NCKU 112.2
Advanced Calculus 2

Eric Liu

Contents

Chapter 1

General Topology

1.1 Equivalent axiomazations

In the first lecture of Topology, we learned that a topological space (X, τ) is a space X with a topology $\tau \subseteq \mathcal{P}(X)$ such that τ satisfy the following three axioms.

$$\begin{cases} X, \varnothing \in \tau \\ \forall O, Y \in \tau, O \cap Y \in \tau \text{ (Closed under finite intersection)} \\ \forall T \subseteq \mathcal{P}(\tau), \bigcup T \in \tau \text{ (Closed under arbitrary union)} \end{cases} \tag{1.1}$$

It is only after the explicit listing of the above three of open sets, we then start defining "closed sets", "neighborhoods", "continuous functions", "compact sets" or "connected sets" based on open sets.

Although this approach, axiomatization via open sets, is mathematically sufficient, in history, there are other axiomatization proved to be equivalent to the traditional axiomaization via open sets.

In this section, we will give other three axiomatizations of topology, via neighborhood systems, via nets and via filters, and show they are equivalent with each other.

In this note, a *neighborhood system* is a function $\mathcal{N} : X \to \mathcal{P}(\mathcal{P}(X))$. We shall first show that there exists an explicit one-to-one correspondence between collection of topologies on X and collection of neighborhood systems that satisfy the following axioms.

Axiom 1.1.1. (Axioms of neighborhood systems)

$$\begin{cases}\forall x \in X, \mathcal{N}(x) \neq \varnothing \text{ (Non empty)}\\ \forall x \in X, \forall M \in \mathcal{N}(x), x \in M \text{ (Around)}\\ \forall x \in X, \forall M \subseteq X, \exists N \in \mathcal{N}(x), N \subseteq M \implies M \in \mathcal{N}(x) \text{(Closed under superset)}\\ \forall x \in X, \forall M, N \in \mathcal{N}(x), M \cap N \in \mathcal{N}(x) \text{(Closed under finite intersection)}\\ \forall x \in X, \forall M \in \mathcal{N}(x), \exists N \in \mathcal{N}(x), N \subseteq M \text{ and } \forall y \in N, M \in \mathcal{N}(y) \text{(Open neighborhood)}\end{cases}$$

Suppose C is the collection of topologies on X and D is the collection of neighborhood systems that satisfy Axioms 1.1.1. Now we wish to prove that the function $f : C \to D$

$$\tau \mapsto \mathcal{N}_\tau \text{ where } \mathcal{N}_\tau(x) = \{A \in \mathcal{P}(X) : \exists O \in \tau, x \in O \subseteq A\} \tag{1.2}$$

is bijective. In order to prove this, we first have to prove that f is well-defined, meaning for each topology τ, the function $\mathcal{N}_\tau$ is indeed a neighborhood system that satisfy the above axioms. This is easy, which we will omit here. It remains that we have to prove two statements: f is one-to-one, and f is onto.

Theorem 1.1.2. (f in Equation 1.2 is one-to-one) As titled.

Proof. Suppose τ and τ' are two different topologies on X. WOLG, suppose $O \in \tau \setminus \tau'$. By definition of f (Equation 1.2), we know

$$\forall x \in O, O \in \mathcal{N}_\tau(x)$$

This means, to prove $\mathcal{N}_\tau \neq \mathcal{N}_{\tau'}$, we only have to prove the existence of some $x \in O$ such that $O \notin \mathcal{N}_\tau(x)$. Assume $\forall x \in O, O \in \mathcal{N}_{\tau'}(x)$. By definition of f (Equation 1.2), we can deduce $\forall x \in O, \exists A_x \in \tau', x \in A_x \subseteq O$. It is easy to verify that $O = \bigcup_{x \in O} A_x$. Then by Axiom 1.1 of open sets , we $O = \bigcup_{x \in O} A_x \in \tau'$ CaC to that $O \in \tau \setminus \tau'$. ■

To prove f is onto is a little bit complicated.

Theorem 1.1.3. (f in Equation 1.2 is onto) As titled.

Proof. Define $g : D \to C$ by

$$\mathcal{N} \mapsto \sigma_\mathcal{N} \text{ where } \sigma_\mathcal{N} = \{O \in \mathcal{P}(X) : \forall x \in O, O \in \mathcal{N}(x)\} \tag{1.3}$$

It is easy to check for each neighborhood system $\mathcal{N} : X \to \mathcal{P}(\mathcal{P}(X))$ satisfying Axioms 1.1.1, the collection $\sigma_\mathcal{N}$ is a topology.

Then, to prove f is onto, it suffice to prove

$$\forall \mathcal{N} \in D, \mathcal{N} = f(\sigma_{\mathcal{N}})$$

In other words, we wish to prove

$$\forall \mathcal{N} \in D, \forall x \in X, \mathcal{N}(x) = \big(f(\sigma_{\mathcal{N}})\big)(x)$$

By definition of f (Equation 1.2), we know

$$\big(f(\sigma_{\mathcal{N}})\big)(x) = \{A \in \mathcal{P}(X) : \exists O \in \sigma_{\mathcal{N}}, x \in O \subseteq A\} \tag{1.4}$$

Suppose $A \in \big(f(\sigma_{\mathcal{N}})\big)(x)$. Let $O \in \sigma_{\mathcal{N}}$ satisfy $x \in O \subseteq A$. By definition of g (Equation 1.3), we know $O \in \mathcal{N}(x)$. Then because Axiom 1.1.1 says that $\mathcal{N}(x)$ is closed under superset, we now see $A \in \mathcal{N}(x)$. We have proved $\big(f(\sigma_{\mathcal{N}})\big)(x) \subseteq \mathcal{N}(x)$.

Suppose $M \in \mathcal{N}(x)$. By the fifth axiom listed in Axiom 1.1.1, we know there exists some $N \in \mathcal{N}(x)$ such that $N \subseteq M$ and $\forall y \in N, M \in \mathcal{N}(y)$. Notice that the union of such N is still a subset of M and M is a neighborhood of each point in the union. This means that there exists a maximum N that satisfy

$$\begin{cases} N \subseteq M \\ \forall y \in N, M \in \mathcal{N}(y) \end{cases}$$

We now prove $N \in \sigma_{\mathcal{N}}$. Arbitrarily pick $z \in N$, by definition of N above, we can deduce $M \in \mathcal{N}(z)$. Then by the fifth axiom listed in Axiom 1.1.1, we know there exists some $N' \in \mathcal{N}(z)$ satisfying $N' \subseteq M$ and $\forall y \in N', M \in \mathcal{N}(y)$. Because N by definition is the maximum of such neighborhood, we see $N' \subseteq N$. Then because Axiom 1.1.1 says that $\mathcal{N}(z)$ is closed under superset, we see $N \in \mathcal{N}(z)$.

We have shown $\forall z \in N, N \in \mathcal{N}(z)$. Then by definition of g (Equation 1.3), we see $N \in \sigma_{\mathcal{N}}$ (done) .

Then because $N \subseteq M$ by definition of N and because $N \in \sigma_{\mathcal{N}}$, we see $M \in \big(f(\sigma_{\mathcal{N}})\big)(x)$, according to Equation 1.4. We have proved $\mathcal{N}(x) \subseteq \big(f(\sigma_{\mathcal{N}})\big)(x)$ (done) ■

Before embarking on the axiomatization via nets, we first have to settle the terminologies. Recall that a set D is **directed** if there exists an ordering $\leq$ on D such that $\leq$ is reflexive, transitive, and we have $\forall i, j \in D, \exists k \in D, i \leq k$ and $j \leq k$. By a **net**, we mean a function w whose domain is directed. We say a subset D' of a directed set is **cofinal** if $\forall d \in D, \exists d' \in D', d \leq d'$. By a **subnet** of $w : D \to X$, we mean a net

$v : E \to X$ such that there exists $h : E \to D$ such that

$$\begin{cases} \forall e, e' \in E, e \leq e' \implies h(e) \leq h(e')(\text{Monotone}) \\ h[E] \text{ is cofinal} \\ v = w \circ h \end{cases}$$

One can check that when w is a sequence x_n and v is the sub-sequence x_{n_k}, the corresponding h is just n_k.

By a **tail** T_d of a directed set D, we mean $T_d = \{e \in D : d \leq e\}$. We say $w : D \to X$ **is eventually in** $A \subseteq X$ if $\exists d \in D, w[T_d] \subseteq A$. We say $w : D \to X$ **is frequently in** $A \subseteq X$ if $\forall d \in D, \exists e \in T_d, w(e) \in A$. Given a topological space (X, τ), we say w **converge** to a point x, if for all neighborhood O around x there exists $d \in D$ such that $w[T_d] \subseteq O$. Notice that if we wish to prove $w \to x$ we only have to verify for all open neighborhoods O around x. Also notice that w can converge to multiple points. A trivial example is when two point are topologically indistinguishable.

1.2 Equivalent definitions

1.3 Product topology

1.4 ABOVE ARE FIXED AND SHOULD BE FOCUSED BEFORE ARRANGEMENT OF THE FOLLOWING

1.5 Manifold

1.6 Quotient topology

1.7 Countability axioms

1.8 Separation axioms

1.9 Path connected

1.10 Basic notion on compact

1.11 Baire space

1.12 Homotopy

1.13 Simply connected

1.14 Fundamental group

Chapter 2

Metric Space and some Linear Algebra

2.1 Pseudo Metric

2.2 Completion

2.3 Bounded and Totally Bounded

2.4 Compactness

2.5 Holder Continuity

2.6 Limit Interchange

Given an arbitrary set X and a complete metric space $(\overline{Y}, d)$, in Section 2.8, we have proved that the set of functions with the following properties

(a) boundedness

(b) unboundedness

are respectively closed under uniform convergence. In next section (Section 2.7), we will prove that the following three properties

(a) continuity

(b) uniform continuity

(c) K-Lipschitz continuity

are again all closed under uniform convergence, where the proof for continuity is closed under uniform convergence use Theorem 2.6.1 as a lemma.

Here, we prove

(a) convergent of sequences

in, of course, complete metric space, is also closed under uniform convergence.

The reason we require the codomain $\overline{Y}$ of sequence to be complete is explained in the last paragraph of Section 2.8. An example of such beautiful closure is lost if the codmain (Y, d) is not complete is $Y = \mathbb{R}^*$ and $a_{n,k} = \frac{1}{n} + \frac{1}{k}$.

Theorem 2.6.1. (Change Order of Limit Operations: Part 1) Given a double sequence $a_{n,k}$ whose codomain is (Y, d). Suppose

(a) $a_{n,k} \to a_{\bullet,k}$ uniformly as $n \to \infty$

(b) $a_{n,k} \to A_n$ pointwise as $k \to \infty$.

(c) $A_n \to A$

Then we can deduce

$$\lim_{k\to\infty} a_{\bullet,k} \text{ exists and } \lim_{k\to\infty} a_{\bullet,k} = A$$

In other words, we can switch the order of limit operations

$$\lim_{k\to\infty} \lim_{n\to\infty} a_{n,k} = \lim_{n\to\infty} \lim_{k\to\infty} a_{n,k}$$

Proof. We wish to prove

$$a_{\bullet,k} \to A \text{ as } k \to \infty$$

Fix ϵ. Because $a_{n,k} \to a_{\bullet,k}$ uniformly and $A_n \to A$ as $n \to \infty$, we know there exists m such that

$$d(A_m, A) < \frac{\epsilon}{3} \text{ and } \forall k \in \mathbb{N}, d(a_{m,k}, a_{\bullet,k}) < \frac{\epsilon}{3} \tag{2.1}$$

Then because $a_{m,k} \to A_m$ as $k \to \infty$, we know there exists K such that

$$\forall k > K, d(a_{m,k}, A_m) < \frac{\epsilon}{3} \tag{2.2}$$

We now claim

$$\forall k > K, d(a_{\bullet,k}, A) < \epsilon$$

The claim is true since by Equation 2.1 and Equation 2.2, we have

$$\forall k > K, d(a_{\bullet,k}, A) \leq d(a_{\bullet,k}, a_{m,k}) + d(a_{m,k}, A_m) + d(A_m, A) < \epsilon \text{ (done)}$$

■

Theorem 2.6.2. (Change Order of Limit Operations: Part 2) Given a double sequence $a_{n,k}$ whose codomain is (Y, d). Suppose

(a) $a_{n,k} \to a_{\bullet,k}$ uniformly as $n \to \infty$

(b) $a_{n,k} \to A_n$ pointwise as $k \to \infty$

(c) $a_{\bullet,k} \to A$ as $k \to \infty$

Then we can deduce

$$A_n \text{ converge and } A_n \to A$$

Proof. Fix ϵ. We wish to find N such that

$$\forall n > N, d(A_n, A) < \epsilon$$

Because $a_{n,k} \to a_{\bullet,k}$ uniformly as $n \to \infty$, we can let N satisfy

$$\forall n > N, \forall k \in \mathbb{N}, d(a_{n,k}, a_{\bullet,k}) < \frac{\epsilon}{3} \tag{2.3}$$

We claim

$$\text{such } N \text{ works}$$

Arbitrarily pick $n > N$. Because $a_{\bullet,k} \to A$, and because $a_{n,k} \to A_n$, we know there exists j such that

$$d(a_{\bullet,j}, A) < \frac{\epsilon}{3} \text{ and } d(a_{n,j}, A_n) < \frac{\epsilon}{3} \tag{2.4}$$

From Equation 2.3 and Equation 2.4, we now have

$$d(A_n, A) \leq d(A_n, a_{n,j}) + d(a_{n,j}, a_{\bullet,j}) + d(a_{\bullet,j}, A) < \epsilon \text{ (done)}$$

■

Corollary 2.6.3. (Change Order of Limit Operation in Complete Metric Space) Given a sequence of function $f_n : E \to (Y, d)$ and a function $f : E \to (Y, d)$ such that

(a) f_n converge uniformly

(b) $\lim_{t \to x} f_n(t)$ exists for all $n \in \mathbb{N}$

(c) (Y, d) is complete

We have

$$\lim_{n\to\infty} \lim_{t\to x} f_n(t) = \lim_{t\to x} \lim_{n\to\infty} f_n(t)$$

Proof. Fix a sequence t_k in E that converge to x. We reduced the problem into proving

$$\lim_{n\to\infty} \lim_{k\to\infty} f_n(t_k) = \lim_{k\to\infty} \lim_{n\to\infty} f_n(t_k)$$

Set

$$a_{n,k} \triangleq f_n(t_k) \tag{2.5}$$

We then reduced the problem into proving

$$\lim_{n\to\infty} \lim_{k\to\infty} a_{n,k} = \lim_{k\to\infty} \lim_{n\to\infty} a_{n,k}$$

Set

$$A_n \triangleq \lim_{t\to x} f_n(t) \text{ and } a_{\bullet,k} \triangleq \lim_{n\to\infty} f_n(t_k)$$

We now prove

$$A_n \text{ converge}$$

Fix ϵ. We wish

$$\text{to find } N \text{ such that } d(A_n, A_m) \leq \epsilon \text{ for all } n, m > N$$

Because $a_{n,k}$ uniformly converge (to $a_{\bullet,k}$) as $n \to \infty$ by our setting, we know there exists N such that

$$\forall n, m > N, \forall k \in \mathbb{N}, d(a_{n,k}, a_{m,k}) < \frac{\epsilon}{3}$$

We claim

$$\text{such } N \text{ works}$$

Fix $n, m > N$. Because $a_{n,k} \to A_n$ and $a_{m,k} \to A_m$, we know there exists $j \in \mathbb{N}$ such that

$$d(a_{n,j}, A_n) < \frac{\epsilon}{3} \text{ and } d(a_{m,j}, A_m) < \frac{\epsilon}{3}$$

We now have

$$\begin{aligned} d(A_n, A_m) &\leq d(A_n, a_{n,j}) + d(a_{n,j}, a_{m,j}) + d(a_{m,j}, A_m) \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \text{ (done)} \end{aligned}$$

Now, because $a_{n,k} \to a_{\bullet,k}$ uniformly, by Theorem 2.6.2, we have

$$\lim_{n\to\infty} \lim_{k\to\infty} a_{n,k} = \lim_{n\to\infty} A_n = \lim_{k\to\infty} a_{\bullet,k} = \lim_{k\to\infty} \lim_{n\to\infty} a_{n,k} \text{ (done)}$$

■

In summary of Theorem 2.6.1 and Theorem 2.6.2, given a double sequence $a_{n,k}$ converging both side

(a) $a_{n,k} \to a_{\bullet,k}$ pointwise as $n \to \infty$

(b) $a_{n,k} \to a_{n,\bullet}$ pointwise as $k \to \infty$

As long as

(a) one side of convergence is uniform

(b) between two sequence $\{a_{\bullet,k}\}_{k\in\mathbb{N}}$ and $\{a_{n,\bullet}\}_{n\in\mathbb{N}}$, one of them converge, say, to A

Then the other sequence also converge, and the limit is also A.

It is at this point, we shall introduce two other terminologies. Suppose f_n is a sequence of functions from an arbitrary set X to a metric space Y. We say f_n is **pointwise Cauchy** if for all fixed $x \in X$, the sequence $f_n(x)$ is Cauchy. We say f_n is **uniformly Cauchy** if for all ϵ, there exists $N \in \mathbb{N}$ such that

$$\forall n, m > N, \forall x \in X, d\big(f_n(x), f_m(x)\big) < \epsilon$$

In last Section (Section 2.8), we define the **uniform metric** d_∞ on X^Y by

$$d_\infty(f,g) = \sup_{x\in X} d\big(f(x), g(x)\big)$$

and say that $f_n \to f$ uniformly if and only if $f_n \to f$ in (X^Y, d_∞). Similar to this clear fact, we have

$$f_n \text{ is uniformly Cauchy} \iff f_n \text{ is Cauchy in } (X^Y, d_\infty)$$

It should be very easy to verify that if f_n uniformly converge, then f_n is uniformly Cauchy, and just like sequences in metric space, the converse hold true if and only if the space $\big(X^Y, d_\infty\big)$ is complete. In Theorem 2.6.4, we give a necessary and sufficient condition for $\big(X^Y, d_\infty\big)$ to be complete.

Theorem 2.6.4. (Space of functions $\big(X^Y, d_\infty\big)$ is Complete iff Y is Complete) Given an arbitrary set X and a metric space (Y,d), we have

$$\text{the extended metric space } \big(X^Y, d_\infty\big) \text{ is complete} \iff Y \text{ is complete}$$

Proof. $(\longleftarrow)$

Suppose f_n is uniformly Cauchy. We wish

to construct a $f : X \to Y$ such that $f_n \to f$ uniformly

Because f_n is uniformly Cauchy, we know that for all $x \in X$, the sequence $f_n(x)$ is Cauchy in (Y,d). Then because Y is complete, we can define $f : X \to Y$ by

$$f(x) = \lim_{n\to\infty} f_n(x)$$

We claim

such f works, i.e. $f_n \to f$ uniformly

Fix ϵ. We wish

to find $N \in \mathbb{N}$ such that for all $n > N$ and $x \in X$ we have $d\big(f_n(x), f(x)\big) < \epsilon$

Because f_n is uniformly Cauchy, we know there exists N such that

$$\forall n,m > N, \forall x \in X, d\big(f_n(x), f_m(x)\big) < \frac{\epsilon}{2} \tag{2.6}$$

We claim

such N works

Assume there exists $n > N$ and $x \in X$ such that $d\big(f_n(x), f(x)\big) \geq \epsilon$. Because $f_k(x) \to f(x)$ as $k \to \infty$, we know

$$\exists m \in \mathbb{N}, d\big(f_m(x), f(x)\big) < \frac{\epsilon}{2} \tag{2.7}$$

Then from Equation 2.6 and Equation 2.7, we can deduce

$$\epsilon \leq d\big(f_n(x), f(x)\big) \leq d\big(f(x), f_m(x)\big) + d\big(f_n(x), f_m(x)\big) < \epsilon \text{ CaC } \text{ (done)}$$

$(\longrightarrow)$

Let K be the set of constant functions in X^Y. We first prove

$$K \text{ is closed}$$

Arbitrarily pick $f \in K^c$. We wish

$$\text{to find } \epsilon \in \mathbb{R}^+ \text{ such that } B_\epsilon(f) \in K^c$$

Because f is not a constant function, we know there exists $x_1, x_2 \in X$ such that

$$d\big(f(x_1), f(x_2)\big) > 0$$

We claim that

$$\epsilon = \frac{d\big(f(x_1), f(x_2)\big)}{3} \text{ works}$$

Arbitrarily pick $g \in B_\epsilon(f)$. We wish

$$\text{to show } g \in K^c$$

Notice the triangle inequality

$$3\epsilon = d\big(f(x_1), f(x_2)\big) \leq d\big(f(x_1), g(x_1)\big) + d\big(g(x_1), g(x_2)\big) + d\big(g(x_2), f(x_2)\big) \tag{2.8}$$

Also, because $g \in B_\epsilon(f)$, we have

$$\forall x \in X, d\big(f(x), g(x)\big) < \epsilon \tag{2.9}$$

Then by Equation 2.8 and Equation 2.9, we see

$$d\big(g(x_1), g(x_2)\big) > \epsilon$$

This then implies g is not a constant function. (done)

Now, Because by premise (X^Y, d_∞) is complete, and we have proved K is closed in (X^Y, d_∞), we know K is complete. Then, we resolve the whole problem into proving

$$Y \text{ is isometric to } K$$

Define $\sigma : Y \to K$ by

$$y \mapsto \tilde{y} \text{ where } \forall x \in X, \tilde{y}(x) = y$$

It is easy to verify σ is an isometry. (done) ■

Corollary 2.6.5. (Space of Bounded functions $\big(B(X,Y), d_\infty\big)$ is Complete iff Y is Complete)

$$\big(B(X,Y), d_\infty\big) \text{ is complete } \iff Y \text{ is complete}$$

Proof. $(\longleftarrow)$

By Theorem 2.6.4, the space (X^Y, d_∞) is complete. Then because $B(X,Y)$ is closed in (X^Y, d_∞), we know $B(X,Y)$ is complete.

$(\longrightarrow)$

Notice that the set of constant function K is a subset of the galaxy $B(X,Y)$. The whole proof in Theorem 2.6.4 works in here too. ■

Remember in the beginning of this section we say we will prove convergent sequences in Y is closed under uniform convergence if Y is complete. The proof of this result relies on Theorem 2.6.4.

Now, before we actually prove convergence sequences are closed under uniform convergence if codomain (Y, d) is complete (Theorem 2.6.7), we will state and prove Weierstrass M-test (Theorem 2.6.6), which concerns the uniform convergence of series of complex functions.

Theorem 2.6.6. (Weierstrass M-test) Given sequences $f_n : X \to \mathbb{C}$, and suppose

$$\forall n \in \mathbb{N}, \forall x \in X, |f_n(x)| \leq M_n \tag{2.10}$$

Then

$$\sum_{n=1}^{\infty} M_n \text{ converge } \implies \sum_{n=1}^{\infty} f_n \text{ uniformly converge}$$

Proof. Because $(\mathbb{C}, \|\cdot\|_2)$ is complete, by Corollary 2.6.5, we only wish to prove

$$\Big\{\sum_{k=1}^{n} f_k\Big\}_{n\in\mathbb{N}} \text{ is uniformly Cauchy}$$

Fix ϵ. We wish

$$\text{to find } N \text{ such that } \forall n,m > N, \forall x \in X, \left|\sum_{k=n}^{m} f_k(x)\right| < \epsilon$$

Because $\sum_{n=1}^{\infty} M_n$ converge, we know there exists N such that

$$\forall n,m > N, \sum_{k=n}^{m} M_k < \epsilon$$

We claim

$$\text{such } N \text{ works}$$

By Premise 2.10, we have

$$\forall n,m > N, \forall x \in X, \left|\sum_{k=n}^{m} f_k(x)\right| \leq \sum_{k=n}^{m} |f_k(x)| \leq \sum_{k=n}^{m} M_k < \epsilon$$

■

Theorem 2.6.7. (Convergent Sequences are Closed under Uniform Convergence if Codomain (Y,d) is Complete) Given a complete metric space (Y,d), let $\mathcal{C}_{\mathbb{N}}^{Y}$ be the set of convergent sequences in Y.

$$Y \text{ is complete } \implies \mathcal{C}_{\mathbb{N}}^{Y} \text{ is closed under uniform convergent}$$

Proof. Let $a_{n,k} \to a_{\bullet,k}$ uniformly as $n \to \infty$ where for all $n,k \in \mathbb{N}, a_{n,k} \in Y$ and let $A_n = \lim_{k\to\infty} a_{n,k}$ for all $n \in \mathbb{N}$.

$$\text{to prove } a_{\bullet,k} \text{ converge}$$

By Theorem 2.6.2, we can reduce the problem to

$$\text{proving } A_n \text{ converge}$$

Then because Y is complete, we can then reduce the problem into proving

$$A_n \text{ is Cauchy}$$

Fix ϵ. We wish to find N such that

$$\forall n, m > N, d(A_n, A_m) < \epsilon$$

Because $a_{n,k} \to a_{\bullet,k}$ uniformly, we can find N such that

$$\forall n, m > N, d_\infty(\{a_{n,k}\}_{k\in\mathbb{N}}, \{a_{m,k}\}_{k\in\mathbb{N}}) < \frac{\epsilon}{3} \tag{2.11}$$

We claim

$$\text{such } N \text{ works}$$

Arbitrarily pick $n, m > N$. We wish to prove

$$d(A_n, A_m) < \epsilon$$

Because $a_{n,k} \to A_n$ and $a_{m,k} \to A_m$ as $k \to \infty$, we can find j such that

$$d(a_{n,j}, A_n) < \frac{\epsilon}{3} \text{ and } d(a_{m,j}, A_m) < \frac{\epsilon}{3} \tag{2.12}$$

Then from Equation 2.11 and Equation 2.12, we can deduce

$$d(A_n, A_m) \leq d(A_n, a_{n,j}) + d(a_{n,j}, a_{m,j}) + d(a_{m,j}, A_m) < \epsilon \text{ (done)}$$

■

2.7 Closed under Uniform Convergence

The end goal for this section is to prove that the following properties

(a) continuity

(b) uniform continuity

(c) K-Lipschitz Continuity

Theorem 2.7.1. (Uniform Limit Theorem) Given a sequence of function f_n from a topological space (X, τ) to a metric space (Y, d), suppose

(a) $f_n \to f$ uniformly as $n \to \infty$

(b) f_n is continuous for all $n \in \mathbb{N}$

Then f is also continuous.

Proof. Fix $x \in X$, and let $x_k \to x$. We wish to prove

$$f(x_k) \to f(x)$$

Because $f_n \to f$ uniformly as $n \to \infty$, we know

$$\Big\{f_n(x_k)\Big\}_{k\in\mathbb{N}} \to \Big\{f(x_k)\Big\}_{k\in\mathbb{N}} \text{ uniformly as } n \to \infty \tag{2.13}$$

Also, because for each $n \in \mathbb{N}$, the function f_n is continuous at x, we know

$$\forall n \in \mathbb{N}, f_n(x_k) \to f_n(x) \text{ as } k \to \infty \tag{2.14}$$

Then because $f_n \to f$ pointwise, we know

$$f_n(x) \to f(x) \tag{2.15}$$

Now, because Equation 2.13, Equation 2.14 and Equation 2.15, by Theorem 2.6.1, we have

$$\lim_{k\to\infty} f(x_k) = \lim_{k\to\infty}\lim_{n\to\infty} f_n(x_k) = \lim_{n\to\infty}\lim_{k\to\infty} f_n(x_k) = \lim_{n\to\infty} f_n(x) = f(x) \text{ (done)}$$

■

Suppose X is a compact Hausdroff space, with Theorem **??**, we can now say that the set $\mathcal{C}(X)$ of complex-valued continuous functions on X

Theorem 2.7.2. (Uniformly Continuous functions are Closed under Uniform Convergence) Given a sequence of functions f_n from a metric space (X, d_X) to metric space (Y, d_Y), suppose

(a) $f_n \to f$ uniformly

(b) f_n is uniformly continuous for all $n \in \mathbb{N}$

Then f is also uniformly continuous

Proof. Fix ϵ. We wish

$$\text{to find } \delta \text{ such that } \forall x, y \in X, d_X(x,y) < \delta \implies d_Y\big(f(x), f(y)\big) < \epsilon$$

Because $f_n \to f$ uniformly, we know there exists $m \in \mathbb{N}$ such that

$$\forall x \in X, d_Y\big(f_m(x), f(x)\big) < \frac{\epsilon}{3} \tag{2.16}$$

Because f_m is uniformly continuous, we know

$$\exists \delta, \forall x, y \in X, d_X(x,y) < \delta \implies d_Y\big(f_m(x), f_m(y)\big) < \frac{\epsilon}{3} \tag{2.17}$$

We claim

$$\text{such } \delta \text{ works}$$

Let $x, y \in X$ satisfy $d_X(x,y) < \delta$. We wish

$$\text{to prove } d_Y\big(f(x), f(y)\big) < \epsilon$$

From Equation 2.16 and Equation 2.17, we have

$$d_Y\big(f(x), f(y)\big) \leq d_Y\big(f(x), f_m(x)\big) + d_Y\big(f_m(x), f_m(y)\big) + d_Y\big(f_m(y), f(y)\big) = \epsilon \text{ (done)}$$

■

Theorem 2.7.3. (K-Lipschitz functions are Closed under Uniform Convergence) Given a sequence of functions f_n from metric space (X, d_X) to metric space (Y, d_Y), suppose

(a) $f_n \to f$ uniformly as $n \to \infty$

(b) f_n is K-Lipschtize continuous for all $n \in \mathbb{N}$

Then f is also K-Lipschtize continuous.

