NCKU 112.2
General Analysis

Eric Liu

Contents

Chapter 1

General Topology

1.1 Equivalent Characterizations of Basic Notions

Abstract

This section give a compact and comprehensive development of some of the most basic notions in the study of topology. In this section, $(X, \mathscr{T})$ is a topological space.

Given an arbitrary subset $E \subseteq X$, we

(a) say $x \in X$ is a **limit point of** E if every open O containing x contain a point $y \in E$ such that $y \neq x$.

(b) say $x \in E$ is an **interior point of** E if there exists $O \in \mathscr{T}$ such that $x \in O \subseteq E$.

(c) define the **interior** E° **of** E to be the union of all open sets contained by E.

(d) say $E \subseteq X$ is a **closed set** if $E^c \in \mathscr{T}$.

(e) define the **closure** $\overline{E}$ **of** E by $\overline{E} \triangleq E \cup E'$ where E' is the set of limit points of E.

(f) define the **boundary** ∂E **of** E by $\partial E \triangleq \overline{E} \setminus E^\circ$

Theorem 1.1.1. (Equivalent Definitions of Interior) The following sets are equivalent

(a) E°

(b) The largest open set contained by E.

(c) The set of interior points of E.

Proof. Check straightforward. ■

Theorem 1.1.2. (Equivalent Definitions of Closed) The following are equivalent.

(a) E is closed.

(b) the set of limit points of E is contained by E.

(c) $\overline{E} = E$.

Proof. The proof of (a) $\implies$ (b) $\implies$ (c) are straight forward. The proof of (c) $\implies$ (a) follows from first noting no $x \in E^c$ is a limit point of E. Then shows $E^c = \bigcup_{x \notin E} O_x$ where O_x is an open set containing x and disjoint with E. ■

Theorem 1.1.3. (Equivalent Definitions of Closure) The following sets are equivalent.

(a) $\overline{E}$

(b) $((E^c)^\circ)^c$

(c) The smallest closed set containing E.

(d) $\{x \in X :$ every open O containing x intersect with $E\}$

Proof. It is straightforward to check $\overline{E}$ is closed. For each closed F containing E, it is straightforward to check $E' \subseteq F' \subseteq F$. This established (a) = (c). It is straightforward to check $(\overline{E})^c = (E^c)^\circ$ using the largest open set and the smallest closed set characterization of interior and closure. This established (b) = (c). It is straightforward to check (a) = (d). ■

Theorem 1.1.4. (Equivalent Definitions of Boundary) The following sets are equivalent.

(a) ∂E

(b) $\overline{E} \cap \overline{E^c}$

(c) $\{x \in X :$ every open O containing x intersect with both E and $E^c\}$

Proof. It is straightforward to check $\partial E = \overline{E} \cap \overline{E^c}$ using $(E^\circ)^c = \overline{E^c}$. This established (a) = (b). It is straightforward to check (b) = (c). ■

The development above starts with a given topology. Indeed, in elementary class of calculus, one only develop the topological theories for metric spaces after given the metric topology by ϵ-language. We now show a different approach.

Suppose X has no topology, and we are given a collection $\mathcal{S}$ of subsets of X such that $\bigcup \mathcal{S} = X$. To uniquely generate a topology $\mathscr{T}$ from $\mathcal{S}$, one has two equivalent approaches.

(a) Let $\mathscr{T}$ be the smallest topology containing $\mathcal{S}$.

(b) Let $\mathscr{T}$ be the collection of unions of finite intersections of elements of $\mathcal{S}$.

More explicitly, for (b), one first construct the set $\mathcal{A}$ of finite intersections of elements of $\mathcal{S}$.

$$\mathcal{A} \triangleq \left\{ \bigcap S : S \subseteq \mathcal{S}, |S| \in \mathbb{N} \right\}$$

And let $\mathscr{T}$ be the collection of unions of elements of $\mathcal{A}$.

$$\mathscr{T} \triangleq \left\{ \bigcup E : E \subseteq \mathcal{A} \right\}$$

It is straightforward to check $\mathscr{T}$ is closed under union and $\mathscr{T}$ contain $\varnothing$ and X. To check $\mathscr{T}$ is closed under finite intersection, utilize

$$\left(\bigcup_{i \in I} A_i \right) \cap \left(\bigcup_{j \in J} B_j \right) = \bigcup_{i \in I, j \in J} A_i \cap B_j$$

To check the two approaches (a) and (b) are equivalent, one note that any topology containing $\mathcal{S}$ must also contain $\mathcal{A}$, and thus must also contain $\mathscr{T}$.

Conversely, if one is first given a topology $\mathscr{T}$, one can study the question: What conditions must $\mathcal{S}$ suffices, so $\mathscr{T}$ is the topology generated by $\mathcal{S}$?

We now define the term **basis** as a sub-collection $\mathcal{B} \subseteq \mathscr{T}$ such that for each $O \in \mathscr{T}$, there exists a sub-collection $\mathcal{B}_0 \subseteq \mathcal{B}$ such that $O = \bigcup \mathcal{B}_0$. Note that in the last paragraph, $\mathcal{A}$ is a basis for $\mathscr{T}$.

Theorem 1.1.5. (Equivalent Definition of Basis) The following are equivalent.

(a) $\mathcal{B}$ is a basis.

(b) For all $O \in \mathscr{T}$ and $x \in O$, there exists $B \in \mathcal{B}$ such that $x \in B \subseteq O$.

Proof. Check straightforward. ∎

In fact, given any collection $\mathcal{B}$ of subsets of X, as long as $\mathcal{B}$ satisfy

(a) For all $x \in X$, there exists $B \in \mathcal{B}$ such that $x \in B$.

(b) For all $x \in X$ and $B_1, B_2 \in \mathcal{B}$ containing x, there exists $B_3 \in \mathcal{B}$ such that $B_3 \subseteq B_1 \cap B_2$ and $x \in B_3$.

We can define a topology $\mathscr{T}_{\mathcal{B}}$ on X by

$$\mathscr{T}_{\mathcal{B}} \triangleq \left\{ \bigcup \mathcal{B}_0 \subseteq X : \mathcal{B}_0 \in 2^{\mathcal{B}} \right\}$$

It is straightforward to check

(a) $\mathscr{T}_{\mathcal{B}}$ is indeed a topology.

(b) $\mathcal{B}$ is a basis of $\mathscr{T}_{\mathcal{B}}$.

(c) $\mathscr{T}_{B}$ is the smallest topology containing $\mathcal{B}$.

Such topology $\mathscr{T}_{\mathcal{B}}$ is called the **topology generated** by $\mathcal{B}$, and if $\mathcal{B}$ satisfy the two conditions listed in the beginning of this paragraph, we say $\mathcal{B}$ is a **pre-basis** of X. Now, given a basis $\mathcal{B}'$ of $\mathscr{T}$, it is again straightforward to check

(a) $\mathcal{B}'$ is a pre-basis of X.

(b) $\mathscr{T}$ is the topology generated by $\mathcal{B}'$.

In particular, (a) shows that given any collection $\mathcal{S}'$ of subsets of X such that $\bigcup \mathcal{S}' = X$, the collection $\mathcal{A}'$ of finite intersections of elements of $\mathcal{S}'$ form a pre-basis of X. If $\mathcal{S}'$ is a collection of subsets of X such that $\bigcup \mathcal{S}' = X$ and $\mathscr{T}$ is the smallest topology containing $\mathcal{S}'$, without ambiguity, we say $\mathcal{S}'$ is a **subbasis** of $\mathscr{T}$ and $\mathscr{T}$ is the topology **generated by the subbasis** $\mathcal{S}'$. Immediately, one should check that a basis of $\mathscr{T}$ is also a subbasis of $\mathscr{T}$.

Corollary 1.1.6. (Equivalent Definitions of Subbasis) The following are equivalent

(a) $\mathcal{S}$ is a subbasis of $\mathscr{T}$.

(b) $\mathscr{T}$ is the collections of unions of finite intersections of $\mathcal{S}$.

We now develop the theory of continuity by first giving a pointwise definition. Given another topological space $(Y, \mathscr{S})$ and a function $f : X \to Y$, we say f is **continuous at** $x \in X$ if for all open O containing $f(x)$, there exists open E containing x such that $f(E) \subseteq O$. We say f is a **continuous (or $(\mathscr{T}, \mathscr{S})$-continuous, if necessary) function** if f is continuous at all $x \in X$.

Theorem 1.1.7. (Equivalent Definitions of Continuous function) The following are equivalent

(a) f is continuous.

(b) $f^{-1}(O) \in \mathscr{T}$ for all $O \in \mathscr{S}$.

(c) $f^{-1}(F)$ is closed for all closed F in Y.

(d) For all $B \subseteq Y$, $f^{-1}(B^\circ) \subseteq (f^{-1}(B))^\circ$

(e) For all $A \subseteq X$, $f(\overline{A}) \subseteq \overline{f(A)}$.

(f) For all $B \subseteq Y, \overline{f^{-1}(B)} \subseteq f^{-1}(\overline{B})$

(g) For all basis $\mathcal{B}$ of Y, $f^{-1}(B) \in \mathscr{T}$ for all $B \in \mathcal{B}$.

Proof. It is straightforward to check (a) $\implies$ (b) $\implies$ (c). Fix $x \in X$ and $O \in \mathscr{S}$ containing $f(x)$. Check $x \in (f^{-1}(O^c))^c$ and $f((f^{-1}(O^c))^c) \subseteq O$. This give us (c) $\implies$ (a). Because, respectively, $f^{-1}(B^\circ) \subseteq f^{-1}(B), A \subseteq f^{-1}(\overline{f(A)})$ and $f^{-1}(B) \subseteq f^{-1}(\overline{B})$, we have, respectively, (b) $\implies$ (d), (c) $\implies$ (e) and (c) $\implies$ (f). It is straightforward to check (d) $\implies$ (b). To check (e) $\implies$ (f), let $A = f^{-1}(B)$. (f) $\implies$ (c) is straightforward. Finally, (b) $\iff$ (g) is straightforward. ∎

Now, one may wonder: Why isn't "For all $A \subseteq X, f(A^\circ) \subseteq$

Now, given a subset $Y \subseteq X$, we define the **subspace topology** $\mathscr{T}_Y$ on Y by $\mathscr{T}_Y \triangleq \{O \cap Y : O \in \mathscr{T}\}$. Immediately, one can check the well-behavior of our definitions for subspace topology: A set $E \subseteq Y$ open in X remains open in $(Y, \mathscr{T}_Y)$.

Theorem 1.1.8. (Equivalent Definition of Subspace Topology) Given a basis $\mathcal{B}$ of $\mathscr{T}$, the following are equivalent.

(a) $\mathscr{T}_Y$

(b) The topology generated by the pre-basis $\mathcal{B}_Y \triangleq \{B \cap Y : B \in \mathcal{B}\}$ in Y.

Proof. Check straightforward. ∎

Theorem 1.1.9. (Equivalent Definitions of Connected)

Theorem 1.1.10. (Equivalent Definition of Finer) Given another topology $\mathscr{T}'$ on X, we say $\mathscr{T}'$ is **finer** than $\mathscr{T}$ if one of the following two equivalent conditions is satisfied

(a) $\mathscr{T} \subseteq \mathscr{T}'$

(b) Given any basis $\mathcal{B}$ of $\mathscr{T}$ and any basis $\mathcal{B}'$ of $\mathscr{T}'$, for all $x \in X$ and basic open $B \in \mathcal{B}$ containing x, there exists basic open $B' \in \mathcal{B}'$ such that $x \in B' \subseteq B$

Proof. From (a) to (b), one find B' by noting $B \in \mathscr{T}'$ and utilizing the definition of basis. From (b) to (a), one show $O \in \mathscr{T}$ belongs to $\mathscr{T}'$ by taking $O = \bigcup_{x \in O} B'_x$. ∎

Chapter 2

Metric Space

2.1 Completion

2.2 Bounded and Totally Bounded

2.3 Compactness

2.4 Limit Interchange

In this section, we

(a) discuss the condition in which we can change the limit order of double sequence in general metric space. (Theorem 2.4.1 and Theorem 2.4.2)

(b) prove that the space of functions is complete if and only if the codomain is complete. (Theorem 2.4.3)

(c) prove that the uniform limit of a sequence of convergent sequences in a complete metric space converge. (Theorem 2.4.5)

Remark on structure of the Theory: The proof of (Theorem 2.4.5: convergent sequences in complete metric space is closed under uniform convergence) relies on (Theorem 2.4.1: exchange limit order), while that of (Theorem 2.4.3: Space of functions (X^Y, d_∞) is complete iff Y is complete) does not.

(Theorem 2.4.1: exchange limit order) will later be used to prove the Uniform Limit Theorem (Theorem 2.5.2) which is a "pointwise" Theorem, and justify abundant of limit exchange, e.g. (Theorem 2.5.1: exchange limit order for functions)

An important consequence of (Corollary 2.4.4: space of bounded functions into complete space is complete) is that $\big(L(\mathbb{R}^n, \mathbb{R}^m), \|\cdot\|_{\text{op}}\big)$ is complete. This will be later shown with extra tools.

Theorem 2.4.1. (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) $A_n \to A$

We have

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

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

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

■

Theorem 2.4.2. (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 $k \to \infty$

We have

$$A_n \to A$$

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

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

■

In summary of Theorem 2.4.1 and Theorem 2.4.2, 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 2.6), 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 2.4.3, we give a necessary and sufficient condition for $\big(X^Y, d_\infty\big)$ to be complete.

Theorem 2.4.3. (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

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

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

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

Fix ϵ. We wish

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

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

$$\forall n, m > N, \forall x \in X, d\big(f_n(x), f_m(x)\big) < \frac{\epsilon}{2} \tag{2.5}$$

We claim

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

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

$$\exists m \in \mathbb{N}, d\big(f_m(x), f(x)\big) < \frac{\epsilon}{2} \tag{2.6}$$

Then from Equation 2.5 and Equation 2.6, 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

$$3\epsilon = d\big(f(x_1), f(x_2)\big) \leq d\big(f(x_1), g(x_1)\big) + d\big(g(x_1), g(x_2)\big) + d\big(g(x_2), f(x_2)\big) \tag{2.7}$$

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

$$\forall x \in X, d\big(f(x), g(x)\big) < \epsilon \tag{2.8}$$

Then by Equation 2.7 and Equation 2.8, 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 2.4.4. (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 2.4.3, 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 2.4.3 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 2.4.3.

Theorem 2.4.5. (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 2.4.2, 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

$$\forall n,m > N, d_\infty(\{a_{n,k}\}_{k\in\mathbb{N}}, \{a_{m,k}\}_{k\in\mathbb{N}}) < \frac{\epsilon}{3} \tag{2.9}$$

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

$$d(a_{n,j}, A_n) < \frac{\epsilon}{3} \text{ and } d(a_{m,j}, A_m) < \frac{\epsilon}{3} \tag{2.10}$$

Then from Equation 2.9 and Equation 2.10, we can deduce

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

■

2.5 Closed under Uniform Convergence

Given $(E, d_E), (Y, d_Y)$ and a sequence of functions $f_n : E \to Y$, converging uniformly to some $f : E \to Y$ such that each f_n has the property

(a) Boundedness

(b) Unboundedness

(c) Continuity

(d) Uniform continuity

(e) K-Lipschitz continuity

on E, then f also has the same property. These fact will later be proved in **??** and **??**. are again all closed under uniform convergence, where the proof for continuity is closed under uniform convergence use Theorem 2.4.1 as a lemma.

