Math 346 Lecture #4
6.4 Properties of the Derivative

Throughout let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces over the same field $\mathbb{F}$, and U an open set in X.

We have already been using the next result, and will continue to use it tacitly.

Lemma 6.4.1. For a function $f : U \to Y$, a point $\mathrm{x} \in Y$, and $L \in \mathscr{B}(X, Y)$, the following are equivalent.

(i) The function f is differentiable at x with derivative L.

(ii) For every $\epsilon > 0$ there exists $\delta > 0$ with $B(\mathrm{x}, \delta) \subset U$ such that for all $h \in B(\mathrm{x}, \delta)$ there holds

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

The only minor wrinkle in the proof of this is the $\leq$ in part (ii). But this follows by replacing ϵ with $\epsilon/2$ when applying the definition of differentiable.

6.4.1. Linearity

The reader is reminded of Theorem 3.5.11 that for Banach spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, the collection $\mathscr{B}(X, Y)$ is a normed linear space, so that linear combinations of finitely many elements of $\mathscr{B}(X, Y)$ belong to $\mathscr{B}(X, Y)$.

Recall that the image $Df(\mathrm{x})v$ of the derivative is not linear in x, but it is linear in v. We now show that this image is linear in f.

Theorem 6.4.2 (Linearity, pointwise). For $f, g : U \to Y$, if f and g are differentiable at $\mathrm{x} \in U$, then for any $a, b \in \mathbb{F}$ the linear combination $af + bg$ is differenitable at x and

$$D(af + bg)(\mathrm{x}) = aDf(\mathrm{x}) + bDg(\mathrm{x}).$$

Proof. By the assumption of differentiability of f and g at x, for any $\epsilon > 0$ there is $\delta > 0$ (the minimum of the δ's for f and g and the radius of a ball centered at x and contained in U) such that for all $h \in B(\mathrm{x}, \delta)$ there holds

$$\|f(\mathrm{x} + h) - f(\mathrm{x}) - Df(\mathrm{x})h\|_Y \leq \frac{\epsilon\|h\|_X}{2(|a| + 1)}$$

and

$$\|g(\mathrm{x} + h) - g(\mathrm{x}) - Dg(\mathrm{x})h\|_Y \leq \frac{\epsilon\|h\|_X}{2(|b| + 1)}.$$

Thus

$$\begin{aligned}
&\|af(\mathrm{x} + h) + bg(\mathrm{x} + h) - af(\mathrm{x}) - bg(\mathrm{x}) - aDf(\mathrm{x})h - bDg(\mathrm{x})h\|_y \\
&\quad \leq |a|\,\|f(\mathrm{x} + h) - f(\mathrm{x}) - Df(\mathrm{x})h\|_Y + |b|\,\|g(\mathrm{x} + h) - g(\mathrm{x}) - Dg(\mathrm{x})h\|_Y \\
&\quad \leq \frac{\epsilon|a|\,\|h\|_X}{2(|a| + 1)} + \frac{\epsilon|b|\,\|h\|_X}{2(|b| + 1)} \\
&\leq \epsilon\|h\|_X.
\end{aligned}$$

Since $Df(\mathrm{x})$ and $Dg(\mathrm{x})$ both belong to $\mathscr{B}(X,Y)$, then $aDf(\mathrm{x}) + bDg(\mathrm{x}) \in \mathscr{B}(X,Y)$.

Thus $af + bg$ is differentiable at x with derivative $aDf(\mathrm{x}) + bDg(\mathrm{x})$. $\square$

Note. The version of Theorem 6.4.2 in the book – if f and g are differentiable on U then $af + bg$ is differentiable on U with derivative $aDf + bDg$ – follows from the pointwise version proved above.

Remark 6.4.3. An immediate consequence of Theorem 6.4.2 is that the set $C^1(U,Y)$ is a vector space. Is there a "natural" norm on this vector space? Yes, and we will learn about it in Section 6.5 (and give the definition of it – something missing in the book).

6.4.2 The Product Rule

We can not multiply $f, g : U \to Y$, i.e., $f(\mathrm{x})g(\mathrm{x})$ might not make sense, for an arbitrary Banach space Y. However if $Y = \mathbb{F}$, then we can.

