Emory University

MATH 411 & 2 Real Analysis

Learning Notes

Jiuru Lyu

April 29, 2025

Contents

1 The Real Line and Euclidean Space

1.1 Algebraic Properties of $\mathbb{R}$ (as a Ordered Field)

Axiom 1.1.1 Field Axioms: Recall the following properties

- Addition Axioms

  (1) *commutativity*: $x + y = y + x$

  (2) *associativity*: $(x + y) + z = x + (y + z)$

  (3) *the zero element*: $x + 0 = x$

  (4) *the negative element*: $x + (-x) = 0$
  This further gives the definition of *subtraction*: $y - x = y + (-x)$.

- Multiplication Axioms

  (5) *commutativity*: $xy = yx$

  (6) *associativity*: $(xy)z = x(yz)$

  (7) *the one element/unit vector*: $x \cdot 1 = x$

  (8) *inverse*: for each $x \neq 0, \exists\, x^{-1}\ s.t.\ x \cdot x^{-1} = 1$
  This further gives the definition of *division*: $y/x = y \cdot x^{-1}$ when $x \neq 0$.

  (9) *distribution*: $x(y + z) = xy + xz$

  (10) $1 \neq 0$

- Order Axioms

  (11) *reflexivity*: $x \leq x$

  (12) *anti-symmetry*: If $x \leq y$ and $y \leq x \implies x = y$.

  (13) *transitivity*: If $x \leq y$ and $y \leq z \implies x \leq z$

  (14) *linear relation*: For each pair x, y, either $x \leq y$ or $y \leq x$.

  (15) *compatibility with addition*: If $x \leq y \implies x + z \leq y + z \quad \forall\, z$

  (16) *compatibility with multiplication*: If $0 \leq x$ and $0 \leq y \implies 0 \leq xy$.

Definition 1.1.2 (Ordered Field). A system (or a set) $\mathcal{F}$ is called an *ordered field* if it satisfies all the above 16 properties.

Remark 1.1 (Examples of Ordered Field) $\mathbb{R}$ *and* $\mathbb{Q}$.

Definition 1.1.3 (Field). A set is called a *field* if satisfies all the addition and multiplication axioms.
Definition 1.1.4 (Ring). A set is a *ring* if it satisfies $(1) - (9)$ except (5) and (8).

Example 1.1.5 $\mathbb{Z}$ as a Ring

$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$, the set of integers, is a commutative ring, but not a field.

Remark 1.2 *There is no division operation in a ring as multiplicative inverse is not defined.*

Definition 1.1.6 (Group). A set is a *group* if it satisfies $(1) - (4)$.

Theorem 1.1.7 Law of Trichotomy

If x and y are elements of an ordered field, then exactly one of the relations $x < y$, $x = y$, or $x > y$ holds.

Proposition 1.1.8 Other Algebraic Properties of $\mathbb{R}$ (as an Ordered Field):

1. *unique identities*: If $a + x = a$ for every a, then $x = 0$. If $a \cdot x = a$ for every a, then $x = 1$.
2. *unique inverses*: If $a + x = 0$, then $x = -a$. If $ax = 1$, then $x = a^{-1}$.
3. *no divisors of zero*: If $xy = 0$, then $x = 0$ or $y = 0$.
4. *cancellation laws for addition*: If $a + x = b + x$, then $a = b$. If $a + x \leq b + x$, then $a \leq b$.
5. *cancellation for multiplication*: If $ax = bx$ and $x \neq 0$, then $a = b$. If $ax \geq bx$ and $x > 0$, then $a \geq b$.
6. $0 \cdot x = 0$ for every x.
7. $-(-x) = x$ for every x.
8. $-x = (-1)x$ for every x.
9. If $x \neq 0$, then $x^{-1} \neq 0$ and $(x^{-1})^{-1} = x$.
10. If $x \neq 0$ and $y \neq 0$, then $xy \neq 0$ and $(xy)^{-1} = x^{-1}y^{-1}$.
11. If $x \leq y$ and $0 \leq z$, then $xz \leq yz$. If $x \leq y$ and $z \leq 0$, then $yz \leq xz$.
12. If $x \leq 0$ and $y \leq 0$, then $xy \geq 0$. If $x \leq 0$ and $y \geq 0$, then $xy \leq 0$.
13. $0 < 1$.
14. For any x, $x^2 \geq 0$.

Proof 1. (Of No. 14)
$\boxed{\text{Case I}}$ If $x \geq 0$, then $x^2 = x \cdot x \geq 0$, by property (16) of Axiom 1.1.
$\boxed{\text{Case II}}$ If $x < 0$, then

$$\begin{aligned} x^2 = x \cdot x &= (-1)(-x) \cdot (-1)(-x) \quad \text{[by property 7 of Proposition 1.7]} \\ &= (-1)^2 \cdot (-x)^2. \end{aligned}$$

Note that $0 = (-1)(-1+1) = (-1)^2 + (-1)$ if we distribute (-1). Then, adding 1 on both sides, we have

$$1 = (-1)^2 + (-1) + 1 = (-1)^2 \quad \text{[by additive inverse]}$$

That is, $(-1)^2 = 1$. So, $x^2 = (-1)^2 \cdot (-x)^2 = 1 \cdot (-x)^2 = (-x)^2 \geq 0$ by Case I.

Q.E.D. ■

Proposition 1.1.9 : $ab \leq \dfrac{a^2+b^2}{2}$.

Proof 2.

$$\begin{aligned} (a-b)^2 &\geq 0 \quad \text{[By property 14 of Proposition 1.7]} \\ a^2+b^2-2ab &\geq 0 \\ 2ab &\leq a^2+b^2 \\ ab &\leq \frac{a^2+b^2}{2}. \end{aligned}$$

Q.E.D. ■

Definition 1.1.10 (Absolute Value (Norm) and Distance (Metric)). For $x, y \in \mathbb{R}$, $|x| = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases}$ is the *absolute value*, and $d(x, y) = |x - y|$ is the *distance*.

Proposition 1.1.11 Properties of Absolute Value and Distance:

- $|x| \geq 0$ for every x.
- $|x| = 0$ if and only if $x = 0$.
- $|xy| = |x||y|$.
- $d(x, y) \geq 0$
- $d(x, y) = 0$ if and only if $x = y$.
- $d(x, y) = d(y, x)$.

Theorem 1.1.12 Triangle Inequalities

$\forall\, x, y, z \in \mathbb{R}$

1. $|x + y| \leq |x| + |y|$
2. $||x| - |y|| \leq |x - y|$
3. $d(x, y) \leq d(x, z) + d(z, y)$

Proof 3. (Of No. 1)

$\boxed{\text{Case I}}$ Suppose $x \geq 0$ and $y \geq 0$. Then, $x + y \geq 0$, and

$$|x + y| = x + y = |x| + |y|. \quad \square$$

$\boxed{\text{Case II}}$ WLOG, suppose $x \geq 0$ and $y < 0$.

- Suppose $x + y \geq 0$, then

$$|x+y| = x+y = |x| - (-y) = |x| - |y| \leq |x| + |y|. \quad \square$$

- Suppose $x + y < 0$, then

$$|x+y| = -(x+y) = -x-y = -|x| + |y| \leq |x| + |y|. \quad \square$$

$\boxed{\text{Case III}}$ Suppose $x < 0$ and $y < 0$. Then, $x + y < 0$, and

$$|x+y| = -(x+y) = -x + (-y) = |x| + |y|$$

Q.E.D. ■

1.2 Construction of $\mathbb{R}$ and Completeness of $\mathbb{R}$

Notation 1.1. Recall the following number systems:

$$\begin{array}{c|c} \mathbb{N} = \mathbb{Z}^+ = \{0,1,2,3,\dots\} & \text{non-negative integers} \\ \mathbb{Z} & \text{integers} \\ \mathbb{Q} = \left\{\dfrac{m}{n} : m,n \in \mathbb{Z},\ n \neq 0\right\} & \text{rational numbers} \end{array}$$

Proposition 1.2.2 Important Properties of Number Systems:

- For $\mathbb{N}$:
  - **Definition 1.2.3 (Principle of mathematical induction).** If S is a subset of $\mathbb{Z}^+$ $s.t.$ $0 \in S$ and $k \in S \implies k+1 \in S$, then $S = \mathbb{Z}^+$.
  - **Definition 1.2.4 (Well-Ordered Property).** Each subset $S \neq \varnothing$ has a smallest element.
    As a consequence of well-ordering property, we have the principle of complete induction:
    Definition 1.2.5 (Principle of Complete Induction). If $S \subset \mathbb{Z}^+$ is a subset $s.t.$ $\{x \in \mathbb{Z}^+ \mid x < n\} \subset S \implies n \in S$, then $S = \mathbb{Z}^+$.
- For $\mathbb{Z}$:
  - Commutative ring with identity
- For $\mathbb{Q}$:
  - **Definition 1.2.6 (Countable).** $\mathbb{Q}$ can be placed in one-to-one correspondence with $\mathbb{N}$ (or a subset of it). The whole $\mathbb{Q}$ can be displayed as a list or sequence.

    Remark 1.3 *A simple way to prove it is to consider the points in the plane with integer coordinates, say (p,q). After assigning fraction $\frac{p}{q}$ (simplified to lowest terms and leave out cases when $q = 0$) to this point, we achieve a one-to-one correspondence.*

– **Definition 1.2.7 (Dense in Itself).** If $x, y \in \mathbb{Q}$ and $x < y \implies \exists\, z \in \mathbb{Q}\; s.t.\; x < z < y$.

