Math 346 Lecture #5
6.5 Mean Value Theorem and FTC

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.

6.5.1 The Mean Value Theorem

Recall the Mean Value Theorem from Calculus I: if $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), then there exists $c \in (a, b)$ such that

$$f(b) - f(a) = f'(c)(b - a).$$

We show how to extend this to the general Banach space setting.

Theorem 6.5.1 (The Mean Value Theorem). Let $f : U \to \mathbb{R}$ be differentiable on U. If, for points $\mathrm{a}, \mathrm{b} \in U$, the entire line segment

$$\ell(\mathrm{a}, \mathrm{b}) = \{(1 - t)\mathrm{a} + t\mathrm{b} : t \in [0, 1]\}$$

lies in U, then there exists $\mathrm{c} \in \ell(\mathrm{a}, \mathrm{b})$ such that

$$f(\mathrm{b}) - f(\mathrm{a}) = Df(\mathrm{c})(\mathrm{b} - \mathrm{a}).$$

Proof. Let $h : [0, 1] \to \mathbb{R}$ be defined by $h(t) = f((1 - t)\mathrm{a} + t\mathrm{b})$.

Since $\ell(\mathrm{a}, \mathrm{b})$ lies entirely in U, the function h is continuous on $[0, 1]$ and differentiable on $(0, 1)$.

By the Calculus I Mean Value Theorem, there exists $t_0 \in (0, 1)$ such that $h(1) - h(0) = h'(t_0)(1 - 0) = h'(t_0)$.

By the Chain Rule we have $h'(t_0) = Df(h(t_0))(\mathrm{b} - \mathrm{a})$.

Setting $\mathrm{c} = h(t_0)$ gives $f(\mathrm{b}) - f(\mathrm{a}) = h(1) - h(0) = h'(t_0) = Df(\mathrm{c})(\mathrm{b} - \mathrm{a})$. □

Remark 6.5.2. A subset U of X is said to be convex if for any $\mathrm{a}, \mathrm{b} \in U$ the line segment $\ell(\mathrm{a}, \mathrm{b})$ lies entirely in U. Open balls $B(\mathrm{x}, r)$ in X are convex because for any $t \in [0, 1]$ we have

$$\begin{aligned} \|(1 - t)\mathrm{a} + t\mathrm{b} - \mathrm{x}\|_X &= \|(1 - t)\mathrm{a} + t\mathrm{b} - (1 - t)\mathrm{x} - t\mathrm{x}\|_X \\ &= \|(1 - t)(\mathrm{a} - \mathrm{x}) + t(\mathrm{b} - \mathrm{x})\|_X \\ &\le (1 - t)\|\mathrm{a} - \mathrm{x}\|_X + t\|\mathrm{b} - \mathrm{x}\|_X \\ &< (1 - t)r + tr \\ &= r. \end{aligned}$$

6.5.2 Single Variable FTC

Technical Note. To extend familiar integral results from Calculus I to functions of the form $f : [a, b] \to X$ (where $[a, b]$ is a compact interval of $\mathbb{R}$ and X is a Banach space), we will use the isomorphism $\mathscr{B}(\mathbb{R}, X) \cong X$ given by the evaluation map $e_1(\phi) = \phi(1)$.

This is an isomorphism because it is

(1) linear: $e_1(a\phi + b\psi) = (a\phi + b\psi)(1) = a\phi(1) + b\psi(1) = ae_1(\phi) + be_1(\psi)$;
(2) injective: $e_1(\phi) = e_1(\psi)$ implies $\phi(1) = \psi(1)$ so that for any $a \in \mathbb{R}$ there holds $\phi(a) = a\phi(1) = a\psi(1) = \psi(a)$, i.e., $\phi = \psi$; and
(3) surjective: for $\mathrm{v} \in X$ the map $\phi(t) = t\mathrm{v}$ belongs to $\mathscr{B}(\mathbb{R}, X)$, i.e., $\|\phi\|_{\mathbb{R},X} = \sup\{\|tv\|_X/|t|_{\mathbb{R}} : t \neq 0\} = \|v\|_X < \infty$, with $e_1(\phi) = \phi(1) = \mathrm{v}$.