Proof. Arbitrarily pick $x, y \in X$, to show f is K-Lipschtize continuous, we wish

$$\text{to show } d_Y\big(f(x), f(y)\big) \leq K d_X(x,y)$$

Fix ϵ. We reduce the problem into proving

$$d_Y\big(f(x), f(y)\big) < K d_X(x,y) + \epsilon$$

Because $f_n \to f$ uniformly as $n \to \infty$, we know there exists m such that

$$\forall z \in X, d_Y\big(f(z), f_m(z)\big) < \frac{\epsilon}{2} \tag{2.18}$$

Because f_m is K-Lispchitz continuous, we know

$$d_Y\big(f_m(x), f_m(y)\big) \leq K d_X(x, y) \tag{2.19}$$

Now, from Equation 2.19 and Equation 2.18, we now see

$$d_Y\big(f(x), f(y)\big) \leq d_Y\big(f(x), f_m(x)\big) + d_Y\big(f_m(x), f_m(y)\big) + d_Y\big(f_m(y), f(y)\big) < K d_X(x, y) + \epsilon$$

$\blacksquare$

An example of sequences of Lipschitz continuous functions with unbounded Lipschitz constant can uniformly converge to a non-Lipschitz continuous function is given below

Example 1 (Lipschitz functions with Unbounded Lipschitz constant Uniformly Converge to a non-Lipschitz function)

$$X = [0, 1] \text{ and } f_n(x) = \sqrt{x + \frac{1}{n}}$$

2.8 Modes of Convergence

This section is the starting point for us to study spaces of function. At first, we will define two modes of convergence for sequence of function and point out some basic properties and the difference between two modes of convergence.

Given an arbitrary set X and a metric space Y, we say a sequence of functions f_n from X to Y **pointwise converge** to f if for all ϵ and x in X, there exists N such that

$$\forall n > N, f_n(x) \in B_\epsilon\big(f(x)\big)$$

In other words, for each fixed x in X, we have $f_n(x) \to f(x)$.

We say f_n **uniformly converge** to f if for all ϵ there exists N such that

$$\forall x \in X, \forall n > N, f_n(x) \in B_\epsilon\big(f(x)\big)$$

The difference between pointwise convergence and uniform convergence is that if we require $f_n(x)$ to be ϵ-close to $f(x)$ for all $n > N$, then

(•) N depend on both ϵ and x if $f_n \to f$ pointwise

(•) N depend on only ϵ if $f_n \to f$ uniformly

A few properties of sequence of functions similar to that of sequences in metric space is obvious. If $f_n \to f$ pointwise, then all sub-sequences $f_{n_k} \to f$ pointwise. If $f_n \to f$ uniformly, then all sub-sequences $f_{n_k} \to f$ uniformly. Suppose $Z \subseteq X$. It is clear that if $f_n \to f$ uniformly (resp: pointwise) the restricts $f_n|_Z \to f|_Z$ uniformly (resp: pointwise). Also, if $f_n \to f$ uniformly, then $f_n \to f$ pointwise.

Suppose we have a family $\mathcal{F}$ of functions $f : X \to (Y, d)$. If we define

$$d_\infty(f, g) = \sup_{x \in X} d\big(f(x), g(x)\big)$$

instead of a metric, d_∞ become an extended metric. If f is bounded and g is unbouneded, we have $d_\infty(f, g) = \infty$. If f, g are both bounded, then $d_\infty(f, g) \in \mathbb{R}^+$. Because of such, for d_∞ to be a metric, one but not the only condition is for $\mathcal{F}$ to be space of bounded functions.

Now, regardless of d_∞ is an extended metric or not, we have

$$f_n \to f \text{ uniformly} \iff d_\infty(f_n, f) \to 0$$

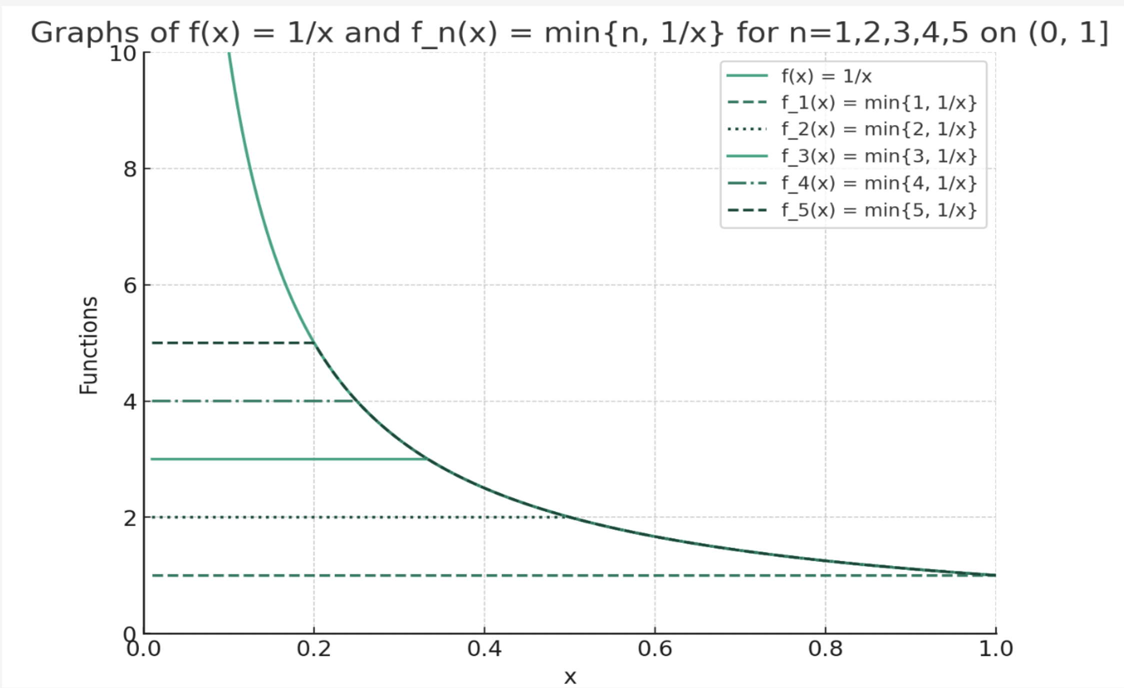
With this in mind, it shall be clear that the uniform limit of bounded (resp: unbounded) functions is always bounded (resp: unbounded).

Examples for bounded (resp: unbounded) function f_n pointwise converge to unbounded (resp: bounded) function f are as follows.

Example 2 (Bounded functions pointwise converge to unbounded function)

$$X = (0, 1], f_n(x) = \min\{n, \frac{1}{x}\}$$

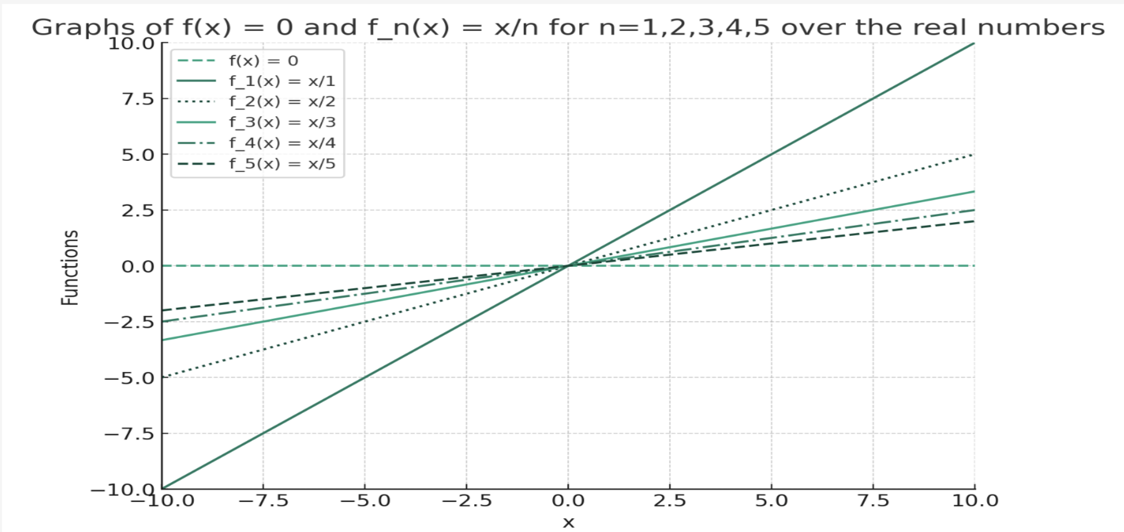
It is clear that $\forall n \in \mathbb{N}, f_n(x) \in [0, n]$, and the limit $f : X \to \mathbb{R}$ is $f(x) = \frac{1}{x}$



Example 3 (Unbounded functions pointwise converge to bounded function)

$$X = \mathbb{R}, f_n(x) = \frac{1}{n}x$$

The limit function is $f(x) = 0$



As pointed out earlier, if $f : X \to (Y, d)$ is bounded and $g : X \to (Y, d)$ is unbounded, then $d_\infty(f, g) = \infty$. This means that if Y is unbounded, the uniform metric d_∞ is extended on X^Y. For this, it is necessary to develop some basic fact concerning extended metric space.

Suppose (X, d) is an extended metric space. If we define $\sim$ on X by $x \sim y \iff d(x, y) < \infty$, then $\sim$ is an equivalence relation. We say each equivalence class is a **galaxy** of (X, d). Suppose T is the collection of the galaxies of (X, d). For each $\mathcal{T} \in T$, the space $(\mathcal{T}, d)$ is just a metric space.

It is easy to see that the way we induce topology from metric space is still valid if the metric is extended. That is

$$\tau = \{Z \in X : \forall z \in Z, \exists \epsilon, B_\epsilon(z) \subseteq Z\}$$

is still a topology, even though d is an extended metric on X.

We can verify that a set Y in X is open if and only if for all $\mathcal{T} \in T$, the set $Y \cap \mathcal{T}$ is open, and the set Y in X is closed if and only if all convergent sequences y_n in Y

converge to points in Y.

Now, suppose we are given an arbitrary set X and a complete metric space $(\overline{Y}, d)$, and on $X^{\overline{Y}}$, we define the uniform metric d_∞. We say a set $\mathcal{F} \subseteq X^{\overline{Y}}$ of functions is **closed under uniform convergence** if for all uniform convergent sequence $f_n \subseteq \mathcal{F}$, the limit function f is also in $\mathcal{F}$. There are justified reasons for us to give the premise that $\overline{Y}$ is complete prior to the definition of the term **closed under uniform convergence**. One reason is that by Theorem 2.6.4, if Y is not complete, then the extended metric space (X^Y, d_∞) is also not complete, which implies the possibility a Cauchy sequence f_n in X^Y converge to a function $f \in X^{\overline{Y}} \setminus X^Y$ where $\overline{Y}$ is the completion of Y. For instance, if we let $Y = \mathbb{R} \setminus \{1\}$ where $X = \mathbb{R}$, and let $f_n(x) = \begin{cases} 0 & \text{if } x \neq 0 \\ 1 + \frac{1}{n} & \text{if } x = 0 \end{cases} \in Y$, we see that the set $\mathcal{F} = \{f_n : n \in \mathbb{N}\}$ is "closed under uniform convergence" in the context of X^Y, but when in fact f_n uniformly converge to $f(x) = \begin{cases} 0 & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}$ which is not in $\mathcal{F}$. This awkward usage of words can be solved if we define the term **closed under uniform convergence** after the premise that Y is complete.

Now, given a set of functions $\mathcal{F} \subseteq X^{\overline{Y}}$, one can verify that

$$\begin{aligned} \mathcal{F} \text{ is closed under uniform convergence } &\iff (\mathcal{F}, d_\infty) \text{ is complete} \\ &\iff \mathcal{F} \text{ is closed with respect to } (X^{\overline{Y}}, d_\infty) \end{aligned}$$

Let $\mathcal{G}$ be a galaxy of $(X^{\overline{Y}}, d_\infty)$. With multiple ways, we can verify that $\mathcal{G}$ is closed with respect to $(X^{\overline{Y}}, d_\infty)$. Then, acknowledging the space of bounded functions $B(X, \overline{Y})$ is a galaxy of $X^{\overline{Y}}$, we see that $B(X, \overline{Y})$ is closed under uniform convergence. The statement that $B(X, \overline{Y})$ is closed under uniform convergence, although already "proved" before as we pointed out the limit of uniform convergent sequence of bounded functions must be bounded, is now in fact actually proved in the sense the term "closed under uniform convergence" is formally given a satisfying definition.

2.9 Arzelà–Ascoli Theorem

In this section, we will give a complete proof of Arzelà–Ascoli Theorem for functions from arbitrary compact topological space to arbitrary metric space. Note that in Baby Rudin, Arzelà–Ascoli Theorem are given for functions from compact metric space to metric space. Because Arzelà–Ascoli Theorem are concerned with family of equicontinuos functions, it is crucial for us to give a definition to equicontinuity for functions from topological space to metric space, for the sake of our generalization.

Let X, Y be metric space. Let Z be topological space. Let $\mathcal{F}_X$ be family of functions from X to Y, and let $\mathcal{F}_Z$ be family of functions from Z to Y. We say $\mathcal{F}_Z$ is **pointwise equicontinuous** if

$$\text{For all } \epsilon \text{ and for all } x, \text{ there exists a neighborhood } U_x \text{ such that } d_Y(f(x), f(y)) < \epsilon \text{ for all } y \in U_x$$

We say $\mathcal{F}_X$ is **equicontinuous** if

$$\text{For all } \epsilon, \text{ there exists } \delta \text{ such that } d_Y(f(x), f(y)) < \epsilon \text{ for all } \delta\text{-close } x, y \in X \text{ and all } f \in \mathcal{F}.$$

It is easy to verify that if $\mathcal{F}_X$ is equicontinuous, then $\mathcal{F}_X$ is pointwise equicontinuous. The converse don't always hold true. Say, $\mathcal{F} = \{n + x^2\}_{n\in\mathbb{N}}$, the set $\{n + x^2\}_{n\in\mathbb{N}}$ is clearly pointwise equicontinuous on $\mathbb{R}$, and is not equicontinuous on $\mathbb{R}$, since no function $n + x^2$ is uniform continuous on $\mathbb{R}$. However, the same set $\mathcal{F} = \{n + x^2\}$ is equicontinous on compact domain $[a, b]$. This is a general result, as we shall prove below.

Theorem 2.9.1. (Pointwise Equicontinous is Uniform on Compact Domain) Given two metric space $(X, d_X), (Y, d_Y)$, and a family $\mathcal{F}$ of functions from X to Y such that

(a) X is compact

(b) $\mathcal{F}$ is pointwise equicontinuous

Then

$$\mathcal{F} \text{ is equicontinuous}$$

Proof. Fix ϵ. We wish to

$$\text{find } \delta \text{ such that } d_X(x, y) < \delta \implies d_Y(f(x), f(y)) \leq \epsilon \text{ for all } f \in \mathcal{F}$$

Because $\mathcal{F}$ is pointwise equicontinuous, we know for each $x \in X$, there exists δ_x such that

$$\forall y \in B_{\delta_x}(x), d_Y\big(f(x), f(y)\big) < \frac{\epsilon}{2} \text{ for all } f \in \mathcal{F} \tag{2.20}$$

It is clear that $\{B_{\frac{\delta_x}{2}}(x) : x \in X\}$ form an open cover of X. Then because X is compact, we know

$$\text{there exists a finite open sub-cover: } \{B_{\frac{\delta_x}{2}}(x) : x \in X_{\text{finite}}\}$$

We claim

$$\delta = \min_{x \in X_{\text{finite}}} \frac{\delta_x}{2} \text{ works}$$

Fix $y, z \in X : d_X(y, z) < \delta$. We have to prove

$$d_Y\big(f(y), f(z)\big) < \epsilon$$

We know y must lie in some $B_{\frac{\delta_x}{2}}(x)$ for some $x \in X_{\text{finite}}$. Because $d_X(y, z) < \frac{\delta_x}{2}$, we see that z must lie in $B_{\delta_x}(x)$. We now know y, z are both in $B_{\delta_x}(x)$. Then from (2.20), we can now deduce

$$d_Y\big(f(y), f(z)\big) \le d_Y\big(f(y), f(x)\big) + d_Y\big(f(x), f(z)\big) < \epsilon \text{ (done)}$$

■

The proof above should be a great example why in the discussion of metric space, instead of using sequential definition of compactness, which leads to the beautiful Bolzano-Weierstrass Theorem, some people prefer the open-cover definitions.

Now, we give proof for the Arzelà–Ascoli Theorem.

Theorem 2.9.2. (Arzelà–Ascoli Theorem) Given a compact topological space (X, τ), a metric space (Y, d_Y), and a family $\mathcal{F} \subseteq C\big(X, Y\big)$ of continuous function

$\mathcal{F}$ is pointwise equicontinuous and $\{f(x) : f \in \mathcal{F}\}$ has compact closure in Y for all $x \in X$
$\implies \mathcal{F}$ has a compact closure in $C(X, Y)$

Proof. Fix a sequence f_n in $\mathcal{F}$. We wish to show

$$f_n \text{ has a sub-sequence } f_{n_k} \text{ uniformly converge to some } f : X \to Y$$

First, we prove

$$\text{there exists a countable set } P \text{ such that } P \text{ works like a dense set}$$

Because $\mathcal{F}$ is pointwise equicontinuous, we know for all $x \in X$

$$\exists U_{x,n}, \forall y \in U_{x,n}, \forall f \in \mathcal{F}, d_Y\big(f(x), f(y)\big) < \frac{1}{n} \text{ for each fixed } n \in \mathbb{N}$$

Now, because X is compact, for each $n \in \mathbb{N}$, there exists a finite subset $P_n \subseteq X$ such that $\left\{U_{x,n} : x \in P_n\right\}$ is a cover of X. Let $P = \bigcup_{n\in\mathbb{N}} P_n$. (done)

Now, we wish to

construct a sub-sequence f_{n_k} pointwise converge on P

Express $P = \{p_k\}_{k\in\mathbb{N}}$. By premise (pointwise image has compact closure), we know there exists a compact set that contain $\{f_n(p_1)\}_{n\in\mathbb{N}}$, so by Bolzano-Weierstrass Theorem, there exists a sub-sequence

$$\left\{f_{g_1(k)}(p_1)\right\}_{k\in\mathbb{N}} \text{ converge to some point in } Y$$

Now, again by premise and Bolzano-Weierstrass Theorem, there exists a sub-sequence

$$\left\{f_{g_2\circ g_1(k)}(p_2)\right\}_{k\in\mathbb{N}} \text{ converge to some point in } Y$$

Repeatedly doing such, we have

$$\begin{array}{cccc}
f_{g_1(1)}(p_1) & f_{g_2\circ g_1(1)}(p_2) & f_{g_3\circ g_2\circ g_1(1)}(p_3) & \cdots \\
f_{g_1(2)}(p_1) & f_{g_2\circ g_1(2)}(p_2) & f_{g_3\circ g_2\circ g_1(2)}(p_3) & \cdots \\
f_{g_1(3)}(p_1) & f_{g_2\circ g_1(3)}(p_2) & f_{g_3\circ g_2\circ g_1(3)}(p_3) & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
\downarrow & \downarrow & \downarrow & \\
y_1 & y_2 & y_3 & \cdots
\end{array}$$

Now, let

$$n_k = g_k \circ \cdots \circ g_1(k)$$

Then

$$n_k \text{ is eventually a sub-sequence of } g_m \circ \cdots \circ g_1(k) \text{ for all } m$$

This then implies

$$f_{n_k}(p_m) \to y_m \text{ for all } p_m \in P \text{ (done)}$$

Next, we show

To prove f_{n_k} uniformly converge on X, it suffice to prove f_{n_k} is uniformly Cauchy on X.

By premise (pointwise image has compact closure), if f_{n_k} is uniformly Cauchy, then we know f_{n_k} pointwise converge to some f.

Fix ϵ. We reduced the problem into

$$\text{finding } N \text{ such that for all } k > N, \text{ we have } d_Y\big(f_{n_k}(x), f(x)\big) \leq \epsilon \text{ for all } x \in X$$

Because f_{n_k} is uniformly Cauchy, we know there exists N such that for all $m, k > M$ $d_Y\big(f_{n_k}(x), f_{n_m}(x)\big) \leq \frac{\epsilon}{2}$ for all $x \in X$. We claim

$$\text{such } N \text{ works}$$

Let $k > N$. Assume $d_Y\big(f_{n_k}(x), f(x)\big) > \epsilon$. We see that

$$d_Y\big(f(x), f_{n_m}(x)\big) \geq d_Y\big(f(x), f_{n_k}(x)\big) - d_Y\big(f_{n_k}(x), f_{n_m}(x)\big) > \frac{\epsilon}{2} \text{ for all } m > N \text{ CaC} \quad \text{(done)}$$

Lastly, we wish to prove

$$f_{n_k} \text{ is uniformly Cauchy}$$

Fix ϵ. We wish

$$\text{to find } N \text{ such that } \forall j, k > N, \forall x \in X, d_Y\big(f_{n_j}(x), f_{n_k}(x)\big) \leq \epsilon$$

Fix $m > \frac{3}{\epsilon}$. Express $P_m = \{p_1^m, \dots, p_u^m\}$. Because $f_{n_k}(p_t^m)$ converge for each $t \in \{1, \dots, u\}$, we know

$$\forall t, \exists N_t, d_Y\big(f_{n_j}(p_t^m), f_{n_k}(p_t^m)\big) < \frac{\epsilon}{3} \text{ for all } j, k > N_t$$

We claim

$$N = \max_t N_t \text{ works}$$

Fix $j, k > N$ and $x \in X$. We have to show

$$d_Y\big(f_{n_j}(x), f_{n_k}(x)\big) \leq \epsilon$$

Because $\{U_{p_t^m,m}\}$ form an open cover of X, we know there exists t such that $x \in U_{p_t^m,m}$. We can now deduce

$$d_Y\big(f_{n_j}(x), f_{n_k}(x)\big) \leq d_Y\big(f_{n_j}(x), f_{n_j}(p_t^m)\big) + d_Y\big(f_{n_j}(p_t^m), f_{n_k}(p_t^m)\big) + d_Y\big(f_{n_k}(p_t^m), f_{n_k}(x)\big) < \epsilon$$

(done)

$\blacksquare$

2.10 Banach Fixed Point Theorem

This section give a complete statement and proof of Banach Fixed Point Theorem. Given two metric spaces $(X, d_X), (Y, d_Y)$ and a function $f : X \to Y$, we say f is a **contraction** if there exists $r \in [0, 1)$ such that

$$d_Y(f(x), f(y)) \leq r d_X(x, y) \qquad (x, y \in X)$$

or equivalently

$$\sup_{x \neq y \in X} \frac{d_Y(f(x), f(y))}{d_X(x, y)} < 1$$

Note that the restriction of a contraction is again a contraction.

Given a function $g : X \to X$, we say g admits a **fixed point** x if we have

$$g(x) = x$$

Theorem 2.10.1. (Banach Fixed Point Theorem) Given a metric space (X, d), if f is a contraction that maps X into X, then

$$f \text{ admits at most one fixed point}$$

Moreover, if (X, d) is complete, then

$$\text{the fixed point exists}$$

And if we use the notation f^n to denote $f \circ f^{n-1}$, then for all $x \in X$,

$$\text{the fixed point can be written in the form } \lim_{n \to \infty} f^n(x)$$

Proof. We first

prove the uniqueness of the fixed point

Suppose x, y are both fixed by f. We have

$$d(f(x), f(y)) = d(x, y)$$

Because f is a contraction mapping, this implies $d(x, y) = 0$. (done)

Suppose (X, d) is complete. We now

prove the existence of the fixed point

Fix $x \in X$. Because we have already prove the uniqueness of the fixed point, we only have to prove

$$\lim_{n\to\infty} f^n(x) \text{ exists and } \lim_{n\to\infty} f^n(x) \text{ is a fixed point of } f$$

Because (X, d) is complete, to prove $\lim_{n\to\infty} f^n(x)$ exists, we only have to prove

$$\{f^n(x)\}_{n\in\mathbb{N}} \text{ is Cauchy}$$

Observe

$$\begin{aligned} d(f^n(x), f^{n+k}(x)) &\leq \sum_{i=0}^{k-1} d(f^{n+i}(x), f^{n+i+1}(x)) \\ &\leq d(x, f(x)) \sum_{i=0}^{k-1} r^{n+i} \\ &\leq \frac{r^n}{1-r} d(x, f(x)) \to 0 \text{ as } n \to \infty \quad \text{(done)} \end{aligned}$$

Note that contraction is Lipschitz thus continuous. This allow us to carry the below limit process

$$f\big(\lim_{n\to\infty} f^n(x)\big) = \lim_{n\to\infty} f(f^n(x)) = \lim_{n\to\infty} f^{n+1}(x) = \lim_{n\to\infty} f^n(x) \text{ (done)}$$

■

Banach Fixed Point Theorem is one of the most important Theorem in Mathematics. It will be used to prove

(a) Inverse Function Theorem (Theorem 4.6.1)

(b) Picard-Lindelof Theorem

(c) Nash-Embedding Theorem

Chapter 3

Calculus

3.1 Equivalent Definitions for Riemann Integral

In this section, we will give a principal Theorem (Theorem 3.1.1) that can serve as a lemma to prove the equivalency of multiple different definitions of Riemann integral on a compact interval. With this approach, we shall diminish the trouble of getting through miscellaneous minor definitions, where they are all equivalent, with only the difference of taking different tags and partitions of certain pattern, which solely serve as a pedagogical tool to give students a concrete idea of integration.

A caveat will be made clear here: this section concern only the proper Riemann Integral. That is, we only consider the integration of a function bounded on a compact interval. For a treatment of inproper integral, see Section **??**.

In this section, by a **partition** P of $[a,b]$, we mean a finite set of values $P = \{a = x_0 \leq x_1 \leq \cdots \leq x_{n_P-1} \leq x_{n_P} = b\}$. We say the partition P' is **finer** than P if $P \subseteq P'$. Given a partition P, we put

$$\begin{cases} M_i = \sup_{[x_{i-1},x_i]} f(x) \\ m_i = \inf_{[x_{i-1},x_i]} f(x) \end{cases} \quad \text{and} \quad \begin{cases} U(P,f) = \sum_{i=1}^{n_P} M_i \Delta x_i \\ L(P,f) = \sum_{i=1}^{n_P} m_i \Delta x_i \end{cases} \quad \text{where } \Delta x_i = x_i - x_{i-1}$$

We shall write n instead of n_P if no confusion will be made.

The word **norm of partition** $\|P\|$ is defined by $\max_{1\leq i\leq n} \Delta x_i$. We say $U(P,f)$ is an **upper sum** of f. We say the **upper integral** $\overline{\int_a^b} f\,dx$ of f on $[a,b]$ is $\inf_P U(P,f)$ where the infimum run through all partitions P of $[a,b]$. The **lower integral** $\underline{\int_a^b} f\,dx$ is defined similarly. We say a function f is **integrable** on $[a,b]$ if $\overline{\int_a^b} f dx = \underline{\int_a^b} f dx$.

Give close attention to the setting that P is finite. This is crucial for making the integration operation possible, since if P is countable and we define $U(P,f)$ by taking limits for sums, the order of addition can make a difference if the sum does not converge absolutely. This fact is backed by Riemann Rearrangement Theorem (Theorem 3.1.7), of which we will later give a proof.