The reason we require the co-domain Y of sequence to be complete is explained in the last paragraph of Section 2.6. 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 2.5.1. (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) $f_n \to f$ uniformly on E

(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

$$a_{n,k} \triangleq f_n(t_k) \tag{2.11}$$

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 2.4.2, we have

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

■

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

(a) continuity

(b) uniform continuity

(c) K-Lipschitz Continuity

Theorem 2.5.2. (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

$$\Big\{f_n(x_k)\Big\}_{k\in\mathbb{N}} \to \Big\{f(x_k)\Big\}_{k\in\mathbb{N}} \text{ uniformly as } n \to \infty \tag{2.12}$$

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

$$\forall n \in \mathbb{N}, f_n(x_k) \to f_n(x) \text{ as } k \to \infty \tag{2.13}$$

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

$$f_n(x) \to f(x) \tag{2.14}$$

Now, because Equation 2.12, Equation 2.13 and Equation 2.14, by Theorem 2.4.1, 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 2.5.3. (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

$$\forall x \in X, d_Y\big(f_m(x), f(x)\big) < \frac{\epsilon}{3} \tag{2.15}$$

Because f_m is uniformly continuous, we know

$$\exists \delta, \forall x, y \in X, d_X(x,y) < \delta \implies d_Y\big(f_m(x), f_m(y)\big) < \frac{\epsilon}{3} \tag{2.16}$$

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 2.15 and Equation 2.16, we have

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

■

Theorem 2.5.4. **(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

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

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

$$\forall z \in X, d_Y\big(f(z), f_m(z)\big) < \frac{\epsilon}{2} \tag{2.17}$$

Because f_m is K-Lispchitz continuous, we know

$$d_Y\big(f_m(x), f_m(y)\big) \le K d_X(x,y) \tag{2.18}$$

Now, from Equation 2.18 and Equation 2.17, we now see

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

■

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)

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

2.6 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

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

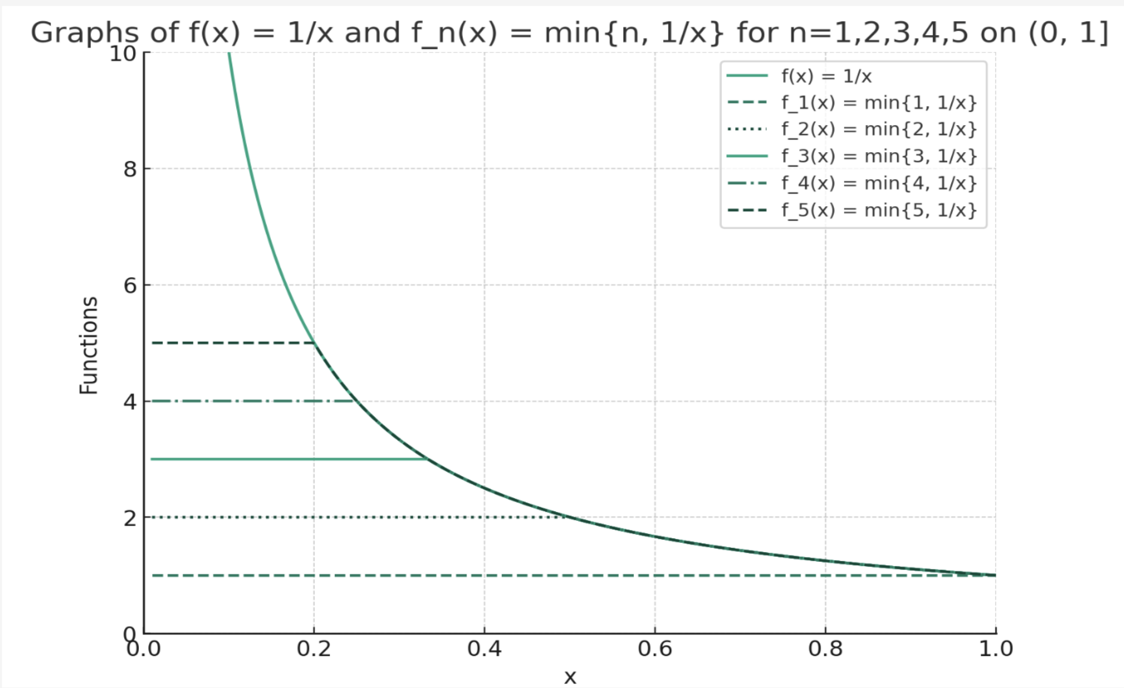
With this in mind, it shall be clear that the uniform limit of bounded (resp: unbounded) functions is always bounded (resp: unbounded).

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)

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

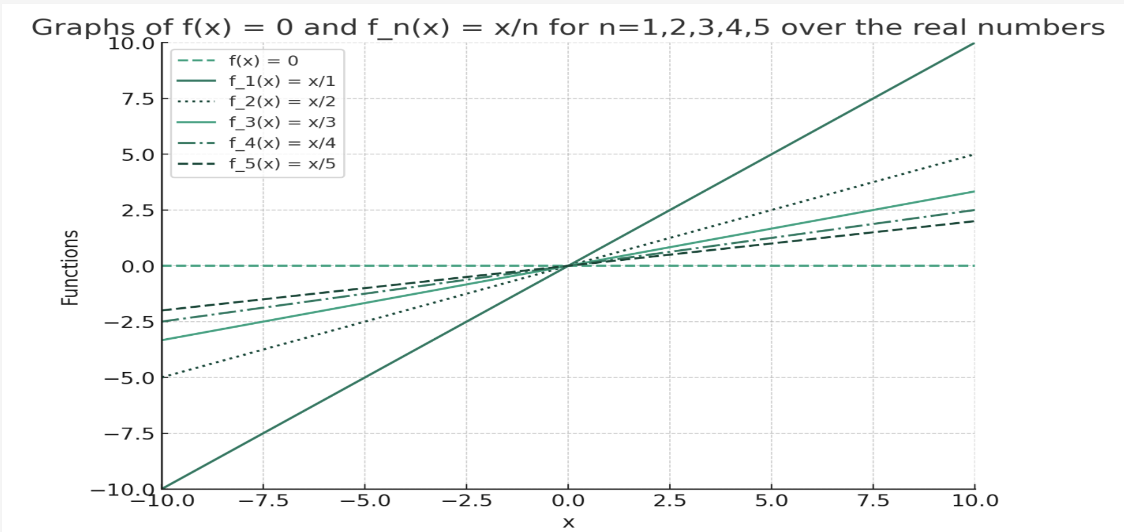
It is clear that $\forall n \in \mathbb{N}, f_n(x) \in [0, n]$, and the limit $f : X \to \mathbb{R}$ is $f(x) = \frac{1}{x}$



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 2.4.3, 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.

2.7 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 2.7.1. (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

$$\forall y \in B_{\delta_x}(x), d_Y\big(f(x), f(y)\big) < \frac{\epsilon}{2} \text{ for all } f \in \mathcal{F} \tag{2.19}$$

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 (2.19), we can now deduce

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

■

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 2.7.2. (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 $\Big\{U_{x,n} : x \in P_n\Big\}$ is a cover of X. Let $P = \bigcup_{n\in\mathbb{N}} P_n$. (done)

Now, we wish to

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

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

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

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

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

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

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

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

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

such N works

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

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

Lastly, we wish to prove

f_{n_k} is uniformly Cauchy

Fix ϵ. We wish

to find N 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)

$\blacksquare$

2.8 Banach Fixed Point Theorem

This section give a complete statement and proof of Banach Fixed Point Theorem. The setting is

(a) a metric space $\left(X, d_X\right)$

(b) a subset $E \subseteq X$

(c) another metric space $\left(Y, d_Y\right)$

(d) a function $f : E \to Y$

(e) another function $g : E \to X$

We say f is a **contraction** on E if there exists $r \in [0, 1)$ such that

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

or equivalently

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

Note that the restriction of a contraction is again a contraction. We say g admits a **fixed point** x if we have

$$g(x) = x$$

Theorem 2.8.1. (Banach Fixed Point Theorem) If g is a contraction that maps E into X, then

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

Moreover, if E is complete and $g(E) \subseteq E$, then

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

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

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

Proof. We first

prove the uniqueness of the fixed point

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

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

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

Suppose E is complete and $g(E) \subseteq E$. We now

prove the existence of the fixed point

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

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

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

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

Observe

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

Note that contraction is Lipschitz thus continuous, and note that $\lim_{n\to\infty} g^n(x) \in E$. This allow us to carry the below limit process

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

■

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

(a) Inverse Function Theorem

(b) Picard-Lindelof Theorem

(c) Nash-Embedding Theorem

Chapter 3

Algebraic Topology

3.1 Fundamental Group

Chapter 4

Differential Calculus

4.1 Basic Technique on Sequence and Series

Abstract

This

Theorem 4.1.1. **(Summation by Part)**

Theorem 4.1.2. **(Dirichlet's Test)**

Theorem 4.1.3. **(Abel's Test)**

Theorem 4.1.4. **(Absolutely Convergent Series Unconditionally Converge)**

Theorem 4.1.5. **(Riemann Rearrangement Theorem)**

Theorem 4.1.6. **(Fubini's Theorem for Double Series)**

Theorem 4.1.7. **(Geometric Series)**

$$|z| < 1 \implies \sum_{n=0}^{\infty} \frac{1}{z^n} = \frac{1}{1-z}$$

Theorem 4.1.8. **(Ratio and Root Test)**

Theorem 4.1.9. **(Root Test is Stronger Than Ratio Test)**

Theorem 4.1.10. (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 4.1.11. (Weierstrass M-test) Given sequences $f_n : X \to \mathbb{C}$, and suppose

$$\forall n \in \mathbb{N}, \forall x \in X, |f_n(x)| \leq M_n \tag{4.1}$$

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 2.4.4, we only wish to prove

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

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 4.1, we have

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

■

4.2 Operator Norm

Abstract

This section

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**. Now, 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

$$T \text{ is bounded } \overset{\triangle}{\iff} \exists M \in \mathbb{R}, \forall x \in \mathcal{X}, \|Tx\| \leq M\|x\|$$

The proof of equivalency is simple. For $(\longrightarrow)$, observe

$$\|Tx\| = \|x\| \cdot \|T\frac{x}{\|x\|}\| \leq \Big(\sup\{\|Ty\| : \|y\| = 1\}\Big)\|x\|$$

For $(\longleftarrow)$, observe

$$\|Tx - Ty\| = \|T(x-y)\| \leq M\|x-y\|$$

We first show that a linear transformation is continuous if and only if it is bounded.

Theorem 4.2.1. (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. If T is bounded, we see that T is Lipschitz.

$$\|Tx - Ty\| \leq M\|x-y\|$$

Now, suppose T is linear and continuous at 0. Let ϵ satisfy

$$\sup_{\|y\| \leq \epsilon} \|Ty\| \leq 1$$

Observe that for all $x \in \mathcal{X}$, we have

$$\|Tx\| = \frac{\|x\|}{\epsilon}\Big\|T\frac{\epsilon x}{\|x\|}\Big\| \leq \frac{\|x\|}{\epsilon}$$

■

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 if V is finite-dimensional, then all norms on V are equivalent. This property will later show its value, as used to prove linear map of finite-dimensional domain is always continuous

Theorem 4.2.2. (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

$$\Big\| \sum \alpha_i e_i \Big\|_\infty \triangleq \max |\alpha_i|$$

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

$$\|x\| = \Big\|\sum \alpha_i e_i\Big\| \le \sum |\alpha_i| \, \|e_i\| \le \|x\|_\infty \sum \|e_i\| = M\|x\|_\infty \text{ (done)}$$

Note that reverse triangle inequality give us

$$\Big|\|x\| - \|y\|\Big| \le \|x - y\| \le M\|x - y\|_\infty \tag{4.2}$$

Then we can check that

(a) $\|\cdot\| : \Big(V, \|\cdot\|_\infty\Big) \to \mathbb{R}$ is Lipschitz continuous because of Equation 4.2.

(b) $S \triangleq \{y \in V : \|y\|_\infty = 1\}$ is sequentially compact and non-empty.

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 $x \in V$ and compute

$$m\|x\|_\infty = \|x\|_\infty(\min_{y\in S} \|y\|) \le \|x\|_\infty \cdot \Big\|\frac{x}{\|x\|_\infty}\Big\| = \|x\| \text{ (done) } \text{ (done)}$$

■

Theorem 4.2.3. (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

$$\text{to find } \delta \text{ such that } \forall h \in \mathcal{X} : \|h\| \le \delta, \|T(x + h) - Tx\| \le \epsilon$$

Let $\{e_1, \dots, e_n\}$ be a basis of $\mathcal{X}$. Note that $\|\sum \alpha_i e_i\|_1 \triangleq \sum |\alpha_i|$ is a norm. Because $\mathcal{X}$ is finite-dimensional, we know $\|\cdot\|$ and $\|\cdot\|_1$ are equivalent. Then, we can fix $M \in \mathbb{R}^+$ such that

$$\|x\|_1 \le M\|x\| \quad (x \in V)$$

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

$$\begin{aligned}\|T(x+h) - Tx\| &= \|\sum \alpha_i Te_i\| \\ &\leq \sum |\alpha_i| \, \|Te_i\| \\ &\leq \|h\|_1 (\max \|Te_i\|) \\ &\leq M\|h\|(\max \|Te_i\|) = \epsilon \text{ (done)}\end{aligned}$$

■

We now see that, because Linear transformation is bounded if and only if it is continuous and Linear map of finite-dimensional domain is always continuous, 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 4 (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

$$\|T\|_{\text{op}} \triangleq \inf\{M \in \mathbb{R}^+ : \forall x \in \mathcal{X}, \|Tx\| \leq M\|x\|\} \tag{4.3}$$

Before we show that our definition is indeed a norm, we first give some equivalent definitions and prove their equivalency.

Theorem 4.2.4. (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 Equation 4.3, we have

$$\|T\|_{\text{op}} = \sup_{x\in\mathcal{X}, x\neq 0} \frac{\|Tx\|}{\|x\|}$$

Proof. Define $J = \{M \in \mathbb{R}^+ : \forall x \in \mathcal{X}, \|Tx\| \leq M\|x\|\}$ and observe

$$J = \{M \in \mathbb{R}^+ : M \geq \frac{\|Tx\|}{\|x\|}, \forall x \neq 0 \in \mathcal{X}\}$$

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

$$\|T\|_{\text{op}} = \sup_{x\in\mathcal{X}, x\neq 0} \frac{\|Tx\|}{\|x\|} \tag{4.4}$$

$$= \sup_{x\in\mathcal{X}, \|x\|=1} \|Tx\| \tag{4.5}$$

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 Equation 4.4 to show $\|T\|_{\text{op}} \geq \frac{\|Tx\|}{\|x\|} > 0$.