For a differentiable $f : (a, b) \to X$, and for each $t \in (a, b)$, the derivative map $Df(t) \in \mathscr{B}(\mathbb{R}, X)$ is identified with the vector $Df(t)(1)$ by the isomorphism $e_1 : \mathscr{B}(\mathbb{R}, X) \to X$, and by abuse of notation we write $Df(t)$ instead of $Df(t)(1)$ to be the vector $Df(t)(1)$.

By way of comparison, for differentiable $f : (a, b) \to \mathbb{R}$, and for each $t \in (a, b)$ the identification is of $Df(t) \in \mathscr{B}(\mathbb{R}, \mathbb{R})$ with $Df(t)(1) = f'(t)1 = f'(t)$.

Lemma 6.5.3. If $f : [a, b] \to X$ is continuous on $[a, b]$, differentiable on (a, b), and $Df(t) = 0$ (i.e., $Df(t)(h) = 0$ for all $h \in \mathbb{R}$) for all $t \in (a, b)$, then f is constant.

Proof. Let $\alpha, \beta \in (a, b)$ with $\alpha < \beta$ and fix an arbitrary $t \in (\alpha, \beta)$.

By hypothesis, for all $\epsilon > 0$ there exists $\delta_t > 0$ such that for all $|h| < \delta_t$ there holds

$$\|f(t + h) - f(t)\|_X = \|f(t + h) - f(t) - 0\|_X = \|f(t + h) - f(t) - Df(t)h\|_X \leq \epsilon|h|.$$

The collection of open intervals $(t - \delta_t, t + \delta_t)$, $t \in (\alpha, \beta)$ is an open covering of the compact $[\alpha, \beta]$, so there is a finite subcovering $(t_i - \delta_{t_i}, t_i + \delta_{t_i})$, $i = 1, \ldots, n$, of $[\alpha, \beta]$.

WLOG we can assume that $\alpha < t_1 < t_2 < \cdots < t_n < \beta$.

We choose points $x_0, x_1, \ldots, x_n \in [\alpha, \beta]$ so that

$$\alpha = x_0 < t_1 < x_1 < t_2 < x_2 < \cdots < t_n < x_n = \beta$$

with $|x_i - t_i| < \delta_{t_i}$ and $|x_i - t_{i-1}| < \delta_{t_{i-1}}$.

We then have

$$\begin{aligned}
\|f(\beta) - f(\alpha)\|_X &= \left\| \sum_{i=1}^n f(x_i) - f(x_{i-1}) \right\|_X \\
&= \left\| \sum_{i=1}^n \big(f(x_i) - f(t_i)\big) + \big(f(t_i) - f(x_{i-1})\big) \right\|_X \\
&\leq \sum_{i=1}^n \|f(x_i) - f(t_i)\|_X + \sum_{i=1}^n \|f(t_i) - f(x_{i-1})\|_X \\
&\leq \sum_{i=1}^n \epsilon\big((x_i - t_i) + (t_i - x_{i-1})\big) \\
&= \epsilon(\beta - \alpha).
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, it follows that f is constant on (α, β).

Since $\alpha, \beta \in (a, b)$, with $\alpha < \beta$, are arbitrary, it follows that f is constant on (a, b).

The continuity of f on $[a, b]$ implies that f is constant on $[a, b]$. □

Theorem 6.5.4 (FTC). (i) If $f \in C([a,b],X)$, then $t \to \int_a^t f(s)\,ds$ is differentiable on (a,b), and for all $t \in (a,b)$ there holds

$$\frac{d}{dt}\int_a^t f(s)\,ds = f(t).$$

(ii) If $F \in C([a,b],X)$ is continuously differentiable on (a,b) and $DF(t)$ extends to a continuous function on $[a,b]$, then

$$\int_a^b DF(s)\,ds = F(b) - F(a).$$

Proof. (i) Since f is continuous on the compact $[a,b]$, the function f is uniformly continuous by Theorem 5.5.9: for every $\epsilon > 0$ there exists $\delta > 0$ such that for all $t \in [a,b]$ and all $|h| < \delta$ with $t+h \in [a,b]$, there holds

