(\mathrm{x}+h)-f(\mathrm{x})\|_Y \leq L\|(\mathrm{x}+h)+\mathrm{x}\|_X = L\|h\|_X.$$

We can make δ smaller if needed so that $\delta \leq \nu$.

For a unit vector $\mathrm{u} \in X$, the vector $h = (\delta/2)\mathrm{u}$ satisfies $0 < \|h\|_X = \delta/2 < \delta$. Thus

$$\|Df(\mathrm{x})\mathrm{u}\|_Y = \frac{\|Df(\mathrm{x})\mathrm{u}\|_Y}{\|\mathrm{u}\|_X} = \frac{(\delta/2)\|Df(\mathrm{x})\mathrm{u}\|_Y}{(\delta/2)\|\mathrm{u}\|_X} = \frac{\|Df(\mathrm{x})h\|_Y}{\|h\|_X}.$$

Since

$$\|Df(\mathrm{x})h\|_Y \leq \|f(\mathrm{x}+h)-f(\mathrm{x})-Df(\mathrm{x})h\|_Y + \|f(\mathrm{x}+h)-f(\mathrm{x})\|_Y,$$

and since $0 < \|h\|_X < \delta$, we obtain

$$\begin{aligned}\|Df(\mathrm{x})u\|_Y &\leq \frac{\|f(\mathrm{x}+h)-f(\mathrm{x})-Df(\mathrm{x})h\|_Y}{\|h\|_X} + \frac{\|f(\mathrm{x}+h)-f(\mathrm{x})\|_Y}{\|h\|_X}\\ &\leq \frac{\|f(\mathrm{x}+h)-f(\mathrm{x})-Df(\mathrm{x})h\|_Y}{\|h\|_X} + \frac{L\|h\|_X}{\|h\|_X}\\ &< \epsilon + L.\end{aligned}$$

This implies that

$$\|Df(\mathrm{x})\|_{X,Y} = \sup_{\|u\|_X=1} \|Df(\mathrm{x})u\|_Y \leq \epsilon + L.$$

Since $\epsilon > 0$ is arbitrary we obtain $\|Df(\mathrm{x})\|_{X,Y} \leq L$. □

Note. Another version of Proposition 6.3.10 is the following: If f is differentiable on U and f is Lipschitz on U with constant L, then $\|Df(\mathrm{x})\|_{X,Y} \leq L$ for all $\mathrm{x} \in U$. This follows from Proposition 6.3.10 and Lipschitz on U implying locally Lipschitz at each point $\mathrm{x} \in U$ with the same constant L. We will see that the Mean Value Theorem (Section 6.5) is a converse of this: if $\|Df(\mathrm{x})\|_{X,Y} \leq L$ for all $\mathrm{x} \in U$, then $\|f(\mathrm{y}) - f(\mathrm{z})\|_X \leq L\|y - z\|_X$ when the line segment $(1-t)\mathrm{y} + t\mathrm{z}$, $t \in [0,1]$, lies completely in U.

6.3.2 Fréchet Derivatives on Cartesian Products

Proposition 6.3.11. For Banach spaces $Y_1, \ldots, Y_m$, a function $f : U \to Y_1 \times \cdots \times Y_m$ defined by $f(\mathrm{x}) = (f_1(\mathrm{x}), \ldots, f_m(\mathrm{x}))$ is differentiable at $\mathrm{x} \in U$ if each f_i is differentiable at x. Moreover, when f is differentiable at x, then for each $h \in X$ we have

$$Df(\mathrm{x})h = (Df_1(\mathrm{x})h, \ldots, Df_m(\mathrm{x})h).$$

The proof of this is similar to that of Proposition 6.1.5 (for curves) and is HW (Exercise 6.15).

Definition 6.3.12. Let $X_1, \ldots, X_n$ be Banach spaces and let $U_1, \ldots, U_n$ be nonempty open subsets of the corresponding Banach spaces. For a function $f : U_1 \times \cdots \times U_n \to Y$, the i^{th} partial derivative of f at $(\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{n}) \in U_1 \times \cdots \times U_n$ is the derivative of the function

$$g_i(\mathrm{z}) = f(\mathrm{x}_1, \ldots, \mathrm{x}_{i-1}, \mathrm{z}, \mathrm{x}_{i+1}, \ldots, \mathrm{x}_\mathrm{n})$$

and is denoted by $D_i f(\mathrm{x}_1, \cdots, \mathrm{x}_n)$, if it exists.

Example 6.3.13. Our familiar notion of partial derivatives for a function $f : U \to \mathbb{R}^m$ for an nonempty open set U in $\mathbb{R}^n$ is a special case of Definition 6.3.12: each X_i is $\mathbb{R}$ and Y is $\mathbb{R}^m$ which matches the definition of partial derivative given in Definition 6.1.13.

Theorem 6.3.14. For Banach spaces $X_1, \ldots, X_n$ and nonempty open subsets $U_1, \ldots, U_n$ in the corresponding X_i, let $f : U_1 \times \cdots \times U_n \to Y$. If f is differentiable at $\mathrm{x} = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$, then its partial derivatives $D_i f(\mathrm{x})$ all exist, and for each $h = (h_1, \ldots, h_n) \in X_1 \times \cdots \times X_n$, we have

$$Df(\mathrm{x})h = \sum_{i=1}^{n} D_i f(\mathrm{x}) h_i.$$

Conversely, if all the partial derivatives $D_i f$ exist and are continuous on $U_1 \times \cdots \times U_n$, then f is continuously differentiable on $U_1 \times \cdots \times U_n$.

The proof of this is straightforward and easy generalization of Theorems 6.2.11 and 6.2.14.