Theorem 3.1.1. (Principal for Proving Equivalency of Definitions for Riemann Integral)

$$\int_a^b f dx \in \mathbb{R} \iff \forall \{P_k\} : \|P_k\| \to 0, U(P_k,f) - L(P_k,f) \to 0$$

Proof. From right to left is obvious. We prove only

$$\int_a^b f dx \in \mathbb{R} \implies \forall \{P_k\} : \|P_k\| \to 0, U(P_k,f) - L(P_k,f) \to 0$$

Fix ϵ. We wish to find a positive number $\beta \in \mathbb{R}^+$ such that $\forall P : \|P\| \leq \beta, U(P,f) - L(P,f) < \epsilon$. Because $\int_a^b f dx \in \mathbb{R}$, we can let W be a partition such that

$$U(W,f) - L(W,f) < \frac{\epsilon}{2}$$

Let $W = \{a = x_0^*, x_1^*, \dots, x_{n_W}^* = b\}$, and let $J = \{1, \dots, n_W\}$ be the set of indices of W. Suppose

$$L = \max_{1 \leq j \leq n_W - 1} \Big(\sup_{[x_{j-1}^*, x_{j+1}^*]} f(x) - \inf_{[x_{j-1}^*, x_{j+1}^*]} f(x) \Big) \tag{3.1}$$

Notice that if $L = 0$, then f must be constant and the proof become trivial, so we can assume $L > 0$. We claim that

$$L\beta n_W \leq \frac{\epsilon}{2} \text{ and } \beta < \min_{j \in J} \Delta x_j$$

suffice so that $\forall P : \|P\| \leq \beta, U(P,f) - L(P,f) < \epsilon$. Let $C = \min_{j \in J} \Delta x_j$. In other words, we now reduce the problem into proving

$$\|P\| \leq \min\{\frac{\epsilon}{2Ln_W}, C\} \implies U(P,f) - L(P,f) < \epsilon$$

Let $I = \{1, \dots, n_P\}$ be the set of indices for P. Suppose

$$P = \{a = x_0, x_1, \dots, x_{n_P} = b\}$$

We partition I into

$$\begin{cases} A = \left\{ i \in I : \exists j \in J, [x_{i-1}, x_i] \subseteq [x^*_{j-1}, x^*_j] \right\} \\ B = I \setminus A \end{cases}$$

We now have

$$U(P,f) - L(P,f) = \sum_{i \in A} \big(M_i - m_i\big)\Delta x_i + \sum_{i \in B} \big(M_i - m_i\big)\Delta x_i \tag{3.2}$$

Because for each $i \in A$, there is a unique corresponding $j \in J$ such that $[x_{i-1}, x_i] \subseteq [x^*_{j-1}, x^*_j]$, we have

$$\sum_{i \in A} \big(M_i - m_i\big)\Delta x_i \leq \sum_{j \in J} \big(M^W_j - m^W_j\big)\Delta x^*_j = U(W,f) - L(W,f) < \frac{\epsilon}{2} \tag{3.3}$$

Because $\|P\| \leq C = \min_{j \in J} \Delta x_j$, we know for each distinct $i \in B$, there exists a distinct $j \in J$ such that $[x_{i-1}, x_i] \subseteq [x^*_{j-1}, x^*_{j+1}]$, so by definition of L (Equation 3.1), we have

$$\sum_{i \in B} \big(M_i - m_i\big)\Delta x_i \leq L \sum_{i \in B} \Delta x_i \leq L n_W \|P\| \leq \frac{\epsilon}{2} \tag{3.4}$$

Combining Equation 3.2, Equation 3.3 and Equation 3.4, we now see

$$U(P,f) - L(P,f) < \epsilon \text{ (done)}$$

■

Recall that we say a series $\sum_{n=1}^{\infty} a_n$ **absolutely converge** if $\sum_{n=1}^{\infty} |a_n|$ converge. We can show that a series converges if it absolutely converges by proving it is Cauchy. In this section, by a **permutation on** $\mathbb{N}$, we mean a bijective function σ from $\mathbb{N}$ to $\mathbb{N}$. Another two important terminologies are the followings. We say that $\sum_{n=1}^{\infty} a_n$ **unconditionally converge** if for all permutation $\sigma : \mathbb{N} \to \mathbb{N}$, the series $\sum_{n=1}^{\infty} a_{\sigma(n)}$ converge to the same number. We say $\sum_{n=1}^{\infty} a_n$ **conditionally converge** if it converge but not unconditionally.

In our treatment, Riemann Rearrangement Theorem will be split into 4 parts. The summary is at Theorem 3.1.7. The first part (Theorem 3.1.2) states that the limit an absolutely convergent series remain the same under any permutations. The proof for the first part (Theorem 3.1.2) may seem technical, but the essence is quite easy to remember. Just "see" that

$$\left|\sum_{k=1}^{\infty} a_{\sigma(k)} - L\right| \leq \left|\sum_{i<M} a_i - L + \sum_{i\geq M} a_i\right|$$
$$\leq \left|\sum_{i<M} a_i - L\right| + \sum_{i\geq M} |a_i| \to 0 \text{ as } M \to 0$$

Theorem 3.1.2. (Riemann Rearrangement Theorem, Part 1)

$$\sum_{k=1}^{\infty} a_k \text{ absolutely converge } \implies \sum_{k=1}^{\infty} a_k \text{ unconditionally converge}$$

Proof. Suppose $\sum_{k=1}^{\infty} |x_k|$ converge. Let $\sum_{k=1}^{\infty} x_k = L$. Fix permutation $\sigma : \mathbb{N} \to \mathbb{N}$. We wish to prove

$$\sum_{k=1}^{\infty} x_{\sigma(k)} = L$$

Fix ϵ. We reduce the problem into

$$\text{finding } N \text{ such that } \forall n > N, \left|\sum_{k=1}^{n} x_{\sigma(k)} - L\right| < \epsilon$$

Because both $\sum_{k=1}^{\infty} x_k$ and $\sum_{k=1}^{\infty} |x_k|$ converge by premise. We know there exists M such that

$$\forall n > M, \left|\sum_{k=1}^{n} x_k - L\right| < \frac{\epsilon}{2} \text{ and } \sum_{k=n}^{\infty} |x_k| < \frac{\epsilon}{2} \tag{3.5}$$

Let

$$I = \sigma^{-1}\big(\{1, \dots, M\}\big) \text{ and } N = \max I$$

We claim

$$\text{such } N \text{ works}$$

To prove our claim, fix $n > N$. We wish to show

$$\left|\sum_{k=1}^{n} x_{\sigma(k)} - L\right| < \epsilon$$

Let $I_n = \{1, \ldots, n\}$. Observe that

$$\left|\sum_{k=1}^{n} x_{\sigma(k)} - L\right| = \left|\sum_{k \in I_n \setminus I} x_{\sigma(k)} + \sum_{k \in I} x_{\sigma(k)} - L\right| \tag{3.6}$$

Notice that $k \notin I \implies \sigma(k) > M$. Then by definition of M (Equation 3.5), we have

$$\left|\sum_{k \in I_n \setminus I} x_{\sigma(k)}\right| \leq \sum_{k \in I_n \setminus I} \left|x_{\sigma(k)}\right| \leq \sum_{j > M}^{\infty} |x_j| < \frac{\epsilon}{2} \tag{3.7}$$

Notice that $\sigma[I] = \{1, \ldots, M\}$. Then also by definition of M (Equation 3.5), we have

$$\left|\sum_{k \in I} x_{\sigma(k)} - L\right| = \left|\sum_{j=1}^{M} x_j - L\right| < \frac{\epsilon}{2} \tag{3.8}$$

Then by inequalities Equation 3.6, Equation 3.7 and Equation 3.8, we now have

$$\begin{aligned}
\left|\sum_{k=1}^{n} a_{\sigma(k)} - L\right| &= \left|\sum_{k \in I_n \setminus I} x_{\sigma(k)} + \sum_{k \in I} x_{\sigma(k)} - L\right| \\
&\leq \left|\sum_{k \in I_n \setminus I} x_{\sigma(k)}\right| + \left|\sum_{k \in I} x_{\sigma(k)} - L\right| \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \text{ (done)}
\end{aligned}$$

■

The second, third and forth parts (Theorem 3.1.4, Theorem 3.1.5 and Theorem 3.1.6), respectively states that if a series converge but not absolutely, then the limit value can be changed to any real number, infinite, negative infinite and even jumping-ly diverges.

The detail of the proof is very tedious and cumbersome, while the essence is easy to understand. The only two tools for proving Theorem 3.1.4, Theorem 3.1.5 and Corollary 3.1.6, is Lemma 3.1.3 and the fact $\sum a_k \to L \implies a_k \to 0$. If any part of the proof can be considered interesting, I believe it lies in that of Lemma 3.1.3, where one split the series $\sum a_k$ into two $\sum a_k^+, \sum a_k^-$, and shows that they must both diverge.

Lemma 3.1.3. (Intrinsic Structure of Series that Converge but not Absolutely) Let $f^+ : \mathbb{N} \to \mathbb{N}$ and $f^- : \mathbb{N} \to \mathbb{N}$ satisfy that $\{a_{f^+(n)}\}$ contain all and only positive terms of $\{a_n\}$ and $\{a_{f^-(n)}\}$ contain all and only negative terms. If $\sum_{k=1}^{\infty} a_k$ converge but not absolutely, then for each $\alpha \in \mathbb{R}^+$ and $n \in \mathbb{N}$, there exists $u_n > n$ and $t_n > n$ such that

$$\sum_{n \leq k \leq u_n} a_{f^+(k)} > \alpha \text{ and } \sum_{n \leq k \leq t_n} a_{f^-(k)} < -\alpha$$

Proof. Let $a_n^+ = \max\{0, a_n\}$ and $a_n^- = \min\{0, a_n\}$. It is easy to check $\forall n \in \mathbb{N}, a_n = a_n^+ + a_n^-$. Because $\sum_{k=1}^{\infty} a_k$ converge but not absolutely, we know

$$\begin{cases} \sum_{k=1}^{\infty} a_k^+ \to \infty \\ \sum_{k=1}^{\infty} a_k^- \to -\infty \end{cases} \tag{3.9}$$

This is true because if both of them converge then $\sum_{k=1}^{\infty} |a_k|$ converges and if only one of them converge them $\sum_{k=1}^{\infty} a_k$ diverges.

Because of Equation 3.9, we know

$$\begin{cases} \sum_{k=1}^{\infty} a_{f^+(k)} \to \infty \\ \sum_{k=1}^{\infty} a_{f^-(k)} \to -\infty \end{cases}$$

The result then follow, since

$$\forall n \in \mathbb{N}, \sum_{k \geq n} a_{f^+(k)} \nearrow \infty \text{ and } \sum_{k \geq n} a_{f^-(k)} \searrow -\infty$$

■

Theorem 3.1.4. (Riemann Rearrangement Theorem, Part 2) If $\sum_{k=1}^{\infty} a_k$ converge but not absolutely, then there exists permutations $\sigma_\infty, \sigma_{-\infty} : \mathbb{N} \to \mathbb{N}$ such that $\sum_{k=1}^{\infty} a_{\sigma_\infty(k)} \to \infty$ and $\sum_{k=1}^{\infty} a_{\sigma_{-\infty}(k)} \to -\infty$.

Proof. We wish

$$\text{to construct } \sigma_\infty : \mathbb{N} \to \mathbb{N} \text{ such that } \sum_{k=1}^{\infty} a_{\sigma_\infty(k)} \to \infty$$

Using Lemma 3.1.3, construct $\sigma_\infty : \mathbb{N} \to \mathbb{N}$ as follows. Let p_n be a sequence of natural number such that for each $n \in \mathbb{N}$, p_{n+1} is the smallest natural number such that

$$\sum_{k=p_n+1}^{p_{n+1}} a_{f^+(k)} > 3 \text{ and } p_1 = 0 \tag{3.10}$$

Similarly, let q_n be a sequence of natural number such that for each $n \in \mathbb{N}, q_{n+1}$ is the smallest natural number such that

$$\sum_{k=q_n+1}^{q_{n+1}} a_{f^-(k)} < -1 \text{ and } q_1 = 0 \tag{3.11}$$

Notice that the definition of p_n and q_n (Equation 3.10, Equation 3.11) are done recursively. Now, recursively define σ_∞ to follow the order

$$\begin{aligned} &f^+(p_1+1), \ldots, f^+(p_2), f^-(q_1+1), \ldots, f^-(q_2) \\ \longrightarrow &f^+(p_2+1), \ldots, f^+(p_3), f^-(q_2+1), \ldots, f^-(q_3), f^+(p_3+1), \ldots \end{aligned}$$

If there exists $k \in \mathbb{N}$ such that $a_k = 0$, which is not in the range $f^+[\mathbb{N}] \cup f^-[\mathbb{N}]$, we can merge these zero term into our σ_∞ by putting them in terms of even order. This way, our σ_∞ then become bijecetive, a permutation.

We claim

$$\text{such } \sigma_\infty \text{ works}$$

Recall the definition of p_n (Equation 3.10) is that for each $n \in \mathbb{N}, p_{n+1}$ is the smallest natural number such that

$$\sum_{k=p_n+1}^{p_{n+1}} a_{f^+(k)} > 3$$

Also recall the similarly defined q_n. This tell us

$$\sum_{k=p_n+q_n+1}^{p_{n+1}+q_{n+1}} a_{\sigma_\infty(k)} \to 2 \text{ as } n \to \infty$$

where

$$\sum_{k=p_n+q_n+1}^{p_{n+1}+q_n} a_{\sigma_\infty(k)} \to 3 \text{ and } \sum_{k=p_{n+1}+q_n+1}^{p_{n+1}+q_{n+1}} a_{\sigma_\infty(k)} \to -1 \text{ as } n \to \infty$$

With this, it is easy to verify $\sum_{k=1}^{\infty} a_{\sigma_\infty(k)} \to \infty$ (done) . The construction of $\sigma_{-\infty}$ and the proof for its validity is done similarly. ∎

Theorem 3.1.5. (Riemann Rearrangement Theorem, Part 3) If $\sum_{k=1}^{\infty} a_k$ converges but not absolutely, then for all $[L, M] \subseteq \mathbb{R}$, there exists a permutation $\sigma : \mathbb{N} \to \mathbb{N}$ such that $\liminf_{n\to\infty} \sum_{k=1}^{n} a_{\sigma(k)} = L$ and $\limsup_{n\to\infty} \sum_{k=1}^{n} a_{\sigma(k)} = M$.

Proof. We wish

to construct a working σ

The construction of σ is similar to that of σ_∞ in Theorem 3.1.4. WOLG, let $M > 0$. Let $p_1 = 0$, and let p_2 be the smallest natural number such that

$$\sum_{k=1}^{p_2} a_{f^+(k)} > M \text{ and } p_1 = 0$$

Next, define $q_1 = 0$ and let q_2 be the smallest natural number such that

$$\sum_{k=1}^{p_2} a_{f^+(k)} + \sum_{k=1}^{q_2} a_{f^-(k)} < L$$

Then, let p_3 be the smallest natural number such that

$$\sum_{k=1}^{p_3} a_{f^+(k)} + \sum_{k=1}^{q_2} a_{f^-(k)} > M$$

Recursively do such. We get two sequences $\{p_n\}, \{q_n\}$ of natural number such that for all $n \in \mathbb{N}, p_{n+1}$ is the smallest natural number such that

$$\sum_{k=1}^{p_{n+1}} a_{f^+(k)} + \sum_{k=1}^{q_n} a_{f^-(k)} > M$$

and for all $n \in \mathbb{N}, q_{n+1}$ is the smallest natural number such that

$$\sum_{k=1}^{p_{n+1}} a_{f^+(k)} + \sum_{k=1}^{q_{n+1}} a_{f^-(k)} < L$$

Them, recursively define σ to follow the order

$$\begin{aligned} &f^+(p_1+1), \ldots, f^+(p_2), f^-(q_1+1), \ldots, f^-(q_2) \\ \longrightarrow &f^+(p_2+1), \ldots, f^+(p_3), f^-(q_2+1), \ldots, f^-(q_3), f^+(p_3+1), \ldots \end{aligned}$$

Again, merge in the zero terms like in Theorem 3.1.4. The proof for the claim such σ works is easy to verify knowing $a_{\sigma(k)} \to 0$ (done) ■

Corollary 3.1.6. (Riemann Rearrangement Theorem, Part 4) If $\sum_{k=1}^\infty a_k$ converges but not absolutely, then for all $L \in \mathbb{R}$, there exists a permutation σ such that $\sum_{k=1}^\infty a_{\sigma(k)} = L$

Theorem 3.1.7. (Summary of Riemann Rearrangement Theorem) If $\sum_{k=1}^{\infty} a_k$ converge, then

$$\sum_{k=1}^{\infty} a_k \text{ absolutely converges} \iff \sum_{k=1}^{\infty} a_k \text{ unconditionally converges}$$

Proof. $(\longrightarrow)$

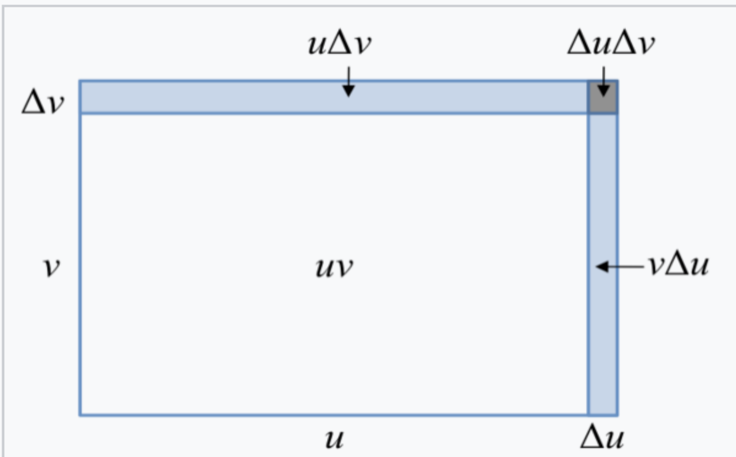
This is Theorem 3.1.2.

$(\longleftarrow)$

The fact that the contraposition of this statement is true is implied by any of Theorem 3.1.4, Theorem 3.1.5 and Corollary 3.1.6. ■

3.2 Product, Quotient and Chain Rule

This section concern mostly the computation of actual value of the derivative and integral of function. With this in mind, we first prove the product and quotient rules for derivative of $\mathbb{R}$ to $\mathbb{R}$ functions taught in most Calculus 1 classes. The proofs for the laws are easy, as it require no ingenious idea but ability to manipulate the limit symbol. However, without philosophical comments, we left an graph for geometric intuition for product rule. There are also graphs for geometric intuition for quotient rule on Internet, but we won't put it here as it require more than subtle work to understand the graph.



Geometric illustration of a proof of the product rule

Theorem 3.2.1. (Product Rule and Quotient Rule for Real to Real Function) Suppose f and g is differentiable at x, and $g'(x) \neq 0$. We have

(a) $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$ (Product Rule)

(b) $(\frac{f}{g})'(x) = \frac{f'(x)g(x)-f(x)g'(x)}{\big(g(x)\big)^2}$ (Quotient Rule)

Proof. Compute

$$\begin{aligned}
(fg)'(x) &= \lim_{h\to 0} \frac{f(x+h)g(x+h)-f(x)g(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x+h)-f(x)g(x+h)+f(x)g(x+h)-f(x)g(x)}{h} \\
&= \lim_{h\to 0} \big(g(x+h)\big)\frac{f(x+h)-f(x)}{h} + \big(f(x)\big)\frac{g(x+h)-g(x)}{h} \\
&= g(x)f'(x) + f(x)g'(x)
\end{aligned}$$

Compute

$$
\begin{aligned}
(\frac{f}{g})'(x) &= \lim_{h\to 0} \frac{\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x) - f(x)g(x+h)}{hg(x+h)g(x)} \\
&= \lim_{h\to 0} (\frac{1}{g(x+h)g(x)}) \cdot (\frac{f(x+h)g(x) - f(x)g(x) + f(x)g(x) - f(x)g(x+h)}{h}) \\
&= \lim_{h\to 0} \Big(\frac{1}{g(x+h)g(x)}\Big) \cdot \Big(\big(g(x)\big)\frac{f(x+h)-f(x)}{h} + \big(f(x)\big)\frac{g(x)-g(x+h)}{h}\Big) \\
&= \Big(\frac{1}{\big(g(x)\big)^2}\Big) \cdot \Big(\big(g(x)f'(x)\big) - f(x)g'(x)\Big) = \frac{f'(x)g(x) - f(x)g'(x)}{\big(g(x)\big)^2}
\end{aligned}
$$

■

Even a year has past, I can still remember what happened in the first class of Vector Analysis last year. The professor asked: "What is derivative?". A lot of answers emerge, from extremely formal and abstract like $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$ to those following geometric intuition like tangent line. Everyone gave a correct answer, but none of them philosophically satisfy the requirement of the question from the professor. Then, he stated: "Derivative is exactly linear approximation", and stated on black board the most general definition:

Definition 3.2.2. (Definition of Differential) Given two normed space V, W and an open subset $U \subseteq V$, we say a function $f : U \to W$ is **differentiable at** x if there exists a bounded linear operator $A : V \to W$ such that

$$\lim_{h\to 0} \frac{\|f(x+h) - f(x) - A(h)\|_W}{\|h\|_V} = 0$$

and we say the bounded linear map A is the **(total) derivative** of f at x.

If one put the key words "proof for chain rule" in Google search box, just like the situation in my classes, lots of rigorous proof emerge, but none of them is philosophical satisfying. For this reason, I shall give a proof of chain rule for real to real function based on the concept of linear approximation.

In Baby Rudin, derivative of a real to real function f is defined by

$$f'(a) = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h}$$

Immediately from this definition, we can derive a linear approximation P of f at a by setting

$$P(x) = f(a) + f'(a)(x - a)$$

Then, we see if we set $R(x) = f(x) - P(x)$ as the error (or the remainder) of the approximation, then trivially we have the behavior

$$R(x) \to 0 \text{ as } x \to a$$

what behavior of $R(x)$ that give P the name approximation is

$$\frac{R(x)}{x-a} \to 0 \text{ as } x \to a$$

The difference between the two behaviors is symbolically apparent, yet without geometric help, it may be difficult to precisely describe how insignificant the first behavior is compared to the second behavior. For this, observe that any function g that converge to $f(a)$ at a satisfy the first behavior, yet only a few satisfy the second. One can easily verify that the only linear $\mathbb{R}$ to $\mathbb{R}$ function that satisfy the second behavior is $P(x) = f(a) + f'(a)(x - a)$. Geometrically, this means that $R(x) = o(f'(x)dx)$ as $x \to a$.

Theorem 3.2.3. (Chain Rule for $\mathbb{R}$ to $\mathbb{R}$ function) Suppose g is differentiable at a and f is differentiable at $g(a)$. We have

$$(f \circ g)'(a) = f'(g(a))g'(a)$$

Proof. Define the remainders $R_{f(g(a))}(x)$ and $R_{g(a)}(x)$ by

$$\begin{cases} R_{f(g(a))}(x) = f(x) - f(g(a)) - f'(g(a))(x - g(a)) \\ R_{g(a)}(x) = g(x) - g(a) - g'(a)(x - a) \end{cases}$$

Compute

$$(f \circ g)'(a) = \lim_{x \to a} \frac{f(g(x)) - f(g(a))}{x - a} \tag{3.12}$$

$$= \lim_{x \to a} \frac{R_{f(g(a))}(g(x)) + f'(g(a))(g(x) - g(a))}{x - a} \tag{3.13}$$

Notice that because $x \to a \implies g(x) \to g(a)$, we have

$$\lim_{x \to a} \frac{R_{f(g(a))}(g(x))}{x - a} = \lim_{x \to a} \frac{R_{f(g(a))}(g(x))}{g(x) - g(a)} \cdot \frac{g(x) - g(a)}{x - a} = 0 \cdot g'(x) = 0 \tag{3.14}$$

Notice that the above deduction (Equation 3.14) is quite informal for two reasons: First, it may happen that $g(x) = g(a)$ locally. Second, for some reader it may require a mini proof to verify that $\frac{R_{f(g(a))}(g(x))}{g(x)-g(a)} \to 0$ as $x \to a$. These two obstacles for advanced readers should be insignificant.

Getting back to Equation 3.12, by Equation 3.14, we now see

$$(f \circ g)'(a) = \lim_{x \to a} \frac{f'(g(a))(g(x) - g(a))}{x - a} = f'(g(a))g'(a)$$

which finish the proof. ■

3.3 IVT, EVT and MVT

If one wish to understand most of the Theorems after this section, one must first know MVT, for what an important role it cast in the sections after. Logically prior to MVT is IVT. Yet, unlike MVT involve the intrinsic nature of field and limit structure of $\mathbb{R}$. IVT can be considered as purely topological in the sense that its proof can be stated almost in the language of topology. There are only two facts (the first are purely topological and the second is very close to purely topological) one need to know to prove IVT.

First, continuous functions map a connected sets to connected set. Second, a set in $\mathbb{R}$ is connected if and only if it is an interval.

Combining the above two facts, we have the following statement:

Theorem 3.3.1. (Continuous Real to Real Function Maps Interval to Interval) as titled.

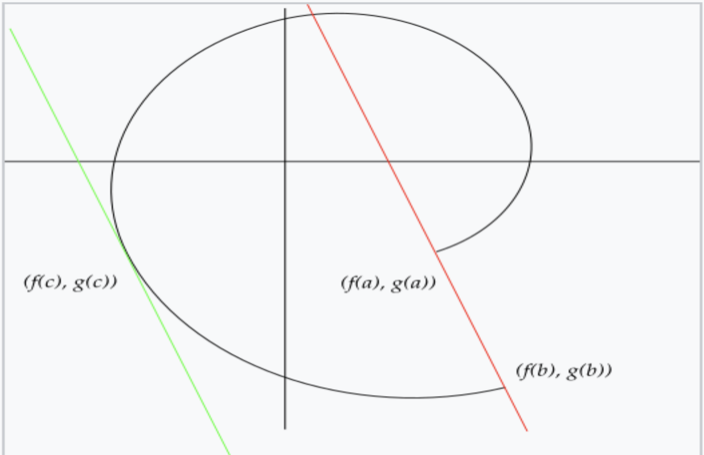
Proof. Consider the fact a continuous function map connected sets to connected sets and the fact a set in $\mathbb{R}$ is connected if and only if it is an interval. ■

Then, given the necessary constraint (the interval considered is compact) to give the conclusion, we have the famous statement:

Theorem 3.3.2. (IVT) Given a continuous function $f : [a, b] \to \mathbb{R}$, for each y that lies between $f(a)$ and $f(b)$, there exists $x \in [a, b]$ such that $f(x) = y$.

Given the simplicity of the logical deduction, we shall not give a rigorous proof here. However, one can notice that the interval considered in IVT "must be" compact, otherwise the Theorem is invalid. This constraint is in some sense a showcase how the concept of compact really match the description of "smallness (bounded) and rigidness (closed)".

Compared to IVT, another famous MVT is richer in both the results and the proof. Clearly for a logical and economic purpose, we shall first prove the Cauchy MVT.



Geometrical meaning of Cauchy's theorem

Theorem 3.3.3. (Cauchy's MVT) Given a function $f : [a,b] \to \mathbb{R}$ such that

(a) f, g are differentiable on (a,b)

(b) f, g are continuous on $[a,b]$

There exists $x \in (a,b)$ such that

$$\big[f(b) - f(a)\big]g'(x) = \big[g(b) - g(a)\big]f'(x)$$

Proof. We wish to find $x \in (a,b)$ such that

$$\big[f(b) - f(a)]g'(x) - [g(b) - g(a)]f'(x) = 0$$

Define h on (a,b) by

$$h(x) = \big[f(b) - f(a)\big]g'(x) - \big[g(b) - g(a)\big]f'(x)$$

We reduced our problem into finding $x \in (a,b)$ such that

$$h(x) = 0$$

Because f, g are both differentiable on (a,b), we know there exists an anti-derivative H of h on (a,b) such that

$$H(x) = \big[f(b) - f(a)\big]g(x) - \big[g(b) - g(a)\big]f(x)$$

We have $h = H'$ on (a,b). This let us reduce our problem into

finding a local extremum of H on (a,b)

Because f, g are both continuous on $[a, b]$, we know H is continuous on $[a, b]$. Then by EVT, we know

$$\exists x \in [a, b], H(x) = \max_{t \in [a,b]} H(t) \text{ and } \exists y \in [a, b], H(y) = \min_{t \in [a,b]} H(t)$$

If such x, y is in (a, b), we are done. If not, says that x, y both are on end points a or b. Compute that

$$H(a) = f(b)g(a) - g(b)f(a) = H(b)$$

We see H is constant on $[a, b]$. Then all points in (a, b) are extremums. (done) ■

Corollary 3.3.4. (Lagrange's MVT) Given a function $f : [a, b] \to \mathbb{R}$ such that

(a) f is differentiable on (a, b)

(b) f is continuous on $[a, b]$

Then there exists $x \in (a, b)$ such that

$$f'(x) = \frac{f(b) - f(a)}{b - a}$$

Proof. Let $g(x) = x$ in Cauchy's MVT (Theorem 3.3.3), and we are done. ■

There are two hypotheses in Lagrange's MVT

(a) f is differentiable on (a, b)

(b) f is continuous on $[a, b]$

They are all necessary. The necessity of differentiablity on (a, b) is clear as shown by the canonical example using absolute value. The necessity of continuity on $[a, b]$ can be shown by the example

$$f(x) = \begin{cases} 1 & \text{if } a < x \leq b \\ 0 & \text{if } x = a \end{cases}$$

Theorem 3.3.5. (First Mean Value Theorem for Definite Integral) Given a function $f : [a, b] \to \mathbb{R}$ such that

(a) f is continuous on (a, b)

There exists $\xi \in (a, b)$ such that

$$\int_a^b f(x)dx = f(\xi) \cdot (b - a)$$

Proof. We wish

$$\text{to find } \xi \in (a,b) \text{ such that } f(\xi) = \frac{\int_a^b f(x)dx}{b-a}$$

Define $\tilde{f} : [a,b] \to \mathbb{R}$ on $[a,b]$ by

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in (a,b) \\ \lim_{t\to a} f(t) & \text{if } x = a \\ \lim_{t\to b} f(t) & \text{if } x = b \end{cases} \tag{3.15}$$

Then, because $\int_a^b f(x)dx = \int_a^b \tilde{f}(x)dx$, we reduce our problem into

$$\text{finding } \xi \in (a,b) \text{ such that } \tilde{f}(\xi) = \frac{\int_a^b \tilde{f}(x)dx}{b-a}$$

Because $\tilde{f}$ is continuous on $[a,b]$ by definition Equation 3.15, by EVT, we know we there exists $\alpha, \beta \in [a,b]$ such that

$$\tilde{f}(\alpha) = \inf_{x\in[a,b]} \tilde{f}(x) \text{ and } \tilde{f}(\beta) = \sup_{x\in[a,b]} \tilde{f}(x) \tag{3.16}$$

WOLG, suppose $\alpha \leq \beta$. Deduce

$$\tilde{f}(\alpha) = \inf_{x\in[a,b]} \tilde{f}(x) \leq \frac{\int_a^b \tilde{f}(x)dx}{b-a} \leq \sup_{x\in[a,b]} \tilde{f}(x) = \tilde{f}(\beta)$$

by IVT, we then know there exists $\xi \in [\alpha, \beta]$ such that

$$\exists \xi \in [\alpha,\beta], \tilde{f}(\xi) = \frac{\int_a^b \tilde{f}(x)dx}{b-a} \tag{3.17}$$

If $a < \alpha$ and $\beta < b$, our proof is done.