(b) For absolute homogenity, use Equation 4.5 and $\|Tcx\| = |c| \cdot \|Tx\|$.

(c) For triangle inequality, use Equation 4.5 and $\|(T_1 + T_2)x\| \leq \|T_1x\| + \|T_2x\|$.

Naturally, and very very importantly, Equation 4.4 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

$$\frac{\|T_1T_2x\|}{\|x\|} \leq \frac{\|T_1\|_{\text{op}} \cdot \|T_2\|_{\text{op}} \cdot \|x\|}{\|x\|} = \|T_1\|_{\text{op}} \cdot \|T_2\|_{\text{op}}$$

Then Equation 4.4 give us

$$\|T_1T_2\|_{\text{op}} \leq \|T_1\|_{\text{op}} \cdot \|T_2\|_{\text{op}}$$

4.3 Directional Derivative and Gradient

Abstract

This short section introduce the idea of directional derivative and gradient. It shall be noted that, although both gradient and directional derivative are defined for real-valued function in this section, the notion of directional derivative can be easily generalized to function between Euclidean space; while the notion of gradient, as the way we define it, is only for real-valued function.

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

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - df_x(h)\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}} = 0 \tag{4.6}$$

Immediately, we should check that the linear approximation is unique. Suppose df_x and df'_x both satisfy Equation 4.6. 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

$$\frac{(df_x - df'_x)th}{t\|h\|_{\mathcal{X}}} \to 0 \text{ as } t \to 0 \tag{4.7}$$

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 Equation 4.7.

It shall be quite clear that a function f differentiable at x must be continuous at x, by noting the nominator of Equation 4.6 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 4.3.1. (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 4.3.1)

(a) The limit on the right hand side is done in $\Big(\mathbb{R}, |\cdot|\Big)$

(b) If f is differentiable at x, then we have

$$\partial_{av+bw} f(x) = df_x(av+bw) = adf_x(v) + bdf_x(w) = a\partial_v f(x) + b\partial_w f(x) \tag{4.8}$$

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 5 (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 4.3.2. (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

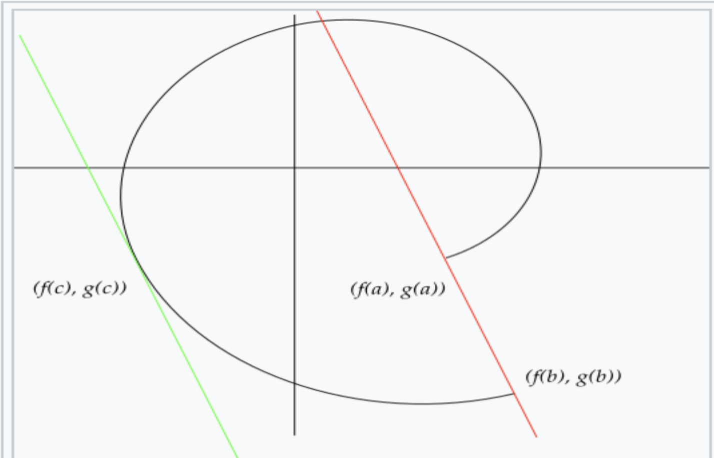
$$\nabla f(x) = \sum_{i=1}^{n} df_x(v_i)v_i = \sum_{i=1}^{n} \partial_{v_i} f(x) v_i$$

This is how you compute $\nabla f(x)$ when you have to.

4.4 MVT

Abstract

This is a short section introducing MVT. It is proved here because it will be used in next section.



Geometrical meaning of Cauchy's theorem

Theorem 4.4.1. **()**

Theorem 4.4.2. **(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. Define h on (a, b) by

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

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

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

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 4.4.3. (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. The proof is done by applying Cauchy's MVT, where $g(x) \triangleq x$. ■

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 differentiability 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

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

4.5 Differentiability Theorem

Abstract

This section prove

(a) The matrix representation of derivative for function between Euclidean spaces

(b) Differentiability Theorem

Note that the proof of Differentiability Theorem use MVT and the fact that all norms on $\mathbb{R}^k$ are equivalent 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 4.5.1), we introduce another notation. If we are provided a normal 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 4.5.1. (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

$$\lim_{t\to 0} \frac{|R_i(t)|}{|t|} = 0 \tag{4.9}$$

Compute

$$\begin{aligned} f_j(x+te_i) - f_j(x) &= \big(f(x+te_i) - f(x)\big) \cdot q_j \\ &= \big(R_i(t) + df_x(te_i)\big) \cdot q_j \\ &= R_i(t) \cdot q_j + t df_x(e_i) \cdot q_j \end{aligned}$$

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 4.9, 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. ∎

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 in Directional Derivative and Gradient. Consider a less trivial one.

Example 6 (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 4.5.1 (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 differentiability. 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 $h \mapsto f'(x)h$. We now give proof to the Differentiability

Theorem, which links between the continuity of total derivative and the continuity of partial derivatives.

Theorem 4.5.2. (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. Because f is differentiable on E, we know $\partial_i f_j$ exists on E by Theorem 4.5.1. 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 the the matrix representation, we know

$$\partial_i f_j(y) - \partial_i f_j(x) = (df_y - df_x)e_i \cdot q_j$$

Then by Cauchy-Inequality, we 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

$$\begin{aligned}
\lim_{h\to 0} \frac{|f(x+h) - f(x) - A(h)|}{|h|} &= \lim_{h\to 0} \frac{\left|\sum_{j=1}^{m} \big(f_j(x+h) - f_j(x) - (df_j)_x(h)\big) q_j\right|}{|h|} \\
&\leq \lim_{h\to 0} \frac{\sum_{j=1}^{m} |f_j(x+h) - f_j(x) - (df_j)_x(h)|}{|h|} = 0 \text{ (done)}
\end{aligned}$$

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

$$|\partial_i f_j(y) - \partial_i f_j(x)| < \frac{\epsilon}{n} \text{ for all } y \in B_\delta(x)$$

We claim

$$\frac{|f_j(y)-f_j(x)-B(y-x)|}{|y-x|}\leq \epsilon \text{ for all } y\in B_\delta(x)$$

Express $y-x=\sum_{k=1}^n h_k e_k$. Define $v_0,\dots,v_n\in\mathbb{R}^n$ by

$$v_0\triangleq 0 \text{ and } v_k\triangleq \sum_{i=1}^k h_i e_i \text{ for all } k\in\{1,\dots,n\}$$

Now observe

$$\begin{aligned}
\frac{|f_j(y)-f_j(x)-B(y-x)|}{|y-x|} &= \frac{|f_j(x+v_n)-f_j(x)-B(\sum_{k=1}^n h_k e_k)|}{|y-x|}\\
&= \frac{\left|\left(\sum_{k=1}^n f_j(x+v_k)-f_j(x+v_{k-1})\right)-\sum_{k=1}^n \partial_k f_j(x)h_k\right|}{|y-x|}\\
&= \frac{|\sum_{k=1}^n f_j(x+v_k)-f_j(x+v_{k-1})-\partial_k f_j(x)h_k|}{|y-x|}\\
&= \frac{|\sum_{k=1}^n f_j(x+v_{k-1}+h_k e_k)-f_j(x+v_{k-1})-\partial_k f_j(x)h_k|}{|y-x|}\\
&= \frac{|\sum_{k=1}^n \partial_k f_j(x+v_{k-1}+t_k e_k)h_k-\partial_k f_j(x)h_k|}{|y-x|}\\
&\leq \frac{\sum_{k=1}^n \left|\left(\partial_k f_j(x+v_{k-1}+t_k e_k)-\partial_k f_j(x)\right)h_k\right|}{|y-x|}\\
&< \frac{\sum_{k=1}^n \frac{\epsilon}{n}|h_k|}{|y-x|}\leq \epsilon \text{ (done)}
\end{aligned}$$

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, 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

$$\big(\partial_i f_j(x+h)\big)^2 - \big(\partial_i f_j(x)\big)^2 < \frac{\epsilon^2}{M^2 n^2} \text{ for all } h \in B_\delta(0)$$

We claim

$$\text{such } \delta \text{ suffices}$$

Let $|y - x| < \delta$. We see

$$\|df_y - df_x\|_{\text{op}} \leq M \|df_y - df_x\|_F < M \sqrt{\sum_{i=1}^n \sum_{j=1}^n \frac{\epsilon^2}{M^2 n^2}} = \epsilon \text{ (done)}$$

$\blacksquare$

4.6 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 4.6.1. (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

$$\frac{\Delta(h,k)}{hk} = \partial_{21} f(x,y) \text{ for some } (x,y) : |x-a| < h \text{ and } \ |y-b| < k \tag{4.10}$$

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

$$\begin{aligned}
\Delta(h,k) &= u(a+h) - u(a) \\
&= hu'(x) \text{ for some } x \in (a, a+h) \text{ by MVT (Corollary 4.4.3)} \\
&= h\big(\partial_1 f(x, b+k) - \partial_1 f(x,b)\big)
\end{aligned}$$

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

$$\text{to find some } \delta \text{ such that for all } h \in (-\delta, \delta) \setminus 0, \left| \lim_{k\to 0} \frac{\Delta(h,k)}{hk} - \partial_{21} f(x,y) \right| \leq \epsilon$$

Because of Equation 4.10 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

$$\left| \lim_{k\to 0} \frac{\Delta(h,k)}{hk} - \partial_{21} f(x,y) \right| \le \left| \lim_{k\to 0} \frac{\Delta(h,k)}{hk} - \frac{\Delta(h,k')}{hk'} \right| + \left| \frac{\Delta(h,k')}{hk'} - \partial_{21} f(a,b) \right| \le \epsilon \text{ (done)}$$

■

Corollary 4.6.2. (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 4.6.1 to g and have

$$\partial_{ij} f(p) = \partial_{12} g(x_0, y_0) = \partial_{21} g(x_0, y_0) = \partial_{ji} f(p) \text{ (done)}$$

■

4.7 Product Rule for real-valued function and Chain Rule

We now prove the Chain Rule for function between normed space.

Theorem 4.7.1. (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

$$\begin{aligned}
\frac{\|dg_{f(x)}\big(k - df_x(h)\big)\|_{\mathcal{Z}}}{\|h\|_{\mathcal{X}}} &= \frac{\|dg_{f(x)}\big(f(x+h) - f(x) - df_x(h)\big)\|_{\mathcal{Z}}}{\|f(x+h) - f(x) - df_x(h)\|_{\mathcal{Y}}} \cdot \frac{\|f(x+h) - f(x) - df_x(h)\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}} \\
&\leq \|dg_{f(x)}\|_{\text{op}} \cdot \frac{\|f(x+h) - f(x) - df_x(h)\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}} \to 0 \text{ (done)}
\end{aligned}$$

■

4.8 Holomorphic Functions

4.9 Uniform Convergence and Differentiation

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 7 (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 8 (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 9 (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 5.8.1) 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 4.9.1. (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

$$f_n \text{ uniformly converge on } [a,b] \tag{4.11}$$

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

$$\begin{cases} |f_n(x_0) - f_m(x_0)| < \frac{\epsilon}{2} \\ \|f_n' - f_m'\|_\infty < \frac{\epsilon}{2(b-a)} \end{cases} \quad \text{for all } n,m > N \tag{4.12}$$

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 4.4.3) and Equation 4.12, 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 4.12, we have

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

We claim

$$f(x) \triangleq \lim_{n\to\infty} f_n(x) \text{ satisfy } f'(x)=\lim_{n\to\infty} f'_n(x) \text{ on } (a,b) \tag{4.13}$$

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

$$\phi_n \text{ uniformly converge on } [a,b] \setminus x \tag{4.14}$$

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

$$\|f'_n - f'_m\|_\infty \leq \epsilon \text{ for all } n,m>N \tag{4.15}$$

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 4.4.3) and Equation 4.15, we can deduce

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

We now show

$$\phi_n \to \phi \text{ pointwise on } [a, b] \setminus x \tag{4.16}$$

Because $f_n \to f$ on $[a, b]$ by definition (Equation 4.13), (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 4.14 and Equation 4.16, 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

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

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 4.9.2. (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 5.8.1).

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

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

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

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

■

4.10 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

$$R = \frac{1}{\limsup_{n\to\infty} \sqrt[n]{c_n}} \tag{4.17}$$

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 10 (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

$$\begin{cases} q < -1 \implies f_q \text{ converge on } S^1 \\ -1 \le q < 0 \implies f_q \text{ converge on } S^1 \setminus \{1\} \\ 0 \le q \implies f_q \text{ diverge on } S^1 \end{cases}$$

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.

$$\left|\sum_{k=1}^{n} z^k\right| = \left|\sum_{k=1}^{n} e^{ik\theta}\right| = \frac{\left|e^{i\theta} - e^{i(n+1)\theta}\right|}{|1-e^{i\theta}|} \le \frac{2}{|1-e^{i\theta}|}$$

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 11 (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 4.10.1. (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

$$f_m \to f^{(k)} \text{ pointwise on } D_R(a) \text{ and } f'_m(z) = \sum_{n=k+1}^{k+m} \frac{n!}{(n-k-1)!} c_n (z-a)^{n-k-1} \tag{4.18}$$

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 4.18, using Theorem 4.9.1 (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

$$\sum_{n=0}^{\infty} \frac{(n+k+1)!}{n!} c_{n+k+1} (z-a)^n \text{ uniformly converge on } \overline{D}_{R-\epsilon}(a) \quad \text{(done)}$$

■

Now by Theorem 4.10.1, 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 12 (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 4.10.1, we see that f is not analytic at 0.

Example 13 (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 4.10.1, 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 4.10.2 (Taylor's Theorem) give satisfying answers to these problems.