– **Proposition 1.2.8 Archimedean Property:**

$$\forall\, x \in \mathbb{Q},\ \exists\, n \in \mathbb{Z}\; s.t.\; n > x.$$

Proof 1. If $x \leq 0$, take $n = 1$. If $x = \dfrac{p}{q}$ with $p, q > 0$, take $n = p + 1$.

Q.E.D. ■

Remark 1.4 *Equivalent formulation of the Archimedean Property:*

- *If $x \in \mathbb{Q}$, then $\exists$ integer n $s.t.$ $x < n$.*
- *If $x, y \in \mathbb{Q}$ and $0 < x < y$, then $\exists$ integer k $s.t.$ $kx > y$.*
- *If $x > 0 \in \mathbb{Q}$, then $\exists$ integer $n > 0$ $s.t.$ $0 < \dfrac{1}{n} < x$.*

– Ordered field.

$\mathbb{Q}$ is already an ordered field, why do we bother to define $\mathbb{R}$ for analysis?

The big idea: $\mathbb{Q}$ is not *quite complete*

- Evidence 1 (Analysis POV): There is no rational whose square is 2. That is, $x^2 = 2$ has no solution in $\mathbb{R}$.

  Proof 2. We will use proof by contradiction. Assume $\exists$ solution $x = \dfrac{m}{n}$ with $m, n \in \mathbb{Z}$ and they have no common factors. Then,

$$\left(\frac{m}{n}\right)^2 = 2 \implies m^2 = 2n^2.$$

  So, m^2 is even, then m is even as well. Suppose $m = 2k,\ k \in \mathbb{Z}$. Then,

$$\begin{aligned} m^2 = (2k)^2 = 4k^2 &= 2n^2 \\ n^2 &= 2k^2. \end{aligned}$$

  So, n^2 is even, and n is even.

  $\divideontimes$ m, n both even, so they have a common factor of 2. This contradict with our assumption. So, $\nexists$ a solution $x \in \mathbb{Q}\; s.t.\; x^2 = 2$.

  Q.E.D. ■

- Evidence 2 (Geometry POV): There is no rational representation of the diagonal of a square of size 1.

Remark 1.5 (Informal Definition of Sequence Limit) *A sequence is said to* converge *to a limit x if we cna guarantee that the points in the sequence are as close as we wish to x by going far enough out in the sequence.*

Definition 1.2.9 (Limit of a Sequence). A sequence $\{x_n\}$ is said to *converge* to x if $\forall\, \varepsilon > 0$, $\exists$ integer N $s.t.$ $|x_n - x| < \varepsilon$ whenever $n \geq N$. (Alternatively, $n \geq N \implies |x_n - x| < \varepsilon$). We denote the *limit* as

$$\lim_{x\to\infty} x_n = x \quad \text{or, simply} \quad x_n \to x \text{ as } n \to \infty.$$

Remark 1.6 *N depends on ε, and the smaller the ε, the bigger the N.*

Example 1.2.10 Show $\lim\limits_{n\to\infty} \dfrac{n+1}{n+2} = 1$**.**

Proof 3. Given $\varepsilon > 0$ [fix ε], we need to find N $s.t.$ $n \geq N \implies |x_n - 1| < \varepsilon$, where $x_n = \dfrac{n+1}{n+2}$. Consider

$$|x_n - 1| = \left|\frac{n+1}{n+2} - 1\right| = \left|\frac{n+1-n-2}{n+2}\right| = \left|\frac{-1}{n+2}\right| = \frac{1}{n+2}.$$

Then, we want

$$\frac{1}{n+2} < \varepsilon \iff n+2 > \frac{1}{\varepsilon} \iff n > \frac{1}{\varepsilon} - 2.$$

By the Archimedean property, choose integer $N > \dfrac{1}{\varepsilon} - 2$. [$N$ is fixed and is what we want to find] Then, based on the arguments, when $n \geq N$ [n is changing], we have

$$|x_n - 1| = \frac{1}{n+2} \leq \frac{1}{N+2} < \varepsilon.$$

That is,

$$\lim_{n\to\infty} x_n = 1.$$

Q.E.D. ■

Theorem 1.2.11 Basic Properties of Limits

- **Sandwich Lemma/Squeeze Theorem**: Suppose $x_n \to L$, $y_n \to L$, and $x_n \leq z_n \leq y_n$ for all n. Then, $z_n \to L$. *It is also enough to assume that* $\exists\, N_0$ $s.t.$ $n > N_0 \implies x_n \leq z_n \leq y_n$
- If $a \leq x_n \leq b$ for every n and $x_n \to x$, then $a \leq x \leq b$.
- **Uniqueness**: If x_n is a sequence in an ordered field and $x_n \to x$ and $x_n \to y$, then $x = y$.
- **Boundedness**: A convergent sequence is bounded.
- **Arithmetic of Sequence and Limits**: Suppose $x_n \to x$ and $y_n \to y$. Then,

$$\{x_n\} + \{y_n\} = \{x_n + y_n\} \implies x_n + y_n \to x + y$$
$$\lambda\{x_n\} = \{\lambda x_n\} \implies \lambda x_n \to \lambda x$$
$$\{x_n\}\{y_n\} = \{x_n y_n\} \implies x_n y_n \to xy$$
$$\{x_n\}/\{y_n\} = \{x_n/y_n\} \implies x_n/y_n \to x/y$$

Definition 1.2.12 (Monotone Sequence Property/MSP). Every *monotone increasing sequence* that is *bounded (bdd) above* converges.

Remark 1.7 *"monotone increasing sequence" refers to a sequence where* $x_n \leq x_{n+1} \quad \forall\, n$*; "bdd above" refers to* $\exists\, x$ *s.t.* $x_n \leq x \quad \forall\, n$*, and we call this* x *an* upper bound.

Definition 1.2.13 (Completeness). An ordered field $\mathcal{F}$ is said to be *complete* if it has the MSP.

Construction of $\mathbb{R}$ (from $\mathbb{Q}$)

Consider set S of sequences,

$$S = \{(x_1, x_2, \dots) \mid x_n \in \mathbb{Q},\ x_n \uparrow \text{ (monotone increasing)},\ x_n \text{ bdd above}\}.$$

Define *equivalence relation* (reflexive, transitive, symmetric) $\sim$ on S:

$$\{x_n\} \sim \{y_n\} \iff x_n \text{ and } y_n \text{ have the } \textit{same upper bounds}.$$

Then, each equivalence class defines a unique real number (as the limit of the representing sequence). Let

$$\mathbb{R} = \{x \mid x \text{ is an equivalence class in } S\}.$$

If $r \in \mathbb{Q}$, then r is represented by the sequence r itself ($\{r\}$). So, $\mathbb{Q} \subseteq \mathbb{R}$.

Claim 1.2.14 $\mathbb{R}$ is a complete ordered field under the following operations: For $x = [\{x_n\}]$ and $y = [\{y_n\}]$,

- Addition: $x + y = [\{x_n + y_n\}]$
- Multiplication: $x \cdot y = [\{x_n \cdot y_n\}]$
- Order: $x \leq y \iff \exists$ upper bd of $\{x_n\}$ that is $\leq$ all upper bd of $\{y_n\}$.

Theorem 1.2.15
$\mathbb{R}$ is the "unique" complete ordered field.

Remark 1.8 *By* unique*, we mean isomorphism. That is, if* $\exists$ *another complete ordered field* $\mathcal{F}$*, we can put* $\mathcal{F}$ *and* $\mathbb{R}$ *into a one-to-one relationship.*

Proposition 1.2.16 Properties of $\mathbb{R}$:

- $\mathbb{R}$ is Archimedean: $\forall\, x \in \mathbb{R},\ \exists$ integer $n > x$.
- $\mathbb{Q}$ is dense in $\mathbb{R}$:
  - If $x, y \in \mathbb{R}$ and $x < y \implies \exists r \in \mathbb{Q}$ *s.t.* $x < r < y$.
  - If $x \in \mathbb{R}$ and $\varepsilon > 0 \implies \exists r \in \mathbb{Q}$ *s.t.* $|x - r| < \varepsilon$.

- The interval $(0,1)$ is uncountable. (Hence, $\mathbb{R}$ is uncountable).

Proof 4. (of uncountability)

Assume $(0,1)$ is countable. Then, it can be put into a one-to-one relationship with $\mathbb{N}$. Say the following list exhauste elements of $\mathbb{R}$:

$$x_1 = 0.a_{11}a_{12}\cdots a_{1n}\cdots,\ x_2 = 0.a_{21}a_{22}\cdots a_{2n}\cdots,\ldots, x_k = 0.a_{k1}a_{k2}\cdots a_{kn}\cdots,\ldots$$

[Goal: find a new number that is not in the list] Define a new number:

$$x = 0.x_1'x_2'\cdots x_k'\cdots,$$

where for each k, $x_k' = \begin{cases} 4 & \text{if } a_{kk} \neq 4 \\ 3 & \text{if } a_{kk} = 4. \end{cases}$ [This construction ensures $x_k' \neq a_{kk}$] Then, $x \in (0,1)$ and $x \neq x_k \quad \forall k$. ⁎ We have constructed a number that is not in the list. So, $(0,1)$ is not countable.