If not, notice that if $\tilde{f}(\alpha) = \tilde{f}(\beta)$, then by definition of α, β (Equation 3.16), the proof is trivial since $\tilde{f}$ is a constant, so we only have to consider when $\tilde{f}(\alpha) < \tilde{f}(\beta)$, and we wish to show

$$\xi \text{ can not happen at } a \text{ nor } b$$

Assume $\xi = a$, WOLG. Because $\xi \in [\alpha, \beta]$, we know $\alpha = a$. Because $\tilde{f}(\beta) > \tilde{f}(\alpha)$, we can find δ such that

$$\inf_{x\in[\beta-\delta,\beta]} \tilde{f}(x) \geq \frac{\tilde{f}(\alpha) + 2\tilde{f}(\beta)}{3} \tag{3.18}$$

We then from Equation 3.17 see that

$$\int_a^b \tilde{f}(x)dx = \tilde{f}(\xi)(b-a) = \tilde{f}(\alpha)(b-a) \tag{3.19}$$

Also, we see from definition of α (Equation 3.16) and Equation 3.18 that

$$\begin{align}
\int_a^b \tilde{f}(x)dx &= \int_a^{\beta-\delta} \tilde{f}(x)dx + \int_{\beta-\delta}^{\beta} \tilde{f}(x)dx + \int_\beta^b \tilde{f}(x)dx \tag{3.20}\\
&\geq (b-\delta-a)\tilde{f}(\alpha) + \delta \cdot (\frac{\tilde{f}(\alpha)+2\tilde{f}(\beta)}{3}) \tag{3.21}\\
&> (b-\delta-a)\tilde{f}(\alpha) + \delta \cdot (\frac{\tilde{f}(\alpha)+\tilde{f}(\beta)}{2}) \tag{3.22}\\
&= \tilde{f}(\alpha)\big(b-a-\frac{\delta}{2}) + \tilde{f}(\beta) \cdot (\frac{\delta}{2}) \tag{3.23}
\end{align}$$

Now, from Equation 3.19 and Equation 3.23, we can deduce

$$\tilde{f}(\alpha)(b-a) > \tilde{f}(\alpha)(b-a-\frac{\delta}{2}) + \tilde{f}(\beta) \cdot (\frac{\delta}{2})$$

Then we can deduce

$$\tilde{f}(\alpha) \cdot (\frac{\delta}{2}) > \tilde{f}(\beta) \cdot (\frac{\delta}{2}) \text{ CaC } \text{ (done)}$$

■

Theorem 3.3.6. (Second Mean Value Theorem for Definite Integral) Given functions $G, \phi : [a,b] \to \mathbb{R}$ such that

(a) G is monotonic

(b) ϕ is Riemann-Integrable

Let $G(a^+) = \lim_{t\to a^+} G(t)$ and $G(b^-) = \lim_{t\to b^-} G(t)$. Then there exists $\xi \in (a,b)$ such that

$$\int_a^b G(t)\phi(t)dt = G(a^+)\int_a^\xi \phi(t)dt + G(b^-)\int_\xi^b \phi(t)dt$$

Proof. Define f on $[a,b]$ by

$$f(x) = G(a^+)\int_a^x \phi(t)dt + G(b^-)\int_x^b \phi(t)dt$$

We then reduce the problem into

$$\text{finding } \xi \in (a,b) \text{ such that } \int_a^b G(t)\phi(t)dt = f(\xi)$$

By Theorem 3.7.1, we know f is continuous on $[a,b]$. Then by IVT, we can reduce the problem into

$$\text{finding an interval } [c,d] \subseteq (a,b) \text{ such that } \int_a^b G(t)\phi(t) \text{ is between } f(c) \text{ and } f(d)$$

Observe that

$$f(a) = G(b^-)\int_a^b \phi(t)dt \text{ and } f(b) = G(a^+)\int_a^b \phi(t)dt$$

■

3.4 Riemann-Stieltjes on Computation

Theorem 3.4.1. (Change of Variable) Given two functions $g, \beta : [A, B] \to \mathbb{R}$, a function $\phi : [A, B] \to [a, b]$ and two functions $f, \alpha : [a, b] \to \mathbb{R}$ such that

(a) $g = f \circ \phi$ for all $x \in [a, b]$

(b) $\beta = \alpha \circ \phi$ for all $x \in [a, b]$

(c) α, β increase respectively on $[a, b]$ and $[A, B]$

(d) $\phi : [A, B] \to [a, b]$ is a homeomorphism

(e) $\int_a^b f d\alpha$ exist

Then

$$\int_A^B g d\beta = \int_a^b f d\alpha \text{ (This implies } \int_A^B g d\beta \text{ exists)}$$

Proof. Fix ϵ. We only wish

$$\text{to find a partition } Q \text{ of } [A, B] \text{ such that } U(Q, g, \beta) - L(Q, g, \beta) < \epsilon$$
$$\text{and such that } \int_a^b f d\alpha \in \Big[L(Q, g, \beta), U(Q, g, \beta)\Big]$$

Because $\int_a^b f d\alpha$ exists, we know

$$\text{there exists a partition } P \text{ of } [a, b] \text{ such that } U(P, f, \alpha) - L(P, f, \alpha) < \epsilon \tag{3.24}$$

where, of course, $\int_a^b f d\alpha \in \Big[L(P, f, \alpha), U(P, f, \alpha)\Big]$.

Let $P = \{a = x_0, x_1, \ldots, x_n = b\}$. Because ϕ is a homeomorphism, we can let ϕ be strictly increasing WOLG.

Define a partition Q on $[A, B]$ by

$$Q = \phi^{-1}[P] = \{A = \phi^{-1}(x_0), \phi^{-1}(x_1), \ldots, \phi^{-1}(x_n) = B\}$$

Now, because $\beta = \alpha \circ \phi$ and $g = f \circ \phi$ for all $x \in [a, b]$ by premise, and because ϕ is a homeomorphism, we have

$$\begin{aligned}U(Q,g,\beta) &= \sum_{k=1}^{n}\big[\sup_{t\in[\phi^{-1}(x_{k-1}),\phi^{-1}(x_k)]} g(t)\big]\big[\beta(\phi^{-1}(x_k))-\beta(\phi^{-1}(x_{k-1}))\big]\\ &= \sum_{k=1}^{n}\big[\sup_{t\in[\phi^{-1}(x_{k-1}),\phi^{-1}(x_k)]} f\circ\phi(t)\big]\big[\alpha\circ\phi\big(\phi^{-1}(x_k)\big)-\alpha\circ\phi\big(\phi^{-1}(x_{k-1})\big)\big]\\ &= \sum_{k=1}^{n}\big[\sup_{t\in[x_{k-1},x_k]} f(t)\big]\big(\alpha(x_k)-\alpha(x_{k-1})\big)=U(P,f,\alpha) \qquad (3.25)\end{aligned}$$

Similarly, we can deduce $L(Q,g,\beta)=L(P,f,\alpha)$. Now, from Equation 3.25 and by definition of P (Equation 3.24), we see

$$U(Q,g,\beta)-L(Q,g,\beta)=U(P,f,\alpha)-L(P,f,\alpha)<\epsilon$$

$$\text{and } \int_a^b f\,d\alpha \in \Big[L(P,f,\alpha),U(P,f,\alpha)\Big]=\Big[L(Q,g,\beta),U(Q,g,\beta)\Big] \text{ (done)}$$

■

Theorem 3.4.2. (Reduction of Riemann-Stieltjes Integral: Part 1) Given two functions $f,\alpha:[a,b]\to\mathbb{R}$ such that

(a) α increase on $[a,b]$

(b) α is differentiable on (a,b)

(c) $\lim_{x\to b^-}\frac{\alpha(x)-\alpha(b)}{x-b}$ exists and $\lim_{x\to a^+}\frac{\alpha(x)-\alpha(a)}{x-a}$ exists

(d) α' is properly Riemann-Integrable on $[a,b]$

(e) f is bounded on $[a,b]$

Then

$$\int_a^b f\,d\alpha \text{ exists} \iff \int_a^b f(x)\alpha'(x)dx \text{ exists and they equal to each other if exists}$$

Proof. We wish to prove

$$\overline{\int_a^b} f\,d\alpha=\overline{\int_a^b} f(x)\alpha'(x)dx$$

Fix ϵ. We reduce the problem into proving

$$\left|\overline{\int_a^b} f\,d\alpha-\overline{\int_a^b} f(x)\alpha'(x)dx\right|<\epsilon$$

Then, because for all partition P of $[a,b]$, we have

$$\left|\overline{\int_a^b} f d\alpha - \overline{\int_a^b} f(x)\alpha'(x)dx\right|$$
$$\leq \left|\overline{\int_a^b} f d\alpha - U(P,f,\alpha)\right| - \left|U(P,f,\alpha) - U(P,f\alpha')\right| - \left|U(P,f\alpha') - \overline{\int_a^b} f(x)\alpha'(x)dx\right|$$

We only wish

$$\text{to find } P \text{ such that} \left|\overline{\int_a^b} f d\alpha - U(P,f,\alpha)\right| < \frac{\epsilon}{3}$$
$$\text{and } \left|U(P,f,\alpha) - U(P,f\alpha')\right| < \frac{\epsilon}{3} \text{ and } \left|\overline{\int_a^b} f(x)\alpha'(x)dx - U(P,f\alpha')\right| < \frac{\epsilon}{3}$$

Because f is bounded on $[a,b]$, we can let $M = \sup_{x\in[a,b]} |f(x)|$. Because $\int_a^b \alpha'(x)dx$ exists, we can let P satisfy

$$U(P,\alpha') - L(P,\alpha') < \frac{\epsilon}{4M} \tag{3.26}$$

By definition of Riemann Upper sum, we can further refine P to let P satisfy

$$\left|\overline{\int_a^b} f d\alpha - U(P,f,\alpha)\right| < \frac{\epsilon}{3} \text{ and } \left|\overline{\int_a^b} f(x)\alpha'(x)dx - U(P,f\alpha')\right| < \frac{\epsilon}{3}$$

It is clear that the statement concerning P (Equation 3.26) remain valid after refinement of P. Fix such P. We now have reduced the problem into proving

$$|U(P,f,\alpha) - U(P,f\alpha')| < \frac{\epsilon}{3}$$

Express P in the form $P = \{a = x_0, x_1, \dots, x_n = b\}$. By MVT (Theorem 3.3.4), we know for all $k \in \{1, \dots, n\}$ there exists $t_k \in [x_{k-1}, x_k]$ such that

$$\Delta\alpha_k = \alpha'(t_k)\Delta x_k \tag{3.27}$$

Then, because $U(P,\alpha') - L(P,\alpha)' < \frac{\epsilon}{3M}$ (Equation 3.26), we now see

$$\sum_{k=1}^n |\alpha'(s_k) - \alpha'(t_k)| \Delta x_k < \frac{\epsilon}{3M} \text{ if } s_k \in [x_{k-1}, x_k] \text{ for all } k \in \{1, \dots, n\} \tag{3.28}$$

Then from Equation 3.27, definition of M and Equation 3.28, we have

$$\begin{aligned}\left|\sum_{k=1}^{n} f(s_k)\Delta\alpha_k - \sum_{k=1}^{n} f(s_k)\alpha'(s_k)\Delta x_k\right| &= \left|\sum_{k=1}^{n} f(s_k)\big(\alpha'(s_k) - \alpha'(t_k)\big)\Delta x_k\right| \\ &\leq \sum_{k=1}^{n} |f(s_k)| \cdot |\alpha'(s_k) - \alpha'(t_k)|\, \Delta x_k \\ &\leq M \sum_{k=1}^{n} |\alpha'(s_k) - \alpha'(t_k)|\, \Delta x_k \\ &< \frac{\epsilon}{4}\end{aligned}$$

Then because $\sum_{k=1}^{m} f(s_k)\alpha'(s_k)\Delta x_k \leq U(P, f\alpha')$, we now have

$$\sum_{k=1}^{n} f(s_k)\Delta\alpha_k < U(P, f\alpha') + \frac{\epsilon}{4} \tag{3.29}$$

Because Equation 3.29 hold true for all choices of s_k, we have

$$U(P, f, \alpha) < U(P, f\alpha') + \frac{\epsilon}{3}$$

Similarly, we can deduce

$$U(P, f\alpha') < U(P, f, \alpha) + \frac{\epsilon}{3} \text{ (done)}$$

■

Theorem 3.4.3. (Substitution Law) Given a function $\phi : [a, b] \to [A, B]$ and a function $f : [A, B] \to \mathbb{R}$ such that

(a) ϕ is a homoeomorphism.

(b) ϕ is differentiable on (a, b)

(c) $\int_a^b \phi'(x)dx$ exists.

(d) f is integrable on $[A, B]$

We have

$$\int_a^b f\big(\phi(x)\big)\phi'(x)dx = \int_A^B f(u)du$$

Proof. Because $f \circ \phi$ and ϕ' is integrable on $[a, b]$, by reduction of Riemann-Stieljes Integral (Tehroem 3.4.2), we know

$$\int_a^b (f \circ \phi)(x)\phi'(x)dx = \int_a^b (f \circ \phi)(x)d\phi$$

Let $\alpha(x) = x$. Let $\beta = \alpha \circ \phi$. Define $g = f \circ \phi$. By Change of Variable (Theorem 3.4.1), we now have

$$\int_a^b (f \circ \phi)(x)d\phi = \int_a^b g(x)d\beta = \int_A^B f(x)dx$$

■

3.5 Weierstrass approximation Theorem: $[a, b] \to \mathbb{R}$

Theorem 3.5.1. (Bernoulli's Inequality) Given $r, x \in \mathbb{R}$, suppose

(a) $r \geq 1$

(b) $x \geq -1$

Then

$$(1 + x)^r \geq 1 + rx$$

Proof. Fix $r \geq 1$. We wish

$$\text{to prove } (1 + x)^r \geq 1 + rx \text{ for all } x \geq -1$$

Define $f : [-1, \infty) \to \mathbb{R}$ by

$$f(x) = (1 + x)^r - (1 + rx) \tag{3.30}$$

We reduced the problem into

$$\text{proving } f(x) \geq 0 \text{ for all } x \geq -1$$

Because $r \geq 1$ by premise, by definition of $f(x)$ (Equation 3.30), we see that

$$f(0) = 0, \text{ and } f(-1) = r - 1 \geq 0$$

Notice that by definition of f (Equation 3.30), $f(x)$ is clearly differentiable on $(-1, \infty)$.

Then, by MVT (Theorem 3.3.4), to prove $f(x) \geq 0$ on $(-1, \infty)$, we only wish

$$\text{to prove } f'(x) \geq 0 \text{ for all } x > 0 \text{ and } f'(x) \leq 0 \text{ for all } x \in (-1, 0)$$

Compute f'

$$\begin{aligned} f'(x) &= r(1 + x)^{r-1} - r \\ &= r\Big((1 + x)^{r-1} - 1\Big) \end{aligned}$$

Because $r \geq 1$, we can deduce

$$x > 0 \implies (1 + x)^{r-1} \geq 1 \implies f'(x) = r\Big((1 + x)^{r-1} - 1\Big) \geq 0$$

and deduce

$$x \in (-1, 0) \implies 1 + x \in (0, 1) \implies (1 + x)^{r-1} \leq 1 \implies f'(x) = r\Big((1 + x)^{r-1} - 1\Big) \leq 0$$

(done) ∎

In this section, notation $\mathcal{C}\big([a,b]\big)$ means the set of **real-valued continuous function on** $[a,b]$.

Theorem 3.5.2. (Weierstrass approximation Theorem: $[a,b] \to \mathbb{R}$**)** Let $\mathbb{R}[x]\big|_{[a,b]}$ be the space of polynomials on $[a,b]$ with real coefficient. We have

$$\mathbb{R}[x]\big|_{[a,b]} \text{ is dense in } \Big(\mathcal{C}\big([a,b]\big), \|\cdot\|_\infty\Big)$$

Proof. WOLG, we can let $[a,b] = [0,1]$. The reason we can assume such is explained at last. Now, let $f : [0,1] \to \mathbb{R}$ be a continuous function. Fix ϵ. We only wish

$$\text{to find } P \in \mathbb{R}[x]\big|_{[0,1]} \text{ such that } \|f - P\|_\infty < \epsilon$$

Define $\tilde{f} \in \mathcal{C}\big([0,1]\big)$ by

$$\tilde{f}(x) = f(x) - f(0) - x\big[f(1) - f(0)\big] \tag{3.31}$$

It is easy to check $\tilde{f}$ is continuous. We first prove that

$$\big(\tilde{f}(x) - f(x)\big) \in \mathbb{R}[x]\big|_{[0,1]}$$

By definition of $\tilde{f}$ (Equation 3.31), we see

$$\tilde{f}(x) - f(x) = (f(0) - f(1))x - f(0) \in \mathbb{R}[x]\big|_{[0,1]} \text{ (done)}$$

This reduce our problem into

$$\text{finding } P \in \mathbb{R}[x]\big|_{[0,1]} \text{ such that } \|\tilde{f} - P\|_\infty < \epsilon$$

Notice that by definition of $\tilde{f}$ (Equation 3.31), we have

$$\tilde{f}(0) = 0 = \tilde{f}(1)$$

Then, we can expand the definition of $\tilde{f}$ by

$$\tilde{f}(x) = \begin{cases} \tilde{f}(x) & \text{if } x \in [0,1] \\ 0 & \text{if } x \notin [0,1] \end{cases} \tag{3.32}$$

This makes $\tilde{f}$ uniformly continuous on $\mathbb{R}$, since $\tilde{f}$ is uniformly continuous on $[0,1]$ and $[0,1]^c$. Now, for each $n \in \mathbb{N}$, define $Q_n \in \mathbb{R}[x]$ by

$$Q_n = c_n(1 - x^2)^n \text{ where } c_n \text{ is chosen to satisfy } \int_{-1}^{1} Q_n(x)dx = 1 \tag{3.33}$$

Define $P_n : [0,1] \to \mathbb{R}$ by

$$P_n(x) = \int_{-1}^{1} \tilde{f}(x+t)Q_n(t)dt$$

We now prove

$$P_n \in \mathbb{R}[x]\big|_{[0,1]}$$

Because $\tilde{f}(x) = 0$ for all $x \notin (0,1)$ by definition of $\tilde{f}$ (Equation 3.32), we see that

$$P_n(x) = \int_{-x}^{1-x} \tilde{f}(x+t)Q_n(t)dt \text{ for all } x \in [0,1] \tag{3.34}$$

Fix $x \in [0,1]$. Now, by change of variable, we see

$$P_n(x) = \int_{-x}^{1-x} \tilde{f}(x+t)Q_n(t)dt = \int_0^1 \tilde{f}(u)Q_n(u-x)du$$

Because Q_n is a polynomial by definition (Equation 3.33), we can express $Q_n(u-x)$ by

$$Q_n(u-x) = \sum_{k=0}^{m} a_k x^k \text{ for some } \{a_0, \dots, a_m\} \text{ depending on } u$$

Then we see

$$P_n(x) = \int_0^1 \tilde{f}(u)Q_n(u-x)du = \sum_{k=0}^{m} x^k \Big(\int_0^1 \tilde{f}(u)a_k du \Big)$$

This shows that $P_n \in \mathbb{R}[x]\big|_{[0,1]}$
(done)

Now, because $\tilde{f}$ is uniformly continuous on $\mathbb{R}$, we can fix $\delta < 1$ such that

$$\forall x, y \in \mathbb{R}, |x-y| < \delta \implies \left| \tilde{f}(x) - \tilde{f}(y) \right| < \frac{\epsilon}{2} \tag{3.35}$$

By definition of $\tilde{f}$ (Equation 3.32), we know $\tilde{f}$ is a bounded function. Then we can set M by

$$M = \sup_{x \in \mathbb{R}} |f(x)|$$

Let n satisfy

$$4M\sqrt{n}(1-\delta^2)^n < \frac{\epsilon}{2} \tag{3.36}$$

Such n exists, because $\delta < 1 \implies \sqrt{n}(1-\delta^2)^n \to 0$. We claim

$$P_n \text{ satisfy } \|\tilde{f} - P_n\|_\infty < \epsilon$$

We first prove

$$c_n < \sqrt{n}$$

By Bernoulli's Inequality (Theorem 3.5.1). Compute

$$\begin{aligned} 1 = \int_{-1}^{1} Q_n(x)dx &= c_n \int_{-1}^{1} (1-x^2)^n dx \\ &= 2c_n \int_0^1 (1-x^2)^n dx \\ &\geq 2c_n \int_0^{\frac{1}{\sqrt{n}}} (1-x^2)^n dx \\ &\geq 2c_n \int_0^{\frac{1}{\sqrt{n}}} 1 - nx^2 dx = c_n(\frac{4}{3\sqrt{n}}) > c_n(\frac{1}{\sqrt{n}}) \end{aligned}$$

This implies

$$\sqrt{n} > c_n \text{ (done)}$$

Because $\sqrt{n} > c_n$, by definition of Q_n (Equation 3.33), we have

$$Q_n(x) < \sqrt{n}(1-x^2)^n \leq \sqrt{n}(1-\delta^2)^n \text{ for all } x \text{ such that } \delta \leq |x| \leq 1$$

Fix $x \in [0,1]$. Finally, because

(a) $\int_{-1}^{1} Q_n(x)dx = 1$ by definition of Q_n (Equation 3.33)

(b) $Q_n(x) = c_n(1-x^2)^n \geq 0$ for all $x \in [-1,1]$

(c) $\left|\tilde{f}(x+t) - \tilde{f}(x)\right| < \frac{\epsilon}{2}$ for all t such that $|t| < \delta$, by definition of δ (Equation 3.36)

(d) $Q_n(x) \leq \sqrt{n}(1-\delta^2)^n$ for all x such that $\delta \leq |x| \leq 1$

(e) $4M\sqrt{n}(1-\delta^2)^n < \frac{\epsilon}{2}$ by definition of n (Equation 3.36) we have

$$\begin{aligned}
\left|P_n(x) - \tilde{f}(x)\right| &= \left|\int_{-1}^{1} \tilde{f}(x+t)Q_n(t)dt - \tilde{f}(x)\right| \\
&= \left|\int_{-1}^{1} \tilde{f}(x+t)Q_n(t)dt - \tilde{f}(x)\int_{-1}^{1} Q_n(t)dt\right| \\
&= \left|\int_{-1}^{1} \tilde{f}(x+t)Q_n(t)dt - \int_{-1}^{1} \tilde{f}(x)Q_n(t)dt\right| \\
&= \left|\int_{-1}^{1} \big[\tilde{f}(x+t) - \tilde{f}(x)\big]Q_n(t)dt\right| \\
&\leq \int_{-1}^{1} \left|\big[\tilde{f}(x+t) - \tilde{f}(x)\big]Q_n(t)\right| dt \\
&= \int_{-1}^{1} \left|\tilde{f}(x+t) - \tilde{f}(x)\right| Q_n(t)dt \\
&\leq \int_{-1}^{-\delta} 2MQ_n(t)dt + \int_{-\delta}^{\delta} \left|\tilde{f}(x+t) - \tilde{f}(x)\right| Q_n(t)dt + \int_{\delta}^{1} 2MQ_n(t)dt \\
&\leq 2M\Big(\int_{-1}^{-\delta} Q_n(t)dt + \int_{\delta}^{1} Q_n(t)dt\Big) + \int_{-\delta}^{\delta} (\frac{\epsilon}{2})Q_n(t)dt \\
&\leq 4M(1-\delta)\sqrt{n}(1-\delta^2)^n + \frac{\epsilon}{2} \\
&\leq 4M\sqrt{n}(1-\delta^2)^n + \frac{\epsilon}{2} < \epsilon
\end{aligned}$$

Because x is arbitrarily picked from $[0,1]$, we now have $\|P_n - \tilde{f}\|_\infty < \epsilon$ (done)

Lastly, we show

our result can be transplanted to arbitrary $\mathcal{C}\big([a,b]\big)$

Let $[a,b]$ be arbitrary. Fix ϵ and $f \in \mathcal{C}\big([a,b]\big)$. We wish

to find $P \in \mathbb{R}[x]\big|_{[a,b]}$ such that $\|f - P\|_\infty \leq \epsilon$

Define $g : [0,1] \to \mathbb{R}$ by

$$g(x) \triangleq f(a + (b-a)x) \tag{3.37}$$

We know there exists $P_n : [0,1] \to \mathbb{R}$ such that

$$\|P_n - g\|_\infty < \epsilon$$

Define $H_n : [a,b] \to \mathbb{R}$ by

$$H_n(x) = P_n(\frac{x-a}{b-a})$$

Because P_n is a real polynomial on $[0,1]$, we know H_n is a real polynomial on $[a,b]$. We now claim

$$\text{such } H_n \text{ works}$$

Fix $x \in [a,b]$. Observe

$$\begin{aligned}\big|f(x) - H_n(x)\big| &= \left|f(x) - P_n(\frac{x-a}{b-a})\right| \\ &= \left|g(\frac{x-a}{b-a}) - P_n(\frac{x-a}{b-a})\right| < \epsilon \text{ (done)}\end{aligned}$$

■

It is at now, we will show that every real-valued continuous functions on $[a,b]$ can be approximated by polynomials with rational coefficient. This fact enable our computer to more easily approximate real-valued continuous function on $[a,b]$.

Note that since $\mathcal{C}\big([a,b]\big)$ is a separable metric space, we can show that $\mathcal{C}\big([a,b]\big)$ has cardinality of at most continuum $\mathfrak{c}$.

Theorem 3.5.3. (The space $\mathbb{Q}[x]|_{[a,b]}$ is dense in $\Big(\mathcal{C}\big([a,b]\big), \|\cdot\|_\infty\Big)$, thus $\mathcal{C}\big([a,b]\big)$ is separable)

$$\Big(\mathcal{C}\big([a,b]\big), \|\cdot\|_\infty\Big) \text{ is separable}$$

Proof. Because $\mathbb{Q}[x]\big|_{[a,b]}$ is countable, to show $\mathcal{C}\big([a,b]\big)$ is separable, we only wish to show

$$\mathbb{Q}[x]\big|_{[a,b]} \text{ is dense in } \mathcal{C}\big([a,b]\big)$$

Because $\mathbb{R}[x]\big|_{[a,b]}$ is dense in $\mathcal{C}\big([a,b]\big)$, we reduce our problem into proving

$$\mathbb{Q}[x]\big|_{[a,b]} \text{ is dense in } \mathbb{R}[x]\big|_{[a,b]}$$

Fix ϵ and $P \in \mathbb{R}[x]\big|_{[a,b]}$. We must

$$\text{find } Q \in \mathbb{Q}[x]\big|_{[a,b]} \text{ such that } \|Q-P\|_\infty \leq \epsilon$$

Express $P(x) = \sum_{k=0}^{n} r_k x^k$. Let $M > \max\{|a|, |b|\}$. Because $\mathbb{Q}$ is dense in $\mathbb{R}$, we know there exists $c_k \in \mathbb{Q}$ such that $|c_k - r_k| < \frac{\epsilon}{(n+1)M^n}$. We claim

$$Q(x) = \sum_{k=0}^{n} c_k x^k \text{ works}$$

Fix $x \in [a, b]$. See

$$\begin{aligned}
|P(x) - Q(x)| &= \left| \sum_{k=0}^{n} (c_k - r_k) x^k \right| \\
&\leq \sum_{k=0}^{n} |c_k - r_k| \cdot |x|^k \\
&\leq \sum_{k=0}^{n} |c_k - r_k| \cdot M^k \\
&\leq (M^n) \sum_{k=0}^{n} |c_k - r_k| \\
&< M^n(n+1)\Big(\frac{\epsilon}{(n+1)M^n}\Big) = \epsilon \text{ (done)}
\end{aligned}$$

■

3.6 The Stone-Weierstrass Theorem

Recall that a **vector space over a field** $\mathbb{F}$ is a set V equipped with **vector addition** $+ : V \times V \to V$ and **scalar multiplication** such that

(a) $(V, +)$ is an abelian group.