$$\|f(t+h) - f(t)\|_X < \epsilon.$$

By using properties of the single-variable Banach-valued integral (or regulated integral), we have for $|h| < \delta$ that

$$\begin{aligned}
&\left\| \int_a^{t+h} f(s)\,ds - \int_a^t f(s)\,ds - f(t)h \right\|_X \\
&= \left\| \int_t^{t+h} f(s)\,ds - f(t)h \right\|_X \\
&= \left\| \int_t^{t+h} \big(f(s) - f(t)\big)\,ds \right\|_X \\
&\leq \left| \int_t^{t+h} \|f(s) - f(t)\|_X\,ds \right| \\
&\leq \left| \int_t^{t+h} \epsilon\,ds \right| \\
&= \epsilon|h|
\end{aligned}$$

[We put the absolute value around the integral in the first $\leq$ to account for $h < 0$.]

This shows that $\int_a^t f(s)\,ds$ is a differentiable function of t on (a,b), and that its derivative is $f(t)$.

(ii) By hypotheses, the function $s \to DF(s)$ from $[a,b]$ to X is continuous.

[Note that here we are using the identification of $DF(s)$ with $DF(s)(1)$ afforded by the isomorphism $e_1 : \mathscr{B}(\mathbb{R},X) \to X$.]

Also by hypothesis, the function F from $[a,b] \to X$ is continuous.

Then the function $G : [a,b] \to X$ defined by

$$G(t) = \int_a^t DF(s)\,ds - F(t)$$

is continuous on $[a,b]$ by Theorem 5.10.12(v), and G is differentiable on (a,b) by part (i) and the assumption that F is differentiable on (a,b).

So by part (i) we have for $t \in (a,b)$ that

$$DG(t) = \frac{d}{dt}\left[\int_a^t DF(s)\ ds - F(t)\right] = DF(t) - DF(t) = 0.$$

By Lemma 6.5.3 the function G is constant on $[a,b]$.

Since

$$G(a) = \int_a^a DF(s)\ ds - F(a) = -F(a)$$

and

$$G(b) = \int_a^b DF(s)\ ds - F(b),$$

the equality $G(b) = G(a)$ implies

$$\int_a^b DF(s)\ ds - F(b) = -F(a)$$

or

$$\int_a^b DF(s)\ ds = F(b) - F(a).$$

This gives the result. □

Corollary 6.5.5 (Integral Mean Value Theorem). If $f \in C^1(U,Y)$ and for $\mathrm{x}_*, \mathrm{x} \in U$ the line segment $\ell(\mathrm{x}_*, \mathrm{x}) = \{(1-t)\mathrm{x}_* + t\mathrm{x} : t \in [0,1]\}$ lies entirely in U, then

$$f(\mathrm{x}) - f(\mathrm{x}_*) = \int_0^1 Df(t\mathrm{x} + (1-t)\mathrm{x}_*)(\mathrm{x} - \mathrm{x}_*)\ dt.$$

[Note here that $Df(t\mathrm{x} + (1-t)\mathrm{x}_*) \in \mathscr{B}(X,Y)$ so that $Df(t\mathrm{x}+(1-t)\mathrm{x}_*)(\mathrm{x}-\mathrm{x}_*) \in Y$, whence the integral value is in Y.] Alternatively, if we write $\mathrm{h} = \mathrm{x} - \mathrm{x}_*$, then we have

$$f(\mathrm{x}_* + \mathrm{h}) - f(\mathrm{x}_*) = \int_0^1 Df(\mathrm{x}_* + t\mathrm{h})\mathrm{h}\ dt.$$

Moreover, we have

$$\|f(\mathrm{x}) - f(\mathrm{x}_*)\|_Y \leq \sup_{\mathrm{c} \in \ell(\mathrm{x}_*, \mathrm{x})} \|Df(\mathrm{c})\|_{X,Y} \|\mathrm{x} - \mathrm{x}_*\|_X.$$

The proof of this is HW (Exercise 6.22).

Corollary 6.5.6 (Change of Variable Formula). Let $f \in C([a,b],X)$. If $g : [c,d] \to [a,b]$ is continuous with g continuously differentiable on (c,d) with Dg continuously extendable to $[c,d]$, then

$$\int_c^d f(g(s))Dg(s)\ ds = \int_{g(c)}^{g(d)} f(t)\ dt.$$

[Here $Dg(s) \in \mathscr{B}(\mathbb{R},\mathbb{R})$ and is therefore identified with $g'(s)$.]