Q.E.D. ■

1.3 Another Approach: Least Upper Bound

Definition 1.3.1 (Upper Bound/Least Upper Bound). Let $S \subset \mathbb{R}$.

- We say b is an *upper bd* for S if $x \leq b \quad \forall x \in S$.
- We say b is a *least upper bd* for S if b is an upper bd and $\leq$ any upper bd of S.

We use $\text{lub}(S) = \sup(S)$ to denote the lease upper bd. ($\sup$ stands for supremum). For sets without an upper bound, we define $\sup(S) = +\infty$.

Remark 1.9 $b = \text{lub}(S) \iff$ *(1) b is an upper bound, and (2) $b \leq$ any upper bound of S.*

Example 1.3.2

Suppose $S_1 = (0,2)$; $S_2 = [0,2]$; $S_3 = \varnothing$; $S_4 = (0,\infty)$. Then, $\text{lub}(S_1) = 2$, $\text{lub}(S_2) = 2$, $\text{lub}(S_3) = +\infty$, $\text{lub}(S_4) = +\infty$.

Definition 1.3.3 (Greatest Lower Bound). We use $\text{glb}(S) = \inf(S)$ to denote the greatest lower bound. It is the largest lower bound of S. For sets without a lower bound, we define $\inf(S) = -\infty$.

Example 1.3.4

Example 1.3.2: $\inf(S_1) = 0$, $\inf(S_2) = 0$, $\inf(S_3) = -\infty$, $\inf(S_4) = 0$, $\inf((-\infty,4)) = -\infty$.

Proposition 1.3.5 : Let $S \subset \mathbb{R}$, $S \neq \varnothing$, then

- $b = \text{lub}(S) \iff b$ is an upper bound and $\forall \varepsilon > 0$, $\exists x \in S$ *s.t.* $x > b - \varepsilon$. *This implies that an element slightly smaller than b is not an upper bound any more.*

- $a = \inf(S) \iff a$ is a lower bound and $\forall\, \varepsilon > 0, \quad \exists\, x \in S\ s.t.\ x < a + \varepsilon$.

Proposition 1.3.6 : Suppose $\varnothing \neq A \subset B \subset \mathbb{R}$. Then,

$$\inf(B) \leq \inf(A) \leq \sup(A) \leq \sup(B).$$

Theorem 1.3.7 Equivalent Condition for Completeness: Least Upper Bound Condition
$\mathbb{R}$ has the following properties:

- LUB property: Every non-empty subset bounded above has the least upper bound.
- GLB property: Every non-empty subset bounded below has the greatest lower bound.

Proof 1. (of the LUB Property)
Set-up: Fix any $S \in \mathbb{R}$ that is bounded above and $S \neq \varnothing$.
[WTS: the existence of $\mathrm{lub}(S) \quad \Longleftarrow$ Tool: MSP (but we need to construct monotone sequence first.)
Step 1 Construction of a Monotone Sequence

Fix an upper bound M for S. For each fixed integer $n \geq 1$, consider $a_k = M - \dfrac{k}{2^n}, k = 1, 2, \ldots$. By the well-ordering property, we can choose an integer k_n who is the 1^{st} integer $k\ s.t.\ a_k$ is not an upper bound.

Let $b_n = M - \dfrac{k_n}{2^n}$. Then, b_n is not an upper bound, but $b_n + \dfrac{1}{2^n}$ is an upper bound (by construction).
Step 2 Apply MSP to $\{b_n\}$

- b_n is monotone increasing:

  Note that
  $$b_{n+1} - b_n = \left(M - \frac{k_{n+1}}{2^{n+1}}\right) - \left(M - \frac{k_n}{2^n}\right) = \frac{2k_n - k_{n+1}}{2^{n+1}}$$
  Suppose, for the sake of contradiction, that $b_{n+1} - b_n < 0$. Then, $b_{n+1} - b_n \leq -\dfrac{1}{2^{n+1}}$. That is,
  $$b_n \geq b_{n+1} + \frac{1}{2^{n+1}}.$$
  $\divideontimes$ However, by construction, b_n is not an upper bound, but $b_{n+1} + \dfrac{1}{2^{n+1}}$ is an upper bound. So, there is a contradiction, and thus $b_{n+1} - b_n > 0$. This contradictions shows that b_n is a monotone increasing sequence.

- b_n is bounded above:

  Note that $b_n \leq M$. So, b_n is bounded above.

By MSP, suppose $b_n \to b$ for some $b \in \mathbb{R}$.
Step 3 Show $b = \mathrm{lub}(S)$

- b is an upper bound:

Fix $x \in S$, we have $x \leq b_n + \dfrac{1}{2^n} \quad \forall n$. When $x \to \infty$, $x \leq b + 0$. So, $x \leq b$.

- b is the least upper bound: [WTS: $\forall \varepsilon > 0, \quad \exists x \in S \; s.t. \; b - \varepsilon < x$.]

  As b is the limit, we can always find a b_n $s.t.$ $|b_n - b| < \varepsilon$. That is, $b - b_n < \varepsilon$, or $b_n > b - \varepsilon$. Hence, b is the least upper bound.

Q.E.D. ■

1.4 Cauchy Sequence and Cauchy Completeness

Definition 1.4.1 (Cauchy Sequence). A sequence $x_n \in \mathbb{R}$ is a *Cauchy Sequence* if $\forall \varepsilon > 0, \; \exists N \; s.t. \; n, m \geq N \implies |x_n - x_m| < \varepsilon$.

Proposition 1.4.2 : Every convergent sequence is Cauchy.

Proof 1. Suppose $x_n \to x \in \mathbb{R}$. Given $\varepsilon < 0$. Consider

$$\begin{aligned} |x_n - x_m| &= |x_n - x + x - x_m| \\ &\leq |x_n - x| + |x - x_m| \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2}. \end{aligned}$$

Q.E.D. ■

> **Theorem 1.4.3 Cauchy Completeness**
> Every Cauchy sequence in $\mathbb{R}$ converges.

Remark 1.10 (Strategy of the Proof) *Cauchy Sequence* $\xRightarrow{\text{Lemma 1.4.4}}$ *Bounded Sequence* $\xRightarrow{\text{Theorem 1.4.5}}$ $\exists$ *convergent subsequence* $+$ *Cauchy sequence* $\xRightarrow{\text{Lemma 1.4.6}}$ *Sequence converges.*

Lemma 1.4.4 : Every Cauchy sequence is Bounded.

> **Theorem 1.4.5**
> Every bounded sequence in $\mathbb{R}$ has a subsequence that converges to some point in $\mathbb{R}$.

Proof 2. Let $\{x_n\}$ be a bounded sequence in $\mathbb{R}$. Fix M $s.t.$ $-M < x_n < M \quad \forall n$.

Divide $[-M, M]$ into subintervals $[-M, 0]$ and $[0, M]$. One of them, called I_0, must contain infinitely many terms of $\{x_n\}$. Choose n_0 $s.t.$ $x_{n_0} \in I_0$.

Divide I_0 into two equal subintervals. One of them, denoted I_1, contains infinitely many elements. Choose $n_1 > n_0$ $s.t.$ $x_{n_1} \in I_1$.

Continuing this process, we obtain subintervals $I_k = [a_k, b_k]$ for $k = 0, 1, \ldots,$ and includes n_k with the following properties:

- $I_0 \supset I_1 \supset I_2 \supset \cdots$
- $b_k - a_k = \dfrac{M}{2^k}$

- $x_{n_k} \in I_k$

[To prove $\{x_{n_k}\}$ converges, we prove $\{a_k\}$ and $\{b_k\}$ converge, and apply the Squeeze Theorem.]

- Show $\{a_k\}$ converges: a_k is monotone increasing and bounded. By MSP, $a_k \to a \in \mathbb{R}$.
- Show $\{b_k\}$ converges: Note that $b_k = a_k + \dfrac{M}{2^k}$. When $k \to 0$,

$$a_k + \frac{M}{2^k} = a + 0 = a.$$

So, $b_k \to a$ when $k \to \infty$.

Hence, as $a_k \le x_{n_k} \le b_k$, $a_k \to a$, $b_k \to a$, it must be that $x_{n_k} \to a$ as well.

Q.E.D. ■

Lemma 1.4.6 : If a subsequence of a Cauchy sequence converges to x, then the sequence itself converges to x.

Proof 3. Given $\{x_n\}$ is Cauchy and $x_{n_k} \to x$, [WTS: $x_n \to x$]. Consider

$$\begin{aligned} |x_n - x| &= |x_n - x_{n_k} + x_{n_k} - x| \\ &\le \underbrace{|x_n - x_{n_k}|}_{\text{Cauchy} \implies \text{small}} + \underbrace{|x_{n_k} - x|}_{\text{Convergent} \implies \text{small}} \end{aligned}$$

Q.E.D. ■

Summary I: Completeness on Ordered Field

Let $\mathcal{F}$ be an ordered field.

Definitions

- **Archimedean Property**: $\forall x \in \mathcal{F}$, $\exists$ integer N $s.t.$ $x < N$.
  (*Equivalently,* $\forall \varepsilon > 0, \quad \exists$ *integer* n $s.t.$ $0 < \dfrac{1}{n} < \varepsilon$).
- **Monotone Sequence Property (MSP)**: Every monotone increasing sequence bounded above converges.
- **Completeness**: We say $\mathcal{F}$ is complete if it has the MSP.
- **LUB Property**: Every set $S \neq \varnothing$ bounded above has a least upper bound.
- **Cauchy Property**: Every Cauchy sequence converges.