(b) Scalar multiplication is compatible with field multiplication: $\Big((ab)v = a(bv)\Big)$

(c) Scalar multiplication is distributive: $\Big((a+b)v = av + bv$ and $a(v+w) = av + aw\Big)$

There are many ways to define the term **algebra over a field** $\mathbb{F}$. One can exhaust all the laws an algebra should obey. In short, an **algebra over a field** $\mathbb{F}$ (or $\mathbb{F}$**-algebra**) is a set $(A, +, \cdot)$ equipped scalar multiplication over $\mathbb{F}$ such that

(a) Multiplication $\cdot$ is <u>distributive</u> with respect to $+$

(b) $(A, +)$ and scalar multiplication form a vector space.

(c) Scalar multiplication and vector multiplication $\cdot$ is compatible: $\Big((av) \cdot (bw) = ab(v \cdot w)\Big)$

Given an arbitrary set E and a field $\mathbb{F}$, let A be the set of all functions from E to $\mathbb{F}$. The following is a list of some algebra

(a) $(\mathbb{R}^3, \text{cross product})$ over $\mathbb{R}$

(b) $(\mathbb{C}, \text{complex multiplication})$ over $\mathbb{C}$

(c) $(\mathbb{Q}[x], \text{function multiplication})$ over $\mathbb{Q}$

(d) (Functions from E to $\mathbb{F}$, function multiplication) over $\mathbb{F}$

(e) (Continuous functions from (E, τ) to $\mathbb{C}$, function multiplication) over $\mathbb{C}$

(f) (Linear transformation from V to V, composition) over $\mathbb{F}$ where V is over $\mathbb{F}$

(g) $\big(M_n(\mathbb{F}), \text{matrix multiplication}\big)$ over $\mathbb{F}$

Note that $B =$ (continuous functions from $\mathbb{C}$ to $\mathbb{C}$, composition) over $\mathbb{C}$ is not an algebra, even though B is both a vector space and a ring. ($\because$ scalar multiplication and multiplication are not compatible).

It is at here we shall introduce some general terminologies. Given an arbitrary set E, a field $\mathbb{F}$ and a point $x \in E$, we say a family $\mathcal{F}$ of functions from E to $\mathbb{F}$ **vanish at** x if for all $f \in \mathcal{F}$, we have $f(x) = 0$. We say $\mathcal{F}$ **separate points** in E if for all $x_2 \neq x_1 \in E$, there exists $f \in \mathcal{F}$ such that $f(x_2) \neq f(x_1)$.

3.7 FTC

Theorem 3.7.1. (Fundamental Theorem of Calculus: Part 1) Suppose a function $f : [a, \infty) \to \mathbb{R}$ satisfy

$$f \text{ is } \underline{\text{proper-Riemann integrable}} \text{ on } [a, b] \text{ for all } b > a$$

If we set $F : [a, \infty) \to \mathbb{R}$

$$F(x) = \int_a^x f(t)dt$$

Then

(a) F is continuous on $[a, \infty)$

(b) F is differentiable at $x_0 \in [a, \infty)$ where $F'(x_0) = f(x_0)$ if f is continuous at x_0

Proof. Fix ϵ and $[a, b]$. We only wish

$$\text{to prove } F \text{ is continuous on } [a, b]$$

To prove F is continuous on $[a, b]$, we only wish

$$\text{to find } \delta \text{ such that } \forall [x, y] \subseteq [a, b], |x - y| < \delta \implies |F(x) - F(y)| < \epsilon$$

Because f is proper-Riemann-Integrable on $[a, b]$, we know f is bounded on $[a, b]$. Let M be an upper bound of $|f|$ on $[a, b]$. We claim

$$\delta = \frac{\epsilon}{M} \text{ works}$$

Because $y - x < \delta = \frac{\epsilon}{M}$, we have

$$\begin{aligned} |F(x) - F(y)| &= \left| \int_x^y f(t)dt \right| \\ &\leq \int_x^y |f(t)| \, dt \\ &\leq (y - x) < \epsilon \text{ (done)} \end{aligned}$$

Now, to prove $F'(x_0) = f(x_0)$, we wish

$$\text{to prove } \lim_{x \to x_0} \frac{F(x) - F(x_0)}{x - x_0} = f(x_0)$$

Fix ϵ. We wish

$$\text{to find } \delta \text{ such that } |x - x_0| < \delta \implies \left| \frac{F(x) - F(x_0)}{x - x_0} - f(x_0) \right| < \epsilon$$

Because f is continuous at x_0, we know

$$\exists \delta, |x - x_0| < \delta \implies |f(x) - f(x_0)| < \epsilon \tag{3.38}$$

We claim

$$\text{such } \delta \text{ in Equation 3.38 works}$$

WOLG, let $x > x_0$. Deduce

$$\begin{aligned}
\left| \frac{F(x) - F(x_0)}{x - x_0} - f(x_0) \right| &= \left| \frac{\int_{x_0}^{x} f(t)dt}{x - x_0} - f(x_0) \right| \\
&= \left| \frac{\int_{x_0}^{x} [f(t) - f(x_0)]dt}{x - x_0} \right| \\
&\leq \frac{\int_{x_0}^{x} |f(t) - f(x_0)|\,dt}{|x - x_0|} \\
&\leq \frac{\int_{x_0}^{x} \epsilon dt}{|x - x_0|} = \epsilon \text{ (done)}
\end{aligned}$$

■

Theorem 3.7.2. (Fundamental Theorem of Calculus: Part 2, Leibniz Rule) Suppose two functions $f, F : [a, \infty) \to \mathbb{R}$ satisfy

(a) f is proper Riemann-Integrable on $[a, b]$ for all $b > a$

(b) $F'(x) = f(x)$ for all $x \in (a, \infty)$

(c) F is continuous on $[a, \infty)$

Then for all $b > a$,

$$\int_a^b f(x)dx = F(b) - F(a)$$

Proof. Fix ϵ. We wish

$$\text{to show that } \left| \Big(F(b) - F(a)\Big) - \int_a^b f(x)dx \right| < \epsilon$$

Because f is proper Riemann-Integrable on $[a,b]$, we know there exists a partition $P = \{a = x_0, x_1, \dots, x_n = b\}$ of $[a,b]$ such that

$$U(P,f) - L(P,f) < \epsilon \tag{3.39}$$

Because $f = F'$ on (a,b), for each $k \in \{1, \dots, n\}$, by MVT (Theorem 3.3.4), we know

$$\exists t_k \in (x_{k-1}, x_k), \frac{F(x_k) - F(x_{k-1})}{x_k - x_{k-1}} = f(t_k)$$

This let us deduce

$$F(b) - F(a) = \sum_{k=1}^{n} F(x_k) - F(x_{k-1}) = \sum_{k=1}^{n} f(t_k)\Delta x_k$$

Now, we have

$$\int_a^b f(x)dx \text{ and } F(b) - F(a) \text{ are both in } \big[L(P,f), U(P,f)\big]$$

Then by Equation 3.39, we can deduce

$$\left| F(b) - F(a) - \int_a^b f(x)dx \right| < \epsilon \text{ (done)}$$

■

Theorem 3.7.3. (Integral By Part) Given four function $f, g, F, G : [a,b] \to \mathbb{R}$ such that

(a) $F'(x) = f(x)$ and $G'(x) = g(x)$ for all $x \in (a,b)$

(b) f, g are properly Riemann-Integrable on $[a,b]$

(c) F, G are continuous on $[a,b]$

We have

$$\int_a^b F(x)g(x)dx = FG\Big|_a^b - \int_a^b f(x)G(x)dx \tag{3.40}$$

Proof. To prove Equation 3.40, we only with

$$\text{to prove } \int_a^b F(x)g(x)dx + \int_a^b f(x)G(x)dx = FG\Big|_a^b$$

We can reduce the problem

$$\text{into proving } \int_a^b (Fg + fG)dx = FG\Big|_a^b$$

Notice that by Chain Rule,

$$\big(FG\big)'(x) = F(x)g(x) + f(x)G(x) \text{ for all } x \in (a, b)$$

Then the result follows from Part 2 of Fundamental Theorem of Calculus (Theorem 3.7.2).
(done) ∎

3.8 Uniform Convergence on Integration and Differentiation

Theorem 3.8.1. (Riemann-Integration and Uniform Convergence) Given a function $\alpha : [a,b] \to \mathbb{R}$ and a sequence of functions $f_n : [a,b] \to \mathbb{R}$ such that

(a) α increase on $[a,b]$

(b) $\int_a^b f_n d\alpha$ exists for all $n \in \mathbb{N}$

(c) $f_n \to f$ uniformly on $[a,b]$

Then

$$\lim_{n\to\infty} \int_a^b f_n d\alpha \text{ exists and } \int_a^b f d\alpha = \lim_{n\to\infty} \int_a^b f_n d\alpha$$

Proof. We first prove

$$\int_a^b f d\alpha \text{ exists}$$

Fix ϵ. We wish to prove

$$\overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha < \epsilon$$

Let $\epsilon_n = \|f_n - f\|_\infty$. Because $f_n \to f$ uniformly, we know

$$\text{there exists } n \in \mathbb{N} \text{ such that } \epsilon_n = \|f_n - f\|_\infty < \frac{\epsilon}{2\big[\alpha(b) - \alpha(a)\big]}$$

Because α increase, by definition of ϵ_n, we see

$$\int_a^b (f_n - \epsilon_n) d\alpha \leq \underline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f d\alpha \leq \int_a^b (f_n + \epsilon_n) d\alpha$$

Because $\epsilon_n < \frac{\epsilon}{2\big[\alpha(b)-\alpha(a)\big]}$, we now see

$$\begin{aligned}\overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha &\leq \int_a^b (f_n + \epsilon_n) d\alpha - \int_a^b (f_n - \epsilon_n) d\alpha \\ &= \int_a^b (2\epsilon_n) d\alpha < 2\epsilon_n \cdot \big[\alpha(b) - \alpha(a)\big] = \epsilon \text{ (done)}\end{aligned}$$

We now prove

$$\int_a^b f_n d\alpha \to \int_a^b f d\alpha \text{ as } n \to \infty$$

Fix ϵ. We wish

$$\text{to find } N \text{ such that } \forall n > N, \left| \int_a^b f_n d\alpha - \int_a^b f d\alpha \right| < \epsilon$$

Recall the definition $\epsilon_n = \|f_n - f\|_\infty$. Because $\epsilon_n \to 0$, we know

$$\text{there exists } N \text{ such that } \forall n > N, \epsilon_n < \frac{\epsilon}{\alpha(b) - \alpha(a)} \tag{3.41}$$

We claim

$$\text{such } N \text{ works}$$

Fix $n > N$. From Equation 3.41, we see

$$\begin{aligned}
\left| \int_a^b f_n d\alpha - \int_a^b f d\alpha \right| &= \left| \int_a^b (f_n - f) d\alpha \right| \\
&\leq \int_a^b |f_n - f| \, d\alpha \\
&\leq \int_a^b \epsilon_n d\alpha = \epsilon_n \big[\alpha(b) - \alpha(a)\big] < \epsilon \text{ (done)}
\end{aligned}$$

■

Before the next Theorem, let's see three examples why this time we don't (can't) use the hypothesis: $f_n \to f$ uniformly.

Example 4 (Differentiable functions are NOT closed under uniform convergence)

$$X = [-1, 1] \text{ and } f(x) = |x|$$

By Weierstrass approximation Theorem, there is a sequence of polynomials (differentiable) that uniformly converge to f, which is not differntiable at 0.

Example 5 (Derivative won't necessarily converge to the right place)

$$X = \mathbb{R} \text{ and } f_n(x) = \frac{\sin nx}{\sqrt{n}}$$

Compute

$$f'(x) = 0 \text{ and } f_n'(x) = \sqrt{n}\cos nx$$

Example 6 (Derivative won't necessarily converge to the right place)

$$X = \mathbb{R} \text{ and } f_n(x) = \frac{x}{1+nx^2}$$

Compute

$$f = \tilde{0} \text{ and } f_n'(0) = 1$$

Informally speaking, these examples together with the fact integral are closed under uniform convergence (Theorem 3.8.1) should give you some ideas that differentiation and integration although are operations inverse to each other, are NOT symmetric. There is a certain hierarchy on continuous functions on a fixed compact interval. Thus, we have the next Theorem in its form. Note that in application, the next Theorem only require us to prove f_n' uniformly converge, and doesn't require us to prove to where does it converge.

Theorem 3.8.2. (Uniform Convergence and Differentiation) Given a sequence of function $f_n : [a,b] \to \mathbb{R}$ such that

(a) f_n' uniformly converge on (a,b)

(b) f_n are continuous on $[a,b]$

(c) $f_n(x_0) \to L$ for some $x_0 \in [a,b]$

Then

(a) f_n uniformly converge on $[a,b]$

(b) and

$$\Big(\lim_{n\to\infty} f_n\Big)'(x_0) = \lim_{n\to\infty} f_n'(x_0) \text{ on } (a,b)$$

Proof. We first prove

$$f_n \text{ uniformly converge on } [a,b] \tag{3.42}$$

Fix ϵ. We wish

$$\text{to find } N \text{ such that } \|f_n - f_m\|_\infty \le \epsilon \text{ for all } n, m > N$$

Because $f_n(x_0)$ converge, and f'_n uniformly converge, we know there exists N such that

$$\begin{cases} |f_n(x_0) - f_m(x_0)| < \frac{\epsilon}{2} \\ \|f'_n - f'_m\|_\infty < \frac{\epsilon}{2(b-a)} \end{cases} \quad \text{for all } n, m > N \tag{3.43}$$

We claim

$$\text{such } N \text{ works}$$

Fix $x \in [a, b]$ and $n, m > N$. We need

$$\text{to show } |f_n(x) - f_m(x)| \le \epsilon$$

We first prove

$$|f_n(x) - f_m(x) - f_n(x_0) + f_m(x_0)| \le \frac{\epsilon}{2}$$

Because $(f_n - f_m)' = f'_n - f'_m$, by MVT (Theorem 3.3.4) and Equation 3.43, we can deduce

$$\begin{aligned} |f_n(x) - f_m(x) - f_n(x_0) + f_m(x_0)| &= |(f_n - f_m)(x) - (f_n - f_m)(x_0)| \\ &= \Big|\big[(f_n - f_m)'(t)\big](x - x_0)\Big| \text{ for some } t \text{ between } x, x_0 \\ &< \frac{\epsilon}{2(b-a)} \cdot |x - x_0| \\ &\le \frac{\epsilon}{2(b-a)} \cdot (b - a) = \frac{\epsilon}{2} \ \big(\because x, x_0 \in [a, b]\big) \text{ (done)} \end{aligned}$$

Now, by Equation 3.43, we have

$$\begin{aligned} |f_n(x) - f_m(x)| &\le |f_n(x) - f_m(x) - f_n(x_0) + f_m(x_0)| + |f_n(x_0) - f_m(x_0)| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \text{ (done)} \end{aligned}$$

We claim

$$f(x) \triangleq \lim_{n\to\infty} f_n(x) \text{ satisfy } f'(x) = \lim_{n\to\infty} f'_n(x) \text{ on } (a, b) \tag{3.44}$$

We first show

$$f \text{ is differentiable on } (a, b)$$

Fix $x \in (a,b)$. We wish to prove

$$\lim_{t\to x} \frac{f(t)-f(x)}{t-x} \text{ exists}$$

Define $\phi : [a,b] \setminus x \to \mathbb{R}$ by

$$\phi(t) \triangleq \frac{f(t)-f(x)}{t-x}$$

We reduce our problem into proving

$$\lim_{t\to x} \phi(t) \text{ exists}$$

Set $\phi_n : [a,b] \setminus x \to \mathbb{R}$ by

$$\phi_n(t) \triangleq \frac{f_n(t)-f_n(x)}{t-x}$$

We first show

$$\phi_n \text{ uniformly converge on } [a,b] \setminus x \tag{3.45}$$

Fix ϵ. We have

$$\text{to find } N \text{ such that } |\phi_n(t) - \phi_m(t)| \leq \epsilon \text{ for all } n,m > N \text{ and } t \in [a,b] \setminus x$$

Because f_n' uniformly converge on $[a,b]$, we know there exists N such that

$$\|f_n' - f_m'\|_\infty \leq \epsilon \text{ for all } n,m > N \tag{3.46}$$

We claim

$$\text{such } N \text{ works}$$

Fix $n,m > N$ and $t \in [a,b] \setminus x$. We wish to prove

$$|\phi_n(t) - \phi_m(t)| \leq \epsilon$$

Because $(f_n - f_m)' = f_n' - f_m'$, by MVT (Theorem 3.3.4) and Equation 3.46, we can deduce

$$\begin{aligned}
|\phi_n(t) - \phi_m(t)| &\leq \left| \frac{f_n(t)-f_n(x)}{t-x} - \frac{f_m(t)-f_m(x)}{t-x} \right| \\
&= \left| \frac{\big(f_n - f_m\big)(t) - \big(f_n - f_m\big)(x)}{t-x} \right| \\
&= \left| \big(f_n' - f_m'\big)(t_0) \right| \text{ for some } t_0 \text{ between } t,x \\
&\leq \epsilon \text{ (done)}
\end{aligned}$$

We now show

$$\phi_n \to \phi \text{ pointwise on } [a,b] \setminus x \tag{3.47}$$

Because $f_n \to f$ on $[a,b]$ by definition (Equation 3.44), (the convergence is in fact uniform as we have shown. This doesn't matter here tho), for each $t \in [a,b] \setminus x$, we can deduce

$$\lim_{n\to\infty} \phi_n(t) = \lim_{n\to\infty} \frac{f_n(t) - f_n(x)}{t-x} = \frac{f(t)-f(x)}{t-x} = \phi(t) \text{ (done)}$$

Now, by Equation 3.45 and Equation 3.47, we know

$$\phi_n \to \phi \text{ uniformly on } [a,b] \setminus x$$

Notice that because $f_n'(x)$ converge, we know

$$\lim_{n\to\infty} \lim_{t\to x} \phi_n(t) = \lim_{n\to\infty} f_n'(x) \text{ exists}$$

Then (Notice that the second equality below hold true because we have known $\lim_{n\to\infty} \lim_{t\to x} \phi_n(t)$ exists), we can finally deduce

$$\begin{aligned}\lim_{t\to x} \phi(t) &= \lim_{t\to x} \lim_{n\to\infty} \phi_n(t) \\ &= \lim_{n\to\infty} \lim_{t\to x} \phi_n(t) \\ &= \lim_{n\to\infty} f_n'(x) \text{ exists} \quad \text{(done)}\end{aligned}$$

Now, notice that $f'(x) = \lim_{t\to x} \phi(t)$, so in fact, we have just proved $f_n' \to f'$ (done) (done)

■

As Rudin remarked, a much shorter (and much more intuitive) proof can be given, if we require f' to be continuous on $[a,b]$.

Theorem 3.8.3. (Uniform Convergence and Differentiation: Weaker Version) Given a sequence of function $f_n : [a,b] \to \mathbb{R}$ such that

(a) f_n' uniformly converge on $[a,b]$

(b) $f_n(x_0) \to L$ for some $x_0 \in [a,b]$

(c) f_n are of class C^1 on $[a,b]$

Then

(a) f_n uniformly converge on $[a,b]$

(b) and

$$\frac{d}{dx}\Big(\lim_{n\to\infty} f_n(x)\Big)\Big|_{x=x_0} = \lim_{n\to\infty} f_n'(x_0) \text{ on } (a,b)$$

Proof. We claim

$$f(x) = \lim_{n\to\infty}\int_{x_0}^{x} f_n'(t)dt + L \text{ works}$$

Note that $\lim_{n\to\infty}\int_{x_0}^{x} f_n'(t)dt$ exists because f_n' uniformly converge (Theorem 3.8.1).

Because f_n' uniformly converge and are continuous on $[a,b]$, by ULT, we know

$$\int_{x_0}^{x}\lim_{n\to\infty} f_n'(t)dt + L \text{ exists}$$

and know

$$f(x) = \int_{x_0}^{x}\lim_{n\to\infty} f_n'(t)dt + L$$

By FTC, we see

$$f'(x) = \lim_{n\to\infty} f_n'(x) \text{ on } (a,b)$$

Such convergence is uniform by premise. To finish the proof, we now only have to prove

$$f_n \to f \text{ uniformly on } [a,b]$$

Fix ϵ. We wish

$$\text{to find } N \text{ such that } |f_n(x) - f(x)| \le \epsilon \text{ for all } n > N \text{ and } x \in [a,b]$$

Because $f_n' \to f'$ uniformly, and $f_n(x_0) \to L = f(x_0)$ (Check $L = f(x_0)$), we know there exists N such that

$$\begin{cases} \|f_n' - f'\|_\infty < \frac{\epsilon}{2(b-a)} \\ |f_n(x_0) - f(x_0)| < \frac{\epsilon}{2} \end{cases} \quad \text{for all } n > N$$

We claim

$$\text{such } N \text{ works}$$

Fix $n > N$ and $x \in [a, b]$. Observe

$$\begin{aligned}
|f(x) - f_n(x)| &= \left| \int_{x_0}^{x} \big(f'(t) - f_n'(t)\big) dt + f(x_0) - f_n(x_0) \right| \\
&\le \int_{x_0}^{x} |f'(t) - f_n'(t)| \, dt + |f(x_0) - f_n(x_0)| \\
&\le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \text{ (done)}
\end{aligned}$$

■

3.9 Fundamental Tests*

Theorem 3.9.1. **(LCT)**

Theorem 3.9.2. **(Geometric Series)**

Theorem 3.9.3. **(p-Series)**

Theorem 3.9.4. **(Limit Superior Lemma 1)**

$$d_n \text{ converge} \implies \limsup_{n\to\infty} c_n d_n = \lim_{n\to\infty} d_n \limsup_{n\to\infty} c_n$$

Theorem 3.9.5. **(Limit Superior Lemma 2)**

$$\limsup_{n\to\infty} \sqrt[n]{c_n} = \limsup_{n\to\infty} \sqrt[n]{c_{n+k}}$$

Theorem 3.9.6. **(Ratio Test)**

Theorem 3.9.7. **(Root Test)**

Theorem 3.9.8. **(Root Test is Stronger Than Ratio Test)**

Theorem 3.9.9. **(Absolute Convergent Series Converge)**

Theorem 3.9.10. **(Absolute Convergent Series Unconditionally Converge)**

Theorem 3.9.11. **(Fubini's Theorem for Infinite Series)**

Theorem 3.9.12. **(Summation by Part)**

Theorem 3.9.13. **(Alternating Series Test)**

Theorem 3.9.14. **(Abel's Test)**

Theorem 3.9.15. **(Dirichlet's Test)**

Theorem 3.9.16. (Merten's Theorem for Cauchy Product) Suppose

(a) $\sum_{n=0}^{\infty} a_n$ converge absolutely

(b) $\sum_{n=0}^{\infty} a_n = A$

(c) $\sum_{n=0}^{\infty} b_n = B$

(d) $c_n = \sum_{k=0}^{n} a_k b_{n-k}$

Then we have

$$\sum_{n=0}^{\infty} c_n = AB$$

Theorem 3.9.17. (Special Sequence) $(n!)^{\frac{1}{n}}$

3.10 Analytic Functions

In this section, by a **real power series**, we mean a pair (a, c_n) where $a \in \mathbb{R}$ is called the **center** of power series, and $c_n \in \mathbb{R}$ are the coefficients sequence. By **radius of convergence**, we mean a unique $R \in \mathbb{R}_0^+ \cup \infty$ such that

$$\sum_{n=0}^{\infty} c_n (x-a)^n \begin{cases} \text{converge absolutely} & \text{if } |x-a| < R \\ \text{diverge} & \text{if } |x-a| > R \end{cases}$$

Such R always exist (and is unique, this fact can be checked without computing the actual value of R) and is exactly

$$R = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{c_n}} \tag{3.48}$$

This result is called **Cauchy-Hadamard Theorem**. It can be directly proved by applying Root Test to $\sum c_n(z-a)^n$. For this, we say $(a-R, a+R)$ is the **interval of convergence**. Note that Cauchy-Hadamard Theorem does not tell us whether a power series converges at points of boundary of disk of convergence. It require extra works to determine if the power series converge at endpoints. Following is an example of such discussion.

Example 7 (Discussion of Convergence on Boundary)

$$f_q(z) = \sum_{n=0}^{\infty} n^q z^n \text{ provided } q \in \mathbb{R}$$

It is clear that f_q has convergence radius 1 for all $q \in \mathbb{R}$. For boundary, we have

$$\begin{cases} q < -1 \implies f_q \text{ converge on } S^1 \\ -1 \leq q < 0 \implies f_q \text{ converge on } S^1 \setminus \{1\} \\ 0 \leq q \implies f_q \text{ diverge on } S^1 \end{cases}$$

At $z = 1$, the discussion is just p-series. On $S^1 \setminus \{1\}$, the discussion use Dirichlet's Test, where boundedness of $\sum_{k=0}^{n} z^k$ is proved by geometric formula.

$$\left| \sum_{k=1}^{n} z^k \right| = \left| \sum_{k=1}^{n} e^{ik\theta} \right| = \frac{|e^{i\theta} - e^{i(n+1)\theta}|}{|1-e^{i\theta}|} \leq \frac{2}{|1-e^{i\theta}|}$$

Notice that the fact $\sum c_n(z-a)^n$ absolutely converge in $(a-R,a+R)$ implies the convergence is uniform on all $[a-R+\epsilon,a+R-\epsilon]$ by M-Test. However, on $(a-R,a+R)$, the convergence is not always uniform.

Example 8 (Failure of Uniform Convergence on $(a-R,a+R)$**)**

$$f(z)=\sum_{n=0}^{\infty} z^n$$

Note $R=1$. Use Geometric Series Formula to show $f(x)=\frac{1}{1-x}$ on $(-1,1)$. It is then clear that f is unbounded on $(-1,1)$ while all partial sums $\sum_{k=0}^{n} z^k$ is bounded on $(-1,1)$.

We now introduce some terminologies. We say a real function f is **real analytic at** $a\in\mathbb{R}$ if there exists a power series (a,c_n) such that f agrees with $\sum_{n=0}^{\infty} c_n(z-a)^n$ on $(a-R,a+R)$ for some R (of course, such R must not be strictly greater than the radius of convergence of (a,c_n)).

It shall be quite clear that if f,g are both analytic at $a\in\mathbb{R}$ with radius $R_f\leq R_g$, then $f+g$ and fg are both analytic at a with radius at least R_f. (the fact fg is analytic at a with radius at least R_f is an immediate consequence of Merten's Theorem)

We now investigate deeper into real analytic functions. We first prove that real analytic functions are smooth, that is, $C^\omega(I)\subseteq C^\infty(I)$ on open $I\subseteq\mathbb{R}$.