Theorem 6.4.4 (Product Rule – pointwise). If $f, g : U \to \mathbb{F}$, are differentiable at $\mathrm{x} \in U$, then the product $h = fg$ is differentiable at x and the derivative of h at x satisfies

$$Dh(\mathrm{x}) = g(\mathrm{x})Df(\mathrm{x}) + f(\mathrm{x})Dg(\mathrm{x}),$$

i.e., for all $\xi \in X$ we have

$$Dh(\mathrm{x})\xi = g(\mathrm{x})(Df(\mathrm{x})\xi) + f(\mathrm{x})(Dg(\mathrm{x})\xi) \in \mathbb{F},$$

because $Df(\mathrm{x}) \in \mathscr{B}(X,\mathbb{F})$ and $Dg(\mathrm{x}) \in \mathscr{B}(X,\mathbb{F})$ so that $Df(\mathrm{x})\xi$ and $Dg(\mathrm{x})\xi$ both belong to $\mathbb{F}$, whence as $g(\mathrm{x})$ and $f(\mathrm{x})$ both belong to $\mathbb{F}$ that $g(\mathrm{x})(Df(\mathrm{x})\xi)$ and $f(\mathrm{x})(Dg(\mathrm{x})\xi)$ both belong to $\mathbb{F}$, so that finally $Dh(\mathrm{x})\xi$ belongs to $\mathbb{F}$.

Proof. Assuming f and g are differentiable at x for each $\epsilon > 0$ there is $\delta_{\mathrm{x}} > 0$ (the minimum of a finite number of positive δ's) with $B(\mathrm{x},\delta) \subset U$, and a constant $L > 0$ (by Proposition 6.3.7) such that for all $0 < \|h\| < \delta_x$ there holds

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

and

$$|f(\mathrm{x}+h) - f(\mathrm{x}) - Df(\mathrm{x})h| \leq \frac{\epsilon\|h\|_X}{3(|g(\mathrm{x})|+1)}$$

and

$$|g(\mathrm{x}+h) - g(\mathrm{x}) - Dg(\mathrm{x})h| \leq \frac{\epsilon\|h\|_X}{3(|f(\mathrm{x})|+L)}.$$

We are going to do an $\epsilon/3$ argument, and this will require the presence of three constraints on the choice of δ.

For $\epsilon > 0$ choose

$$\delta = \min\left\{1, \delta_x, \frac{\epsilon}{3L(\|Dg(\mathrm{x})\|_{X,\mathbb{F}}+1)}\right\}.$$

Each constraint on δ will be used for one of the $\epsilon/3$ parts.

When $0 < \|h\|_X < \delta$ that

$$\begin{aligned}
&|f(\mathrm{x}+h)g(\mathrm{x}+h) - f(\mathrm{x})g(\mathrm{x}) - g(\mathrm{x})Df(\mathrm{x})h - f(\mathrm{x})Dg(\mathrm{x})h| \\
&\quad = |f(\mathrm{x}+h)g(\mathrm{x}+h) - f(\mathrm{x}+h)g(\mathrm{x}) + f(\mathrm{x}+h)g(\mathrm{x}) - f(\mathrm{x})g(\mathrm{x}) \\
&\qquad + f(\mathrm{x}+h)Dg(\mathrm{x})h - f(\mathrm{x}+h)Dg(\mathrm{x})h - g(\mathrm{x})Df(\mathrm{x})h - f(\mathrm{x})Dg(\mathrm{x})h| \\
&\quad \le |f(\mathrm{x}+h)|\,|g(\mathrm{x}+h) - g(\mathrm{x}) - Dg(\mathrm{x})h| \\
&\qquad + |g(\mathrm{x})|\,|f(\mathrm{x}+h) - f(\mathrm{x}) - Df(\mathrm{x})h| \\
&\qquad + |f(\mathrm{x}+h) - f(\mathrm{x})|\,\|Dg(\mathrm{x})\|_{X,\mathbb{F}}\|h\|_X \\
&\quad \le (|f(\mathrm{x})| + L)\frac{\epsilon\|h\|_X}{3(|f(\mathrm{x})|+L)} + |g(\mathrm{x})|\frac{\epsilon\|h\|_X}{3(|g(\mathrm{x})|+1)} + \delta L\|Dg(\mathrm{x})\|_{X,\mathbb{F}}\|h\|_X \\
&\quad \le \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon
\end{aligned}$$

where we have made use of the implication

$$|f(\mathrm{x}+h) - f(\mathrm{x})| \le L\|h\|_X \;\Rightarrow\; |f(\mathrm{x}+h)| \le |f(\mathrm{x})| + L\|h\|_X$$

and the implication $\delta < 1 \Rightarrow \|h\|_X < 1$. □

We now look at other product-like differentiation rules. One of these involve matrix functions. We say that a matrix function is differentiable at a point in its domain if every entry in the matrix function is differentiable at that point. The derivative of a differentiable matrix function is the entry-wise derivative of the matrix function.