The proof of this is HW (Exercise 6.23).

6.5.3 Uniform Convergence and Derivatives

Recall that since the single-variable regulated integral is a bounded linear transformation from the closed subspace $\overline{S([a,b],X)}$ of $(L^\infty([a,b],X),\|\cdot\|_\infty)$ to X, the integral commutes with uniform limits: if $(f_n)_{n=1}^\infty$ converges to f in $\overline{S([a,b],X)}$ then

$$\lim_{n\to\infty}\int_a^b f_n\ dt = \int_a^b \left(\lim_{n\to\infty} f_n\right)\ dt = \int_a^b f\ dt.$$

Although differentiation is an unbounded linear transformation, we will use the Mean Value Theorem to prove that the differentiation commutes with uniform limits provided the necessary limits exist.

Recall that you have seen this already in Math 341: suppose (f_n) is a sequence of real-valued differentiable functions on the compact interval $[a,b]\in\mathbb{R}$; if

(i) there is $x_0\in[a,b]$ such that $f_n(x_0)$ converges and

(ii) (f_n') converges uniformly to a function g on $[a,b]$,

then (f_n) converges uniformly and the limit function $f=\lim f_n$ is differentiable and satisfies $f'=g$ on $[a,b]$, i.e.,

$$\left(\lim_{n\to\infty} f_n\right)' = f' = g = \lim_{n\to\infty} f_n'.$$

The proof of this uses the Mean Value Theorem reviewed at the beginning of this lecture note.

There are two technicalities in extending this Math 341 result to the general Banach space setting. The first is that differentiation is defined over open subsets while uniform convergence is best applied over compact subsets. We overcome this technicality with what is called the "compact-open topology" on $C(U,Y)$ which we address after this paragraph. The second technicality is that to use the Integral Mean Value Theorem requires the functions in the sequence have continuous derivatives as expressed in the statement of Theorem 6.5.11.

For $f\in C(U,Y)$ and K a compact subset of the open U, the restriction $f|_K$ belongs to $(C(K,Y),\|\cdot\|_\infty)$ with norm $\|f|_K\|_\infty=\sup\{\|f(\mathrm{x})\|_Y : \mathrm{x}\in K\}$ (because continuous on compact implies bounded).

Note that $(C(K,Y),\|\cdot\|_\infty)$ is a Banach space because it is a closed subset of the Banach space $(L^\infty(U,Y),\|\cdot\|_\infty)$ (see Theorem 5.7.5; also recall that closed subsets of complete metric spaces are complete).

Definition 6.5.7. We say that a sequence $(f_n)_{n=1}^\infty$ in $C(U,Y)$

(i) is Cauchy in $C(U,Y)$ if for every compact $K\subset U$ the sequence $(f_n|_K)_{n=1}^\infty$ is Cauchy in $(C(K,Y),\|\cdot\|_\infty)$, and

(ii) converges uniformly on compact subsets of U to $f\in C(U,Y)$ if for every compact $K\subset U$ the sequence $(f_n|_K)_{n=1}^\infty$ converges to $f|_K$ in $(C(K,Y),\|\cdot\|_\infty)$.

Note. A Cauchy sequence $(g_n)_{n=1}^\infty$ in $(C(K,Y),\|\cdot\|_\infty)$ converges uniformly to a unique $g \in C(K,Y)$ because $(C(K,Y),\|\cdot\|_\infty)$ is a Banach space.

Remark 6.5.8. To check that a sequence (f_n) in $C(U,Y)$, when U is an open subset of a finite dimensional Banach space X, is Cauchy or that if converges uniformly on compact subsets of U, it suffices to verify the condition on compact subsets that are closed balls of the form $\overline{B(\mathrm{x},r)}$.

Moreover, when $U = B(\mathrm{x}_0,R)$ it suffices to check the condition on closed balls $\overline{B(\mathrm{x}_0,r)}$ for all $0<r<R$ (part of the proof is HW Exercise 6.24).