Theorem 3.10.1. (Analytic functions are Smooth) Given a power series (a,c_n) of convergence radius R, if we define $f:D_R(a)\to\mathbb{C}$ by

$$f(z)=\sum_{n=0}^{\infty} c_n(z-a)^n$$

Then

(a) f is of class C^∞ on $D_R(a)$

(b) $f^{(k)}(z)=\sum_{n=k}^{\infty}\frac{n!}{(n-k)!}c_n(z-a)^{n-k}$

Proof. We prove by induction. Base case $k=0$ is trivial. Fix $k\geq 0$. Suppose we have

$$f^{(k)}(z)=\sum_{n=k}^{\infty}\frac{n!}{(n-k)!}c_n(z-a)^{n-k} \text{ on } D_R(a)$$

We are required to prove

$$f^{(k+1)}(z) = \sum_{n=k+1}^{\infty} \frac{n!}{(n-k-1)!} c_n (z-a)^{n-k-1} \text{ on } D_R(a)$$

Set f_m

$$f_m(z) \triangleq \sum_{n=k}^{k+m} \frac{n!}{(n-k)!} c_n (z-a)^{n-k}$$

We have

$$f_m \to f^{(k)} \text{ pointwise on } D_R(a) \text{ and } f'_m(z) = \sum_{n=k+1}^{k+m} \frac{n!}{(n-k-1)!} c_n (z-a)^{n-k-1} \tag{3.49}$$

We abstract our problem into proving

$$f'_m \to f^{(k+1)} \text{ pointwise on } D_R(a)$$

Fix $z_0 \in D_R(a)$. We only wish to prove

$$(f^{(k)})'(z_0) = \lim_{m\to\infty} f'_m(z_0)$$

Fix ϵ such that $|z_0 - a| < R - \epsilon$. By Equation 3.49, using Theorem 3.8.2 (Uniform Convergence and Differentiaiton). We only have to prove

$$f'_m \text{ uniformly converge on } \overline{D}_{R-\epsilon}$$

Note that

$$f'_m(z) = \sum_{n=0}^{m-1} \frac{(n+k+1)!}{n!} c_{n+k+1} (z-a)^n$$

so we can compute the radius of convergence for f'_m

$$\begin{aligned}\limsup_{n\to\infty} \sqrt[n]{\frac{(n+k+1)!}{n!} |c_{n+k+1}|} &= \limsup_{n\to\infty} \sqrt[n]{|c_{n+k+1}|} \\ &= \limsup_{n\to\infty} \sqrt[n]{|c_n|} = R\end{aligned}$$

Together by Cauchy-Hadamrd (absolute convergent on $a + R - \epsilon$) and M-test show that

$$\sum_{n=0}^{\infty} \frac{(n+k+1)!}{n!} c_{n+k+1} (z-a)^n \text{ uniformly converge on } \overline{D}_{R-\epsilon}(a) \quad \text{(done)}$$

$\blacksquare$

Now by Theorem 3.10.1, if we are given a real function f analytic at a, the power series representation $(a, c_n) : \sum_{n=0}^{\infty} c_n(z-a)^n = f$ must be unique, since f is proved to be infinitely differentiable at a and proved to satisfy $c_k = \frac{f^{(k)}(a)}{k!}$.

Notice that the arguments above are all based on the hypothesis that f is analytic, and that smoothness does not imply analytic. See the following examples.

Example 9 (Smooth but not Analytic Function)

$$f(x) = \begin{cases} e^{\frac{-1}{x^2}} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Use induction to show that

$$f^{(k)}(x) = P_k(\frac{1}{x})e^{-(\frac{1}{x})^2} \quad \exists P_k \in \mathbb{R}[x^{-1}], \forall k \in \mathbb{Z}_0^+, \forall x \in \mathbb{R}^*$$

and again use induction to show that

$$f^{(k)}(0) = 0 \quad \forall k \in \mathbb{Z}_0^+$$

The trick to show $f^{(k)}(0) = 0$ is let $u = \frac{1}{x}$.

Now, with Theorem 3.10.1, we see that f is not analytic at 0.

Example 10 (Bump Function)

$$f(x) = \begin{cases} e^{\frac{-1}{1-x^2}} & \text{if } |x| < 1 \\ 0 & \text{otherwise} \end{cases}$$

Use the same trick (but more advanced) to show f is smooth, and note that f is not analytic at ± 1.

Now, it comes an interesting question. Given a real function f analytic at a with radius R, and suppose $b \in (a-R, a+R)$.

(a) Is f also analytic at b?

(b) What do we know about the radius of convergence of f at b?

(c) Suppose f is indeed analytic at b. It is trivial to see that the power series $(a, c_{a;n})$ and $(b, c_{b;n})$ must agree on the common convergence interval, and because f is given, we by Theorem 3.10.1, have already known the value of $c_{b;n}$. Can we verify that the power series $(a, c_{a;n})$ and $(a, c_{b;n})$ do indeed agree with each other on the common convergence interval?

Theorem 3.10.2 (Taylor's Theorem) give satisfying answers to these problems.

Theorem 3.10.2. (Taylor's Theorem) Given a real function f analytic at a with radius R, and suppose $b \in (a - R, a + R)$. Then

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(b)}{k!}(x-b)^k \text{ on } |x-b| < R - |b-a|$$

Proof. WOLG, let $a = 0$. Suppose x satisfy $|x - b| < R - |b|$. Compute

$$\begin{aligned}
f(x) &= \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} x^k \\
&= \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-b+b)^k \\
&= \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} \sum_{n=0}^{k} \binom{k}{n} (x-b)^n b^{k-n} \\
&= \sum_{k=0}^{\infty} \sum_{n=0}^{k} \frac{f^{(k)}(a)}{k!} \binom{k}{n} (x-b)^n b^{k-n}
\end{aligned}$$

Note that

$$\begin{aligned}
\sum_{k=0}^{\infty} \left| \sum_{n=0}^{\infty} \frac{f^{(k)}(a)}{k!} \binom{k}{n} (x-b)^n b^{k-n} \right| &\le \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \left| \frac{f^{(k)}(a)}{k!} \right| \binom{k}{n} |x-b|^n \cdot |b|^{k-n} \\
&= \sum_{k=0}^{\infty} \left| \frac{f^{(k)}(a)}{k!} \right| \sum_{n=0}^{\infty} \binom{k}{n} |x-b|^n \cdot |b|^{k-n} \\
&= \sum_{k=0}^{\infty} \left| \frac{f^{(k)}(a)}{k!} \right| \big(|x-b| + |b| \big)^k
\end{aligned}$$

converge, by Cauchy-Hadamard Theorem and $|x - b| + |b| < R$.

Now, using Fubini's Theorem for Infinite Series, we have

$$\begin{aligned}
\sum_{k=0}^{\infty}\sum_{n=0}^{k}\frac{f^{(k)}(a)}{k!}\binom{k}{n}(x-b)^n b^{k-n} &= \sum_{k=0}^{\infty}\sum_{n=0}^{\infty}\frac{f^{(k)}(a)}{k!}\binom{k}{n}(x-b)^n b^{k-n} \\
&= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\frac{f^{(k)}(a)}{k!}\binom{k}{n}(x-b)^n b^{k-n} \\
&= \sum_{n=0}^{\infty}\sum_{k=n}^{\infty}\frac{f^{(k)}(a)}{k!}\binom{k}{n}(x-b)^n b^{k-n} \\
&= \sum_{n=0}^{\infty}\Big[\sum_{k=n}^{\infty}\frac{f^{(k)}(a)}{k!}\binom{k}{n}b^{k-n}\Big](x-b)^n
\end{aligned}$$

We have reduced the problem into proving

$$\sum_{k=n}^{\infty}\frac{f^{(k)}(a)}{k!}\binom{k}{n}b^{k-n} = \frac{f^{(n)}(b)}{n!}$$

Using the formula in Theorem 3.10.1, because b is in $(a-R, a+R)$, we can compute

$$\begin{aligned}
f^{(n)}(b) &= \sum_{k=n}^{\infty}\frac{k!}{(k-n)!}c_{a;k}(b)^{k-n} \\
&= \sum_{k=n}^{\infty}\frac{k!}{(k-n)!}\cdot\frac{f^{(k)}(a)}{k!}\cdot b^{k-n} \\
&= \sum_{k=n}^{\infty}\frac{f^{(k)}(a)}{(k-n)!}b^{k-n}
\end{aligned}$$

This now implies

$$\begin{aligned}
\frac{f^{(n)}(b)}{n!} &= \sum_{k=n}^{\infty}\frac{f^{(k)}(a)}{n!(k-n)!}b^{k-n} \\
&= \sum_{k=n}^{\infty}\frac{f^{(k)}(a)}{k!}\binom{k}{n}b^{k-n} \text{ (done)}
\end{aligned}$$

■

3.11 Abel's Theorem and its application

In this section, we use the notation $\mathbb{S}_M(R)$ to denote **stolz region**

$$\mathbb{S}_M(R) \triangleq \{z \in \mathbb{C} : \frac{|R-z|}{R-|z|} \in (0, M)\}$$

Theorem 3.11.1. (Abel's Theorem for Power Series) Given a complex Maclaurin series $f(z) = \sum_{n=0}^{\infty} c_n z^n$ of convergence radius R such that

$$\sum_{n=0}^{\infty} c_n R^n \text{ converge}$$

Then for all $M > 1$, we have

$$f|_{\mathbb{S}_M(R)}(z) \to \sum_{n=0}^{\infty} c_n R^n = f(R) \text{ as } z \to R$$

Proof. We first

$$\text{prove when } R = 1$$

Fix ϵ. We wish

$$\text{to find } \delta \text{ such that } \left| \sum_{n=0}^{\infty} c_n z^n - c_n \right| < \epsilon \text{ for all } z \in \mathbb{S}_M(1) \cap D_\delta(1)$$

To use summation by part, we first fix

$$s_n \triangleq \sum_{k=0}^{n} c_k \text{ and } s \triangleq \lim_{n\to\infty} s_n$$

Now Use summation by part

$$\begin{aligned}
\sum_{n=0}^{k} c_n z^n &= \sum_{n=0}^{k} (s_n - s_{n-1}) z^n \\
&= \sum_{n=0}^{k} s_n z^n - \sum_{n=0}^{k-1} s_n z^{n+1} \\
&= s_k z^k + (1-z) \sum_{n=0}^{k-1} s_n z^n
\end{aligned}$$

Note that

$$(1-z)\sum_{n=0}^{\infty} z^n = 1 \quad (|z|<1)$$

This give us

$$\begin{aligned}\lim_{z\to 1^-}\left(\sum_{n=0}^{\infty} c_n z^n - \sum_{n=0}^{\infty} c_n\right) &= \lim_{z\to 1^-}\left(\lim_{k\to\infty} s_k z^k + (1-z)\sum_{n=0}^{k-1} s_n z^n - s\right) \\ &= \lim_{z\to 1^-}(1-z)\sum_{n=0}^{\infty}(s_n - s)z^n \quad (\because \forall z\in\mathbb{C}: |z|<1, \lim_{k\to\infty} s_k z^k = 0)\end{aligned}$$

We reduce the problem into

$$\text{finding } \delta \text{ such that } \left|(1-z)\sum_{n=0}^{\infty}(s_n-s)z^n\right| \leq \epsilon \text{ for all } z\in\mathbb{S}_M(1)\cap D_\delta(1)$$

Because $s_n \to s$, we know there exists N such that $|s_n - s| < \frac{1}{2M}$ for all $n > N$. We claim

$$\delta = \frac{\epsilon}{2\sum_{n=0}^{N}|s_n - s|} \text{ suffices}$$

Note that $\sum_{n=0}^{\infty}(s_n - s)z^n$ absolutely converges by direct comparison test. Then we can deduce

$$\begin{aligned}\left|(1-z)\sum_{n=0}^{\infty}(s_n-s)z^n\right| &= |1-z|\cdot\left|\sum_{n=0}^{N}(s_n-s)z^n + \sum_{n=N+1}^{\infty}(s_n-s)z^n\right| \\ &\leq |1-z|\left(\sum_{n=0}^{N}|s_n-s| + \frac{\epsilon}{2M}\sum_{n=N+1}^{\infty}|z|^n\right) \\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2M}\cdot\frac{|1-z|}{1-|z|}\cdot|z|^{N+1} \leq \epsilon \quad (\because |z|<1 \text{ and } \frac{|1-z|}{1-|z|} < M) \text{ (done)}\end{aligned}$$

We now prove

$$\text{when } R\in\mathbb{R}^+$$

Fix ϵ. We wish

$$\text{to find } \delta \text{ such that } \left|\sum_{n=0}^{\infty} c_n z^n - c_n R^n\right| < \epsilon \text{ for all } z\in\mathbb{S}_M(R)\cap D_\delta(R)$$

Fix

$$a_n = c_n R^n \text{ and } g(z) \triangleq \sum_{n=0}^{\infty} a_n z^n = \sum_{n=0}^{\infty} c_n R^n z^n \quad (|z| < 1)$$

By premise and our result, we know

$g(1)$ exists and there exists δ' such that $|g(z) - g(1)| < \epsilon$ for all $z \in \mathbb{S}_M(1) \cap D_{\delta'}(1)$

We claim

$$\delta = R\delta' \text{ suffices}$$

First note that

$$\frac{|R - z|}{R - |z|} \in (0, M) \implies \frac{\left|1 - \frac{z}{R}\right|}{1 - \left|\frac{z}{R}\right|} \in (0, M)$$

This tell us

$$z \in \mathbb{S}_M(R) \implies \frac{z}{R} \in \mathbb{S}_M(1)$$

Fix $z \in \mathbb{S}_M(R) \cap D_\delta(R)$. We now have

$$\frac{z}{R} \in \mathbb{S}_M(1) \cap D_{\delta'}(1)$$

This then let us conclude

$$\left| \sum_{n=0}^{\infty} c_n z^n - c_n R^n \right| = \left| g(\frac{z}{R}) - g(1) \right| < \epsilon \text{ (done)}$$

∎

Example 11 (Identity of ln derived from Abel's Theorem)

$$\ln(1 + x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n} \text{ for all } x \in (-1, 1]$$

Check that both side satisfy $y' = \frac{1}{1+x}$, and $y(0) = 0$. This tell us that two sides equal on $(-1, 1)$. Now using Abel's Theorem and the continuity of ln, we have

$$\ln 2 = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}$$

Example 12 ($1 - 1 + 1 - 1 + \cdots = \frac{1}{2}$)

$$1 - 1 + 1 - 1 + \cdots = \frac{1}{2} \text{ is WRONG!!!}$$

When people say: "$1 - 1 + 1 - 1 + \cdots = \frac{1}{2}$", they mean the sum of the series in the sense of Abel. Compute the Macularin series of $\frac{1}{1+r}$

$$\frac{1}{1+r} = \sum_{n=0}^{\infty} (-1)^n r^n$$

Check both side do equal on $(-1, 1)$ by direct computation. Apply Abel's Theorem to see the magic.

Chapter 4

Calculus in Euclidean Space

4.1 Operator Norm

In this section, and particularly in functional analysis, we say a function T between two metric space is a **bounded operator** if T always map bounded set to bounded set. In particular, if T is a linear transformation between two normed space, we say T is a **bounded linear operator**.

Suppose $\mathcal{X}, \mathcal{Y}$ are two normed space over $\mathbb{R}$ or $\mathbb{C}$. In space $L(\mathcal{X}, \mathcal{Y})$, alternatively, we can define the boundedness for each linear transformation T by

$$T \text{ is bounded } \overset{\triangle}{\iff} \exists M \in \mathbb{R}, \forall x \in \mathcal{X}, \|Tx\| \leq M\|x\|$$

The proof of equivalency is simple. For $(\longrightarrow)$, let $E \triangleq \{y \in \mathcal{X} : \|y\| = 1\}$ is non-empty. Clearly, E is bounded. Let $M = \sup_{y \in E} \|Ty\|$. We now have

$$\|Tx\| = \|x\| \cdot \|T\frac{x}{\|x\|}\| \leq M\|x\|$$

For $(\longleftarrow)$, use $\|Tx - Ty\| = \|T(x-y)\| \leq M\|x-y\|$.

We first show that a linear transformation is continuous if and only if it is bounded. (Theorem 4.1.1)

Theorem 4.1.1. (Liner Operator is Bounded if and only if it is Continuous) Given two normed space $\mathcal{X}, \mathcal{Y}$ over $\mathbb{R}$ or $\mathbb{C}$ and $T \in L(\mathcal{X}, \mathcal{Y})$, we have

$$T \text{ is a bounded operator } \iff T \text{ is continuous on } \mathcal{X}$$

Proof. $(\longrightarrow)$ We show

$$T \text{ is Lipschitz continuous on } V$$

Because T is bounded, we can let $M \in \mathbb{R}^+$ satisfy $\|Tx\| \leq M\|x\|$. We see

$$\|Tx - Ty\| \leq \|T(x-y)\| \leq M\|x-y\| \text{ (done)}$$

$(\longleftarrow)$

Because T is linear and continuous at 0, we know there exists ϵ such that

$$\sup_{\|y\|\leq\epsilon} \|Ty\| \leq 1$$

We claim

$$\|Tx\| \leq \frac{1}{\epsilon}\|x\| \qquad (x \in \mathcal{X})$$

Fix $x \in V$. Compute

$$\|Tx\| = \frac{\|x\|}{\epsilon}\Big\|T\frac{\epsilon x}{\|x\|}\Big\| \leq \frac{\|x\|}{\epsilon} \text{ (done)}$$

∎

Here, we introduce a new terminology, which shall later show its value. Given a set X, we say two metrics d_1, d_2 on X are **equivalent**, and write $d_1 \sim d_2$, if we have

$$\exists m, M \in \mathbb{R}^+, \forall x, y \in X, md_1(x,y) \leq d_2(x,y) \leq Md_1(x,y)$$

Now, given a fixed vector space V, naturally, we say two norms $\|\cdot\|_1, \|\cdot\|_2$ on V are **equivalent** if

$$\exists m, M \in \mathbb{R}^+, \forall x \in X, m\|x\|_1 \leq \|x\|_2 \leq M\|x\|_1$$

We say two metric d_1, d_2 on X are **topologically equivalent** if the topology they induce on X are identical.

A few properties can be immediately spotted.

(a) Our definition of $\sim$ between metrics of a fixed X is an equivalence relation.

(b) Our definition of $\sim$ between norms on a fixed V is an equivalence relation.

(c) Equivalent norms induce equivalent metrics.

(d) Equivalent metrics are topologically equivalent.

We now prove that if V is finite-dimensional, then all norms on V are equivalent (Theorem 4.1.2). This property will later show its value, as used to prove that linear map of finite-dimensional domain is alwasy continuous (Theorem 4.1.3)

Theorem 4.1.2. (All Norms on Finite-dimensional space are Equivalent) Suppose V is a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$. Then

$$\text{all norms on } V \text{ are equivalent}$$

Proof. Let $\{e_1, \dots, e_n\}$ be a basis of V. Define ∞-norm $\|\cdot\|_\infty$ on V by

$$\left\| \sum \alpha_i e_i \right\|_\infty \triangleq \max |\alpha_i|$$

It is easily checked that $\|\cdot\|_\infty$ is indeed a norm. Fix a norm $\|\cdot\|$ on V. We reduce the problem into

$$\text{finding } m, M \in \mathbb{R}^+ \text{ such that } m\|x\|_\infty \leq \|x\| \leq M\|x\|_\infty$$

We first claim

$$M = \sum \|e_i\| \text{ suffices}$$

Compute

$$\|x\| = \left\| \sum \alpha_i e_i \right\| \leq \sum |\alpha_i| \, \|e_i\| \leq \|x\|_\infty \sum \|e_i\| = M\|x\|_\infty \text{ (done)}$$

Reverse triangle inequality give us

$$\Big| \|x\| - \|y\| \Big| \leq \|x - y\| \leq M\|x\|_\infty$$

This implies that $\|\cdot\| : \Big(V, \|\cdot\|_\infty\Big) \to \mathbb{R}$ is Lipschitz continuous.

Note that $S \triangleq \{y \in V : \|y\|_\infty = 1\}$ is non-empty. Check that S is compact in $\|\cdot\|_\infty$ by checking S is sequentially compact using the fact $\mathbb{R}^{n-1}$ is locally compact.

Now, by EVT, we know $\min_{y \in S} \|y\|$ exists. Note that $\min_{y \in S} \|y\| > 0$, since $0 \notin S$.

We claim

$$m = \min_{y \in S} \|y\| \text{ suffices}$$

Fix $x \in V$. Compute

$$m\|x\|_\infty = \|x\|_\infty (\min_{y \in S} \|y\|) \leq \|x\|_\infty \cdot \left\| \frac{x}{\|x\|_\infty} \right\| = \|x\| \text{ (done)} \quad \text{(done)}$$

∎

Theorem 4.1.3. (Linear map of Finite-dimensional Domain is always Continuous) Given a finite-dimensional normed space $\mathcal{X}$ over $\mathbb{R}$ or $\mathbb{C}$, an arbitrary normed space $\mathcal{Y}$ over $\mathbb{R}$ or $\mathbb{C}$ and a linear transformation $T : \mathcal{X} \to \mathcal{Y}$, we have

$$T \text{ is continuous}$$

Proof. Fix $x \in \mathcal{X}, \epsilon$. We wish

$$\text{to find } \delta \text{ such that } \forall h \in \mathcal{X} : \|h\| \leq \delta, \|T(x+h) - Tx\| \leq \epsilon$$

Let $\{e_1, \dots, e_n\}$ be a basis of $\mathcal{X}$. Note that $\|\sum \alpha_i e_i\|_1 := \sum |\alpha_i|$ is a norm. By Theorem 4.1.2, we know $\|\cdot\|$ and $\|\cdot\|_1$ are equivalent. Then, we can fix $M \in \mathbb{R}^+$ such that

$$\|x\|_1 \leq M\|x\| \quad (x \in V)$$

We claim

$$\delta = \frac{\epsilon}{M(\max \|Te_i\|)} \text{ suffices}$$

Fix $\|h\| \leq \delta$ and express $h = \sum \alpha_i e_i$. Compute using linearity of T

$$\begin{aligned}\|T(x+h) - Tx\| &= \|\sum \alpha_i Te_i\| \\ &\leq \sum |\alpha_i| \, \|Te_i\| \\ &\leq \|h\|_1 (\max \|Te_i\|) \\ &\leq M\|h\|(\max \|Te_i\|) \leq \epsilon \text{ (done)}\end{aligned}$$

∎

As a corollary of Theorem 4.1.1 and Theorem 4.1.3, we now see that, if $\mathcal{X}$ is finite-dimensional, then all linear map of domain $\mathcal{X}$ are bounded. A counter example to the generalization of this statement is followed.

Example 13 (Differentiation is an Unbounded Linear Operator)

$$\mathcal{X} = \Big(\mathbb{R}[x]|_{[0,1]}, \|\cdot\|_\infty\Big), D(P) \triangleq P'$$

Note that $\{x^n\}_{n\in\mathbb{N}}$ is bounded in $\mathcal{X}$ and $\{D(x^n)\}_{n\in\mathbb{N}}$ is not.

Now, suppose $\mathcal{X}, \mathcal{Y}$ are two fixed normed spaces over $\mathbb{R}$ or $\mathbb{C}$. We can easily check that the set $BL(\mathcal{X}, \mathcal{Y})$ of bounded linear operators from $\mathcal{X}$ to $\mathcal{Y}$ form a vector space over whichever field $\mathcal{Y}$ is over.

Naturally, our definition of boundedness of linear operator derive us a norm on $BL(\mathcal{X}, \mathcal{Y})$, as followed

$$\|T\|_{\text{op}} \triangleq \inf\{M \in \mathbb{R}^+ : \forall x \in \mathcal{X}, \|Tx\| \leq M\|x\|\} \tag{4.1}$$

Before we show that our definition is indeed a norm, we first give some equivalent definitions and prove their equivalency.

Theorem 4.1.4. (Equivalent Definitions of Operator Norm) Given two fixed normed space $\mathcal{X}, \mathcal{Y}$ over $\mathbb{R}$ or $\mathbb{C}$, a bounded linear operator $T : \mathcal{X} \to \mathcal{Y}$, and define $\|T\|_{\text{op}}$ as in (4.1), we have

$$\|T\|_{\text{op}} = \sup_{x \in \mathcal{X}, x \neq 0} \frac{\|Tx\|}{\|x\|}$$

Proof. Define $J = \{M \in \mathbb{R}^+ : \forall x \in \mathcal{X}, \|Tx\| \leq M\|x\|\}$, so that we have $\|T\|_{\text{op}} = \inf J$. Now, observe

$$\begin{aligned} J &= \{M \in \mathbb{R}^+ : M \geq \frac{\|Tx\|}{\|x\|}, \forall x \neq 0 \in \mathcal{X}\} \\ &= \left\{M \in \mathbb{R}^+ : M \geq \sup_{x \in \mathcal{X}, x \neq 0} \frac{\|Tx\|}{\|x\|}\right\} \end{aligned}$$

This let us conclude

$$\|T\|_{\text{op}} = \inf J = \sup_{x \in \mathcal{X}, x \neq 0} \frac{\|Tx\|}{\|x\|}$$

∎

It is now easy to see

$$\begin{aligned} \|T\|_{\text{op}} &= \sup_{x \in \mathcal{X}, x \neq 0} \frac{\|Tx\|}{\|x\|} && (4.2) \\ &= \sup_{x \in \mathcal{X}, \|x\|=1} \|Tx\| && (4.3) \end{aligned}$$

It is not all in vain to introduce the equivalent definitions. See that the verification of $\|\cdot\|_{\text{op}}$ being a norm on $BL(\mathcal{X}, \mathcal{Y})$ become simple by utilizing the equivalent definitions.

(a) For positive-definiteness, fix non-trivial T and fix $x \in \mathcal{X} \setminus N(T)$. Use (4.2) to show $\|T\|_{\text{op}} \geq \frac{\|Tx\|}{\|x\|} > 0$.

(b) For absolute homogenity, use (4.3) and $\|Tcx\| = |c| \cdot \|Tx\|$.

(c) For triangle inequality, use (4.3) and $\|(T_1 + T_2)x\| \leq \|T_1 x\| + \|T_2 x\|$.

Naturally, and very very importantly, (4.2) give us

$$\|Tx\| \leq \|T\|_{\text{op}} \cdot \|x\| \qquad (x \in \mathcal{X})$$

This inequality will later be the best tool to help analyze the derivatives of functions between Euclidean spaces, and perhaps better, it immediately give us

$$\frac{\|T_1 T_2 x\|}{\|x\|} \leq \frac{\|T_1\|_{\text{op}} \cdot \|T_2\|_{\text{op}} \cdot \|x\|}{\|x\|} = \|T_1\|_{\text{op}} \cdot \|T_2\|_{\text{op}}$$

Then (4.2) give us

$$\|T_1 T_2\|_{\text{op}} \leq \|T_1\|_{\text{op}} \cdot \|T_2\|_{\text{op}}$$

4.2 $\partial_u f$ and ∇f of general $f : \mathbb{R}^n \to \mathbb{R}$

Given two normed space $\mathcal{X}, \mathcal{Y}$, suppose f maps an open neighborhood O around x in $\mathcal{X}$ into $\mathcal{Y}$. We say f is **differentiable at** x if there exists a bounded linear transformation $A_x : \mathcal{X} \to \mathcal{Y}$ (from now, A_x will be denoted df_x) such that

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - df_x(h)\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}} = 0 \tag{4.4}$$

Immediately, we should check that the linear approximation is unique. Suppose df_x and df'_x both satisfy (4.4). We are required to show $(df_x - df'_x)h = 0$ for all $\|h\|_{\mathcal{X}} = 1$. Fix $h \in \mathcal{X}$ such that $\|h\|_{\mathcal{X}} = 1$. Note that

$$\frac{(df_x - df'_x)th}{t} \text{ is a constant in } t \text{ for } t \neq 0$$

This then reduced the problem into showing

$$\frac{(df_x - df'_x)th}{t\|h\|_{\mathcal{X}}} \to 0 \text{ as } t \to 0 \tag{4.5}$$

Observe

$$(df_x - df'_x)th = \Big(f(x+th) - f(x) - df'_x(th)\Big) - \Big(f(x+th) - f(x) - df_x(th)\Big)$$

which implies

$$\|(df_x - df'_x)th\|_{\mathcal{Y}} \leq \|f(x+th) - f(x) - df'_x(th)\|_{\mathcal{Y}} + \|f(x+th) - f(x) - df_x(th)\|_{\mathcal{Y}}$$

and thus implies (4.5).

It shall be quite clear that a function f differentiable at x must be continuous at x, by noting the nominator of (4.4) must tend to 0.

For clarity, we here specify the notation. By $\mathbb{R}$, we mean a field equipped with the usual norm $\|x\| \triangleq |x|$. By $\mathbb{R}^n$ we mean the set of functions from $\{1, \dots, n\}$ to $\mathbb{R}$ equipped with the usual vector addition, scalar multiplication, dot product and induced norm.

Definition 4.2.1. (Definition of Directional Derivative of Scalar function) Given a normal vector $v \in \mathbb{R}^n$ and a function f that maps an open-neighborhood E around $x \in \mathbb{R}^n$

into $\mathbb{R}$, by the **directional derivative** $\partial_v f(x)$ of f with respect to v at x, we mean

$$\partial_v f(x) \triangleq \lim_{t\to 0} \frac{f(x+tv)-f(x)}{t} \text{ if exists}$$

Something to note in our definition for directional derivative (Definition 4.2.1)

(a) The limit on the right hand side is done in $\Big(\mathbb{R}, |\cdot|\Big)$

(b) If f is differntiable at x, then we have

$$\partial_v f(x) = df_x(v) \tag{4.6}$$

regardless if v is normal or not, as long as $v \neq 0$.

(c) In our definition of directional derivative, we required v to be normal, even though the definition remain well-defined if the hypothesis that v is normal is dropped. A reason to require v to be normal in our definition is to preserve the spirit of the term "directional". Suppose $\partial_v f(x)$ exists, we have

$$\partial_{2v} f(x) = 2\partial_v f(x) \tag{4.7}$$

In (4.7), we see that the value of directional derivative vary if the magnitude of v change.