Proposition 6.4.6 (pointwise version). (i) For an open set U of $\mathbb{R}^n$, let $\mathrm{u}, \mathrm{v} : U \to \mathbb{R}^m$, and define $f : U \to \mathbb{R}$ by

$$f(\mathrm{x}) = \mathrm{u}(\mathrm{x})^{\mathrm{T}}\mathrm{v}(\mathrm{x}).$$

If u and v are differentiable at a point $\mathrm{x} \in U$, then f is differentiable at x and the derivative of f at x satisfies

$$Df(\mathrm{x}) = \mathrm{u}(\mathrm{x})^{\mathrm{T}}D\mathrm{v}(\mathrm{x}) + \mathrm{v}(\mathrm{x})^{\mathrm{T}}D\mathrm{u}(\mathrm{x}),$$

i.e., for all $h \in \mathbb{R}^n$ we have

$$Df(\mathrm{x})\mathrm{h} = \mathrm{u}(\mathrm{x})^{\mathrm{T}}(D\mathrm{v}(\mathrm{x})h) + \mathrm{v}(\mathrm{x})^{\mathrm{T}}(D\mathrm{u}(\mathrm{x})h) \in \mathbb{R},$$

because $D\mathrm{v}(\mathrm{x})$ and $D\mathrm{u}(\mathrm{x})$ both belong to $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ so that $D\mathrm{v}(\mathrm{x})h$ and $D\mathrm{u}(\mathrm{x})h$ both belong to $\mathbb{R}^m$, whence as $\mathrm{u}(\mathrm{x})$ and $\mathrm{v}(\mathrm{x})$ both belong to $\mathbb{R}^m$ that $\mathrm{u}(\mathrm{x})^{\mathrm{T}}(Dv(\mathrm{x})h)$ and $\mathrm{v}(\mathrm{x})^{\mathrm{T}}(Du(\mathrm{x})h)$ both belong to $\mathbb{R}$, so finally that $Df(\mathrm{x})h$ belongs to $\mathbb{R}$.

(ii) For a matrix $A \in M_n(\mathbb{R})$ the function $g : \mathbb{R}^n \to \mathbb{R}$ defined by $g(\mathrm{x}) = \mathrm{x}^{\mathrm{T}}A\mathrm{x}$ is differentiable at every $\mathrm{x} \in \mathbb{R}^n$ with

$$Dg(\mathrm{x}) = \mathrm{x}^{\mathrm{T}}(A + A^{\mathrm{T}}),$$

i.e., for all $h \in \mathbb{R}^n$ we have

$$Dg(\mathrm{x})h = \mathrm{x}^{\mathrm{T}}(A + A^{\mathrm{T}})h \in \mathbb{R},$$

because $A+A^{\mathrm{T}} \in M_n(\mathbb{R})$ and $h \in \mathbb{R}$ so that $(A+A^{\mathrm{T}})h \in \mathbb{R}^n$, so that $\mathrm{x}^{\mathrm{T}}(A+A^{\mathrm{T}})h \in \mathbb{R}$.

(iii) For an open subset U of $\mathbb{R}^n$ let $\mathrm{w} : U \to \mathbb{R}^m$ and $B : U \to M_{k\times m}(\mathbb{R})$ and define $H : U \to \mathbb{R}^k$ by $H(\mathrm{x}) = B(\mathrm{x})\mathrm{w}(\mathrm{x})$. If w and B are differentiable at $\mathrm{x} \in U$, then the function H is differentiable at x with

$$DH(\mathrm{x}) = B(\mathrm{x})D\mathrm{w}(\mathrm{x}) + \begin{bmatrix} \mathrm{w}(\mathrm{x})^{\mathrm{T}} D\mathrm{b}_1^{\mathrm{T}}(\mathrm{x}) \\ \mathrm{w}(\mathrm{x})^{\mathrm{T}} D\mathrm{b}_2^{\mathrm{T}}(\mathrm{x}) \\ \vdots \\ \mathrm{w}(\mathrm{x})^{\mathrm{T}} D\mathrm{b}_k^{\mathrm{T}}(\mathrm{x}) \end{bmatrix},$$