Theorem 4.10.2. (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 4.10.1, 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

$$\begin{aligned}
\frac{f^{(n)}(b)}{n!} &= \sum_{k=n}^{\infty}\frac{f^{(k)}(a)}{n!(k-n)!}b^{k-n} \\
&= \sum_{k=n}^{\infty}\frac{f^{(k)}(a)}{k!}\binom{k}{n}b^{k-n} \text{ (done)}
\end{aligned}$$

■

4.11 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 4.11.1. (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

$$\begin{aligned}\lim_{z\to 1^-}\Big(\sum_{n=0}^{\infty} c_n z^n - \sum_{n=0}^{\infty} c_n\Big) &= \lim_{z\to 1^-}\Big(\lim_{k\to\infty} s_k z^k + (1-z)\sum_{n=0}^{k-1} s_n z^n - s\Big)\\ &= \lim_{z\to 1^-}(1-z)\sum_{n=0}^{\infty}(s_n - s)z^n \quad (\because \forall z\in\mathbb{C}: |z|<1, \lim_{k\to\infty} s_k z^k = 0)\end{aligned}$$

We reduce the problem into

$$\text{finding } \delta \text{ such that } \left|(1-z)\sum_{n=0}^{\infty}(s_n - s)z^n\right| \le \epsilon \text{ for all } z\in\mathbb{S}_M(1)\cap D_\delta(1)$$

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

$$\begin{aligned}\left|(1-z)\sum_{n=0}^{\infty}(s_n-s)z^n\right| &= |1-z|\cdot\left|\sum_{n=0}^{N}(s_n-s)z^n + \sum_{n=N+1}^{\infty}(s_n-s)z^n\right|\\ &\le |1-z|\Big(\sum_{n=0}^{N}|s_n-s| + \frac{\epsilon}{2M}\sum_{n=N+1}^{\infty}|z|^n\Big)\\ &\le \frac{\epsilon}{2} + \frac{\epsilon}{2M}\cdot\frac{|1-z|}{1-|z|}\cdot|z|^{N+1} \le \epsilon \quad (\because |z|<1 \text{ and } \frac{|1-z|}{1-|z|} < M) \text{ (done)}\end{aligned}$$

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

$$\left| \sum_{n=0}^{\infty} c_n z^n - c_n R^n \right| = \left| g(\frac{z}{R}) - g(1) \right| < \epsilon \text{ (done)}$$

$\blacksquare$

Example 14 (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 15 ($1 - 1 + 1 - 1 + \cdots = \frac{1}{2}$)

$$1 - 1 + 1 - 1 + \cdots = \frac{1}{2} \text{ is WRONG!!!}$$

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 5

Measure Theory

5.1 σ-algebra

5.2 Carathéodory's extension theorem

Definition 5.2.1. (Definition of pre-measure) Given a R ring of subsets of X, we say $\mu : R \to [0, \infty]$ is a **pre-measure** of (X, R) if

(a) $\mu(\varnothing) = 0$

(b) $\mu\big(\bigcup E_n\big) = \sum \mu(E_n)$ for each countable disjoint $\{E_n\}$ such that $\bigcup E_n \in R$. (countably additive, or σ-additive)

It is straightforward to check

(a) $\mu(A_1 \cup \cdots \cup A_n) = \mu(A_1) + \cdots + \mu(A_n)$ if each A_i are disjoint.

(b) $A \subseteq B \implies \mu(A) \leq \mu(B)$

Theorem 5.2.2. (Basic property of pre-measure) Given an increasing sequence of measurable set A_n

$$\mu(A_n) \nearrow \mu(A)$$

where $A = \bigcup A_n$. If A_n is decreasing, then

$$\mu(A_n) \searrow \mu(A)$$

where $A = \bigcap A_n$.

Proof. If A_n is increasing, fix $B_n \triangleq A_n \setminus A_{n-1}$ for each $n \geq 2$ and $B_1 = A_1$. The proof then follows from checking

$$\mu(A_n) = \sum_{k=1}^{n} \mu(B_k) \text{ and } \mu(A) = \sum_{k=1}^{\infty} \mu(B_k)$$

The proof for another statement is similar. ∎

Using the same trick $B_n \triangleq A_n \setminus (A_{n-1} \cup \cdots \cup A_1)$, we have

$$\mu(\bigcup_{n \in \mathbb{N}} A_n) \leq \sum_{n \in \mathbb{N}} \mu(A_n) \text{ for arbitrary } A_n$$

Definition 5.2.3. (Definition of outer measure) Given a set X, by an **outer measure**, we mean a function $\mu^* : 2^X \to [0, \infty]$ such that

(a) $\mu^*(\varnothing) = 0$

(b) Given arbitrary subset $A, B_1, B_2, \ldots$ such that $A \subseteq \bigcup_n B_n$, we have $\mu^*(A) \leq \sum_n \mu^*(B_n)$ (countably subadditive)

Equivalently, one can replace countably sub-additive with

$$A \subseteq B \implies \mu^*(A) \leq \mu^*(B) \text{ and } \mu^*\Big(\bigcup_n B_n\Big) \leq \sum_n \mu^*(B_n)$$

Theorem 5.2.4. (Pre-measure induces outer measure) Given a pre-measure μ on some ring R of subsets of X, if we define $\mu^* : 2^X \to [0,\infty]$ by

$$\mu^*(T) \triangleq \inf \Big\{ \sum_n \mu(S_n) : T \subseteq \bigcup_n S_n \text{ and } S_1, S_2, \dots \in R \Big\} \text{ where } \inf \varnothing = \infty$$

Then

$$\mu^* \text{ is an outer measure}$$

Proof. It is clear $\mu^*(\varnothing) = 0$ and $A \subseteq B \implies \mu^*(A) \leq \mu^*(B)$. It remains to prove for arbitrary B_n we have

$$\mu^*\Big(\bigcup_n B_n\Big) \leq \sum_n \mu^*(B_n)$$

If $\sum_n \mu^*(B_n) = \infty$, the proof is trivial. We from now suppose $\sum_n \mu^*(B_n) < \infty$.

Fix ϵ. We prove

$$\mu^*\Big(\bigcup_n B_n\Big) \leq \sum_n \mu^*(B_n) + \epsilon$$

Because $\mu^*(B_n) < \infty$ for each $n \in \mathbb{N}$, we know for each $n \in \mathbb{N}$ there exists an countable cover of B_n consisting elements $S_{n,k}$ of R such that

$$\sum_k \mu(S_{n,k}) \leq \mu^*(B_n) + \epsilon(2^{-n})$$

It is clear that $\{S_{n,k} : n, k \in \mathbb{N}\}$ is a countable cover of $\bigcup_n B_n$, we now see

$$\begin{aligned} \mu^*\Big(\bigcup_n B_n\Big) &\leq \sum_n \sum_k \mu(S_{n,k}) \\ &\leq \sum_n \mu^*(B_n) + \epsilon(2^{-n}) = \sum_n \mu^*(B_n) + \epsilon \text{ (done)} \end{aligned}$$

∎

Note that

$$\mu(A) = \mu^*(A) \text{ if } A \in R$$

Definition 5.2.5. (Definition of measure)

Theorem 5.2.6. (Outer measure induce measure) We say a set $A \in 2^X$ is measurable if

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \setminus E) \text{ for all } E \in 2^X$$

Proof. ■

Theorem 5.2.7. (Measure induce complete measure)

5.3 Lebesgue measure

In this section, we develop the definition of Lebesgue measure.

5.4 Riesz–Markov–Kakutani representation theorem

5.5 bounded variation

This section prove some key properties of functions of bounded variations. These properties are worthy of discuss, as they make the set $BV\big([a,b]\big)$ of function of bounded variation on $[a,b]$ a natural candidate for the class of Riemann-Stieltjes integrator. The key properties include

(a) Functions of bounded variation must be continuous almost everywhere. (Corollary 5.5.9)

(b) Functions of bounded variation can be expressed as difference of two increasing functions. (Theorem 5.5.5)

(c) Functions of bounded variation can only have jump discontinuity. (Corollary 5.5.6)

Definition 5.5.1. (Definition of variation and function of bounded variation) Given a compact interval $[a,b]$, by a **partition** P of $[a,b]$, we mean a finite set $\{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\}$ contain by $[a,b]$ and containing a,b. If f is a complex-valued function defined on $[a,b]$, we say the **total variation** $V_f(a,b)$ of f on $[a,b]$ is

$$\sup_P \sum_{k=1}^n |f(x_k) - f(x_{k-1})|$$

We say f is of **bounded variation** on $[a,b]$ if $V_f(a,b) < \infty$, and we denote the set of real-valued function from $[a,b]$ of bounded variation on $[a,b]$ by $BV\big([a,b]\big)$.

It is straightforward to check

(a) $f \in BV\big([a,b]\big)$ must be bounded on $[a,b]$

(b) real-valued f monotone on $[a,b]$ is of bounded variation on $[a,b]$

(c) $BV\big([a,b]\big)$ form a vector space over $\mathbb{R}$

In fact, $BV\big([a,b]\big)$ also form a commutative algebra, as below proved.

Theorem 5.5.2. (Bounded variation is closed under multiplication) Given two real-valued (or more generally complex-valued) f,g defined on $[a,b]$

$$V_{fg}(a,b) \le AV_f(a,b) + BV_f(a,b)$$

where

$$A = \sup_{[a,b]} |g| \text{ and } B = \sup_{[a,b]} |f|$$

Proof. For every partition P, we have

$$\begin{aligned}\sum_{k=1}^{n}|fg(x_k)-fg(x_{k-1})| &= \sum_{k=1}^{n}|(f(x_k)-f(x_{k-1}))g(x_k)+f(x_{k-1})g(x_k)-fg(x_{k-1})| \\ &\leq \sum_{k=1}^{n}|f(x_k)-f(x_{k-1})|\,|g(x_k)|+\sum_{k=1}^{n}|f(x_{k-1})|\,|g(x_k)-g(x_{k-1})| \\ &\leq AV_f(a,b)+BV_f(a,b)\end{aligned}$$

■

Note that the proof above only consider when g, f are both bounded on $[a, b]$. If not, the statement hold trivially. For the brevity of the proof of the next Theorem, if we are given a partition $P = \{a = x_0 < \cdots < x_n = b\}$ of $[a, b]$, we denote $\sum_{k=1}^{n}|f(x_k)-f(x_{k-1})|$ by $\sum(P)$.

Theorem 5.5.3. (Additive property of total variation) Given a real-valued function f defined $[a, b]$, and $c \in (a, b)$

$$V_f(a,b)=V_f(a,c)+V_f(c,b)$$

Proof. We first prove

$$V_f(a,b)\geq V_f(a,c)+V_f(c,b)$$

Note that it is possible $V_f(a,c)=\infty$. We only prove when $V_f(a,c)<\infty$, since the proof for the statement when $V_f(a,c)=\infty$ is similar. Fix ϵ. We reduce the problem into proving

$$V_f(a,b)\geq V_f(a,c)+V_f(c,b)-\epsilon$$

Let Y, Z respectively be the set of all partitions of $[a, c]$ and $[a, b]$. Fix $P_y \in Y$ and $P_z \in Z$ such that

$$\sum(P_y)>V_f(a,c)-\frac{\epsilon}{2}\text{ and }\sum(P_z)>V_f(c,b)-\frac{\epsilon}{2}$$

It is clear that $P_y \cup P_z$ is a partition of $[a, b]$. Observe

$$V_f(a,b)\geq\sum(P_y\cup P_z)=\sum(P_y)+\sum(P_z)>V_f(a,c)+V_f(c,b)-\epsilon\text{ (done)}$$

It remains to prove

$$V_f(a,b)\leq V_f(a,c)+V_f(c,b)$$

Fix a partition P of $[a,b]$. We are required to prove

$$\sum(P) < V_f(a,c) + V_f(c,b)$$

Let $P_c = P \cup \{c\}$. Observe

$$\sum(P) \le \sum(P_c) \le V_f(a,c) + V_f(c,b) \text{ (done)}$$

■

Corollary 5.5.4. (Additive property of total variation) Given a real-valued function f defined $[a,b]$, and $c \in (a,b)$

$$f \in BV\big([a,b]\big) \iff f \in BV\big([a,c]\big) \text{ and } f \in BV\big([c,b]\big)$$

Perhaps, the best property of bounded variation are the following.

Theorem 5.5.5. (Function of bounded variation can be expressed as difference of two increasing functions) Given real-valued f defined on $[a,b]$

$$f \in BV\big([a,b]\big) \iff \exists \text{ increasing } g,h : [a,b] \to \mathbb{R}, f = g - h$$

Proof. From right to left is trivial. From left to right, we claim

$$g(x) \triangleq V_f(a,x) \text{ and } h(x) \triangleq V_f(a,x) - f(x) \text{ suffices}$$

It is clear that $V_f(a,x)$ is increasing. Fix $a \le x \le y \le b$. We prove

$$h(x) \le h(y)$$

Use Theorem 5.5.3

$$\begin{aligned} h(y) - h(x) &\ge V_f(a,y) - f(y) - V_f(a,x) + f(x) \\ &= \big(V_f(a,y) - V_f(a,x)\big) - \big(f(y) - f(x)\big) \\ &= V_f(x,y) - \big(f(y) - f(x)\big) \ge 0 \text{ (done)} \end{aligned}$$

■

Corollary 5.5.6. (Function of bounded variation can only have jump discontinuity) Given $f : [a,b] \to \mathbb{R}$, if $f \in BV\big([a,b]\big)$, then f can only have jump discontinuity.

Corollary 5.5.7. (Function of bounded variation can be expressed as difference of two strictly increasing functions) Given real-valued f defined on $[a,b]$

$$f \in BV\big([a,b]\big) \iff \exists \text{ strictly increasing } g,h : [a,b] \to \mathbb{R}, f = g - h$$

Proof. From right to left is again trivial. If $g(x) \triangleq V_f(a,x)$ and $h(x) \triangleq V_f(a,x) - f(x)$ is not strictly increasing, define $g' \triangleq g + (x-a)$ and $h' \triangleq h + (x-a)$. Such g', h' suffice. ■

The reason we showed functions of bounded variation can be expressed as difference of two increasing functions is the following (Theorem 5.5.9). Theorem 5.5.9 make function of bounded variation continuous almost everywhere. This make functions of bounded variation a natural candidate of the class of Riemann-Stieltjes integrators, in the perspective of Lebesgue-Riemann Criterion. (Theorem **??**)

Theorem 5.5.8. (Function monotone on $[a,b]$ must be continuous almost everywhere on $[a,b]$) If f is monotone on $[a,b]$, then the set of discontinuities of f is countable.

Proof. WOLG, suppose f is increasing on $[a,b]$. Because f is increasing on $[a,b]$, we know that every discontinuities of f on $[a,b]$ is a jump discontinuity. Define

$$S_n \triangleq \{x \in [a,b] : f(x+) - f(x-) > \frac{1}{n}\}$$

Then the set of discontinuity of f is exactly $\bigcup_n S_n$. Note that each S_n must be finite, otherwise $f(b) = \infty$. This conclude the proof. ∎

Corollary 5.5.9. (Function of bounded variation on $[a,b]$ must be continuous almost everywhere on $[a,b]$) If f is of bounded variation on $[a,b]$, then the set of discontinuities of f on $[a,b]$ is countable.

If we only consider real-valued continuous function of bounded variation on $[a,b]$, the structure is even richer.