Facts in any ordered field

- MSP $\Longrightarrow$ Archimedean Property

  Remark 1.11 *In general, the converse is not true. For example,* $\mathbb{Q}$ *has the Archimedean property but not MSP.*

- MSP $\Longleftrightarrow$ LUB Property.
- MSP $\Longrightarrow$ Cauchy Property

  Remark 1.12 *The converse is true when Archimedean property is true.*

Facts in $\mathbb{R}$

- MSP $\Longleftrightarrow$ LUB Property $\Longleftrightarrow$ Cauchy Property

1.5 $\liminf$ and $\limsup$

Example 1.5.1 Cluster Points of a Sequence

Consider the sequence

$$a_n = (-1)^n + \frac{1}{n}.$$

Then, $a_1 = 0,\ a_2 = 1 + \frac{1}{2},\ a_3 = -1 + \frac{1}{3},\ a_4 = 1 + \frac{1}{4}, \cdots$. This sequence does not converge. However, its terms seem to "cluster" around 1 and -1.

Definition 1.5.2 (Cluster Points). A point x is called a *cluster point* of a sequence $\{x_n\}$ if $\forall\, \varepsilon > 0$, $\exists$ infinitely many values of n $s.t.$ $|x_n - x| < \varepsilon$.

Remark 1.13 *This definition is weaker than that of limits.*

Proposition 1.5.3 Relation Between Limits and Cluster Points: Suppose $x_n \in \mathbb{R}$ and $x \in \mathbb{R}$. Then,

1. x is a cluster point of $\{x_n\} \iff \forall\, \varepsilon > 0$ and $\forall$ integer N, $\exists\, n > N$ $s.t.$ $|x_n - x| < \varepsilon$.
2. x is a cluster point of $\{x_n\} \iff \exists$ subsequence $x_{n_k} \to x$.
3. $x_n \to x \iff$ every subsequence converges to x.
4. $x_n \to x \iff$ the sequence is bounded and x is the only cluster point.
5. $x_n \to x \iff$ every subsequence has a further sequence that converges to x.

Proof 1. (of some claims)

1. Follows from Definition.
2. ($\Rightarrow$) Assume x is a cluster point. [WTS: $\exists$ subsequence $x_{n_k} \to x$].

   Given $\varepsilon_1 = 1$ and $N = 1$, by (1), $\exists\, n_1 > 1$ $s.t.$ $|x_{n_1} - x| < \varepsilon = 1$.

   Given $\varepsilon_2 = \frac{1}{2}$ and $N = n_1$, by (1), $\exists\, n_2 > n_1$ $s.t.$ $|x_{n_2} - x| < \varepsilon = \frac{1}{2}$.

   So, in general, given $\varepsilon_k = \frac{1}{k}$ and $N = n_{k-1}$,

   $$\exists\, n_k > n_{k-1} = N_k\ s.t.\ |x_{n_k} - x| < \varepsilon_k = \frac{1}{k}.$$

   Then, $x_{n_k} \to x$ as $k \to \infty$.
3. ($\Leftarrow$) [Prove by contrapositive/contradiction] Assume every subsequence converges. For the sake of contradiction, assume x_n does not converge to x. Then we need to construct a subsequence x_{n_k} $s.t.$ $x_{n_k} \not\to x$.
4. ($\Leftarrow$) [Prove by contrapositive/contradiction]

5. ($\Leftarrow$) Use (4). Every subsequence has its own subsequence that converges to x. So, x is a cluster point of every subsequence. Then, we just need to show x is the only cluster point of $\{x_n\}$.

Q.E.D. ■

Definition 1.5.4 ($\liminf$ **and** $\limsup$)**.** Given a sequence $x_n \in \mathbb{R}$. For each integer $k \geq 1$, let

$$a_k = \inf \underbrace{\{x_{k+1}, x_{k+2}, \dots\}}_{\text{Set } S_k} \quad \text{and} \quad b_k = \sup \{x_{k+1}, x_{k+2}, \dots\} = \sup S_k.$$

Then,

$$\liminf x_n = \sup \{a_k\} \quad \text{and} \quad \limsup x_n = \inf \{b_k\}.$$

Remark 1.14

- $a_k \leq b_k$, a_k *is monotone increasing sequence, and* b_k *is monotone decreasing sequence. Thus,*

$$\liminf x_n = \lim_{k\to\infty} a_k \quad \textit{and} \quad \limsup x_n = \lim_{k\to\infty} b_k.$$

  Also, $\liminf x_n \leq \limsup x_n$.

- $\limsup x_n = +\infty \iff b_k = +\infty \quad \forall\, k \iff x_n$ *is not bounded above.*

  $\liminf x_n = -\infty \iff a_k = -\infty \quad \forall\, k \iff x_n$ *is not bounded below.*

Proposition 1.5.5 : $\limsup x_n = b \in \mathbb{R} \iff \forall\, \varepsilon > 0$,

1. $\exists\, N\ s.t.\ n \geq N \implies x_n < b + \varepsilon$, and
2. $\forall\, M,\ \exists\, n \geq M\ s.t.\ x_n > b - \varepsilon$.

Proof 2. (of forward direction) By definition, we know $\lim_{k\to\infty} b_k = b$, which implies $\forall \varepsilon > 0, \exists N\ s.t.\ k \geq N \implies |b_k - b| < \varepsilon$. That is, $-\varepsilon < b_k - b < \varepsilon$. As be is monotone decreasing, $b_k - b \geq 0$. So, $\boxed{0 \leq b_k - b < \varepsilon}$.

1. Note that $b_k = \sup \{x_{k+1}, x_{k+2}, \dots\}$. So, if $n > k$, $x_n \leq b_k < b + \varepsilon \quad \forall k \geq N$. Therefore,

$$n \geq N + 1 \implies x_n < b + \varepsilon.$$

2. We have $0 \leq b_k - b$, or $b_k \geq b \quad \forall\, k$. Given any integer M. [We need to find $n \geq M\ s.t.\ x_n > b - \varepsilon$] Then,

$$b_M = \sup \{x_{M+1}, x_{M+2}m, \dots\} \geq b.$$

   So, by definition of supremum, we can find $n > M\ s.t.\ x_n > b_M - \varepsilon \geq b - \varepsilon$.

Q.E.D. ■

Proposition 1.5.6 : $\limsup x_n = b \in \mathbb{R} \implies \exists$ subsequence $x_{n_k} \to b$.

Proof 3. We will construct a subsequence n_k inductively such that

$$b - \varepsilon_k < x_{n_k} < b + \varepsilon_k, \quad \varepsilon_k = \frac{1}{k}.$$

Given $\varepsilon = 1$, by Proposition 1.5.5(1), $\exists N_1$ $s.t.$ $n \geq N_1 \implies x_n < b + \varepsilon_1$. Further, by Proposition 1.5.5(2), for $M = N_1$, $\exists\, n_1 > N_1$ $s.t.$ $x_{n_1} > b - \varepsilon_1$. Therefore,

$$b - \varepsilon_1 < x_{n_1} < b + \varepsilon_1.$$

Claim *Given k_n, we can find n_{k+1} $s.t.$ $n_{k+1} > n_k$, and*

$$b - \frac{1}{k+1} < x_{n_{k+1}} < b + \frac{1}{k+1}.$$

After $\{x_{n_k}\}$ is constructed, use the sandwich lemma to prove $x_{n_k} \to b$.

Q.E.D. ■

Remark 1.15 *Similar arguments hold for* $\liminf x_n = a$.

Proposition 1.5.7 Relation Between Cluster Points and Limit: Let $x_n \in \mathbb{R}$ be a given sequence.

1. If x is a cluster point $\implies \liminf x_n \leq x \leq \limsup x_n$.
2. If $a = \liminf x_n$ is finite $\implies$ a is the smallest cluster point.
3. If $b = \limsup x_n$ is finite $\implies$ b is the largest cluster point.
4. $x_n \to x \in \mathbb{R} \iff \liminf x_n = \limsup x_n = x$.

Proof 4. (of (1)) Suppose x is a cluster point. Then, $\exists$ subsequence $x_{n_k} \to x$ as $k \to \infty$.
[WTS: $a_n \leq x \leq b_n \quad \forall n$]

For each n, $b_n = \sup\{x_{n+1}, x_{n+2}, \dots\} \geq x_{n_k}$ for large enough k. Let $k \to \infty$, we have $b_n \geq x$. Similarly, $a_n = \inf\{x_{n+1}, x_{n+2}, \dots\} \leq x_{n_k}$ for large enough k. As $k \to \infty$, $a_n \leq x$.

So, $a_n \leq x \leq b_n$. Take the limit as $n \to \infty$:

$$\lim_{n\to\infty} a_n \leq x \leq \lim_{n\to\infty} b_n \implies \liminf x_n \leq x \leq \limsup x_n.$$

Q.E.D. ■

1.6 Euclidean Space $\mathbb{R}^n$ and General Metric Space

Notation 1.1. $\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) \mid x_1, \dots, x_n \in \mathbb{R}\}$.