Of course, one can "fix" the (4.7) by defining $\partial_v f(x)$ to be $\lim_{t\to 0} \frac{f(x+tv)-f(x)}{t|v|}$. This definition is not desired, because we will no longer have (4.6). With what we observed, one can immediately see that if a function f is differentiable at x, then f has directional derivative with respect to any direction at x. The converse is not true. It is possible that f has directional derivative with respect to all directions, and yet f is still not differentaible. Consider

Example 14 (Discontinuous function such that all directional derivatives exist)

$$f : \mathbb{R}^2 \to \mathbb{R}, f(x,y) = \begin{cases} \frac{x^2 y}{x^4+y^2} & \text{if } x^2+y^2 \neq 0 \\ 0 & \text{if } x = y = 0 \end{cases}$$

Definition 4.2.2. (Definition of Gradient of $\mathbb{R}^n \to \mathbb{R}$ function) Given a point $x \in \mathbb{R}^n$ with open neighborhood E, a function $f : E \to \mathbb{R}$ differentiable at x, we define the **gradient** $\nabla f(x) \in \mathbb{R}^n$ of f at x to be the unique vector that satisfy

$$\nabla f(x) \cdot v = df_x(v) \text{ for all } v \in \mathbb{R}^n$$

We should immediately discuss whether our definition of gradient is well-defined. The proof of existence and uniqueness follows from generating an orthogonal basis $\{v_1, \dots, v_n\}$ and noting $\nabla f(x)$ must equal to $\sum_{i=1}^n df_x(v_i)v_i$.

A few things you must know about gradient is as followed

(a) $\nabla f(x)$ is only defined when f is differentiable at x.

(b) gradient $\nabla f(x)$ "points toward" the direction at which $f : \mathbb{R}^n \to \mathbb{R}$ grow the fastest. Suppose v is normal. See

$$\nabla f(x) \cdot v = df_x(v) = \partial_v f(x)$$

Using Cauchy-Schwarz Inequality, we see that $\partial_v f(x)$ is of largest value when $v = \frac{\nabla f(x)}{|\nabla f(x)|}$. If $v = \frac{\nabla f(x)}{|\nabla f(x)|}$, then

$$\partial_v f(x) = |\nabla f(x)|$$

(c) It is possible $\nabla f(x) = 0$. This is true if and only if df_x maps $\mathbb{R}^n$ into 0. This fact echos with the fact gradient points toward the fastest growing direction. See (b).

(d) Given an orthogonal basis $\{v_1, \dots, v_n\}$, we have

$$\nabla f(x) = \sum_{i=1}^n df_x(v_i)v_i = \sum_{i=1}^n \partial_{v_i} f(x)v_i$$

This is how you compute $\nabla f(x)$ when you have to.

(e) Given an open set $E \subseteq \mathbb{R}^n$, differentaible $f : E \subseteq \mathbb{R}^n \to \mathbb{R}$, a value $y \in \mathbb{R}$ such that $S \triangleq f^{-1}[y]$ has non-empty interior and $\nabla f(p) \neq 0$ for all $p \in S$, we have

$$\nabla f(p) \perp S \text{ for all } p \in S$$

More precisely this means that all differentiable curves $\alpha : (-\epsilon, \epsilon) \to S$ that lies in S satisfy

$$\alpha'(0) \cdot \nabla f(\alpha(0)) = 0$$

This can be easily proved after we learn how to use the Chain Rule. Note that at here, we utilize the set-theoretic nature of $\mathbb{R}^n$ to present the statement. By $\alpha'(0)$, we mean $(\alpha_1'(0), \dots, \alpha_n'(0))$, where $\alpha(t)$ is in fact a function from $\{1, \dots, n\}$ to $\mathbb{R}$, and $\alpha_1(t) = (\alpha(t))(1)$ for all t.

4.3 Differentiability Theorem

We shall now include functions between $\mathbb{R}^n$ and $\mathbb{R}^m$ into our discussion. The goal of this section is to prove

(a) If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at x, then all partial derivatives exists and the total derivative df_x is a linear combination of the partial derivatives. (Theorem 4.3.1)

(b) $\partial_i f_j$ all exist and continuous around x if and only if f is continuously differentiable at x. (Theorem 4.3.3)

The proof of the latter in continuously differentiable part use the fact that all norms on $\mathbb{R}^k$ are equivalent (Theorem 4.1.2), where $k = nm$, and utilize the Frobenius norm.

Given an orthonormal basis $\{q_1, \dots, q_m\}$ of $\mathbb{R}^m$ and a function $f : \mathbb{R}^n \to \mathbb{R}^m$, we let $f_j(x)$ be a real number

$$f_j(x) = f(x) \cdot q_j$$

It shall be clear that

$$f(x) = \sum_{j=1}^{m} f_j(x) q_j$$

which explain why we require $\{q_1, \dots, q_m\}$ to be orthonormal in the first place.

For brevity of the statement of the next theorem (Theorem 4.3.1), we introduce more a notation. If we are provided a basis $\{e_1, \dots, e_n\}$ of $\mathbb{R}^n$, we denote $\partial_{e_i} f_j(x)$ by $\partial_i f_j(x)$

Theorem 4.3.1. (Derivative is Jacobian) Suppose $\alpha = \{e_1, \dots, e_n\}$ is a basis of $\mathbb{R}^n$, and $\beta = \{q_1, \dots, q_m\}$ is an orthonormal basis of $\mathbb{R}^m$. Suppose f maps an open neighborhood O around $x \in \mathbb{R}^n$ to $\mathbb{R}^m$. Then

$$f \text{ is differentiable at } x \implies \begin{cases} \partial_i f_j(x) \text{ exists for all } i, j \\ [df_x]_\alpha^\beta = \begin{bmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(x) & \cdots & \partial_n f_m(x) \end{bmatrix} \end{cases}$$

Proof. Suppose $e_1, \dots, e_n$ are all normal. Fix i, j. We wish to show

$$\partial_i f_j(x) \text{ exists}$$

Because f is differentiable at x, by definition of df_x, we have

$$\lim_{t\to 0}\frac{|f(x+te_i)-f(x)-df_x(te_i)|}{|te_i|}=0$$

Set $R_i:\mathbb{R}\to\mathbb{R}^m$ by $R_i(t)\triangleq f(x+te_i)-f(x)-df_x(te_i)$. We have

$$\lim_{t\to 0}\frac{|R_i(t)|}{|t|}=0 \tag{4.8}$$

Compute

$$\begin{aligned}f_j(x+te_i)-f_j(x)&=\big(f(x+te_i)-f(x)\big)\cdot q_j\\&=\big(R_i(t)+df_x(te_i)\big)\cdot q_j\\&=R_i(t)\cdot q_j+tdf_x(e_i)\cdot q_j\end{aligned}$$

This then give us

$$\frac{f_j(x+te_i)-f_j(x)}{t}=\frac{R_i(t)\cdot q_j}{t}+df_x(e_i)\cdot q_j$$

and

$$df_x(e_i)\cdot q_j-\frac{|R_i(t)\cdot q_j|}{|t|}\leq\frac{f_j(x+te_i)-f_j(x)}{t}\leq df_x(e_i)\cdot q_j+\frac{|R_i(t)\cdot q_j|}{|t|}$$

By Cauchy-Schwarz Inequality, we now have

$$df_x(e_i)\cdot q_j-\frac{|R_i(t)|}{|t|}\leq\frac{f_j(x+te_i)-f_j(x)}{t}\leq df_x(e_i)\cdot q_j+\frac{|R_i(t)|}{|t|}$$

Now applying Squeeze Theorem and Equation 4.8, we have

$$\partial_i f_j(x)=\lim_{t\to 0}\frac{f_j(x+te_i)-f_j(x)}{t}=df_x(e_i)\cdot q_j \text{ (done)}$$

Using the fact β is orthonormal, we now have

$$df_x(e_i)=\sum_{j=1}^m\Big(df_x(e_i)\cdot q_j\Big)q_j=\sum_{j=1}^m\partial_i f_j(x)q_j$$

and suggest the matrix representation. Now use Equation 4.7 on f_j to prove the case when $e_1,\dots,e_n$ are not all normal. ■

Note that the converse is not always true. It is possible that a function f has all partial derivatives with respect to a given basis, or even all directions, and yet f is still discontinuous. We have given an example already. Consider a less trivial one.

Example 15 (Non-differentiable Continuous Funciton with Partial Derivative)

$$f : \mathbb{R}^2 \to \mathbb{R} \text{ and } f(x,y) = \begin{cases} \frac{x|y|}{\sqrt{x^2+y^2}} & \text{if } (x,y) \neq 0 \\ 0 & \text{if } (x,y) = 0 \end{cases}$$

We have

$$\partial_x f(0) = \partial_y f(0) = 0$$

By Theorem 4.3.1 (Derivative is Jacobian), if f is differentiable at 0, then df_0 must be trivial. Yet

$$\frac{|f(h,h) - f(0) - df_0(h,h)|}{|(h,h)|} = \frac{h}{2|h|} \not\to 0$$

Note that f is continuous at 0, by observing

$$x^2 + y^2 - 2|xy| = \big(|x| - |y|\big)^2 \geq 0 \implies \frac{x^2+y^2}{2} \geq |xy|$$

which implies

$$|f| \leq \frac{\sqrt{x^2+y^2}}{2}$$

We now introduce a property of function between normed space that are stronger than differentiablity.

Given two normed space $\mathcal{X}, \mathcal{Y}$, and an open $E \subseteq \mathcal{X}$, we say $f : E \to \mathcal{Y}$ is **continuously differentiable** on $\mathcal{Y}$ if the map $D : (E, \|\cdot\|_{\mathcal{X}}) \to \Big(BL(\mathcal{X},\mathcal{Y}), \|\cdot\|_{\text{op}}\Big)$ defined by

$$D(x) = df_x$$

is continuous. Note that the definition of the term "continuously differentiable" coincide when $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ and df_x is just $h \mapsto f'(x)h$.

Before we prove the differentiability Theorem (Theorem 4.3.3), we first prove a small lemma.

Lemma 4.3.2. (Concave functions is Subaddivity if $f(0) \geq 0$) If f is concave on $[0, \infty)$ and $f(0) \geq 0$, then we have

$$f(a) + f(b) \geq f(a+b) \text{ for all } a, b \in \mathbb{R}^+$$

Proof. Let $t \in [0, 1]$. We see

$$f(tx) = f(tx + (1-t)0) \geq tf(x) + (1-t)f(0) \geq tf(x)$$

This give us

$$f(a) + f(b) = f\big((a+b)\frac{a}{a+b}\big) + f\big((a+b)\frac{b}{a+b}\big) \geq \frac{a}{a+b}f(a+b) + \frac{b}{a+b}f(a+b) = f(a+b)$$

■

Theorem 4.3.3. (Differentiability Theorem) Suppose $\alpha = \{e_1, \dots, e_n\}$ is an orthonormal basis of $\mathbb{R}^n$, and $\beta = \{q_1, \dots, q_m\}$ is an orthonormal basis of $\mathbb{R}^m$. Suppose f maps an open set $E \subseteq \mathbb{R}^n$ to $\mathbb{R}^m$. Then

$$f \text{ is continuously differentiable on } E \iff \partial_i f_j \text{ exists and is continuous on } E \text{ for all } i, j$$

Proof. $(\longrightarrow)$

Fix i, j. We wish to show

$$\partial_i f_j \text{ exists and is continuous on } E$$

Because f is differentiable on E, we know $\partial_i f_j$ exists on E. (Theorem 4.3.1). Fix $x \in E$. We only have to show

$$\partial_i f_j \text{ is continuous at } x$$

Fix ϵ. We wish

$$\text{to find } \delta \text{ such that } |\partial_i f_j(y) - \partial_i f_j(x)| \leq \epsilon \text{ for all } |y - x| < \delta$$

Because f is continuously differentaible at x, we know there exists δ such that

$$\|df_y - df_x\|_{\text{op}} < \epsilon \text{ for all } |y - x| \leq \delta$$

We claim

$$\text{such } \delta \text{ suffices}$$

By Theorem 4.3.1 (Derivative is Jacobian), we know

$$\partial_i f_j(y) - \partial_i f_j(x) = (df_y - df_x)e_i \cdot q_j$$

By Cauchy-Inequality, we then have

$$\begin{aligned}|\partial_i f_j(y) - \partial_i f_j(x)| &\leq |(df_y - df_x)e_i| \\ &\leq \|df_y - df_x\|_{\text{op}} < \epsilon \text{ (done)}\end{aligned}$$

$(\longleftarrow)$

We first show

$$f \text{ is differentiable on } E$$

We first prove

$$\forall j \in \{1, \dots, m\}, f_j : \mathbb{R}^n \to \mathbb{R} \text{ is differentiable on } E \implies f \text{ is differentiable on } E$$

Fix $x \in E$. We wish to prove

$$f \text{ is differentiable at } x$$

Define $A : E \to \mathbb{R}^m$ by

$$A(h) \triangleq \sum_{j=1}^{m} (df_j)_x(h) q_j$$

We claim

$$A \text{ suffices to be the } df_x$$

Using the fact q_j are orthonormal, we have

$$f(x+h) - f(x) - A(h) = \sum_{j=1}^{m} \big(f_j(x+h) - f_j(x) - (df_j)_x(h)\big) q_j$$

This give us

$$\begin{aligned}\lim_{h \to 0} \frac{|f(x+h) - f(x) - A(h)|}{|h|} &= \lim_{h \to 0} \frac{\left|\sum_{j=1}^{m} \big(f_j(x+h) - f_j(x) - (df_j)_x(h)\big) q_j\right|}{|h|} \\ &= \lim_{h \to 0} \frac{\sum_{j=1}^{m} |f_j(x+h) - f_j(x) - (df_j)_x(h)|}{|h|} = 0 \text{ (done)}\end{aligned}$$

Fix $j \in \{1, \dots, m\}$. We can now reduce the problem into

$$f_j : \mathbb{R}^n \to \mathbb{R} \text{ is differentiable on } E$$

Fix $x \in E$. We wish to prove

$$f_j \text{ is differentiable at } x$$

Express $h = \sum_{i=1}^n h_i e_i$. Define $B : E \to \mathbb{R}$ by

$$B(h) = \sum_{i=1}^n \partial_i f_j(x) h_i$$

We claim

$$B \text{ suffices to be } (df_j)_x$$

By continuity of each $\partial_i f_j$ on E, we can let δ satisfy

$$|\partial_i f_j(y) - \partial_i f_j(x)| < \epsilon \text{ for all } y \in B_\delta(x)$$

We claim

$$\frac{|f_j(y) - f_j(x) - B(y - x)|}{|y - x|} \leq \epsilon \text{ for all } y \in B_\epsilon(x)$$

Express $y - x = \sum_{k=1}^n h_k e_k$. Define $v_0, \dots, v_n \in \mathbb{R}^n$ by

$$v_0 \triangleq 0 \text{ and } v_k \triangleq \sum_{i=1}^k h_i e_i \text{ for all } k \in \{1, \dots, n\}$$

Note that the squared-root $x \mapsto \sqrt{x}$ is a concave function on $[0, \infty)$. Then Lemma 4.3.2 give us

$$\sum_{k=1}^n |h_k| \leq \sqrt{\sum_{k=1}^n h_k^2} = |y - x|$$

Now observe

$$\begin{aligned}
\frac{|f_j(y)-f_j(x)-B(y-x)|}{|y-x|} &= \frac{|f_j(x+v_n)-f_j(x)-B(\sum_{k=1}^n h_k e_k)|}{|y-x|}\\
&= \frac{\left|\left(\sum_{k=1}^n f_j(x+v_k)-f_j(x+v_{k-1})\right)-\sum_{k=1}^n \partial_k f_j(x)h_k\right|}{|y-x|}\\
&= \frac{|\sum_{k=1}^n f_j(x+v_k)-f_j(x+v_{k-1})-\partial_k f_j(x)h_k|}{|y-x|}\\
&= \frac{|\sum_{k=1}^n f_j(x+v_{k-1}+h_k e_k)-f_j(x+v_{k-1})-\partial_k f_j(x)h_k|}{|y-x|}\\
&= \frac{|\sum_{k=1}^n \partial_k f_j(x+v_{k-1}+t_k e_k)h_k-\partial_k f_j(x)h_k|}{|y-x|}\\
&\leq \frac{\sum_{k=1}^n \left|\left(\partial_k f_j(x+v_{k-1}+t_k e_k)-\partial_k f_j(x)\right)h_k\right|}{|y-x|}\\
&< \frac{\sum_{k=1}^n \epsilon\,|h_k|}{|y-x|} \leq \epsilon \text{ (done)}
\end{aligned}$$

We now prove

$$f \text{ is continuously differentiable on } E$$

Fix ϵ and $x \in E$. We are required

$$\text{to find } \delta \text{ such that } \|df_y - df_x\|_{\text{op}} \leq \epsilon \text{ for all } y \in B_\delta(x)$$

Note that one can define a norm $\|\cdot\|_F$ called "Forbenius Norm" on $BL(\mathbb{R}^n,\mathbb{R}^n)$ by

$$\|A\|_F \triangleq \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_{i,j}^2} \text{ where } [A]_\alpha^\beta = \begin{bmatrix} a_{1,1} & \cdots & a_{n,1} \\ \vdots & \ddots & \vdots \\ a_{1,n} & \cdots & a_{n,n} \end{bmatrix}$$

Because all norms on finite-dimensional real vector spaces are equivalent (Theorem 4.1.2), we know there exists M such that for all $x \in E$, we have

$$\|df_x\|_{\text{op}} \leq M\|df_x\|_F$$

Because the partial derivatives are all continuous by definition, we can let δ satisfy

$$\left(\partial_i f_j(x+h)\right)^2 - \left(\partial_i f_j(x)\right)^2 < \frac{\epsilon^2}{M^2 n^2} \text{ for all } h \in B_\delta(0)$$

We claim

$$\text{such } \delta \text{ suffices}$$

Let $|y-x|<\delta$. We see

$$\|df_y - df_x\|_{\text{op}} \leq M\|df_y - df_x\|_F < M\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\epsilon^2}{M^2n^2}} = \epsilon$$

■

4.4 Smoothness of $f : \mathbb{R}^n \to \mathbb{R}^m$

Given the set of all real-valued functions on an open set E of $\mathbb{R}^n$, we can define the differentaible class by saying a function $f : E \to \mathbb{R}$ is of class C^k if

$$\frac{\partial^k f}{(\partial x_1)^{\alpha_1} \cdots (\partial x_n)^{\alpha_n}} \text{ is continuous on } E \text{ for all } \alpha_1 + \cdots + \alpha_n = k$$

Alternatively, we can say a function $f : E \to \mathbb{R}$ is of C^2 if the function $D : E \to \big(L(\mathbb{R}^n, \mathbb{R}), \|\cdot\|_{\text{op}}\big)$ defined by

$$D(x) = df_x$$

is again differentaible on E.

Theorem 4.4.1. (Structure of Mixed Partial Derivative) Given an open set $E \subseteq \mathbb{R}^2$, a point $p \in E$, a basis $\{e_1, e_2\}$ of $\mathbb{R}^2$, a function $f : E \to \mathbb{R}$ such that

(a) $\partial_1 f$ exists on E

(b) $\partial_2 f$ exists on E

(c) $\partial_{21} f$ exists on E and is continuous at p

Then

$$\partial_{12} f(p) = \partial_{21} f(p)$$

Proof. Express elements of E in the basis $\{e_1, e_2\}$, and express $p = (a, b)$. We are required to prove

$$\lim_{h \to 0} \frac{\partial_2 f(a+h, b) - \partial_2 f(a, b)}{h} = \partial_{21} f(a, b)$$

Define $\Delta(h, k)$ on $E \setminus p$ by

$$\Delta(h, k) \triangleq f(a+h, b+k) - f(a, b+k) - f(a+h, b) + f(a, b)$$

Note that because $\partial_2 f$ exists on E, for all $h \neq 0$, we have

$$\lim_{k \to 0} \frac{\Delta(h, k)}{hk} = \frac{\partial_2 f(a+h, b) - \partial_2 f(a, b)}{h}$$

This let us reduce the problem into proving

$$\lim_{h \to 0} \lim_{k \to 0} \frac{\Delta(h, k)}{hk} = \partial_{21} f(a, b)$$

We first show

$$\frac{\Delta(h,k)}{hk} = \partial_{21} f(x,y) \text{ for some } (x,y) : |x-a| < h \text{ and } |y-b| < k \tag{4.9}$$

Define $u(t)$ by

$$u(t) \triangleq f(t, b+k) - f(t,b)$$

Compute

$$u'(t) = \partial_1 f(t, b+k) - \partial_1 f(t,b)$$

Then we have

$$\begin{aligned}
\Delta(h,k) &= u(a+h) - u(a) \\
&= hu'(x) \text{ for some } x \in (a, a+h) \text{ by MVT (Corollary 3.3.4)} \\
&= h\big(\partial_1 f(x, b+k) - \partial_1 f(x,b)\big)
\end{aligned}$$

Define $v(t)$ by

$$v(t) \triangleq \partial_1 f(x,t)$$

Compute

$$v'(t) = \lim_{h\to 0} \frac{\partial_1 f(xe_1 + (t+h)e_2) - \partial_1 f(xe_1 + te_2)}{h} = \partial_{21} f(x,t)$$

Then we have

$$\begin{aligned}
\Delta(h,k) &= h(\partial_1 f(x, b+k) - f(x,b)) \\
&= h\big(v(b+k) - v(b)\big) \\
&= hkv'(y) \text{ for some } y \in (b, b+k) \\
&= hk\partial_{21} f(x,y) \text{ (done)}
\end{aligned}$$

Fix ϵ. We wish

$$\text{to find some } \delta \text{ such that for all } h \in (-\delta,\delta) \setminus 0, \left| \lim_{k\to 0} \frac{\Delta(h,k)}{hk} - \partial_{21} f(x,y) \right| \le \epsilon$$

Because of Equation 4.9 and $\partial_{21} f$ is continuous at p, we know there exists δ such that for all $h,k \in (-\delta,\delta) \setminus 0$, we have

$$\left| \frac{\Delta(h,k)}{hk} - \partial_{21} f(a,b) \right| < \frac{\epsilon}{2}$$

We claim

$$\text{such } \delta \text{ works}$$

Fix $h \in (-\delta, \delta) \setminus 0$. Note that $\lim_{k\to 0} \frac{\Delta(h,k)}{hk} = \frac{\partial_2 f(a+h,b) - \partial_2 f(a,b)}{h}$ exists, so we can find small enough k' such that

$$0 < |k'| < \delta \text{ and } \left| \lim_{k\to 0} \frac{\Delta(h,k)}{hk} - \frac{\Delta(h,k')}{hk'} \right| < \frac{\epsilon}{2}$$

Now observe

$$\left| \lim_{k\to 0} \frac{\Delta(h,k)}{hk} - \partial_{21} f(x,y) \right| \leq \left| \lim_{k\to 0} \frac{\Delta(h,k)}{hk} - \frac{\Delta(h,k')}{hk'} \right| + \left| \frac{\Delta(h,k')}{hk'} - \partial_{21} f(a,b) \right| \leq \epsilon \text{ (done)}$$

■

Corollary 4.4.2. (Clairaut's Theorem on equality of mixed partial) Given a basis $\{e_1, \dots, e_n\}$ of $\mathbb{R}^n$, an open set $E \subseteq \mathbb{R}^n$, a function $f : E \to \mathbb{R}$ such that

$$\partial_{ij} f \text{ exist and is continuous on } E \text{ for all } i, j \in \{1, \dots, n\}$$

Then

$$\partial_{ij} f = \partial_{ji} f \text{ on } E \text{ for all } i, j \in \{1, \dots, n\}$$

Proof. Fix $p \in E$, and fix $i, j \in \{1, \dots, n\}$. We are required to prove

$$\partial_{ij} f(p) = \partial_{ji} f(p)$$

Express p in the form $p = x_0 e_i + y_0 e_j + v$ where $v \in \text{span}\Big(\{e_1, \dots, e_n\} \setminus \{e_i, e_j\}\Big)$.

Because E is open, we can let $B_\epsilon(p) \subseteq \mathbb{R}^n$ be contained by E. Define $U \subseteq \mathbb{R}^2$ by

$$U = B_\epsilon(x_0, y_0)$$

Check that

$$\{x e_i + y e_j + v \in \mathbb{R}^n : (x,y) \in U\} \subseteq E$$

Define $g : U \to \mathbb{R}$ by

$$g(x,y) \triangleq f(x e_i + y e_j + v)$$

Check that

$$\partial_1 g(x,y) = \partial_i f(x e_i + y e_j + v) \text{ and } \partial_2 g(x,y) = \partial_j f(x e_i + y e_j + v)$$

We can now apply Theorem 4.4.1 to g and have

$$\partial_{ij} f(p) = \partial_{12} g(x_0, y_0) = \partial_{21} g(x_0, y_0) = \partial_{ji} f(p) \text{ (done)}$$

■

4.5 Product Rule and Chain Rule

We now prove the Chain Rule for function between normed space.

Theorem 4.5.1. (Chain Rule) Given three normed space $\mathcal{X},\mathcal{Y},\mathcal{Z}$, a point $x\in\mathcal{X}$, a function g that map an open set $U\subseteq\mathcal{Y}$ containing $f(x)$ into $\mathcal{Z}$, a function f that map an open-neighborhood around x into U such that

(a) f is differentiable at x

(b) g is differentiable at $f(x)$

we have

$$d(g\circ f)_x = dg_{f(x)}\circ df_x$$

Proof. For brevity, we use $F\triangleq g\circ f$. We wish to prove

$$\lim_{h\to 0}\frac{\|F(x+h)-F(x)-dg_{f(x)}df_x(h)\|_{\mathcal{Z}}}{\|h\|_{\mathcal{X}}}=0$$

Fix $k\triangleq f(x+h)-f(x)$. Observe

$$F(x+h)-F(x)-dg_{f(x)}df_x(h)=\Big(g\big(f(x)+k\big)-g\big(f(x)\big)-dg_{f(x)}(k)\Big)+dg_{f(x)}\big(k-df_x(h)\big)$$

This now implies

$$\frac{\|F(x+h)-F(x)-dg_{f(x)}df_x(h)\|_{\mathcal{Z}}}{\|h\|_{\mathcal{X}}}\text{ is smaller than}$$

$$\frac{\|g\big(f(x)+k\big)-g\big(f(x)\big)-dg_{f(x)}(k)\|_{Z}+\|dg_{f(x)}\big(k-df_x(h)\big)\|_{\mathcal{Z}}}{\|h\|_{\mathcal{X}}}$$

This let us reduce the problem into proving

$$\lim_{h\to 0}\frac{\|g\big(f(x)+k\big)-g\big(f(x)\big)-dg_{f(x)}(k)\|_{Z}}{\|h\|_{\mathcal{X}}}=0$$

$$\text{and }\lim_{h\to 0}\frac{\|dg_{f(x)}\big(k-df_x(h)\big)\|_{\mathcal{Z}}}{\|h\|_{\mathcal{X}}}=0$$

We first prove

$$\lim_{h\to 0}\frac{\|g\big(f(x)+k\big)-g\big(f(x)\big)-dg_{f(x)}(k)\|_{Z}}{\|h\|_{\mathcal{X}}}=0$$

Note that if $\|k\|_{\mathcal{Y}} = 0$, we have

$$\frac{\|g\big(f(x)+k\big) - g\big(f(x)\big) - dg_{f(x)}(k)\|_Z}{\|h\|_{\mathcal{X}}} = 0$$

Now, observe that

$$\frac{\|g\big(f(x)+k\big) - g\big(f(x)\big) - dg_{f(x)}(k)\|_Z}{\|h\|_{\mathcal{X}}} = \frac{\|g\big(f(x)+k\big) - g\big(f(x)\big) - dg_{f(x)}(k)\|_Z}{\|k\|_{\mathcal{Y}}} \cdot \frac{\|k\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}}$$

Because $h \to 0 \implies k \to 0$, we can now reduce the problem into proving

$$\limsup_{h\to 0} \frac{\|k\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}} \text{ exists}$$

Observe

$$\begin{aligned}
\frac{\|k\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}} &= \frac{\|f(x+h) - f(x) - df_x(h) + df_x(h)\|_{\mathcal{X}}}{\|h\|_{\mathcal{X}}} \\
&\leq \frac{\|f(x+h) - f(x) - df_x(h)\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}} + \frac{\|df_x(h)\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}} \\
&\leq \frac{\|f(x+h) - f(x) - df_x(h)\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}} + \|df_x\|_{\text{op}} \text{ (done)}
\end{aligned}$$

We now prove

$$\lim_{h\to 0} \frac{\|dg_{f(x)}\big(k - df_x(h)\big)\|_{\mathcal{Z}}}{\|h\|_{\mathcal{X}}} = 0$$

Note that if $f(x+h) - f(x) - df_x(h) = 0$, then $\|dg_{f(x)}\big(k - df_x(h)\big)\|_{\mathcal{Z}} = 0$. Now, observe

$$\begin{aligned}
\frac{\|dg_{f(x)}\big(k - df_x(h)\big)\|_{\mathcal{Z}}}{\|h\|_{\mathcal{X}}} &= \frac{\|dg_{f(x)}\big(f(x+h) - f(x) - df_x(h)\big)\|_{\mathcal{Z}}}{\|f(x+h) - f(x) - df_x(h)\|_{\mathcal{Y}}} \cdot \frac{\|f(x+h) - f(x) - df_x(h)\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}} \\
&\leq \|dg_{f(x)}\|_{\text{op}} \cdot \frac{\|f(x+h) - f(x) - df_x(h)\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}} \to 0 \text{ (done)}
\end{aligned}$$

■

Interestingly, if $f : \big(\mathbb{R}, \|\cdot\|_2\big) \to \big(\mathbb{R}^n, \|\cdot\|_2\big)$ is a curve in $\mathbb{R}^n$

$$f(t) = (f_1(t), \cdots, f_n(t))$$

and we define

$$f'(t) \triangleq (f_1'(t), \cdots, f_n'(t))$$

We have

$$|f'(t)| = \|df_t\|_{\text{op}}$$

This give us the following expected result (Corollary 4.5.3).