Theorem 5.5.10. (Continuous function of bounded variation) Given $f : [a,b] \to \mathbb{R}$ such that $f \in BV\big([a,b]\big)$. If we define $V : [a,b] \to \mathbb{R}$ by

$$V(x) \triangleq V_f(a,x)$$

Then for each $x \in [a,b]$

$$V \text{ is continuous at } x \iff f \text{ is continuous at } x$$

Proof. $(\longrightarrow)$

Assume f is discontinuous at x. WOLG, suppose that there exists $x_n \searrow x$ such that

$$\lim_{n\to\infty} f(x_n) \in \mathbb{R} \text{ and } c \triangleq \left|\lim_{n\to\infty} f(x_n) - f(x)\right| > 0$$

Now, observe

$$|V(x_n) - V(x)| = |V_f(x,x_n)| \geq |f(x_n) - f(x)|$$

This give us

$$\begin{aligned}\liminf_{n\to\infty}|V(x_n)-V(x)| &\geq \liminf_{n\to\infty}|f(x_n)-f(x)| \\ &= \lim_{n\to\infty}|f(x_n)-f(x)| = \left|\lim_{n\to\infty}f(x_n)-f(x)\right| = c>0 \text{ CaC}\end{aligned}$$

$(\longleftarrow)$

Fix ϵ. Let $P=\{x=x_0<x_1<\cdots<x_n=b\}$ be a partition of $[x,b]$ such that

$$V_f(x,b)-\frac{\epsilon}{2}<\sum(P)$$

Let δ satisfy

$$|f(y)-f(x)|<\frac{\epsilon}{4n}\text{ for all }y\in[x,x+\delta]$$

We claim

$$\text{such }\delta\text{ satisfy }|V(y)-V(x)|\leq\epsilon\text{ for all }y\in[x,x+\delta]$$

Let $y\triangleq x+\delta$. Because V is increasing on $[a,b]$, we only have to prove

$$|V(y)-V(x)|\leq\epsilon$$

Theorem 5.5.3 allow us to reduce the problem into proving

$$V_f(x,y)\leq\epsilon$$

Denote $P\cup\{y\}$ by P' and express $P'=\{x=x_0'<\cdots<x_r'\}$. Express $y=x_m'$. Note that $m<r\leq n+1$ and observe

$$\begin{aligned}V_f(x,b)-\frac{\epsilon}{2}<\sum(P)\leq\sum(P') \\ &=\sum_{k=1}^{m}|f(x_k')-f(x_{k-1}')|+\sum_{k=m+1}^{n}|f(x_k')-f(x_{k-1}')| \\ &\leq\frac{m\epsilon}{2n}+V_f(y,b)\leq\frac{\epsilon}{2}+V_f(y,b)\end{aligned}$$

Theorem 5.5.3 now give us

$$V_f(x,y)=V_f(x,b)-V_f(y,b)\leq\epsilon\text{ (done)}$$

Proof for $V(x-)=V(x)$ is similar, and when $x=a$ or b, some trivial modifications are needed. ■

Give very careful attention to the statement of Theorem 5.5.10. Note that we require to the domain of f to be $[a,b]$. If the domain of f contain a or b as interior point, the statement isn't always true.

Corollary 5.5.11. (Continuous function of bounded variation can be expressed as difference of two continuous strictly increasing functions) Given continuous real-valued f defined on $[a,b]$

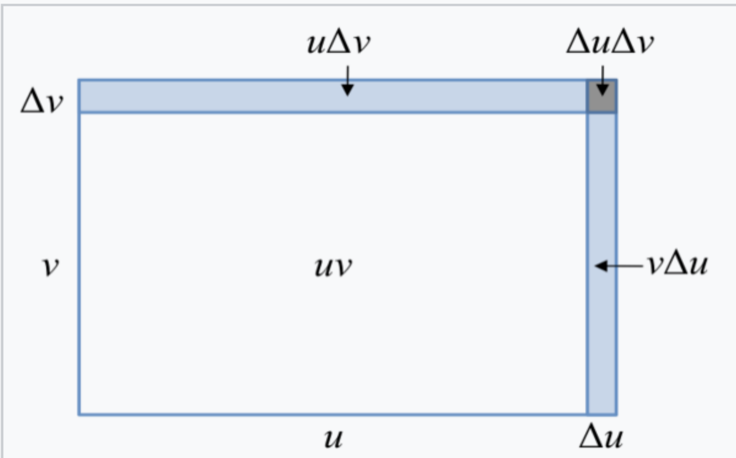
$$f \in BV\big([a,b]\big) \iff \exists \text{ continuous strictly increasing } g,h:[a,b]\to\mathbb{R}, f=g-h$$

Proof. From right to left is again trivial. If $g(x)\triangleq V_f(a,x)$ and $h(x)\triangleq V_f(a,x)-f(x)$ is not strictly increasing, define $g'\triangleq g+(x-a)$ and $h'\triangleq h+(x-a)$. Such g',h' suffice. ∎

5.6 Construction of Riemann-Stieltjes integral

5.7 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 5.7.1. (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 5.7.2. (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 5.7.3. (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

$$\begin{aligned} (f \circ g)'(a) &= \lim_{x \to a} \frac{f(g(x)) - f(g(a))}{x - a} && (5.1) \\ &= \lim_{x \to a} \frac{R_{f(g(a))}(g(x)) + f'(g(a))(g(x) - g(a))}{x - a} && (5.2) \end{aligned}$$

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

$$\lim_{x \to a} \frac{R_{f(g(a))}(g(x))}{x - a} = \lim_{x \to a} \frac{R_{f(g(a))}(g(x))}{g(x) - g(a)} \cdot \frac{g(x) - g(a)}{x - a} = 0 \cdot g'(x) = 0 \qquad (5.3)$$

Notice that the above deduction (Equation 5.3) 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 5.1, by Equation 5.3, 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. ■

Theorem 5.7.4. (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

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in (a, b) \\ \lim_{t \to a} f(t) & \text{if } x = a \\ \lim_{t \to b} f(t) & \text{if } x = b \end{cases} \tag{5.4}$$

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 5.4, by EVT, we know we there exists $\alpha, \beta \in [a, b]$ such that

$$\tilde{f}(\alpha) = \inf_{x \in [a,b]} \tilde{f}(x) \text{ and } \tilde{f}(\beta) = \sup_{x \in [a,b]} \tilde{f}(x) \tag{5.5}$$

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

$$\exists \xi \in [\alpha,\beta], \tilde{f}(\xi) = \frac{\int_a^b \tilde{f}(x)dx}{b-a} \tag{5.6}$$

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 5.5), 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

$$\inf_{x\in[\beta-\delta,\beta]} \tilde{f}(x) \geq \frac{\tilde{f}(\alpha) + 2\tilde{f}(\beta)}{3} \tag{5.7}$$

We then from Equation 5.6 see that

$$\int_a^b \tilde{f}(x)dx = \tilde{f}(\xi)(b-a) = \tilde{f}(\alpha)(b-a) \tag{5.8}$$

Also, we see from definition of α (Equation 5.5) and Equation 5.7 that

$$\begin{aligned}
\int_a^b \tilde{f}(x)dx &= \int_a^{\beta-\delta} \tilde{f}(x)dx + \int_{\beta-\delta}^{\beta} \tilde{f}(x)dx + \int_{\beta}^{b} \tilde{f}(x)dx && (5.9)\\
&\geq (b-\delta-a)\tilde{f}(\alpha) + \delta \cdot \big(\frac{\tilde{f}(\alpha)+2\tilde{f}(\beta)}{3}\big) && (5.10)\\
&> (b-\delta-a)\tilde{f}(\alpha) + \delta \cdot \big(\frac{\tilde{f}(\alpha)+\tilde{f}(\beta)}{2}\big) && (5.11)\\
&= \tilde{f}(\alpha)\big(b-a-\frac{\delta}{2}\big) + \tilde{f}(\beta)\cdot\big(\frac{\delta}{2}\big) && (5.12)
\end{aligned}$$

Now, from Equation 5.8 and Equation 5.12, 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 5.7.5. (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 5.12.1, 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

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

■

5.8 Uniform Convergence and Riemann Integration

Theorem 5.8.1. (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

$$\text{there exists } N \text{ such that } \forall n > N, \epsilon_n < \frac{\epsilon}{\alpha(b) - \alpha(a)} \tag{5.13}$$

We claim

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

Fix $n > N$. From Equation 5.13, we see

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

■

5.9 Riemann-Stieltjes on Computation

Theorem 5.9.1. (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

$$\text{there exists a partition } P \text{ of } [a, b] \text{ such that } U(P, f, \alpha) - L(P, f, \alpha) < \epsilon \tag{5.14}$$

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

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

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

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

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

$$\begin{aligned}
U(Q,g,\beta) &= \sum_{k=1}^{n} \Big[\sup_{t\in[\phi^{-1}(x_{k-1}),\phi^{-1}(x_k)]} g(t) \Big] \Big[\beta(\phi^{-1}(x_k)) - \beta(\phi^{-1}(x_{k-1})) \Big] \\
&= \sum_{k=1}^{n} \Big[\sup_{t\in[\phi^{-1}(x_{k-1}),\phi^{-1}(x_k)]} f\circ\phi(t) \Big] \Big[\alpha\circ\phi\big(\phi^{-1}(x_k)\big) - \alpha\circ\phi\big(\phi^{-1}(x_{k-1})\big) \Big] \\
&= \sum_{k=1}^{n} \Big[\sup_{t\in[x_{k-1},x_k]} f(t) \Big] \big(\alpha(x_k) - \alpha(x_{k-1}) \big) = U(P,f,\alpha) \qquad (5.15)
\end{aligned}$$

Similarly, we can deduce $L(Q,g,\beta) = L(P,f,\alpha)$. Now, from Equation 5.15 and by definition of P (Equation 5.14), we see

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

■

Theorem 5.9.2. (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

$$U(P,\alpha') - L(P,\alpha') < \frac{\epsilon}{4M} \tag{5.16}$$

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 5.16) 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 4.4.3), we know for all $k \in \{1,\dots,n\}$ there exists $t_k \in [x_{k-1}, x_k]$ such that

$$\Delta\alpha_k = \alpha'(t_k)\Delta x_k \tag{5.17}$$

Then, because $U(P,\alpha') - L(P,\alpha)' < \frac{\epsilon}{3M}$ (Equation 5.16), we now see

$$\sum_{k=1}^n |\alpha'(s_k) - \alpha'(t_k)| \Delta x_k < \frac{\epsilon}{3M} \text{ if } s_k \in [x_{k-1}, x_k] \text{ for all } k \in \{1,\dots,n\} \tag{5.18}$$

Then from Equation 5.17, definition of M and Equation 5.18, 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

$$\sum_{k=1}^{n} f(s_k)\Delta\alpha_k < U(P, f\alpha') + \frac{\epsilon}{4} \tag{5.19}$$

Because Equation 5.19 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 5.9.3. (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 5.9.2), we know

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

Let $\alpha(x) = x$. Let $\beta = \alpha \circ \phi$. Define $g = f \circ \phi$. By Change of Variable (Theorem 5.9.1), we now have

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

■

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

Theorem 5.10.1. (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

$$f(x) = (1+x)^r - (1+rx) \tag{5.20}$$

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 5.20), we see that

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

Notice that by definition of f (Equation 5.20), $f(x)$ is clearly differentiable on $(-1, \infty)$.

Then, by MVT (Theorem 4.4.3), 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 5.10.2. (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

$$\tilde{f}(x)=f(x)-f(0)-x\big[f(1)-f(0)\big] \tag{5.21}$$

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 5.21), 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 5.21), we have

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

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

$$\tilde{f}(x)=\begin{cases}\tilde{f}(x) & \text{if } x\in[0,1]\\ 0 & \text{if } x\notin[0,1]\end{cases} \tag{5.22}$$

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

$$Q_n=c_n(1-x^2)^n \text{ where } c_n \text{ is chosen to satisfy } \int_{-1}^1 Q_n(x)dx=1 \tag{5.23}$$

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 5.22), we see that

$$P_n(x) = \int_{-x}^{1-x} \tilde{f}(x+t)Q_n(t)dt \text{ for all } x \in [0,1] \tag{5.24}$$

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 5.23), 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

$$\forall x, y \in \mathbb{R}, |x-y| < \delta \implies \left| \tilde{f}(x) - \tilde{f}(y) \right| < \frac{\epsilon}{2} \tag{5.25}$$

By definition of $\tilde{f}$ (Equation 5.22), 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

$$4M\sqrt{n}(1-\delta^2)^n < \frac{\epsilon}{2} \tag{5.26}$$

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 5.10.1). 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 5.23), 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 5.23)

(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 5.26)

(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 5.26) 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

$$g(x) \triangleq f(a + (b-a)x) \tag{5.27}$$

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 5.10.3. (The space $\mathbb{Q}[x]\big|_{[a,b]}$ 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

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

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

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

■

5.11 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: $\Big((a+b)v = av + bv \text{ and } a(v+w) = av + aw\Big)$

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) $(\text{Functions from } E \text{ to } \mathbb{F}, \text{function multiplication})$ 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 $B = (\text{continuous functions from } \mathbb{C} \text{ to } \mathbb{C}, \text{composition})$ 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 $f(x_2) \neq f(x_1)$.

5.12 FTC

Theorem 5.12.1. (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

$$\exists \delta, |x - x_0| < \delta \implies |f(x) - f(x_0)| < \epsilon \tag{5.28}$$

We claim

$$\text{such } \delta \text{ in Equation 5.28 works}$$

WOLG, let $x > x_0$. Deduce

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

■

Theorem 5.12.2. (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

$$U(P,f) - L(P,f) < \epsilon \tag{5.29}$$

Because $f = F'$ on (a,b), for each $k \in \{1,\dots,n\}$, by MVT (Theorem 4.4.3), 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 5.29, we can deduce

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

■

Theorem 5.12.3. (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

$$\int_a^b F(x)g(x)dx = FG\Big|_a^b - \int_a^b f(x)G(x)dx \tag{5.30}$$

Proof. To prove Equation 5.30, 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 5.12.2).
(done) ■

Example 16 (Discontinuous Derivative)

$$f(x) = \begin{cases} \frac{x^2}{\sin x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Chapter 6

Riemann Calculus

Chapter 7

Complex Analysis

Chapter 8

Lebesgue Calculus

8.1 Basic Property of Measurable Functions

In this section, we give two equivalent general definition (Theorem 8.1.1) of measurable function and show

(a) Measurable $f : X \to \mathbb{R}$ or $\mathbb{C}$ is closed under addition and multiplication. (Theorem 8.1.3)

(b) The (superior) limit of a sequence of measurable ($\mathbb{R} \cup \{\pm\infty\}$ or $\mathbb{C} \cup \{\infty\}$)-valued functions on general (X, Σ_X) is measurable. (Theorem 8.1.4)

(c) Positive and negative part of a measurable $f : X \to [-\infty, \infty]$ are measurable. (Corollary 8.1.6)

If we are given a function $f : (X, \Sigma_X) \to (Y, \Sigma_Y)$ such that for all $E \in \Sigma_Y$, we have $f^{-1}(E) \in \Sigma_X$, then we say f is a **measurable function**. Immediately, one can check that the composition of two measurable functions must be measurable. We now introduce an equivalent definition of measurable function, which makes checking if a function is measurable much easier.

Theorem 8.1.1. (Equivalent Definition of measurable function) If a function $f : (X, \Sigma_X) \to (Y, \sigma(T))$ satisfy

$$f^{-1}(E) \in \Sigma_X \text{ for all } E \in T$$

then

$$f \text{ is measurable}$$

Proof. Define

$$\mathcal{A} \triangleq \{E \subseteq Y : f^{-1}(E) \in \Sigma_X\}$$

Check that $\mathcal{A}$ is a σ-algebra on Y. By premise, $T \subseteq \mathcal{A}$. It then follows from definition that $\sigma(T) \subseteq \mathcal{A}$. This conclude that f is $(\Sigma_X, \sigma(T))$-measurable. ∎

There are two important consequences of Theorem 8.1.1

(a) If X, Y are both Borel, then a continuous function $f : X \to Y$ must also be a measurable function.

(b) If there exists T such that $\Sigma_Y = \sigma(T)$, then we only have to check that $f^{-1}(E) \in \Sigma_X$ for all $E \in T$ to show f is measurable. In particular, if $Y = \mathbb{R}^n$, then T can be just the set of all open boxes.