Remark 1.16 ($\mathbb{R}^n$ is a Vector Space) *We can write its standard bases as* $\{e_1, e_2, \ldots, e_n\}$*, and the general representation of* x *will be*

$$x = \sum_{j=1}^{n} x_j e_j.$$

Definition 1.6.2 (Norm and Metric). For $x, y \in \mathbb{R}^n$, define *norm* (or length) as

$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}$$

and the *metric* (distance) as

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

Definition 1.6.3 (Inner Product). We define the *inner product* (or dot product) as

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i.$$

Geometrically, if θ is the angle between x and y, then

$$\langle x, y \rangle = \|x\| \cdot \|y\| \cdot \cos\theta.$$

So, if $x \perp y$, $\langle x, y \rangle = 0$.

Proposition 1.6.4 Properties of Inner Product: Suppose $\langle \cdot, \cdot \rangle$ is an inner product, then

- Positive definite: $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0 \iff x = 0$.
- Linearity: $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ and $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$.
- Symmetry: $\langle x, y \rangle = \langle y, x \rangle$.

Proposition 1.6.5 Properties of Norm: Suppose $\|\cdot\|$ is a norm, then

- Positive definite: $\|x\| \geq 0$ and $\|x\| = 0 \iff x = 0$.
- Linearity: $\|\alpha x\| = |\alpha| \cdot \|x\|$.
- Triangle Inequality: $\|x + y\| \leq \|x\| + \|y\|$.

Proposition 1.6.6 Properties of Metric: Suppose $d(\cdot, \cdot)$ is a metric, then

- Positive definite: $d(x, y) \geq 0$ and $d(x, y) = 0 \iff x = y$.
- Symmetry: $d(x, y) = d(y, x)$.
- Triangle Inequality: $d(x, y) \leq d(x, z) + d(z, y)$.

Remark 1.17 *Inner product always induces a norm. Norm always induced a metric.*

Theorem 1.6.7 Cauchy-Schwarz Inequality

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|.$$

Example 1.6.8 Use Cauchy-Schwarz Inequality to Prove Triangle Inequality of Norms

Proof 1. Note that

$$\begin{aligned}
\|x+y\|^2 &= \langle x+y, x+y\rangle && \text{[Definition]} \\
&= \langle x+y, x\rangle + \langle x+y, y\rangle && \text{[Distribution]} \\
&= \langle x, x\rangle + \langle y, x\rangle + \langle x, y\rangle + \langle y, y\rangle && \text{[Dsitributionl]} \\
&= \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle && \text{[Symmetry]} \\
&\leq \|x\|^2 + \|y\|^2 + 2 \cdot \|x\| \cdot \|y\| && \text{[Cauchy-Schwarz]} \\
&= (\|x\| + \|y\|)^2.
\end{aligned}$$

Q.E.D. ■

Definition 1.6.9 (General Metric Space). A *metric space* (M, d) is a set M and a function $d : M \times m \to \mathbb{R}$ $s.t.$ $\forall x, y, z \in M$, the following conditions hold:

- Positive definite: $d(x, y) \geq 0$ and $d(x, y) = 0 \iff x = y$.
- Symmetry: $d(x, y) = d(y, x)$.
- Triangle Inequality: $d(x, y) \leq d(x, z) + d(z, y)$.

Definition 1.6.10 (General Normed Space). A *normed space* $(V, \|\cdot\|)$ is a vector space V together with a function $\|\cdot\| : V \to \mathbb{R}$ $s.t.$ $\forall x, y \in V$ and $\forall \alpha \in \mathbb{R}$,

- Positive definite: $\|x\| \geq 0$ and $\|x\| = 0 \iff x = 0$.
- Linearity: $\|\alpha x\| = |\alpha| \cdot \|x\|$
- Triangle Inequality: $\|x + y\| \leq \|x\| + \|y\|$

Definition 1.6.11 (General Inner Product Space). An *inner product space* $(V, \langle\cdot,\cdot\rangle)$ is a vector space V and a function $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ $s.t.$ $\forall x, y, z \in V$ and $\forall \alpha \in \mathbb{R}$:

- Positive definite: $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0 \iff x = 0$.
- Symmetry: $\langle x, y\rangle = \langle y, x\rangle$.
- Linearity: $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$ and $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$.

Example 1.6.12

- $\mathbb{R}^n$ is a metric space with $d(x, y) = \|x - y\|$.
- *Discrete Metric*: Given any set M, define

$$d(x, y) = \begin{cases} 0, & x = y \\ 1, & x \neq y. \end{cases}$$

- *Bounded Metric*: Given metric space (M, d), define $\rho : M \times M \to \mathbb{R}$:

$$\rho(x, y) = \frac{d(x, y)}{1 + d(x, y)}.$$

  Claim 1.6.13 (M, ρ) is also a metric space.

- $\mathbb{R}^2$ is a metric space under the taxicab metric $d_1 : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$:

$$d_1\big((x_1, y_1), (x_2, y_2)\big) = |x_2 - x_1| + |y_2 - y_1|.$$

- Let $\mathcal{C}([0, 1])$ be the collection of all continuous function $f : [0, 1] \to \mathbb{R}$. Define

$$\langle f, g\rangle = \int_0^1 f(x)g(x)\,\mathrm{d}x.$$

  Then, $\mathcal{C}$ is an inner product space.

Remark 1.18 (Relation Among Inner Product, Normed, and Metric Space)

$$\textit{Inner Product} \implies \textit{Norm} \implies \textit{Metric}$$

- *An inner product $\langle\cdot, \cdot\rangle$ induces a norm:*

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

- *A norm $\|\cdot\|$ always induces a metric:*

$$d(x, y) = \|x - y\|.$$

Theorem 1.6.14 General Cauchy-Schwarz Inequality
In an inner product space $(V, \langle\cdot,\cdot\rangle)$, we have $\forall\, v, w \in V$,

$$|\langle v, w\rangle| \leq \langle v, v\rangle^{\frac{1}{2}} \cdot \langle w, w\rangle^{\frac{1}{2}}.$$

Proof 2. If $v = 0$ or $w = 0$, it is trivial.
Assume $v \neq 0$ and $w \neq 0$. For any $t \in \mathbb{R}$, consider

$$\langle tv + w, tv + w\rangle$$

Then,

$$0 \leq \langle tv + w, tv + w\rangle = t^2 \underbrace{\langle v, v\rangle}_{a} + 2t \underbrace{\langle v, w\rangle}_{b} + \underbrace{\langle w, w\rangle}_{c}$$

Let $f(t) = at^2 + 2bt + c$ be a 2$^{\text{nd}}$ order polynomial of t. Note that $f(t) \geq 0 \quad \forall\, t \in \mathbb{R}$. On the other hand (OTOH), since $a = \langle v, v\rangle > 0$, $f(t)$ has minimum where $f'(t) = 0$.

$$\begin{aligned} f'(t) = 2at + 2b &= 0 \\ t &= -\frac{b}{a}. \end{aligned}$$

So, $f\left(-\frac{b}{a}\right) \geq 0$, or

$$\begin{aligned} \left(-\frac{b}{a}\right)^2 a + 2b\left(-\frac{b}{a}\right) + c &\geq 0 \\ \frac{b^2}{a} - 2\frac{b^2}{a} + c &\geq 0 \\ c &\geq \frac{b^2}{a} \\ b^2 &\leq ac \\ (\langle v, w\rangle)^2 &\leq \langle v, v\rangle \cdot \langle w, w\rangle \\ |\langle v, w\rangle| &\leq \langle v, v\rangle^{\frac{1}{2}} \cdot \langle w, w\rangle^{\frac{1}{2}}. \end{aligned}$$

Q.E.D. ■

2 Topology of Euclidean Space

2.1 Open Set

Definition 2.1.1 (Neighborhood & Open Set). Let (M,d) be a metric space. Fix $x \in M$ and $\varepsilon > 0$.

- *Neighborhood (nbdd)*:
$$D(x,\varepsilon) = \{y \in M \mid y(x,y) < \varepsilon\}.$$
It is also referred as ε*-nbdd*, ε*-disk*, or ε*-ball*.

- *Open Set*: A set $A \subset M$ is *open* if $\forall\, x \in A,\ \exists \varepsilon > 0\ s.t.\ D(x,\varepsilon) \subset A$.

Example 2.1.2 Open Set

- The unit disk $D = \left\{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 > 1\right\}$ is open in $\mathbb{R}^2$.
- The interval $(0,1) \subset \mathbb{R}^1$ is open.
- Given any metric space (M,d) and $x_0 \in M$. The disk
$$D(x_0,r) = \{x \in M \mid d(x,x_0) < r\}$$
is open $\forall r > 0$.

Proof 1. Fix $x \in D(x_0,r)$. [WTS: $\exists\, \varepsilon > 0\ s.t.\ D(x,\varepsilon) \subset D(x_0,r)$.]

Since $x \in D(x_0,r)$, by definition, $d(x,x_0) < r$. Hence, $\varepsilon = r - d(x,x_0) > 0$.

Claim 2.1.3 $D(x,\varepsilon) \subset D(x_0,r)$.