Theorem 4.5.2. (Basic Property of Derivative) Suppose f maps a convex open set $E \subseteq \Big(\mathbb{R}^n, \|\cdot\|_2\Big)$ into $\Big(\mathbb{R}^m, \|\cdot\|_2\Big)$, f is differentiable on E, and there exists $M \in \mathbb{R}$ such that

$$\|df_x\|_{\text{op}} \leq M \qquad (x \in E)$$

Then for all $a, b \in E$, we have

$$|f(b) - f(a)| \leq M\,|b - a|$$

Proof. Define $\gamma : [0,1] \to E$ by

$$\gamma(t) \triangleq a + (b-a)t$$

Now, note that

$$\begin{aligned}
|f(b) - f(a)| &= |(f \circ \gamma)(1) - (f \circ \gamma)(0)| \\
&= \left| \int_0^1 (f \circ \gamma)'(t) dt \right| \\
&\leq \int_0^1 |(f \circ \gamma)'(t)|\, dt \\
&= \int_0^1 \|d(f \circ \gamma)_t\|_{\text{op}} dt \\
&\leq \int_0^1 \|df_{\gamma(t)}\|_{\text{op}} \cdot \|d\gamma_t\|_{\text{op}} dt \\
&= \int_0^1 M \cdot |b - a|\, dt = M\,|b - a|
\end{aligned}$$

■

Corollary 4.5.3. (Basic Property of Derivative) Suppose f maps a convex open set $E \subseteq \Big(\mathbb{R}^n, \|\cdot\|_2\Big)$ into $\Big(\mathbb{R}^m, \|\cdot\|_2\Big)$, f is differentiable on E, and $df_x = 0$ for all $x \in E$, then

$$f \text{ stay constant on } E$$

4.6 Inverse Function Theorem

Theorem 4.6.1. (Inverse Function Theorem) Given a function f that maps an open subset E of $\mathbb{R}^n$ into $\mathbb{R}^n$ such that f is continuously differentiable on E, if a point $a \in E$ has invertible derivative df_a, then there exists an open set $U \subseteq E$ containing a such that

$$f(U) \text{ is open in } \mathbb{R}^n \text{ and } f|_U : U \to f(U) \text{ is a diffeomorphism}$$

Proof. Fix

$$\lambda \triangleq \frac{1}{2\|(df_a)^{-1}\|_{\text{op}}} \tag{4.10}$$

Because f is continuously differentaible at x, we know there exists δ such that

$$\|df_x - df_a\|_{\text{op}} < \lambda \qquad (x \in B_\delta(a)) \tag{4.11}$$

We claim

$$U \triangleq B_\delta(a) \text{ suffices}$$

For each $y \in \mathbb{R}^n$, define $\phi_y : U \to \mathbb{R}^n$ by

$$\phi_y(x) \triangleq x + (df_a)^{-1}(y - f(x))$$

Before anything, we first prove

$$\text{for all } y \in \mathbb{R}^n, \phi_y : U \to \mathbb{R}^n \text{ is a contraction}$$

Fix $y \in \mathbb{R}^n$, and $x_1, x_2 \in U$. We claim

$$|\phi_y(x_1) - \phi_y(x_2)| \leq \frac{1}{2}|x_1 - x_2|$$

Because U is convex, Theorem 4.5.2 allow use to reduce the problem into proving

$$\|d(\phi_y)_x\|_{\text{op}} \leq \frac{1}{2} \text{ for all } x \in U$$

Fix $x \in U$. Use Chain Rule and the fact the derivative of a bounded linear transformation is itself to compute $d(\phi_y)_x$

$$\begin{aligned} d(\phi_y)_x &= I + (df_a)^{-1}(-df_x) \\ &= (df_a)^{-1}(df_a - df_x) \end{aligned}$$

This together with Equation 4.10 and Equation 4.11 give us

$$\|d(\phi_y)_x\|_{\text{op}} \leq \|(df_a)^{-1}\|_{\text{op}}\|df_a - df_x\|_{\text{op}} < \frac{1}{2} \text{ (done)}$$

We now prove

$$f \text{ is one-to-one in } U$$

Fix y in $f(U)$. We wish to show

$$\text{there exists at most one } x \in U \text{ such that } f(x) = y$$

Because $f(x) = y \iff x$ is a fixed point of ϕ_y, we can reduce the problem into

$$\phi_y \text{ has at most one fixed point in } U$$

Suppose $x, x' \in U$ are both fixed point of ϕ_y. Observe

$$|x - x'| = |\phi_y(x) - \phi_y(x')| \leq \frac{1}{2}|x - x'|$$

This implies $|x - x'| = 0$. (done)

We now prove

$$f(U) \text{ is open in } \mathbb{R}^n$$

Fix $y_0 \in f(U)$. Let $x_0 = f^{-1}(y_0)$. Because U is open, we know there exists r such that

$$\overline{B_r(x_0)} \subseteq U$$

We claim

$$B_{\lambda r}(y_0) \subseteq f(U)$$

Fix $y \in B_{\lambda r}(y_0)$. We are required to prove

$$y \in f(U)$$

Because

$$y = f(x) \iff x \text{ is a fixed point of } \phi_y$$

We then can use Banach Fixed Point Theorem to reduce the problem into proving

$$\phi_y \text{ is a contraction that maps some complete subset of } U \text{ into itself}$$

We claim

$$\overline{B_r(x_0)} \text{ suffices}$$

We have already ϕ_y is a contraction on U, and it is clear that $\overline{B_r(x_0)}$ is complete. We reduce the problem into proving

$$\phi_y\big(\overline{B_r(x_0)}\big) \subseteq \overline{B_r(x_0)}$$

Using

(a) definition of ϕ_y

(b) $|y - y_0| < \lambda r$

(c) $\|(df_a)^{-1}\|_{\text{op}} = \frac{1}{2\lambda}$

We can deduce

$$\begin{aligned}|\phi_y(x_0) - x_0| &= \left|(df_a)^{-1}(y - f(x_0))\right| \\ &\leq \|(df_a)^{-1}\|_{\text{op}} \, |y - y_0| < \frac{r}{2}\end{aligned}$$

Fix $x \in \overline{B_r(x_0)}$. We then see

$$\begin{aligned}|\phi_y(x) - x_0| &\leq |\phi_y(x_0) - \phi_y(x)| + |x_0 - \phi_y(x_0)| \\ &\leq \frac{1}{2}\,|x_0 - x| + \frac{r}{2} \leq r \text{ (done)}\end{aligned}$$

Lastly, it remains to prove

$$f^{-1} : f(U) \to U \text{ is continuously differentiable}$$

Because $f^{-1} \circ f = I$. We know $d(f^{-1})_{f(x)}$ if exists must satisfy

$$d(f^{-1})_{f(x)} = (df_x)^{-1}$$

■

4.7 Implicit Function Theorem

4.8 Appendix: Linear Algebra

This section contains

(a) definition and basic properties of the term **norm**

(b) definition and basic properties of the term **inner product**

(c) definition and basic properties of the term **positive semi-definite Hermitian form**

(d) full statement and proof of **Cauchy Schwarz Inequality** for both inner product space and positive semi-definite Hermitian form

(e) statement and proof of **SVD** (singular value decomposition).

(Norm Axiom Part)

Recall that by a **normed space** V, we mean a vector space over a sub-field $\mathbb{F}$ of $\mathbb{C}$ equipped with $\|\cdot\| : V \to \mathbb{R}_0^+$ satisfying the following axioms:

(a) $\|x\| = 0 \implies x = 0$ (positive-definiteness)

(b) $\|sx\| = |s| \cdot \|x\|$ for all $s \in \mathbb{F}$ and $x \in V$ (absolute-homogenity)

(c) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$ (triangle inequality)

Observe

$$\|0\| = \|0 + x\| \leq \|0\| + \|x\| \text{ for all } x \in V$$

This shows that $\|x\| \geq 0$ for all $x \in V$. Also observe

$$\|0\| = \|0(x)\| = |0| \cdot \|x\| = 0$$

We can now rewrite the normed space axioms into

(a) $\|x\| = 0 \iff x = 0$ (positive-definiteness)

(b) $\|sx\| = |s| \cdot \|x\|$ for all $s \in \mathbb{F}$ and $x \in V$ (absolute-homogeneity)

(c) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$ (triangle inequality)

(d) $\|x\| \geq 0$ for all $x \in V$ (non-negativity)

(Inner Product Axiom Part)

Recall that by an **inner product space** V, we mean a vector space over $\mathbb{R}$ or $\mathbb{C}$ equipped with $\langle \cdot, \cdot \rangle : V^2 \to \mathbb{R}$ or $\mathbb{C}$ satisfying the following axioms

(a) $\langle x, x \rangle > 0$ for all $x \neq 0$ (Positive-definiteness)

(b) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ (Conjugate symmetry)

(c) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ and $\langle cx, z \rangle = c\langle x, z \rangle$ (Linearity in the first argument)

Note that conjugate symmetry let us deduce

$$\langle x, x \rangle = \overline{\langle x, x \rangle} \implies \langle x, x \rangle \in \mathbb{R}$$

Also, one can easily use linearity in first argument to deduce

$$\langle 0, 0 \rangle = 2\langle 0, 0 \rangle \implies \langle 0, 0 \rangle = 0$$

This now let us rewrite the inner product space over $\mathbb{C}$ axioms into

(a) $\langle x, x \rangle \geq 0$ for all $x \in V$ (non-negativity)

(b) $\langle x, x \rangle = 0 \iff x = 0$ (positive-definiteness)

(c) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ (conjugate symmetry)

(d) $\langle cx + y, z \rangle = c\langle x, z \rangle + \langle y, z \rangle$ and $\langle x, cy + z \rangle = \overline{c}\langle x, y \rangle + \langle x, z \rangle$ (Linearity)

Note that using $c = 1$ and $y = 0$, ($\because \langle 0, z \rangle = 0\langle x, z \rangle = 0$) one can check that the latter expression of linearity implies the first expression.

If the scalar field is $\mathbb{R}$, then conjugate symmetry is just symmetry and we also have linearity in the second argument.

This now let us rewrite the inner product space over $\mathbb{R}$ axioms into

(a) $\langle x, x \rangle \geq 0$ for all $x \in V$ (non-negativity)

(b) $\langle x, x \rangle = 0 \iff x = 0$ (positive-definiteness)

(c) $\langle x, y \rangle = \langle y, x \rangle$ (symmetry)

(d) Linearity in both arguments

If we do not require $\langle \cdot, \cdot \rangle$ to be positive-definite, but only non-negative, i.e. $\langle x, x \rangle \geq 0$ for all $x \in V$, then we have a **positive semi-definite Hermitian form**. Formally speaking, a positive semi-definite Hermitian form $\langle \cdot, \cdot \rangle : V^2 \to \mathbb{R}$ or $\mathbb{C}$ satisfy the following axioms

(a) $\langle x, x \rangle \geq 0$ for all $x \in V$ (non-negativity)

(b) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ (conjugate symmetry)

(c) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ and $\langle cx, z \rangle = c\langle x, z \rangle$ (Linearity in the first argument)

Example 16 (Example of Positive semi-definite Hermitian form)

arbitrary V over $\mathbb{R}$ or $\mathbb{C}$ $\quad \langle x, y \rangle \triangleq 0$ for all x, y

(Norm Induce Part)

Given a vector space V over $\mathbb{R}$ or $\mathbb{C}$, one can check that if V is equipped with an inner product $\langle \cdot, \cdot \rangle : V^2 \to \mathbb{R}$ or $\mathbb{C}$, then we can induce a norm on V by

$$\|x\| \triangleq \sqrt{\langle x, x \rangle} \qquad (x \in V)$$

Note that

$$\|x\| = 0 \iff \langle x, x \rangle = 0$$

This implies that if $\langle \cdot, \cdot \rangle$ is an inner product (satisfy positive-definiteness), then $\| \cdot \|$ is also positive-definite. And if $\langle \cdot, \cdot \rangle$ is not positive-definite, then there exists $x \neq 0 \in V$ such that $\|x\| = 0$, which make $\| \cdot \|$ a **semi-norm**.

Absolute homogeneity follows from the linearity of inner product.

To check triangle inequality, we first have to prove Cauchy-Schwarz inequality.

Theorem 4.8.1. (Basic Property of Positive semi-definite Hermitian form) Given a positive semi-definite Hermitian form $\langle \cdot, \cdot \rangle : V^2 \to \mathbb{R}$ or $\mathbb{C}$ and $x, y \in V$, we have

$$\langle x, x \rangle = 0 \implies \langle x, y \rangle = 0$$

Proof. Assume $\langle x, y\rangle \neq 0$. Fix $t > \frac{\|y\|^2}{2|\langle x,y\rangle|^2}$. Compute

$$\begin{aligned}\|y - t\langle y, x\rangle x\|^2 &= \|y\|^2 + \|(-t)\langle y, x\rangle x\|^2 + \langle -t\langle y, x\rangle x, y\rangle + \langle y, -t\langle y, x\rangle x\rangle \\ &= \|y\|^2 + t^2 \left|\langle x, y\rangle\right|^2 \|x\|^2 - t\langle y, x\rangle\langle x, y\rangle - t\langle x, y\rangle\langle y, x\rangle \\ &= \|y\|^2 - 2t \left|\langle x, y\rangle\right|^2 < 0 \text{ CaC}\end{aligned}$$

■

Theorem 4.8.2. (Cauchy-Schwarz Inequality) Given a positive semi-definite Hermitian form $\langle\cdot,\cdot\rangle : V^2 \to \mathbb{C}$ on vector space V over $\mathbb{C}$, we have

(a) $|\langle x, y\rangle| \leq \|x\| \cdot \|y\| \quad (x, y \in V)$

(b) the equality hold true if x, y are linearly dependent

(c) the equality hold true if and only if x, y are linearly dependent (provided $\langle\cdot,\cdot\rangle$ is an inner product)

Proof. We first prove

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\| \qquad (x, y \in V)$$

Fix $x, y \in V$. Theorem 4.8.1 tell us $\|x\| = 0 \implies \langle x, y\rangle = 0$. Then we can reduce the problem into proving

$$\frac{|\langle x, y\rangle|^2}{\|x\|^2} \leq \|y\|^2$$

Set $z \triangleq y - \frac{\langle y,x\rangle}{\|x\|^2}x$. We then have

$$\langle z, x\rangle = \langle y - \frac{\langle y, x\rangle}{\|x\|^2}x, x\rangle = \langle y, x\rangle - \frac{\langle y, x\rangle}{\|x\|^2}\langle x, x\rangle = 0$$

Then from $y = z + \frac{\langle y,x\rangle}{\|x\|^2}x$, we can now deduce

$$\begin{aligned}\langle y, y\rangle &= \langle z + \frac{\langle y, x\rangle}{\|x\|^2}x, z + \frac{\langle y, x\rangle}{\|x\|^2}x\rangle \\ &= \langle z, z\rangle + \left|\frac{\langle y, x\rangle}{\langle x, x\rangle}\right|^2 \langle x, x\rangle \\ &= \langle z, z\rangle + \frac{|\langle x, y\rangle|^2}{\langle x, x\rangle}\end{aligned}$$

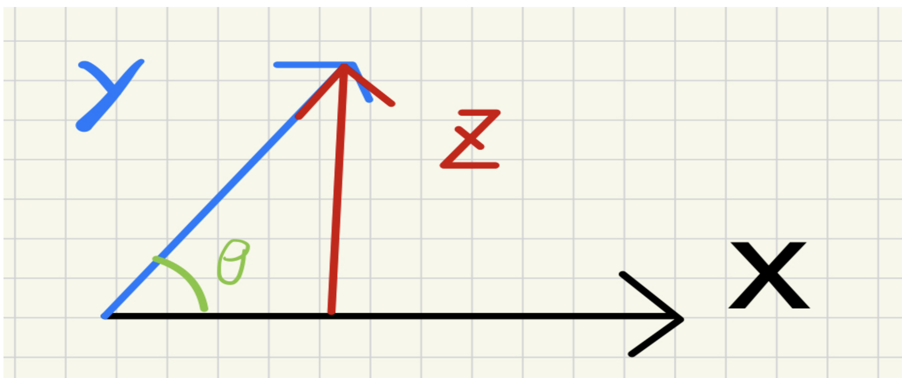
Because $\langle z, z\rangle \geq 0$, we now have

$$\langle y, y\rangle = \langle z, z\rangle + \frac{|\langle x, y\rangle|^2}{\langle x, x\rangle} \geq \frac{|\langle x, y\rangle|^2}{\langle x, x\rangle} \text{ (done)}$$

The equality hold true if and only if $\langle z, z\rangle = 0$. This explains the other two statements regarding the equality. ■

The proof is clearly geometrical. If one wish to remember the proof, one should see the trick we use is exactly

$$z \triangleq y - |y| (\cos\theta)\hat{x} \text{ is the projection of } y \text{ onto } x^{\perp}$$



Then all we do rest is just expanding $|y|^2 = |z + \tilde{x}|^2$, where $\tilde{x} = y - z = |y| (\cos\theta)\hat{x}$, which give the answer and is easy to compute since $z \cdot \tilde{x} = 0$.

Now, with Cauchy-Schwarz Inequality, we can check the triangle inequality

$$\begin{aligned}
\|x + y\|^2 &= \langle x + y, x + y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&= \langle x, x\rangle + \langle y, y\rangle + 2 \text{ Re } \langle x, y\rangle \\
&\leq \|x\|^2 + \|y\|^2 + 2\,|\langle x, y\rangle| \\
&\leq \|x\|^2 + \|y\|^2 + 2\|x\| \cdot \|y\| = \Big(\|x\| + \|y\|\Big)^2
\end{aligned}$$

(Euclidean Space Abstract Part)

By a **concrete Euclidean Space**, we mean some space of n-tuple $(x_1, \dots, x_n)$ over $\mathbb{R}$, equipped with inner product $\langle \cdot, \cdot \rangle_E$ defined by

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle_E = \sqrt{\sum_{k=1}^{n} (y_k - x_k)^2}$$

By an **Euclidean Space**, we simply mean a finite dimensional vector space V over $\mathbb{R}$, equipped with an inner product $\langle \cdot, \cdot \rangle$ such that there exists a concrete Euclidean space E and an isomorphism $\phi : V \to E$ such that

$$\langle x, y \rangle = \langle \phi(x), \phi(y) \rangle_E \qquad (x, y \in V)$$

Note that if you define $\langle \cdot, \cdot \rangle$ on the space of n-tuples $(x_1, \ldots, x_n)$ over $\mathbb{R}$ by

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = 2\sqrt{\sum_{k=1}^{n} (y_k - x_k)^2}$$

Then, the space of n-tuple is clearly not a concrete Euclidean space, and clearly an Euclidean space.

(SVD)

Chapter 5

Beauty

5.1 Euler's Formula

Suppose that we define

$$\exp(z) \triangleq \sum_{n=0}^{\infty} \frac{z^n}{n!} \qquad (z \in \mathbb{C})$$

$$\sin(z) \triangleq \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1} \qquad (z \in \mathbb{C})$$

$$\cos(z) \triangleq \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} \qquad (z \in \mathbb{C})$$

Some properties we are familiar with is now easily seen using basic technique we learned in Chapter 3.

(a) $\exp(z), \sin(z), \cos(z)$ are well defined on $\mathbb{C}$ by Cauchy-Hadamard 3.48.

(b) $\begin{cases} \exp(x) \\ \sin(x) \\ \cos(x) \end{cases} \in \mathbb{R}$ provided $x \in \mathbb{R}$.

(c) $\begin{cases} \exp(0) = 1 \\ \sin(0) = 0 \\ \cos(0) = 1 \end{cases}$

(d) $\exp(x)$ is strictly increasing on $\mathbb{R}$ $\quad (\because |y^n - x^n| \leq |y^n| + |x^n|)$

(e) $\exp(x) \nearrow \infty$ as $x \to \infty \quad (x \in \mathbb{R})$

(f) $\exp(x) \cdot \exp(y) = \exp(x+y) \quad (x \in \mathbb{C})$ by Merten's Theorem of Cauchy Product 3.9.16

(g) $\exp(x) \searrow 0$ as $x \to -\infty \quad (x \in \mathbb{R})$

(h) $\begin{cases} \frac{d}{dz}\exp(z) = \exp(z) \\ \frac{d}{dz}\sin(z) = \cos(z) \\ \frac{d}{dz}\cos(z) = -\sin(z) \end{cases} \quad (z \in \mathbb{C})$, using Term-by-Term Differentiation 3.10.1.

(i) $\exp(x)$ is convex on $\mathbb{R} \quad (\because (e^x)'' = e^x > 0)$

(j) $\exp(nz) = \big(\exp(z)\big)^n \quad (z \in \mathbb{C}, n \in \mathbb{Z})$, by induction and Merten's Theorem of Cauchy Product 3.9.16.

In particular, we have **Euler's Formula**.

Theorem 5.1.1. (Euler's Formula)

$$\exp(iz) = \cos(z) + i\sin(z) \qquad (z \in \mathbb{C})$$

Proof. Define

$$I_i(n) \triangleq \begin{cases} 1 & \text{if } n \equiv 0 \pmod 4 \\ i & \text{if } n \equiv 1 \pmod 4 \\ -1 & \text{if } n \equiv 2 \pmod 4 \\ -i & \text{if } n \equiv 3 \pmod 4 \end{cases}$$

Compute

$$\begin{aligned} \exp(iz) &= \sum_{n=0} \frac{(iz)^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{I_i(n) z^n}{n!} \qquad (5.1) \\ &= \sum_{n=0}^{\infty} \frac{I_i(2n)}{(2n)!} z^{2n} + \frac{I_i(2n+1)}{(2n+1)!} z^{2n+1} \quad \big(\because \text{ this is a sub-sequence of (5.1)}\big) \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} + i \cdot \frac{(-1)^n}{(2n+1)!} z^{2n+1} \end{aligned}$$

Now, we can conclude

$$\begin{aligned}
\cos(z) + i\sin(z) &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} + i \cdot \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1} \\
&= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} + \sum_{n=0}^{\infty} i \cdot \frac{(-1)^n}{(2n+1)!} z^{2n+1} \\
&= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} + i \cdot \frac{(-1)^n}{(2n+1)!} z^{2n+1} = \exp(iz)
\end{aligned}$$

■

5.2 Alternative Definitions of e^x

Theorem 5.2.1. (First Characterization) for all $x \in \mathbb{R}$, the sequence $\{(1+\frac{x}{n})^n\}_{n\in\mathbb{N}}$ has limit

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = \sum_{k=0}^{\infty}\frac{x^k}{k!}$$

Proof. The proof is trivial if $x = 0$. First suppose $x \in \mathbb{R}^+$. Set

$$t_n \triangleq \left(1+\frac{x}{n}\right)^n \text{ and } s_n \triangleq \sum_{k=0}^{n}\frac{x^k}{k!}$$

We wish to show

$$\limsup_{n\to\infty} t_n \le \lim_{n\to\infty} s_n \le \liminf_{n\to\infty} t_n$$

We first prove

$$\limsup_{n\to\infty} t_n \le \lim_{n\to\infty} s_n$$

Use Binomial Theorem to compute

$$\begin{aligned} t_n = \left(1+\frac{x}{n}\right)^n &= \sum_{k=0}^{n}\binom{n}{k}\left(\frac{x}{n}\right)^k \\ &= \sum_{k=0}^{n}\frac{x^k}{k!}\cdot\frac{n(n-1)\cdots(n-k+1)}{n^k} \le \sum_{k=0}^{n}\frac{x^k}{k!} = s_n \text{ (done)} \end{aligned}$$

We now prove

$$\lim_{n\to\infty} s_n \le \liminf_{n\to\infty} t_n$$

Fix ϵ. Because $s_n \nearrow$, we know there exists m such that

$$s_m > \lim_{n\to\infty} s_n - \epsilon$$

Fix such m. Observe

$$t_n = \sum_{k=0}^{n}\binom{n}{k}\left(\frac{x}{n}\right)^k \ge \sum_{k=0}^{m}\frac{x^k}{k!}\cdot\frac{n(n-1)\cdots(n-k+1)}{n^k} \qquad (n \ge m)$$

Clearly, there exists N such that

$$\forall n > N, \left| \Big(\sum_{k=0}^{m} \frac{x^k}{k!} \cdot \frac{n(n-1)\cdots(n-k+1)}{n^k} \Big) - s_m \right| \leq s_m - (\lim_{n\to\infty} s_n - \epsilon)$$

Then, we see for all $n > \max\{N, m\}$

$$\begin{aligned} t_n \geq \sum_{k=0}^{m} \frac{x^k}{k!} \cdot \frac{n(n-1)\cdots(n-k+1)}{n^k} &\geq s_m - (s_m - \lim_{n\to\infty} s_n - \epsilon) \\ &= \lim_{n\to\infty} s_n - \epsilon \text{ (done)} \quad \text{(done)} \end{aligned}$$

For negative real, we only wish to show

$$\sum_{n=0}^{\infty} \frac{(-x)^n}{n!} = \Big(\sum_{n=0}^{\infty} \frac{x^n}{n!} \Big)^{-1} \text{ and } \lim_{n\to\infty} \left(1 - \frac{x}{n}\right)^n = \left(\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n \right)^{-1}$$

Because the convergence is absolute, we can use Merten's Theorem for Cauchy product (Theorem 3.9.16)to compute

$$\begin{aligned} \Big(\sum_{n=0}^{\infty} \frac{(-x)^n}{n!} \Big) \cdot \Big(\sum_{n=0}^{\infty} \frac{x^n}{n!} \Big) &= \sum_{n=0}^{\infty} \Big(\sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} \Big) \frac{x^n}{n!} \\ &= \sum_{n=0}^{\infty} (1-1)^n \frac{x^n}{n!} = 1 \text{ (done)} \end{aligned}$$

We first prove

$$\lim_{n\to\infty} \left(1 - \frac{x^2}{n^2}\right)^n = 1$$

Computing term by term, it is clear that

$$\limsup_{n\to\infty} \left(1 - \frac{x^2}{n^2}\right)^n \leq 1$$

Using Bernoulli's Inequality (Theorem 3.5.1), we see that for large enough n, we have

$$\left(1 - \frac{x^2}{n^2}\right)^n \geq 1 - \frac{x^2}{n}$$

This then implies

$$\liminf_{n\to\infty} \left(1 - \frac{x^2}{n^2}\right)^n \geq 1 \text{ (done)}$$

Compute

$$\lim_{n\to\infty}\left(1-\frac{x}{n}\right)^n = \lim_{n\to\infty}\left(\frac{1-\frac{x^2}{n^2}}{1+\frac{x}{n}}\right)^n$$
$$= \frac{\lim_{n\to\infty}(1-\frac{x^2}{n^2})^n}{\lim_{n\to\infty}(1+\frac{x}{n})^n} = \frac{1}{\lim_{n\to\infty}(1+\frac{x}{n})^n} \text{ (done)}$$

■

$$\ln(x) \triangleq \int_1^x \frac{1}{t}dt$$

By FTC (Theorem 3.7.1), it is easy to see that

$$\frac{d}{dx}\ln(x) = \frac{1}{x} \qquad (x \in \mathbb{R}^+)$$

To see

$$\ln(xy) = \ln(x) + \ln(y)$$

Fix $y \in \mathbb{R}^+$ and set

$$f(x) \triangleq \ln(x) \text{ and } g(x) \triangleq \ln(xy)$$

Conclude $f'(x) = g'(x)$, and use FTC (Theorem 3.7.2) to conclude $f - g$ is some fixed constant k. Now, see that

$$g(1) = f(1) + k \implies k = \ln(y)$$

Then, we have

$$\ln(xy) = g(x) = f(x) + k = \ln(x) + \ln(y)$$

Using induction, it is now easy to see

$$\ln(x^n) = n\ln(x) \qquad (n \in \mathbb{Z}_0^+)$$

Theorem 5.2.2. (Second Characterization)

5.3 Alternative Definition for $\sin$ and $\cos$

5.4 Fundamental Theorem of Algebra

Theorem 5.4.1. (Fundamental Theorem of Algebra)