These two consequences give the following Lemma, which later prove Theorem 8.1.3, Theorem that make checking if $\mathbb{C}$ or $\mathbb{R}$-valued function is measurable easier.

Lemma 8.1.2. (Computational Lemma) Given two measurable function $u, v : (X, \Sigma_X) \to \mathbb{R}$ and a continuous function $\Phi : \mathbb{R}^2 \to (Y, \tau_Y)$, define $h : (X, \Sigma_X) \to (Y, \mathcal{B}_Y)$ by

$$h(x) \triangleq \Phi(u(x), v(x))$$

We can deduce

$$h : (X, \Sigma_X) \to (Y, \mathcal{B}_Y) \text{ is measurable}$$

Proof. Define $f : X \to \mathbb{R}^2$ by $f(x) \triangleq (u, v)(x)$. Because $h = \Phi \circ f$, we can reduce the problem into proving

$$f \text{ is measurable}$$

Fix an open-rectangle $I_1 \times I_2 \subseteq \mathbb{R}^2$. Because $\mathcal{B}_{\mathbb{R}^2}$ can be generated by the set of all open rectangles, we can reduce the problem into proving

$$f^{-1}(I_1 \times I_2) \in \Sigma_X$$

Because u, v are measurable, we know

$$f^{-1}(I_1 \times I_2) = u^{-1}(I_1) \cap v^{-1}(I_2) \in \Sigma_X \text{ (done)}$$

∎

Theorem 8.1.3. (Basic tools to show a real/complex-valued function is measurable) Given a measurable set X, and two real-valued function $u, v : X \to \mathbb{R}$

(a) $u + iv : X \to \mathbb{C}$ is measurable if and only if u, v are measurable.

(b) If $f, g : X \to \mathbb{R}$ or $\mathbb{C}$ are measurable, so are $f + g, fg$ and $|f|$.

(c) If $f : X \to \mathbb{R}$ or $\mathbb{C}$ is measurable and $g : X \to \mathbb{R}$ or $\mathbb{C}$ isn't, then $f+g$ is not measurable.

(d) If $f : X \to \mathbb{C}$ is measurable, then there exists $\alpha : X \to \mathbb{C}$ such that $|\alpha| = 1$ and $f = \alpha\,|f|$

Proof. (a) follows from Lemma 8.1.2 and noting $u = \text{Re}(u+iv), v = \text{Im}(u+iv)$, since Re,Im : $\mathbb{C} \to \mathbb{R}$ are continuous.

(b) follows from the fact $\mathbb{R}, \mathbb{C}$ are topological field and Lemma 8.1.2.

It is clear that the set of all function from X to $\mathbb{R}$ or $\mathbb{C}$ form a group under addition. (b) shows that the set of measurable functions from a subgroup, thus giving (c).

It remains to prove (d). Define

$$E \triangleq \{x \in X : f(x) = 0\} \text{ and } \phi(z) \triangleq \frac{z}{|z|}$$

We claim

$$\alpha \triangleq \phi \circ (f + \mathbf{1}_E) \text{ suffices}$$

Because f is measurable, we know E is measurable. This implies that $\mathbf{1}_E : X \to \mathbb{C}$ is measurable. It follows that $f + \mathbf{1}_E$ is $(X, \mathcal{B}_{\mathbb{C}})$-measurable. Note that $f + \mathbf{1}_E$ is never 0 on X.

Now by Theorem **??**, we see $f + \mathbf{1}_E$ is $(X, \mathcal{B}_{\mathbb{C}^*})$-measurable. It then follows from the fact $\phi : \mathbb{C}^* \to \mathbb{C}$ is continuous that α is $(X, \mathcal{B}_{\mathbb{C}})$ measurable.

Observe that α maps E into $\{1\}$, and when $x \notin E$, we have $\alpha(x) = \frac{f(x)}{|f(x)|}$. (done) ∎

Note that Theorem 8.1.3 does not consider function whose range include ∞. This will be later addressed using approximation of simple functions.

Theorem 8.1.4. (Superior limit of measurable $f_n : X \to [-\infty, \infty]$ is measurable) Given a sequence $f_n : X \to [-\infty, \infty]$ of measurable functions

$$g \triangleq \sup f_n \text{ and } f \triangleq \limsup_{n\to\infty} f_n \text{ are both measurable}$$

Proof. It is straightforward to check

$$g^{-1}(\alpha, \infty] = \bigcup_n f_n^{-1}(\alpha, \infty]$$

It is straightforward to check

$$\mathcal{B}_{[-\infty,\infty]} = \sigma\Big(\{(\alpha,\infty] \subseteq [-\infty,\infty] : \alpha \in \mathbb{R}\}\Big)$$

These two facts and Theorem 8.1.1 shows g is measurable. The same arguments shows that $\inf g_k$ is measurable if g_k are all measurable.

It is straight forward to check

$$f = \inf_{n\geq 1} \sup_{k\geq n} f_k$$

It then follows f is measurable. ■

Corollary 8.1.5. (Pointwise limit of measurable $f_n : X \to [-\infty,\infty]$ is measurable) If the sequence $f_n : X \to [-\infty,\infty]$ pointwise converge to $f : X \to [-\infty,\infty]$, then f is measurable.

Corollary 8.1.6. (Positive and Negative parts of a measurable function are measurable) If we are given measurable $f : X \to [-\infty,\infty]$, and we define the **positive and negative part $f^+, f^- : X \to [0,\infty]$ of f** by

$$f^+ \triangleq \max\{f,0\} \text{ and } f^- \triangleq -\min\{f,0\}$$

Then

$$f^+ \text{ and } f^- \text{ are measurable}$$

Proof. Define

$$h_n \triangleq \begin{cases} f & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{cases}$$

then we have

$$\limsup_{n\to\infty} h_n = f^+ \text{ and } \liminf_{n\to\infty} h_n = -f^-$$

Now, because 0 is a measurable function, by Theorem 8.1.4, f^+ is measurable. Moreover, because -1 is a measurable function, by Theorem 8.1.3, f^- is measurable. ■

8.2 Egorov's and Lusin's Theorem

8.3 Abstract integration

We say a function is a **simple function** if its range is of finite cardinality. Simple function is the cornerstone of the development of Lebesgue Integral Theory, as we shall see.

Suppose s is a simple function defined on some measurable $E \subseteq (X, \Sigma_X)$ with range $\{c_1, \cdots, c_n\} \subseteq [0, \infty]$. Define

$$E_j \triangleq \{x \in X : s(x) = c_j\}$$

It is clear that $\{E_j\}$ is a finite disjoint decomposition of E, and s is measurable if and only if all E_j are measurable.

We can write

$$s = \sum_{j=1}^{n} c_i \mathbf{1}_{E_j}$$

Then if s is measurable, we can define

$$\int_E s d\mu \triangleq \sum_{j=1}^{n} c_j \mu(E_j)$$

and if s is defined on some larger domain while $s|_E$ is measurable, we can define

$$\int_E s d\mu \triangleq \sum_{j=1}^{n} c_j \mu(E_j \cap E)$$

Note that if s is defined on some measurable F containing E and s is measurable on F, then s is surely measurable on E. It is straightforward to verify our definition so far is consistent.

We now expand our definitions to the class of all measurable functions range in $[0, \infty]$.

Given a function f range in $[0, \infty]$ and measurable on E, we define

$$\int_E f d\mu \triangleq \sup_{s \leq f \text{ on } E} \int_E s d\mu$$

If f is defined on some large measurable domain F, we see that

$$\int_E f d\mu = \int_F f \mathbf{1}_E d\mu$$

It is at this point one should verify our definition is so far consistent. That is, for each simple s, we have

$$\int_E s d\mu = \sup_{s' \leq s \text{ on } E} \int_E s' d\mu$$

The trick is to decompose each E_j' into $\bigcup E_j' \cap E_i$.

We now expand our definitions to the class of all measurable functions range in $[-\infty, \infty]$, but before such, we have to first introduce the idea of **Lebesgue integrable**. For either s or f, range in either $[0, \infty)$ or $[0, \infty]$, it is always possible that $\int_E s$ or $f d\mu = \infty$.

If we have $\int_E f d\mu = \infty$, we say f is **not Lebesgue integrable**, and if $\int_E f d\mu < \infty$, we say f is **Lebesgue integrable on** E. Because $\mathbb{R}$ is complete in order, we know f is either Lebesgue integrable or not Lebesgue integrable on E.

Given a function range in $[-\infty, \infty]$ and measurable on E, if both f^+, f^- are Lebesgue integrable on E, we say f is Lebesgue integrable on E, and write

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$$

If either f^+ or f^- are not Lebesgue integrable on E, we say f is not Lebesgue integrable on E.

Notice that

$$f \text{ is Lebesgue integrable on } E \iff \int_E |f| \, d\mu < \infty$$

It is clear that our definition is again so far consistent.

We now

8.4 Basic property of abstract integration

This section prove some basic properties of Lebesgue integral over general measure space (X, Σ_X, μ). From now when we use the notation X, it shall be understood X is equipped with an σ-algebra Σ_X and a measure μ. We will prove

(a) Lebesgue Monotone Convergence Theorem (Theorem 8.4.1)

(b) Fatou's Lemma (Theorem 8.4.3)

(c) Reverse Fatou's Lemma (Theorem 8.4.4)

(d) Dominated Convergence Theorem (Theorem 8.4.5)

Theorem 8.4.1. (Lebesgue Monotone Convergence Theorem) Given a sequence of measurable $f_n : X \to [0, \infty]$ such that $\{f_n(x)\}_{n\in\mathbb{N}}$ is an increasing sequence for each $x \in X$, then

$$\int_X f d\mu = \lim_{n\to\infty} \int_X f_n d\mu$$

Proof. f is measurable by Corollary 8.1.5. Because $f_n \nearrow f$ on X, we know

$$\lim_{n\to\infty} \int_X f_n d\mu = \sup_n \int_X f_n d\mu \leq \int_X f d\mu$$

It remains to prove

$$\int_X f d\mu \leq \lim_{n\to\infty} \int_X f_n d\mu$$

Fix simple $0 \leq s \leq f$ on X. We reduce the problem into proving

$$\int_X s d\mu \leq \lim_{n\to\infty} \int_X f_n d\mu$$

Fix $c \in (0, 1)$. We reduce the problem into proving

$$c \int_X s d\mu \leq \lim_{n\to\infty} \int_X f_n d\mu$$

Define

$$E_n \triangleq \{x \in X : f_n(x) \geq cs(x)\}$$

E_n are measurable because $f_n - cs$ are measurable. Now because f_n are non-negative on X, we have

$$\int_X f_n d\mu \geq \int_{E_n} f_n d\mu \geq c \int_{E_n} s d\mu$$

Taking limit, we see

$$\lim_{n\to\infty}\int_X f_n d\mu \geq \lim_{n\to\infty} c\int_{E_n} s d\mu$$

It is straightforward to check E_n is increasing and $\bigcup E_n = X$. Then if we decompose $s=\sum_j c_j \mathbf{1}_{F_j}$, by Theorem 5.2.2, we can take limit

$$\lim_{n\to\infty}\mu(F_j\cap E_n)=\mu(F_j)$$

It then follows that

$$\lim_{n\to\infty}\int_X f_n d\mu \geq \lim_{n\to\infty} c\int_{E_n} s d\mu = c\int_X s d\mu \text{ (done)}$$

■

It is worth pointing out in our proof for Lebesgue Monotone Convergence Theorem, instead of proving $\int_X s d\mu \leq \lim_{n\to\infty}\int_X f_n d\mu$, we proved $c\int_X s d\mu \leq \lim_{n\to\infty}\int_X f_n d\mu$. Multiplying $\int_X s d\mu$ with $c\in(0,1)$ is not just a random limit technique. Our action play a much more profound role. Consider the Example.

Example 17 (Why we take $c\int_X s d\mu$?)

$$X=[0,1] \text{ and } f_n = 1-\frac{1}{n}$$

We can take $s=f$, and see $E_n=\varnothing$ for all n, which renders our proceeding proof invalid.

Corollary 8.4.2. (Monotone Convergence Theorem for general functions) Given a sequence of measurable $f_n : X\to[0,\infty]$ such that

(a) $\{f_n(x)\}_{n\in\mathbb{N}}$ is an increasing sequence on N^c

(b) $f:X\to[0,\infty]$ is the limit of f_n on N^c

(c) $\mu(N)=0$

(d) μ is complete

We have

$$\int_X f d\mu = \lim_{n\to\infty}\int_X f_n d\mu$$

Proof. Let

$$g(x) \triangleq \begin{cases} f(x) & \text{if } x \in N^c \\ 0 & \text{if } x \in N \end{cases}$$

Note that

$$\int_X f d\mu = \int_{N^c} f d\mu = \lim_{n\to\infty} \int_{N^c} f d\mu = \lim_{n\to\infty} \int_X f d\mu$$

■

Theorem 8.4.3. (Fatou's Lemma) Given measurable $f_n : X \to [0, \infty]$

$$\int_X \liminf_{n\to\infty} f_n d\mu \le \liminf_{n\to\infty} \int_X f_n d\mu$$

Proof. Since $\inf_{k\ge n} f_k \le f_n$ for each n, x, we see

$$\int_X \inf_{k\ge n} f_k d\mu \le \int_X f_n d\mu \text{ for all } n$$

Because $\inf_{k\ge n} f_k \nearrow \liminf_{m\to\infty} f_m$ as $n \to \infty$, by (Theorem 8.4.1: Lebesgue Monotone Convergence Theorem), we can take limit

$$\int_X \liminf_{n\to\infty} f_n d\mu = \lim_{n\to\infty} \int_X \inf_{k\ge n} f_k d\mu \le \liminf_{n\to\infty} \int_X f_n d\mu$$

■

Example 18 (Fatou's Lemma strict inequality)

$$f_{2k}(x) = \begin{cases} 0 & \text{if } x \in [0, \frac{1}{2}] \\ 1 & \text{if } x \in (\frac{1}{2}, 1] \end{cases} \text{ and } f_{2k+1}(x) = \begin{cases} 1 & \text{if } x \in [0, \frac{1}{2}) \\ 0 & \text{if } x \in [\frac{1}{2}, 1] \end{cases}$$

From now, we use $L^1(\mu)$ to denote the set of all function defined and μ-Lebesgue-integrable on X, and we say a sequence of function $f_n : X \to [0, \infty]$ is **dominated** by g, if g is a $[0, \infty]$-valued function defined on X such that

$$\sup_n |f_n(x)| \le g(x) \text{ for all } x \in X$$

Theorem 8.4.4. (Reverse Fatou's Lemma) Given measurable $f_n : X \to [0, \infty]$ dominated by some $g \in L^1(\mu)$

$$\limsup_{n\to\infty} \int_X f_n d\mu \le \int_X \limsup_{n\to\infty} f_n d\mu$$

Proof. By (Theorem 8.4.3: Fatou Lemma)

$$\int_X \liminf_{n\to\infty}(g-f_n)d\mu \leq \liminf_{n\to\infty}\int_X (g-f_n)d\mu$$

Multiplying both side with -1, we have

$$\limsup_{n\to\infty}\int_X (f_n-g)d\mu \leq \int_X \limsup_{n\to\infty}(f_n-g)d\mu$$

Then adding both side the constant $\int_X gd\mu$, we reach to the conclusion. ■

Note that in our proof above, when we "pull" the negative sign out from $\liminf_{n\to\infty}$, it changed to $\limsup_{n\to\infty}$. This is a standard technique, which can be justified using the sub-sequence definition of limit superior.