Example (in lieu of 6.5.9). For the open interval $U=(0,1)$, the function $f(x)=1/x$ belongs to $C(U,\mathbb{R})$, but $\|f\|_\infty=\infty$ so that $f \notin L^\infty(U,\mathbb{R})$.

However, the sequence $f_n(x)=1/(x+1/n)$ in $C(U,\mathbb{R})$ converges uniformly on compact subsets to f.

Any compact subset K of U lies in a compact subinterval $[a,b]$ of $(0,1)$.

On the compact $[a,b]$ we have

$$\begin{aligned}\left\| f_n|_{[a,b]} - f|_{[a,b]} \right\|_\infty &= \sup_{t\in[a,b]} \left| \frac{1}{x+1/n} - \frac{1}{x} \right| \\ &= \sup_{t\in[a,b]} \left| \frac{1/n}{x(x+1/n)} \right| \\ &= \frac{1/n}{a(a+1/n)} \\ &= \frac{1}{a(na+1)}.\end{aligned}$$

This has limit 0 as $n\to\infty$, and hence $f_n \to f$ uniformly on compact subsets of U.

Theorem 6.5.11. For a finite dimensional Banach space X, an open set $U=B(\mathrm{x}_*,r)\subset X$ for $\mathrm{x}_* \in X$ and $r>0$, and a sequence $(f_n)\in C^1(U,Y)$, if

(i) $(f_n(\mathrm{x}_*))_{n=1}^\infty$ converges in Y and
(ii) $(Df_n)_{n=1}^\infty$ converges uniformly on compact subsets of U to $g \in C(U,\mathscr{B}(X,Y))$,

then $(f_n)_{n=1}^\infty$ converges uniformly on compact subsets of U to $f \in C^1(U,Y)$ for which $Df = g$.

Proof. By hypothesis, the sequence $(f_n(\mathrm{x}_*))_{n=1}^\infty$ converges to, say, $\mathrm{z}\in Y$.

For each $\mathrm{x}\in B(\mathrm{x}_*,r)=U$ set $\mathrm{h}=\mathrm{x}-\mathrm{x}_*$.

A candidate for the limit function $f\in C(U,Y)$ is

$$f(\mathrm{x}) = \mathrm{z} + \int_0^1 g(\mathrm{x}_*+t\mathrm{h})\mathrm{h}\,dt$$

where $g\in C(U,\mathscr{B}(X,Y))$ is what $(Df_n)_{n=1}^\infty$ converges uniformly on compact sets to by hypothesis.

Since $U = B(\mathrm{x}_*, r)$ is convex, the function $g(\mathrm{x}_* + t\mathrm{h})$ is defined for all $t \in [0, 1]$.

When $\mathrm{h} = 0$, i.e., $\mathrm{x} = \mathrm{x}_*$, then

$$f(\mathrm{x}_*) = \mathrm{z} + \int_0^1 g(\mathrm{x}_* + t0)0\ dt = \mathrm{z} = \lim_{n\to\infty} f_n(\mathrm{x}_*).$$

We show that $(f_n)_{n=1}^{\infty}$ converges to f uniformly on compact subsets of $B(\mathrm{x}_*, r)$.

To this end it suffices to show uniform convergence on the arbitrary compact set $K = \overline{B(\mathrm{x}_*, \rho)} \subset U$, i.e., $0 < \rho < r$. [The case of $\rho = 0$ we already have.]

On K we have, by applying the Integral Mean Value Theorem to f_n and using the definition of f, that

$$\begin{aligned}
&\|f_n(\mathrm{x}) - f(\mathrm{x})\|_Y \\
&\quad = \left\| f_n(\mathrm{x}_*) + \int_0^1 Df_n(\mathrm{x}_* + t\mathrm{h})\mathrm{h}\ dt - \mathrm{z} - \int_0^1 g(\mathrm{x}_* + t\mathrm{h})\mathrm{h}\ dt \right\|_Y \\
&\quad \le \|f_n(\mathrm{x}_*) - \mathrm{z}\|_Y + \left\| \int_0^1 \big(Df_n(\mathrm{x}_* + t\mathrm{h}) - g(\mathrm{x}_* + t\mathrm{h})\big)\mathrm{h}\ dt \right\|_Y \\
&\quad \le \|f_n(\mathrm{x}_*) - \mathrm{z}\|_Y + \sup_{\mathrm{c}\in K} \|Df_n(\mathrm{c}) - g(\mathrm{c})\|_{X,Y} \|h\|_X.
\end{aligned}$$