Proof. Let $y \in D(x,\varepsilon)$. Then,
$$\begin{aligned} d(y,x) &\le d(y,x_0) + d(x_0,x) \\ &< \varepsilon + d(x_0,x) \\ &= r - \cancel{d(x_0,x)} + \cancel{d(x_0,x)} \\ &= r. \end{aligned}$$
So, $d(y,x) < r$. By definition, $y \in D(x_0,r)$. $\square$

So, $D(x,\varepsilon) \subset D(x_0,r)$. By definition, $d(X_0,r)$ is open.

Q.E.D. ■

- The set $S = \left\{(x,y) \in \mathbb{R}^2 \mid xy > 1\right\}$ is open.

Proof 2. Givene $(x,y) \in S$. [WTS: $\exists\, \varepsilon > 0\ s.t.\ D\big((x,y),\varepsilon\big) \subset S$.]

Since $xy > 1$, $\lambda = \dfrac{1}{2}\left(1 - \dfrac{1}{xy}\right) > 0$.

WLOG, assume $x > 0$ and $y > 0$.

Let $\varepsilon = \min\{\lambda x, \lambda y\}$. Then, for $(u, v \in D\big((x,y), \varepsilon\big)$, we have

$$\begin{aligned} d\big((u,v),(x,y)\big) &< \varepsilon \\ \sqrt{(x-u)^2 + (y-v)^2} &< \varepsilon. \end{aligned}$$

So, $|x-u| < \varepsilon$ and $|y-v| < \varepsilon$. Then,

$$\begin{aligned} x\left|q - \frac{u}{x}\right| &< \varepsilon \\ \frac{u}{x} &> 1 - \frac{\varepsilon}{x} \geq 1 - \frac{\lambda x}{x} = 1 - \lambda. \end{aligned}$$

Similarly,

$$\frac{v}{y} > 1 - \lambda.$$

Then,

$$\begin{aligned} u \cdot v = \frac{u}{x} \cdot \frac{v}{y} \cdot (xy) &> (1-\lambda)^2 (xy) \\ &> (1-2\lambda)(xy) = 1. \end{aligned}$$

So, as $uv > 1$, $(u,v) \in S$. Hence, S is open.

Sketch. Given $xy > 0$; Want $uv > 1$. Note that

$$\begin{aligned} uv &= \underbrace{\frac{u}{x}}_{(1-\lambda)} \cdot \underbrace{\frac{v}{y}}_{(1-\lambda)} \cdot xy \\ &= (1-\lambda)^2 (xy) \\ &> (1 - 2\lambda + \lambda^2)(xy) \\ &> (1-2\lambda)(xy) \\ &\geq 1 \\ \implies 1 - 2\lambda &\geq \frac{1}{xy}. \end{aligned}$$

Q.E.D. ■

Remark 2.1

- *In the above definition, ε depends on the point x.*
- *The open set is defined $w.r.t.$ the underline metric space.*

Example 2.1.4
$A = (0, 1)$. Then, A is an open set as a subset in $\mathbb{R}^1$. However, A is not open as a subset in $\mathbb{R}^2$.

Proposition 2.1.5 Properties of Open Set: Let (M, d) be a metric space. Then,

- The intersection of a finite number of open sets is open.
- The union of any number of open sets is open.
- $\varnothing$ and M are open.

Proof 3. (of ①) Suppose $A = \bigcap_{j=1}^{n} A_j$. Fix $x \in A$. By definition, $x \in A_j \quad \forall\, j = 1, \ldots, n$. Then, we can find $\varepsilon_j > 0$ $s.t.$ $D(x, \varepsilon_j) \in A_j$. As A_j is open. Take $\varepsilon = \min\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$. We know

$$D(x, \varepsilon) \in A_j \quad \forall\, j = 1, \ldots, n.$$

Hence, $D(x, \varepsilon) \in \bigcap_{j=1}^{n} A_j$. So, A is open.

Q.E.D. ■

Remark 2.2 *The intersection of infinitely many number of open sets may not be open.*

Definition 2.1.6 (Interior Point). Let $A \subset M$. A point $x \in A$ is called an *interior point* of A if $\exists\, \varepsilon > 0$ $s.t.$ $D(x, \varepsilon) \subset A$. The *interior of* A is the collection of all interior points, denoted by $\operatorname{int}(A)$.

Example 2.1.7

- $A = \{x_0\} \subset \mathbb{R}^n$, $\operatorname{int}(A) = \varnothing$ as there is no nbdd around the point x_0.
- $A = (0, 1) \subset \mathbb{R}^1$, $\operatorname{int}(A) = A$.

  Remark 2.3 *A set is open if every point in A is an interior point of A.*

- $B = [0, 1] \subset \mathbb{R}^1$, $\operatorname{int}(B) = (0, 1)$.

Proposition 2.1.8 Properties of $\operatorname{int}(A)$**:**

- $\operatorname{int}(A)$ is open.
- $\operatorname{int}(A)$ is the union of all open subsets of A.

  Remark 2.4 *Or,* $\operatorname{int}(A)$ *is the largest open subset of A.*

- A is open $\iff A = \operatorname{int}(A)$.

2.2 Closed Sets

Definition 2.2.1 (Closed Set). A set $A \subset M$ is *closed* if its complement, $A^C = M \backslash A$, is open.

Example 2.2.2

- $A = [0, 1] \subset \mathbb{R}^1$.
  $A^C = (-\infty, 0) \cup (1, +\infty)$.
  A^C is open $\implies$ A is closed.
- $B = \{(x, y) \in \mathbb{R}^2 \mid 1 < x^2 + y^2 \leq 4\}$.
  $B^C = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1 \text{ or } x^2 + y^2 > 4\}$.
  B is not open and not closed.
- A single point set is closed.
- $B(x, \varepsilon) = \{y \in M \mid y(y, x) \leq \varepsilon\}$ is closed.

Proposition 2.2.3 Basic Properties of Closed Sets: Given (M, d), then

- Union of finite number of closed set is closed.
- Intersection of any number of closed set is closed.
- $\varnothing$ and M are always closed.

Remark 2.5 *In property ①, one cannot replace "finite number" by "countably many."*

Definition 2.2.4 (Accumulation Point). A point $x \in M$ is an *accumulation point* of the set A if $\forall \varepsilon > 0$, $\exists y \in A$ *s.t.* $y \neq x$ and $y \in D(x, \varepsilon)$. The collection of accumulation points of A is denoted as $\mathrm{ac}(A)$.

Remark 2.6 *x does not need to be in A.*

Definition 2.2.5 (Closure/$\mathrm{cl}(A)$).

$$\begin{aligned}\mathrm{cl}(A) &= \text{intersection of all closed sets containing } A \\ &= A \cup \mathrm{ac}(A).\end{aligned}$$

Definition 2.2.6 (Boundary of A/∂A/$\mathrm{bd}(A)$).

$$\begin{aligned}\mathrm{bd}(A) = \partial A &= \mathrm{cl}(A) \cap \mathrm{cl}(M \backslash A) \\ &= \mathrm{cl}(A) \backslash \mathrm{int}(A).\end{aligned}$$

Theorem 2.2.7 Equivalent Conditions of Closed Sets
Let $A \subset M$, the following are equivalent (TFAE):

- A is closed.
- $\mathrm{ac}(A) \subset A$.
- $A = \mathrm{cl}(A)$.
- $\mathrm{bd}(A) \subset A$.

Proof 1.
(① $\implies$ ②): Let $A \subset M$ be closed and $x \in \mathrm{ac}(A)$. [WTS: $x \in A$.] Assume $x \notin A$. Then, $x \in M \backslash A$. [Proof by contradiction.]

Since A is closed, $M \backslash A$ is open, which means $x \in \mathrm{int}(M \backslash A)$. That is, $\exists \varepsilon > 0$ $s.t.$ $D(x, \varepsilon) \subset M \backslash A$. Hence, $D(x, \varepsilon)\mathcal{A} = \varnothing$. ※ This contradicts with the assumption that $x \in \mathrm{ac}(A)$. As $D(x, \varepsilon) \cap A = \varnothing$, $\nexists\, y \in A$ $s.t.$ $y \in D(x, \varepsilon)$. Hence, $x \in A$. □

(② $\iff$ ③): We have $\mathrm{cl}(A) = A \cup \mathrm{ac}(A)$.

(⇒): If ② is true, $\mathrm{ac}(A) \subset A$. Then, $\mathrm{cl}(A) = A$.

(⇐): If ③ is true $\mathrm{cl}(A) = A$. Then, $A \cup \mathrm{ac}(A) = A$, so $\mathrm{ac}(A) \subset A$. □

(③ $\implies$ ④): Note that $\mathrm{bd}(A) = \mathrm{cl}(A) \cap \mathrm{cl}(M \backslash A)$. Then, $\mathrm{bd}(A) \subset \mathrm{cl}(A)$. If $A = \mathrm{cl}(A)$, then $\mathrm{bd}(A) \subset \mathrm{cl}(A) = A$. □

(④ $\implies$ ①): Suppose $\mathrm{bd}(A) \subset A$. Assume A is not closed, then $M \backslash A$ is not open. [Proof by contradiction.] So, $\exists\, x_0 \in M \backslash A$ that is not an interior point. Hence, $\forall\, \varepsilon > 0$, $D(x_0, \varepsilon) \not\subset M \backslash A$. So, $D(x_0, \varepsilon) \cap A \neq \varnothing$. Hence, $\exists\, y \in D(x_0, \varepsilon) \cap A$. Note that $x_0 \in M \backslash A$ but $y \in D(x_0, \varepsilon) \cap A$. So, $x_0 \neq y$. By definition, $x_0 \in \mathrm{ac}(A)$. ※ As $x_0 \in \mathrm{ac}(A) \subset \mathrm{bd}(A)$, but $x_0 \notin A$, this contradicts with the assumption that $\mathrm{bd}(A) \subset A$. Hence, A must be closed.