where b_i is the k^{th} row of B, i.e., for each $\xi \in \mathbb{R}^n$ we have

$$DH(\mathrm{x})\xi = B(\mathrm{x})(D\mathrm{w}(\mathrm{x})\xi) + \begin{bmatrix} \mathrm{w}(\mathrm{x})^{\mathrm{T}} (D\mathrm{b}_1^{\mathrm{T}}(\mathrm{x})\xi) \\ \mathrm{w}(\mathrm{x})^{\mathrm{T}} (D\mathrm{b}_2^{\mathrm{T}}(\mathrm{x})\xi) \\ \vdots \\ \mathrm{w}(\mathrm{x})^{\mathrm{T}} (D\mathrm{b}_k^{\mathrm{T}}(\mathrm{x})\xi) \end{bmatrix} \in \mathbb{R}^k,$$

because $D\mathrm{w}(\mathrm{x}) \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$, whence $D\mathrm{w}(\mathrm{x})\xi \in \mathbb{R}^m$ so that $B(\mathrm{x})(D\mathrm{w}(\mathrm{x})\xi) \in \mathbb{R}^k$, and because $\mathrm{b}_i^{\mathrm{T}} : U \to \mathbb{R}^m$ so that $D\mathrm{b}_i^{\mathrm{T}}(\mathrm{x}) \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$, whence $D\mathrm{b}_i^{\mathrm{T}}(\mathrm{x})\xi \in \mathbb{R}^m$, so that $\mathrm{w}(\mathrm{x})^{\mathrm{T}}(D\mathrm{b}_i^{\mathrm{T}}(\mathrm{x})\xi) \in \mathbb{R}$.

The proof of Proposition 6.4.6 is HW (Exercise 6.16). Hint: for part (i) write u(x) and v(x) in terms of standard coordinates and apply Theorem 6.4.4; for part (ii) put the guess for the derivative in the definition and see what happens; for part (iii) write w(x) and $B(\mathrm{x})$ in standard coordinates for $n = 2$, $m = 2$, and $k = 2$ and see what happens, keeping in mind that Fréchet derivatives are linear transformations.

6.4.3 The Chain Rule

Recall from Theorem 3.5.14 that for normed linear spaces $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$, and $(Z, \|\cdot\|_Z)$, if $T \in \mathscr{B}(X,Y)$ and $S \in \mathscr{B}(Y,Z)$, then the composition $ST \in \mathscr{B}(X,Z)$.

Theorem 6.4.7 (The Chain Rule, pointwise version). Suppose $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$, and $(Z, \|\cdot\|_Z)$ are Banach spaces, that U is open in X and V is open in Y, and $f : U \to Y$ and $g : V \to Z$ with $f(U) \subset V$. If f is Fréchet differentiable at $\mathrm{x} \in U$ and g is Fréchet differentiable at $\mathrm{y} = f(\mathrm{x}) \in V$, then $h = g \circ f : U \to Z$ is Fréchet differentiable at x with

$$Dh(\mathrm{x}) = Dg(f(\mathrm{x}))Df(\mathrm{x}),$$

i.e., for all $\xi \in X$ we have

$$Dh(\mathrm{x})\xi = Dg(f(\mathrm{x}))\big(Df(\mathrm{x})\xi\big) \in Z$$

because $Df(\mathrm{x}) \in \mathscr{B}(X,Y)$ so that $Df(\mathrm{x})\xi \in Y$, and because $Dg(f(\mathrm{x})) \in \mathscr{B}(Y,Z)$ so that $Dg(f(\mathrm{x}))(Df(\mathrm{x})\xi) \in Z$.

Proof. Choose $\epsilon > 0$.

By the assumed differentiability of f at x and the assumed differentiability of g at $\mathrm{y} = f(\mathrm{x})$, there is $\delta_1 > 0$ such that $B(\mathrm{x}, \delta_1) \subset U$ and for all $\xi \in X$ satisfying $0 < \|\xi\|_X < \delta_1$ there holds

$$\|f(\mathrm{x}+\xi) - f(\mathrm{x}) - Df(\mathrm{x})\xi\|_Y \leq \frac{\epsilon\|\xi\|_X}{2(\|Dg(\mathrm{y})\|_{Y,Z}+1)}.$$

By Proposition 6.3.7, the assumed differentiability of f at x implies that f is locally Lipschitz at x, i.e., there exists $\delta_2 > 0$ and $L > 0$ such that $B(\mathrm{x}, \delta_2) \subset U$ and for all $\xi \in X$ satisfying $0 < \|\xi\|_X < \delta_2$ there holds

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

Set $\delta_{\mathrm{x}} = \min\{\delta_1, \delta_2\}$.