Theorem 8.4.5. (Dominate Convergence Theorem) Given a sequence $f_n : X \to \mathbb{C}\cup\{\infty\}$ of measurable function such that

$$f \triangleq \lim_{n\to\infty} f_n \text{ exists on } X$$

If there exists $g \in L^1(\mu)$ dominating f_n, then

$$\lim_{n\to\infty}\int_X |f_n-f|\,d\mu = 0 \text{ and } \int_X fd\mu = \lim_{n\to\infty}\int_X f_n d\mu \text{ exists in } \mathbb{C}\cup\{\infty\}$$

Proof. Because f is measurable by Corollary 8.1.5 and $|f|\leq g$ on X, $f\in L^1(\mu)$.

We first prove

$$\lim_{n\to\infty}\int_X |f_n-f|\,d\mu = 0$$

We relax the problem into proving

$$\limsup_{n\to\infty}\int_X |f_n-f|\,d\mu = 0$$

Note that $|f_n-f|\leq 2g$. We can now apply (Theorem 8.4.3: Fatou lemma) to $2g-|f_n-f|$ and see

$$\begin{aligned}
\int_X 2gd\mu &= \int_X \lim_{n\to\infty}(2g-|f_n-f|)d\mu \\
&\leq \liminf_{n\to\infty}\int_X (2g-|f_n-f|)d\mu \\
&= \int_X 2gd\mu + \liminf_{n\to\infty}\Big(-\int_X |f_n-f|\,d\mu\Big) \\
&= \int_X 2gd\mu - \limsup_{n\to\infty}\int_X |f_n-f|\,d\mu
\end{aligned}$$

Then because $g \in L^1(\mu)$, we can subtract it and obtain

$$\limsup_{n\to\infty} \int_X |f_n - f| \, d\mu = 0 \text{ (done)}$$

It then follows that

$$\limsup_{n\to\infty} \left| \int_X (f_n - f) d\mu \right| \leq \limsup_{n\to\infty} \int_X |f_n - f| \, d\mu = 0$$

which implies

$$\lim_{n\to\infty} \int_X (f_n - f) d\mu = 0$$

and because $f \in L^1(\mu)$, we have

$$\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu$$

■

Example 19 (Counterexample for Dominate Convergence Theorem)

$$f_n(x) = \begin{cases} \frac{1}{n} & \text{if } |x| < n \\ 0 & \text{if } |x| \geq n \end{cases}$$

8.5 Equivalent Definitions of Lebesgue Measurable Functions and Integral

Chapter 9

Harmonic Analysis

Chapter 10

Calculus in Euclidean Space

10.1 Inverse Function Theorem

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 10.1.2).

Theorem 10.1.1. (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

$$\begin{aligned}
|f(b)-f(a)| &= |(f\circ\gamma)(1)-(f\circ\gamma)(0)| \\
&= \left|\int_0^1 (f\circ\gamma)'(t)dt\right| \\
&\le \int_0^1 |(f\circ\gamma)'(t)|\,dt \\
&= \int_0^1 \|d(f\circ\gamma)_t\|_{\text{op}}dt \\
&\le \int_0^1 \|df_{\gamma(t)}\|_{\text{op}}\cdot\|d\gamma_t\|_{\text{op}}dt \\
&\le \int_0^1 M\cdot|b-a|\,dt = M\,|b-a|
\end{aligned}$$

■

Corollary 10.1.2. (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 $df_x=0$ for all $x\in E$, then

$$f\text{ stay constant on }E$$

In this section, we will give a local statement and the proof for Inverse Function Theorem in $\mathbb{R}^n$ (Theorem 10.1.4). Let $L(\mathbb{R}^n)$ be the set of linear transformation that maps $\mathbb{R}^n$ into itself, and let Ω be the set of all invertibles in $L(\mathbb{R}^n)$. We will first prove that Ω is open (Theorem 10.1.3).

Theorem 10.1.3. (Ω is Open) Suppose $A\in\Omega$. If we define $\epsilon\triangleq\frac{1}{\|A^{-1}\|_{\text{op}}}$, then

$$B_\epsilon(A)\overset{\text{def}}{=}\{T\in L(\mathbb{R}^n):\|T-A\|_{\text{op}}<\epsilon\}\subseteq\Omega$$

Proof. Fix $T\in B_\epsilon(A)$ and $x\neq 0\in\mathbb{R}^n$. We are required to show

$$|Tx|>0$$

If $T=A$, then the proof is trivial. We therefore suppose $T\neq A$. Define

$$\beta\triangleq\|T-A\|_{\text{op}}$$

Note that $T\neq A\in B_\epsilon(A)$ implies $0<\beta<\epsilon$. We claim

$$(\epsilon-\beta)\,|x|\le|Tx|$$

Observe

$$\epsilon |x| = \epsilon \left|A^{-1}Ax\right| \leq |Ax| \tag{10.1}$$

Observe

$$|Ax| \leq |(A-T)x| + |Tx| \leq \beta |x| + |Tx| \tag{10.2}$$

Equation 10.1 and Equation 10.2 implies

$$\epsilon |x| \leq \beta |x| + |Tx|$$

which implies

$$(\epsilon - \beta) |x| \leq |Tx| \text{ (done)}$$

■

Theorem 10.1.3 is essential for proving Inverse Function Theorem (Theorem 10.1.4). Think about what happen if f is linear. If f is linear, then f^{-1} is linear, and we will have

$$df^{-1} = f^{-1} = (f)^{-1} = (df)^{-1}$$

Because derivative is unique, it is reasonable to guess that if f is not linear and f^{-1} is differentiable, we would have

$$df^{-1} = (df)^{-1}$$

Now, if we wish df^{-1} to exists everywhere on $f(U)$, we must guarantee that df is invertible on U, and this is when Theorem 10.1.3 kick in. Note that in our proof of Inverse Function Theorem (Theorem 10.1.4), our selection of U guarantee df_x is invertible for all $x \in U$, by Theorem 10.1.3.

The rest of the proof boiled down to a fixed point argument (Theorem 2.8.1) to show f is one-to-one in U and $f(U)$ is open.

Theorem 10.1.4. (Inverse Function Theorem) Given a function f that maps an open neighborhood $E \subseteq \mathbb{R}^n$ around a into $\mathbb{R}^n$ such that

(a) f is differentiable on E

(b) df_a is invertible

(c) f is continuously differentiable at a

Then there exists open $U \subseteq E$ containing a such that

(a) f is one-to-one in U

(b) $f(U)$ is open

(c) The inverse of $f|_U$ is differentaible at $f(a)$.

Proof. Fix

$$\lambda \triangleq \frac{1}{2\|(df_a)^{-1}\|_{\text{op}}} \tag{10.3}$$

Because f is continuously differentaible at a, we know there exists δ such that

$$\|df_x - df_a\|_{\text{op}} < \lambda \qquad (x \in B_\delta(a)) \tag{10.4}$$

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

$$\text{for all } y \in \mathbb{R}^n, \phi_y : U \to \mathbb{R}^n \text{ is a contraction of } U$$

Fix $y \in \mathbb{R}^n$, and $x_1, x_2 \in U$. We claim

$$|\phi_y(x_1) - \phi_y(x_2)| \leq \frac{1}{2}|x_1 - x_2| \tag{10.5}$$

Because U is convex, Theorem 10.1.1 allow us to reduce the problem into proving

$$\|d(\phi_y)_x\|_{\text{op}} \leq \frac{1}{2} \text{ for all } x \in U$$

Fix $x \in U$. Using Chain Rule (Theorem 4.7.1) and the fact that the derivative of a bounded linear transformation is itself, we can 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 10.3 and Equation 10.4 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

$$\phi_y \text{ has at most one fixed point}$$

Because ϕ_y is a contraction of U, Banach Fixed Point Theorem (Theorem 2.8.1) tell us ϕ_y has at most one fixed point. (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 (Theorem 2.8.1) 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 known 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 can now deduce

$$\begin{aligned}|\phi_y(x) - x_0| &\leq |\phi_y(x_0) - \phi_y(x)| + |x_0 - \phi_y(x_0)| \\ &\leq \frac{1}{2}\,|x_0 - x| + \frac{r}{2} \leq r \text{ (done)}\end{aligned}$$

We now prove

$$df_x \text{ is invertible for all } x \in U$$

Fix $x \in U$.

$$\|df_x - df_a\|_{\text{op}} \cdot \|(df_a)^{-1}\|_{\text{op}} < \frac{1}{2}$$

Theorem 10.1.3 now implies df_x is invertible. (done)

Lastly, it remains to prove

$$f^{-1} : f(U) \to U \text{ is differentiable on } f(U)$$

Fix $y \in f(U)$, and $x \triangleq f^{-1}(y)$. We are required to prove

$$\lim_{k \to 0} \frac{\left|f^{-1}(y + k) - x - (df_x)^{-1}k\right|}{|k|} = 0$$

Fix $h(k) \triangleq f^{-1}(y+k) - f(x)$. In other words, $h \in \mathbb{R}^n$ is fixed to be the unique vector such that

$$f(x + h) = y + k$$

We now see

$$\begin{aligned}f^{-1}(y + k) - x - (df_x)^{-1}k &= h - (df_x)^{-1}k \\ &= -(df_x^{-1})(f(x + h) - f(x) - df_x h)\end{aligned}$$

and see

$$\left|f^{-1}(y+k)-x-(df_x)^{-1}k\right| \leq \|(df_x)^{-1}\|_{\text{op}} \left|f(x+h)-f(x)-df_x h\right|$$

which give us

$$\frac{\left|f^{-1}(y+k)-x-(df_x)^{-1}k\right|}{|k|} \leq \|(df_x)^{-1}\|_{\text{op}} \frac{|f(x+h)-f(x)-df_x h|}{|h|} \cdot \frac{|h|}{|k|}$$

This allow us to reduce the problem into proving

$$\limsup_{k\to 0} \frac{|h|}{|k|} \in \mathbb{R}$$

We claim

$$\frac{|h|}{|k|} \leq \lambda^{-1} \text{ for all } k \text{ such that } y+k \in f(U)$$

Compute

$$\phi_y(x+h)-\phi_y(x) = h-(df_a)^{-1}k$$

Equation 10.5 let us deduce

$$\left|h-(df_a)^{-1}k\right| = |\phi_y(x+h)-\phi_y(x)| \leq \frac{|h|}{2}$$

This with triangle inequality implies

$$\|(df_a)^{-1}\|_{\text{op}} |k| \geq \left|(df_a)^{-1}k\right| \geq \frac{|h|}{2} \text{ (done)}$$

■

The following is a technical recap of our proof for the Inverse Function Theorem (Theorem 10.1.4).

1: Let $\lambda \triangleq \frac{1}{2\|(df_a)^{-1}\|_{\text{op}}}$

2: Claim $B_\delta(a)$ suffices to be U, where $\|df_x - df_a\|_{\text{op}} < \lambda$

3: For each $y \in \mathbb{R}^n$, define $\phi_y : U \to \mathbb{R}^n$ by $\phi_y(x) \triangleq x + (df_a)^{-1}(y)$.

4: Prove that ϕ_y is a contraction of U by taking derivative and utilize step 1 and 2.

5: Prove that ϕ_y fix $x \iff f(x) = y$

6: Prove f is one-to-one in U using step 4,5.

7: Prove $f(U)$ is open by proving $B_{\lambda r}(y_0) \subseteq f(U)$, while $\overline{B_r(x_0)} \subseteq U$. The proof use step 4,5, some computation and ultimately claim that ϕ_y admits a fixed point as ϕ_y maps $\overline{B_r(x_0)}$ into itself.

8: Prove df_x is invertible in U by Theorem 10.1.3

9: Algebraically prove $df^{-1} = (df)^{-1}$, using $\left|h - (df_a)^{-1}k\right| = |\phi_y(x+h) - \phi_y(x)| \leq \frac{|h|}{2}$

Theorem 10.1.5. (Inversion is Continuous)

$$\text{The mapping } A \to A^{-1} \text{ is continuous on } \Omega$$

Proof. Fix $A \in \Omega$ and let $T \in \Omega$. We are required to prove

$$\lim_{T \to A} \|T^{-1} - A^{-1}\|_{\text{op}} = 0$$

We know

$$T^{-1} - A^{-1} = T^{-1}(A - T)A^{-1}$$

This implies

$$\|T^{-1} - A^{-1}\|_{\text{op}} \leq \|T^{-1}\|_{\text{op}} \|A - T\|_{\text{op}} \|A^{-1}\|_{\text{op}}$$

This allow us to reduce the problem into proving

$$\limsup_{T \to A} \|T^{-1}\|_{\text{op}} \in \mathbb{R}$$

Fix $\epsilon \triangleq \frac{1}{\|A^{-1}\|_{\text{op}}}$, $T \in B_\epsilon(A)$ and $\beta \triangleq \|T - A\|_{\text{op}} < \epsilon$. We claim

$$\|T^{-1}\|_{\text{op}} \leq (\epsilon - \beta)^{-1}$$

Following the proof of Theorem 10.1.3, we have

$$(\epsilon - \beta)\,|x| \leq |Tx| \text{ for all } x \in \mathbb{R}^n$$

This implies

$$\frac{\left|T^{-1}x\right|}{|x|} \leq (\epsilon - \beta)^{-1} \text{ for all } x \neq 0 \in \mathbb{R}^n \quad \text{(done)}$$

■

Corollary 10.1.6. (Continuously Differentaible Version of Inverse Function Theorem) Given a function f that maps open $E \subseteq \mathbb{R}^n$ containing a into $\mathbb{R}^n$ such that

(a) f is differentiable on E

(b) df_a is invertible

(c) f is continuously differentiable on E

Then there exists open $U \subseteq E$ containing a such that

$$f(U) \text{ is open and } f|_U : U \to f(U) \text{ is a diffeomorphism}$$

10.2 Implicit Function Theorem

Some notations first. If $A \in L(\mathbb{R}^{n+m}, \mathbb{R}^k)$. We define $A|_{\mathbb{R}^n} : L(\mathbb{R}^n, \mathbb{R}^k)$ by

$$A|_{\mathbb{R}^n}(x) \triangleq A(x, 0)$$

Theorem 10.2.1. (Implicit Function Theorem) Suppose a function f that maps open $E \subseteq \mathbb{R}^n \times \mathbb{R}^m$ containing (a, b) into $\mathbb{R}^n$ satisfy

(a) $f(a, b) = 0$

(b) $(df_{(a,b)})|_{\mathbb{R}^n}$ is invertible

(c) f is continuously differentiable on E

Then there exists open $U \subseteq E$ containing (a, b) and open $W \subseteq \mathbb{R}^m$ containing b such that there exists a unique function g from W to $\mathbb{R}^n$ such that

(a) $\big(g(y), y\big) \in U$ for all $y \in W$

(b) $f(g(y), y) = 0$ for all $y \in W$

Moreover, g satisfy

(a) g is continuously differentiable on W

(b) g satisfy $dg_b = -(df_{(a,b)}|_{\mathbb{R}^n})^{-1} \circ df_{(a,b)}|_{\mathbb{R}^m}$

Proof. Define $F : E \to \mathbb{R}^n \times \mathbb{R}^m$ by

$$F(x, y) \triangleq \Big(f(x, y), y\Big)$$

Because f is continuously differentiable on E, using Differentaibility Theorem (Theorem 4.5.2), we can deduce

$$F \text{ is continuously differentiable on } E$$

Again using Differentiability Theorem (Theorem 4.5.1), we can write down $dF_{(a,b)}$ in the matrix form with respect to standard basis

$$[dF_{(a,b)}] = \begin{bmatrix} df_{(a,b)}|_{\mathbb{R}^n} & O \\ df_{(a,b)}|_{\mathbb{R}^m} & I \end{bmatrix}$$

Now, because $df_{(a,b)}|_{\mathbb{R}^n}$ is invertible, we know $dF_{(a,b)}$ is invertible.