Q.E.D. ■

Proposition 2.2.8 :

- $\mathrm{cl}(A) \cap A = A$.
- If A is open, then $\mathrm{bd}(A) \subset M \backslash A$.

Definition 2.2.9 (Limit Point of a Set). A point $x \in M$ is called a limit point of A if $U \cap A \neq$ for every open set U containing x.

Proposition 2.2.10 :

- If $x \in \mathrm{ac}(A)$, then x is a limit point.
- If x is a limit point of A and $x \notin A$, then $x \in \mathrm{ac}(A)$.
- If x is a limit point of A, $\exists$ a sequence $x_n \in A$ with $x_n \to x$.
- A is closed $\iff$ A contains all of its limit points.

Summary II: Definitions on Point Set Topology

Let M be a metric space and $A \subset M$.

- $x \in A$ is an *interior point* of A if $\exists\, \varepsilon > 0$ with $D(x, \varepsilon) \subset A$.
- A is said to be *open* if every point of A is an interior point, or equivalently, $\operatorname{int}(A) = A$.
- A *neighborhood* of a point x is any open set U containing x.
- A is *closed* if its complement $M \backslash A$ is open.
- A point $x \in M$ is an *accumulation point* of A is $\forall\, \varepsilon > 0$, $\exists\, y \in A$ with $y \neq x$ and $y \in D(x, \varepsilon)$.
- *Closure* of A: $\operatorname{cl}(A) = A \cup \operatorname{ac}(A)$.
- *Boundary* of A: $\partial A = \operatorname{bd}(A) = \operatorname{cl}(A) \cap \operatorname{cl}(A \backslash M) = \operatorname{cl}(A) \backslash \operatorname{int}(A)$.

2.3 Convergence

Definition 2.3.1 (Convergence of a Sequence). Let (M, d) be a metric space. Let $x_k \in M$ be a sequence and $x \in M$. We say that x_k *converges* to x (write $x_k \to x$) if $\forall\, \varepsilon > 0,\ \exists\, N\ s.t.\ d(x_k, x) < \varepsilon \quad \forall k \geq N$.

Theorem 2.3.2 Equivalent Definitions of Convergence

- $x_k \to x \iff \forall$ open set U containing x, $\exists\, N\ s.t.\ x_k \in U \quad \forall\, k \geq N$.

  Remark 2.7 *This definition replaces ε- neighborhood by an arbitrary neighborhood.*

- $x_k \to x \iff d(x_k, x) \to 0$.

Theorem 2.3.3 Equivalent Definition of Convergence in $\mathbb{R}^n$

In $\mathbb{R}^n$, write

$$v_k = \left(v_k^{(1)}, v_k^{(2)}, \ldots, v_k^{(n)}\right) \quad \text{and} \quad v = \left(v^{(1)}, v^{(2)}, \ldots, v^{(n)}\right).$$

Then,

$$d(v_k, v)^2 = \|v_k - v\|^2 = \sum_{i=1}^{n} \left|v_k^{(i)} - v^{(i)}\right|^2.$$

Thus, $v_k \to v \iff v_k^{(1)} \to v^{(i)} \quad \forall\, i = 1, \ldots, n$

Proposition 2.3.4 : Let $v_k, w_k \in \mathbb{R}^n$ and $\lambda_k, \lambda \in \mathbb{R}$ with $v_k \to v, w_k \to w, \lambda_k \to \lambda$. Then,

- $v_k + w_k \to v + w$
- $\lambda v_k \to \lambda v$
- $\lambda_k v_k \to \lambda v$

Theorem 2.3.5 Convergence and Closedness

Let (M, d) be a metric space and $A \subset M$.

- A is closed $\iff$ for every sequence $x_k \in A$ that converges in M, the limit lies in A.
- $x \in \mathrm{cl}(A) \iff \exists x_k \in A\ s.t.\ x_k \to x$.

Proof 1. (of ①, sketch):

($\Rightarrow$) Assume $A \subset M$ is closed. Let $x_k \in A$ be a sequence with $x_k \to x \in M$. [WTS: $x \in A$.] Suppose $x \notin A$. Then, $x \in M \backslash A. A$ is closed $\implies M \backslash A$ is open $\implies \exists \varepsilon > 0$ with $D(x, \varepsilon) \subset M \backslash A$. As $x_k \to x$, some $x_k \in D(x, \varepsilon) \subset M \backslash A$. ※ This contradicts with our assumption that $x_k \in A$. So, $x \in A$. □

($\Leftarrow$): Suppose $x_k \in A$ with $x_k \to x \in A$. Assume $A \subset M$ is not closed. Then, $M \backslash A$ is not open $\implies \exists x \in M \backslash A\ s.t. \forall\, \varepsilon > 0,\ D(x, \varepsilon) \not\subset M \backslash A$. For $\varepsilon = \dfrac{1}{k}$, $\exists x_k \in D\left(x, \dfrac{1}{k}\right) \cap A$. Then, ※ $x_k \to x \notin A$, contradicting with the assumption $x_k \to x \in A$. Hence, A must be closed.

Q.E.D. ■

2.4 Completeness

Definition 2.4.1 (Cauchy Sequence). $\{x_k\} \in M$ is a *Cauchy sequence* if $\forall\, \varepsilon > 0$, $\exists N\ s.t.\ \forall m, n \geq N$, $d(x_n, x_m) < \varepsilon$.

Definition 2.4.2 (Bounded Sequence). A sequence $\{x_k\} \in M$ is *bounded* if $\exists\, x_0 \in M$ and $\exists\, R > 0\ s.t.$

$$d(x_0, x_k) \leq R \quad \forall\, k.$$

Or, $x_k \in B(x_0, R) \quad \forall k$, where $B(x_0, R)$ denotes a closed call centered at x_0 with radius R.

Definition 2.4.3 (Completeness). (M, d) is *complete* if every Cauchy sequence in M converges.

Example 2.4.4

- $\mathbb{R}^1$ and $\mathbb{R}^n$ are complete
- $M = \mathbb{R}^1 \backslash \{0\}$ is not complete. For example, $x_k = \dfrac{1}{k}$ does not converge in $\mathbb{R}^1 \backslash \{0\}$.
- $\mathbb{Q}$ is not complete.

Proposition 2.4.5 Basic Properties of Cauchy Sequence:

- Cauchy sequence is always bounded.
- Any converging sequence is always Cauchy.
- If a subsequence of a Cauchy sequence converges, then the original sequence converges.

Proof 1. (of ①): Suppose $\{x_k\}$ is Cauchy sequence. [WTS: $\exists\, x_0$ and $\exists\, R\ s.t.\ x_k \in B(x_0, R) \quad \forall\, k$.]
Then, fix $\varepsilon = 1$. By Cauchy sequence, $\exists\, N\ s.t.\ m, n \geq N \implies d(x_m, x_n) < \varepsilon = 1$. Define

$$\begin{aligned} R &= \max\{\varepsilon, d(x_N, x_1), d(x_N, x_2), \ldots, d(x_N, x_{N-1})\} \\ &= \max\{1, d(x_N, x_k) : k = 1, \ldots, N-1\} \end{aligned}$$

Then, we have $d(x_k, x_N) \leq R \quad \forall\, k$, which implies that Cauchy sequence is bounded.

Q.E.D. ■

Theorem 2.4.6 Closedness and Completeness
Let (M, d) be a metric space.

- $N \subset M$ is complete $\implies$ N is closed. *[Completeness is stronger than closedness]*
- $N \subset M$ is closed and M is complete $\implies$ N is complete.

Remark 2.8 *If (M, d) is a metric space and $N \subset M$, then (N, d) is also a metric space.*

Proof 2.

- (of ①): Suppose $N \subset M$ is complete. [WTS: every sequence $x_k \in N$ that converges, the limit is in N.]

  Given $\{x_k\} \in N$ with $x_k \to x \in M$. [WTS: $x \in N$.]

  Since $\{x_k\} \in M$ converges, it is Cauchy. Further, as $N \subset M$ is complete, by definition, $x_k \to x \in N$ as desired. □

- (of ②): Suppose $N \subset M$ is closed and M is complete. [WTS: Cauchy sequence $x_k \to x \in N$.]

  Given $x_k \in N$ is a Cauchy sequence. Then, $x_k \in M$ as $N \subset M$. Since M is complete, we know $x_k \to x \in M$. Further, as N is closed, we know $x_k \to x \in N$. Hence, every Cauchy sequence converges in N. By definition, N is complete.

Q.E.D. ■

Definition 2.4.7 (Cluster Point). x is a *cluster point* of $\{x_k\}$ if $\forall \varepsilon > 0$, $\exists$ infinitely many indices k $s.t.$ $d(x_k, x) < \varepsilon$.

Proposition 2.4.8 Properties of Cluster Points:

- x is a cluster point $\iff$ $\forall \varepsilon > 0, \forall N, \exists k > N$ $s.t.$ $d(x_k, x) < \varepsilon$.
- x is a cluster point $\iff$ $\exists$ subsequence $x_{n_k} \to x$.
- $x_k \to x \iff$ each subsequence $x_{n_k} \to x$.
- $x_k \to x \iff$ each subsequence has a further subsequence that converges to x.