By the assumed differentiability of g at y there exists $\delta_y > 0$ such that $B(\mathrm{y}, \delta_{\mathrm{y}}) \subset V$ and for all $0 < \|\eta\|_Y < \delta_{\mathrm{y}}$ there holds

$$\|g(\mathrm{y}+\eta) - g(\mathrm{y}) - Dg(\mathrm{y})\eta\|_Z \leq \frac{\epsilon\|\eta\|_Y}{2L}.$$

Set $\delta = \min\{\delta_{\mathrm{x}}, \delta_{\mathrm{y}}/L\}$. (Hence $L\delta \leq \delta_{\mathrm{y}}$ which we will use in a moment.)

If we set $\eta(\xi) = f(\mathrm{x}+\xi) - f(\mathrm{x}) = f(\mathrm{x}+\xi) - \mathrm{y}$, then $h = g \circ f$ satisfies

$$h(\mathrm{x}+\xi) - h(\mathrm{x}) = g(f(\mathrm{x}+\xi)) - g(f(\mathrm{x})) = g(y+\eta(\xi)) - g(\mathrm{y}).$$

Thus for all $\xi \in X$ satisfying $\|\xi\|_X < \delta$ we have

$$\|\eta(\xi)\|_Y = \|f(\mathrm{x}+\xi) - f(\mathrm{x})\|_Y \leq L\|\xi\|_X < L\delta \leq \delta_{\mathrm{y}}$$

so that

$$\begin{aligned}
&\|h(\mathrm{x}+\xi) - h(\mathrm{x}) - Dg(\mathrm{y})Df(\mathrm{x})\xi\|_Z \\
&\quad = \|g(\mathrm{y}+\eta(\xi)) - g(\mathrm{y}) - Dg(\mathrm{y})\eta(\xi) + Dg(\mathrm{y})\eta(\xi) - Dg(\mathrm{y})Df(\mathrm{x})\xi\|_Z \\
&\quad \leq \|g(\mathrm{y}+\eta(\xi)) - g(\mathrm{y}) - Dg(\mathrm{y})\eta(\xi)\|_Z + \|Dg(\mathrm{y})\eta(\xi) - Dg(\mathrm{y})Df(\mathrm{x})\xi\|_Z \\
&\quad \leq \|g(\mathrm{y}+\eta(\xi)) - g(\mathrm{y}) - Dg(\mathrm{y})\eta(\xi)\|_Z + \|Dg(\mathrm{y})\|_{Y,Z}\|\eta(\xi) - Df(\mathrm{x})\xi\|_Y \\
&\quad = \|g(\mathrm{y}+\eta(\xi)) - g(\mathrm{y}) - Dg(\mathrm{y})\eta(\xi)\|_Z \\
&\qquad + \|Dg(\mathrm{y})\|_{Y,Z}\|f(\mathrm{x}+\xi) - f(\mathrm{x}) - Df(\mathrm{x})\xi\|_Y \\
&\quad \leq \frac{\epsilon\|\eta(\xi)\|_Y}{2L} + \|Dg(\mathrm{y})\|_{Y,Z}\frac{\epsilon\|\xi\|_X}{2(\|Dg(\mathrm{y})\|_{Y,Z}+1)} \\
&\quad \leq \frac{\epsilon L\|\xi\|_X}{2L} + \frac{\epsilon\|\xi\|_X}{2} \\
&\quad = \epsilon\|\xi\|_X.
\end{aligned}$$

Since $Df(\mathrm{x}) \in \mathscr{B}(X, Y)$ and $Dg(\mathrm{y}) \in \mathscr{B}(Y, Z)$, then $Dg(\mathrm{y})Df(\mathrm{x}) \in \mathscr{B}(X, Z)$.

This shows that $h = g \circ f$ is differentiable at x with derivative is $Dg(f(\mathrm{x}))Df(\mathrm{x})$. $\square$