We can now apply Inverse Function Theorem (Theorem 10.1.4) to $F : E \to \mathbb{R}^n \times \mathbb{R}^m$. This give us

(a) an open $U \subseteq E \subseteq \mathbb{R}^n \times \mathbb{R}^m$ containing (a,b)

(b) open $V \triangleq F(U) \subseteq \mathbb{R}^n \times \mathbb{R}^m$ containing $(0,b)$

(c) $F|_U : U \to V$ is a diffeomorphism.

Define W by

$$W \triangleq \{y \in \mathbb{R}^m : (0,y) \in V\}$$

We claim

$$\text{such } U, W \text{ suffices}$$

Note that it is easy to check W is open, utilizing V is open and the same ϵ.

Now, because $F|_U : U \to V$ is bijective, we know for each $y \in W$, there exists unique $(x,y) \in U$ such that

$$F(x,y) = (0,y)$$

We can now well define a function $g : W \to \mathbb{R}^n$ such that

$$\Big(g(y), y\Big) \in U \text{ and } f(g(y), y) = 0 \text{ for all } y \in W$$

It remains to show

(a) g is continuously differnetiable on W

(b) $dg_b = -(df_{(a,b)})|_{\mathbb{R}^n}^{-1}(df_{(a,b)})|_{\mathbb{R}^m}$

Fix $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$. We wish to prove

$$\partial_i g_j \text{ exists and is continuous on } W$$

Express

$$g(y_1, \dots, y_m) = \Big(g_1(y_1, \dots, y_m), \dots, g_n(y_1, \dots, y_m)\Big)$$

Express

$$F^{-1}(z_1, \dots, z_{n+m}) = \Big(F_1^{-1}(z_1, \dots, z_{n+m}), \dots, F_{n+m}^{-1}(z_1, \dots, z_{n+m})\Big)$$

Because F^{-1} is continuously differentiable on V, we reduce the problem into proving

$$\partial_i g_j(y) = \partial_{n+i} F_j^{-1}(0,y) \text{ for all } y \in W$$

Because $F^{-1}(0,y) = (g(y),y)$ for all $y \in W$, we know

$$F_j^{-1}(0,\dots,0,y_1,\dots,y_m) = g_j(y_1,\dots,y_m) \text{ for all } y \in W \tag{10.6}$$

Fix arbitrary $y = (y_1,\dots,y_m) \in W$. Because W is open, we can see from Equation 10.6 that

$$\begin{aligned}
\partial_i g_j(y) &= \lim_{t\to 0} \frac{g_j(y_1,\dots,y_i+t,\dots,y_m)}{t} \\
&= \lim_{t\to 0} \frac{F_j^{-1}(0,\dots,0,y_1,\dots,y_i+t,\dots,y_m)}{t} \\
&= \partial_{n+i} F_j^{-1}(y) \text{ (done)}
\end{aligned}$$

Define $\Phi : W \to U$ by

$$\Phi(y) = \Big(g(y), y\Big)$$

By definition of g, we have

$$f \circ \Phi = 0 \text{ on } W$$

This by Chain Rule give us

$$df_{\Phi(y)} \circ d\Phi_y = 0 \text{ on } W$$

In particular

$$df_{(a,b)} \circ d\Phi_b = 0$$

Now, compute

$$d\Phi_b = \begin{bmatrix} dg_b \\ I \end{bmatrix} \text{ and } df_{(a,b)} = \begin{bmatrix} df_{(a,b)}|_{\mathbb{R}^n} & df_{(a,b)}|_{\mathbb{R}^m} \end{bmatrix}$$

This then give us

$$df_{(a,b)}|_{\mathbb{R}^n} \circ dg_b + df_{(a,b)}|_{\mathbb{R}^m} = 0$$

and of course

$$dg_b = -\big(df_{(a,b)}|_{\mathbb{R}^n}\big)^{-1} \circ df_{(a,b)}|_{\mathbb{R}^m} \text{ (done) (done)}$$

■

Example 20 (Unit Circle Example)

$$f(x,y) \triangleq x^2 + y^2 - 1 \text{ and } (a,b) \triangleq (1,1)$$

We have

$$g(y) = \sqrt{2 - y^2} \text{ on } y \in (1-\epsilon, 1+\epsilon)$$

Compute

$$df_{(a,b)} = \begin{bmatrix} 2 & 2 \end{bmatrix} \text{ and } dg_1 = \begin{bmatrix} -1 \end{bmatrix}$$

This established

$$dg_a = -(df_{(a,b)}|_{\mathbb{R}^1})^{-1} \circ df_{(a,b)}|_{\mathbb{R}^1}$$

Example 21 (Implicit Function Theorem Implies Inverse Function Theorem)

Given continuously differentaible $h : E \overset{\text{open}}{\subseteq} \mathbb{R}^n \ni a \to \mathbb{R}^n$ such that dh_a is invertible

Define $f : E \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$f(x,y) \triangleq h(x) - y$$

It is easily checked that $f(x, h(x)) = 0$ and the rest of the condition is satisfied. Now by Implicit Function Theorem (Theorem 10.2.1), we see that there exists $g : W \overset{\text{open}}{\subseteq} \mathbb{R}^n \to E$ such that

$$f(g(y), y) = 0 \text{ for all } y \in W$$

In other words,

$$h(g(y)) = y \text{ for all } y \in W$$

10.3 Feynman's Trick

In this section

Theorem 10.3.1. (Feynman's Trick) Given a real-valued function $f(x,t)$ defined on $[a,b]\times[c,d]$, and an real-valued function α of bounded variation on $[a,b]$, such that

(a) $\partial_2 f(x,t)$ exists on $[a,b]\times[c,d]$

(b) For all $t\in[c,d]$, the integral $\int_a^b f(x,t)d\alpha(x)$ exists.

(c) For all $s\in[c,d]$ and ϵ, there exists δ such that

$$|\partial_2 f(x,t)-\partial_2 f(x,s)|<\epsilon \text{ for all } x\in[a,b] \text{ and all } t\in(s-\delta,s+\delta)$$

Then we have

$$\frac{d}{dt}\int_a^b f(x,t)d\alpha(x)=\int_a^b \frac{\partial}{\partial t}f(x,t)d\alpha(x)$$

In other words, if we define $g(t)\triangleq\int_a^b f(x,t)d\alpha(x)$, then we have

$$g'(t)=\int_a^b \partial_2 f(x,t)d\alpha(x)$$

Proof. Fix $s\in[c,d]$. We are required to prove

$$g'(s)=\int_a^b \partial_2 f(x,s)dx$$

Note that, for all $t\neq s\in[c,d]$, we have

$$\frac{g(t)-g(s)}{t-s}=\int_a^b \frac{f(x,t)-f(x,s)}{t-s}d\alpha(x)$$

This allow us to reduce the problem into proving

$$\frac{f(x,t)-f(x,s)}{t-s}\to\partial_2 f(x,s) \text{ uniformly for all } x\in[a,b] \text{ as } t\to s$$

By MVT (Corollary 4.4.3), we know for all $t\neq s\in[c,d]$, there exists u_t between t and s such that

$$\frac{f(x,t)-f(x,s)}{t-s}=\partial_2 f(x,u_t)$$

The proof now follows from (c). (done) ∎

Example 22 (Introductive application of Feynman's Trick)

$$\text{What is the value of } \int_0^1 \frac{x-1}{\ln x} dx \text{ ?}$$

Define $f(x,t) \triangleq \frac{x^t-1}{\ln x}$ on $[0,1] \times [0,1]$. Observe $\partial_2 f(x,t) = x^t$, and observe

$$\int_0^1 f(x,0)dx = 0 \text{ and } \int_0^1 \partial_2 f(x,t)dx = \frac{1}{t}$$

We can con compute

$$\begin{aligned}\int_0^1 \frac{x-1}{\ln x} dx &= \int_0^1 f(x,1)dx \\ &= \int_0^1 f(x,0)dx + \int_0^1 \Big(\int_0^1 \partial_2 f(x,t)dx \Big) dt = 0\end{aligned}$$

Example 23 (Dirichlet's Integral)

$$\text{What is the value of } \int_0^\infty \frac{\sin t}{t} dt \text{ ?}$$

Define the Laplace transformation

$$f(s,t) \triangleq e^{-st} \frac{\sin t}{t} \text{ on } \mathbb{R}_0^+ \times \mathbb{R}^+$$

Observe

$$\partial_1 f(s,t) = -e^{-st} \sin t$$

Now compute

$$\begin{aligned}\int_0^\infty -e^{-st} \sin t dt &= \frac{1}{-2i} \int_0^\infty e^{-st}(e^{it} - e^{-it})dt \\ &= \frac{1}{-2i} \Big(\frac{e^{t(i-s)}}{i-s} - \frac{e^{t(-i-s)}}{-i-s} \Big) \Big|_{t=0}^\infty \\ &= \frac{-1}{1+s^2} = \frac{d}{ds}(-\arctan s)\end{aligned}$$

It is clear that

$$\lim_{s \to \infty} \int_0^\infty f(s,t)dt = 0$$

We now have

$$\begin{aligned}
\int_0^\infty \frac{\sin t}{t} dt &= \int_0^\infty f(0,t) dt \\
&= \lim_{s\to\infty} \int_0^\infty f(s,t) dt - \int_0^\infty \int_0^\infty \partial_1 f(s,t) dt ds \\
&= \int_0^\infty \frac{1}{1+s^2} ds = \frac{\pi}{2}
\end{aligned}$$

10.4 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 <u>axioms</u>:

(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 24 (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 10.4.1. (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 10.4.2. (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 10.4.1 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

$$\begin{aligned}\langle y, y\rangle &= \langle z + \frac{\langle y, x\rangle}{\|x\|^2}x, z + \frac{\langle y, x\rangle}{\|x\|^2}x\rangle \\ &= \langle z, z\rangle + \left|\frac{\langle y, x\rangle}{\langle x, x\rangle}\right|^2 \langle x, x\rangle \\ &= \langle z, z\rangle + \frac{|\langle x, y\rangle|^2}{\langle x, x\rangle}\end{aligned}$$

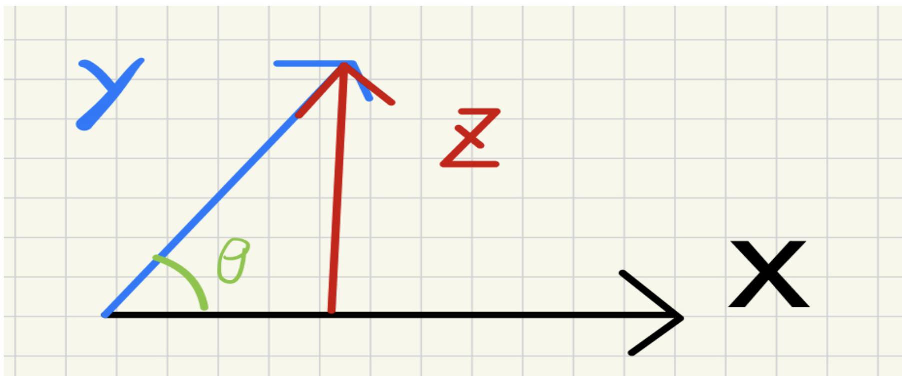
Because $\langle z, z \rangle \geq 0$, we now have

$$\langle y, y \rangle = \langle z, z \rangle + \frac{|\langle x, y \rangle|^2}{\langle x, x \rangle} \geq \frac{|\langle x, y \rangle|^2}{\langle x, x \rangle} \text{ (done)}$$

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 $(x_1, \ldots, x_n)$ 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 11

Beauty

11.1 Fundamental Theorem of Algebra

Theorem 11.1.1. (Fundamental Theorem of Algebra)

11.2 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 4.17.

(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 4.1.10

(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 4.10.1.

(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 4.1.10.

In particular, we have **Euler's Formula**.

Theorem 11.2.1. (Euler's Formula)

$$\exp(iz) = \cos(z) + i\sin(z) \qquad (z \in \mathbb{C})$$

Proof. Define

$$I(n) \triangleq \begin{cases} 1 & \text{if } n \equiv 0 \pmod 4 \\ i & \text{if } n \equiv 1 \pmod 4 \\ -1 & \text{if } n \equiv 2 \pmod 4 \\ -i & \text{if } n \equiv 3 \pmod 4 \end{cases}$$

Compute

$$\begin{aligned} \exp(iz) &= \sum_{n=0}^{\infty} \frac{(iz)^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{I(n)z^n}{n!} \qquad (11.1) \\ &= \sum_{n=0}^{\infty} \frac{I(2n)}{(2n)!} z^{2n} + \frac{I(2n+1)}{(2n+1)!} z^{2n+1} \qquad \big(\because \text{ this is a sub-sequence of } (11.1)\big) \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} + i \cdot \frac{(-1)^n}{(2n+1)!} z^{2n+1} \end{aligned}$$

Now, we can conclude

$$\begin{aligned} \cos(z) + i\sin(z) &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} + i \cdot \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} + \sum_{n=0}^{\infty} i \cdot \frac{(-1)^n}{(2n+1)!} z^{2n+1} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} + i \cdot \frac{(-1)^n}{(2n+1)!} z^{2n+1} = \exp(iz) \end{aligned}$$

■

11.3 Equivalent Definitions of Exponential Functions

Theorem 11.3.1. (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 4.1.10)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 5.10.1), 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 5.12.1), 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 5.12.2) 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 11.3.2. (Second Characterization)

11.4 Equivalent Definitions of Trigonometric Functions

11.5 Equivalent Definitions of Gamma and Beta Functions

11.6 Prime Number Theorem

Chapter 12

and the Beast

12.1 Topologist's Sine Curve

12.2 Long Line

12.3 Bugged Eye Line

12.4 Weierstrass Function

12.5 Fabius Function

12.6 Vitali Set

12.7 Cantor Set

12.8 Cantor-Lebesgue Function

12.9 Volterra's Function

12.10 Peano Space-filling Curve