Since $f_n(_*) \to \mathrm{z}$, for $\epsilon > 0$ there is $N \in \mathbb{N}$ such that for all $n \ge N$ we have

$$\|f_n(\mathrm{x}_*) - \mathrm{z}\|_Y < \frac{\epsilon}{2},$$

and since $Df_n \to g$ uniformly on the compact K, there is $M \ge N$ such that for all $n \ge M$ we have

$$\sup_{\mathrm{c}\in K} \|Df_n(\mathrm{c}) - g(\mathrm{c})\|_Y < \frac{\epsilon}{2r}.$$

Thus for all $n \ge M$, and with $\|h\|_X < r$, we have

$$\|f_n(\mathrm{x}) - f(\mathrm{x})\|_Y < \frac{\epsilon}{2} + \frac{\epsilon}{2r} r = \epsilon.$$

The choice of N and M is independent of the point $\mathrm{x} \in K$ chosen, so for all $n \ge N$ we have

$$\|f_n - f\|_\infty = \sup_{\mathrm{x}\in K} \|f_n(\mathrm{x}) - f(\mathrm{x})\|_Y \le \epsilon.$$

Thus $f_n \to f$ uniformly on the compact K.

Last we show that $Df(\mathrm{x}) = g(\mathrm{x})$ for all $\mathrm{x} \in U$.

For a fixed $\mathrm{x} \in B(\mathrm{x}_*, r) = U$ there exists $a > 0$ such that $K = \overline{B(\mathrm{x}, a)} \subset U$.

For $\|\mathrm{h}\|_X \le a$ we have by the Integral Mean Value Theorem that

$$f_n(\mathrm{x} + h) - f_n(\mathrm{x}) = \int_0^1 Df_n(\mathrm{x} + t\mathrm{h})\mathrm{h}\ dt.$$

The left-hand side of this converges uniformly on the compact set K to $f(\mathrm{x}+\mathrm{h})-f(\mathrm{x})$, while the integrand on the right-hand side converges uniformly on K to $g(\mathrm{x}+t\mathrm{h})\mathrm{h}$.

By the commutativity of integration with uniform limits we thus get

$$\begin{aligned} f(\mathrm{x}+\mathrm{h})-f(\mathrm{x}) &= \lim_{n\to\infty}\big(f_n(\mathrm{x}+\mathrm{h})-f_n(\mathrm{x})\big) \\ &= \lim_{n\to\infty}\int_0^1 Df_n(\mathrm{x}+t\mathrm{h})\mathrm{h}\ dt \\ &= \int_0^1 \lim_{n\to\infty} Df_n(\mathrm{x}+t\mathrm{h})\mathrm{h}\ dt \\ &= \int_0^1 g(\mathrm{x}+t\mathrm{h})\mathrm{h}\ dt. \end{aligned}$$

Hence

$$\begin{aligned} \|f(\mathrm{x}+\mathrm{h})-f(\mathrm{x})-g(\mathrm{x})\mathrm{h}\|_Y &= \left\|\int_0^1 \big(g(\mathrm{x}+t\mathrm{h})\mathrm{h}-g(\mathrm{x})\mathrm{h}\big)\ dt\right\|_Y \\ &= \left\|\int_0^1 \big(g(\mathrm{x}+t\mathrm{h})-g(\mathrm{x})\big)\mathrm{h}\ dt\right\|_Y. \end{aligned}$$

Continuity of g on the compact $K=\overline{B(\mathrm{x},a)}$ implies g is uniformly continuous on K.

For $\epsilon>0$ there then exists $0<\delta<a$ such that for all $|\mathrm{h}|_X<\delta$ there holds

$$\|g(\mathrm{x}+t\mathrm{h})-g(\mathrm{x})\|_{X,Y}<\epsilon.$$

We thus arrive at

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

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