3 Compactness and Connectedness

3.1 Compactness

Definition 3.1.1 (Cover and Subcover). Let $A \subset M$.

- A *cover* of a set $A \subset M$ is a collection $\{U_i\}$ of sets $U_i \subset M$ such that

$$\bigcup_i U_i \supset A.$$

- We say $\{U_i\}$ of A is an *open cover* if each U_i is open.
- A *subcover* of a given cover is a subcollection of $\{U_i\}$ whose union contains A.
- We say a cover is a *finite cover* if the subcollection contains finite number of sets.

Example 3.1.2

Suppose $A = [0,1] \subset \mathbb{R}^1$. Consider

$$U_1 = (-1, 0.1), \quad U_2 = (0, 0.5), \quad U_3 = (0.5, 1).$$

$$U_4 = (0.2, 0.6), \quad U_5 = (0.8, 2), \quad U_6 = (0, 1).$$

Then,

- $\{U_1, \dots, U_6\}$ is a finite cover of A.
- It is also an open cover.
- $\{U_1, U_5, U_6\}$ is a subcover.

Definition 3.1.3 (Compactness). A set $A \subset M$ is called *compact* if every open cover of A has a finite subcover.

Definition 3.1.4 (Sequencially Compact). A set $A \subset M$ is *sequencially compact* if every sequence in A has a subsequence that converges to a point in A.

Definition 3.1.5 (Totally Bounded). A set $A \subset M$ is *totally bounded* if $\forall\, \varepsilon > 0$, $\exists$ finite set $\{x_1, x_2, \dots, x_N\} \subset M$ $s.t.$

$$A \subset \bigcup_{i=1}^{N} D(x_i, \varepsilon).$$

Remark 3.1

- *A is sequencially compact $\implies$ A is closed and bounded.*

  Proof 1. *Suppose A is unbounded. Fix $x_0 \in M$. For any $n \geq 1$, $\exists x_n \in A$ $s.t.$*

$$d(x_n, x_0) \geq n.$$

By sequential compactness, $\exists$ subsequence $x_{n_k} \to x \in A$ such that

$$\begin{aligned} d(x_{n_k}, x_0) &\leq d(x_{n_k}, x) + d(x, x_0) \\ &< \varepsilon + d(x, x_0). \end{aligned}$$

Take $\varepsilon = 1$, $d(x_{n_k}, x_0) < 1 + d(x, x_0)$ is a finite number. However, $d(x_{n_k}, x_0) \geq n_k$. ※ As $n_k \to \infty$, $1 + d(x, x_0)$ is a finite number, we reach a contradiction. Hence, A must be bounded.

Q.E.D. ■

- *A is totally bounded $\implies$ A is bounded.*

Theorem 3.1.6 Bolzano-Weirstrass Theorem (B-W Thm.)
$A \subset M$ is compact $\iff$ A is sequentially compact.

Proof 2.

Lemma 3.1.7 : $A \subset M$ is compact $\implies$ A is closed.

Proof. [WTS: $M \backslash A$ is open.]

Fix $x \in M \backslash A$. For $n = 1, 2, \ldots$, let $U_n = \left\{ y \mid d(x, y) > \dfrac{1}{n} \right\}$.

Claim *$\{U_n \mid n = 1, 2, \ldots\}$ is an open cover of A.*

Proof. In fact, let $a \in A$. Then, $d(a, x) > 0$. By Archimedean, $\exists\, n\ s.t.$

$$\frac{1}{n} < d(a, x).$$

This implies that $a \in U_n$. So, $a \in \bigcup_{i=1}^{\infty} U_i$. That is, $A \subset \bigcup_{i=1}^{\infty} U_i$. □

By the compactness, $\exists$ finite subcover, say $\{U_1, \ldots, U_N\}$. Thus,

$$A \subset \bigcup_{i=1}^{N} U_i = U_N = \left\{ y \mid d(y, x) > \frac{1}{N} \right\}.$$

Therefore,

$$D\left(x, \frac{1}{N}\right) = \left\{ y \mid d(y, x) < \frac{1}{N} \right\} \subset M \backslash A.$$

Hence, by definition, $M \backslash A$ is open, and so A must be closed. □

Lemma 3.1.8 (When is the converse of Lemma 3.1.7 true?):

$B \subset M$ is closed and M is compact $\implies$ B is compact.

Proof. Given an open cover $\{V_i \mid i \in I\}$ of B. [WTS: $\exists$ a finite subcover of B.]

Since B is closed, $M \backslash B$ is open. Then,

$$\{V_i \mid i \in I\} \cup \{M \backslash B\} \quad \text{is an open cover of } M.$$

Since M is compact, $\exists$ a finite subcover of M:

$$\{V_1, V_2, \ldots, V_N\} \cup \{M \backslash B\}.$$

Note that

$$\bigcup_{i=1}^{N} V_i \supset B,$$

we know

$$\{V_1, V_2, \ldots, V_N\} \quad \text{is a finite subcover of } B.$$

Hence, by definition, B is compact. $\square$

($\Rightarrow$): Now, we prove the forward direction of the B-W Theorem. Let $A \subset M$ be compact. [WTS: A is sequentially compact]

- Set Up: Given a sequence $\{x_k\} \in A$. [WTS: $\exists x_{n_k} \to x \in A$]

  By Lemma 3.1.7, compactness $\implies$ closedness. Since A is closed, all converging sequence converges to some point in A. Hence, we only need to show $\exists$ converging subsequence.

- Reduction: To this end, we may assume that $\{x_k\}$ contains a subsequence of distinct terms. Denote this subsequence by $\{y_k\}$. [WTS: $\{y_k\}$ has a convergent subsequence]

  If $\{x_k\}$ does not contain subsequence of distinct terms, then $\{x_k\}$ is a constant sequence after sufficient terms. Therefore, it must converge and is trivial in this discussion.

- Suppose, for the sake of contradiction, $\{y_k\}$ does not have a convergent subsequence.

- **Claim** *y_k's are "isolated:" For each $k = 1, 2, \ldots$, $\exists$ neighborhood U_k of y_k s.t. $y_j \notin U_k$ for any $j \neq k$.*

  Proof. Suppose, for the sake of contradiction, that the claim does not hold. Then, $\exists\, k$ with the property $\forall\, \varepsilon > 0$, $\exists\, j \neq k$ $s.t.$ $y_j \in U_k = D(y_k, k)$. Take $\varepsilon = \dfrac{1}{m}$. We obtain subsequence $y_{j_m} \in D\left(y_k, \dfrac{1}{m}\right), \quad m = 1, 2, \ldots$. Hence, when $m \to \infty$, $y_{j_m} \to y_k$.

  This implies $\{y_k\}$ has a convergent subsequence. ※ This contradicts with our assumption that $\{y_k\}$ does not have a convergent subsequence. Hence, the claim must be true. $\square$

- Now, proceed with the assumption that this claim is true.

  Consider the set formed by elements in $\{y_k\}$:

  $$B = \{y_1, y_2, \ldots\}$$

  Since $\{y_n\}$ has no convergent subsequence, B has no accumulation point, and so $\mathrm{cl}(B) = B$, which implies B is closed.

  By Lemma 3.1.8, B is compact.

On the other hand, $\{U_k\}$ is an open cover of B. But by claim, $\exists$ no finite subcover. ⨳ This contradicts with the fact that B is compact. Thus, $\{y_k\}$ has a convergent subsequence, which converges to a point because A is closed.

□

(⇐): Now, let's consider the backward direction. Suppose $A \subset M$ is sequentially compact. [WTS: A is compact]

Let $\{u_i\}$ be an open cover of A. [WTS: $\exists$ a finite subcover]

Claim (1) $\exists\, r > 0$ *s.t. for each* $y \in A,\ D(y,r) \subset U_i$ *for some* i. $\implies$ *Each point has a neighborhood of fixed size that is contained in some* U_i.

Proof. Suppose otherwise. Then,

$$\forall\, r = \frac{1}{n} > 0,\ \exists\, y_n \in A\ s.t.\ D\left(y_n, \frac{1}{n}\right) \text{ is not contained in any } U_i.$$

By assumption, A is sequentially compact. Then, $\{y_n\}$ has a convergent subsequence $z_n \to z \in A$.

On the other hand, U_i is an open cover of A, then $z_n \in U_{i_0}$ for some i_0. Further, since U_{i_0} is open, $\exists\, \varepsilon > 0\ s.t.\ D(z, \varepsilon) \subset U_{i_0}$.

Fix large N $s.t.$

$$d(z_N, z) < \frac{\varepsilon}{2}.$$

So,

$$D\left(z, \frac{\varepsilon}{2}\right) \subset D(z, \varepsilon) \subset U_{i_0}.$$

⨳ This is a contradiction with our assumption that $D\left(y_n, \frac{1}{n}\right)$ is not contained in any U_i. Hence, the original claim is true. □

Claim (2) *A is totally bounded.*

Proof. Suppose otherwise. Then, $\exists\, \varepsilon > 0\ s.t.\ A$ cannot be covered by finite number of balls of radius ε. Choose $y_1 \in A$ and $y_2 \in A \backslash D(y_1, \varepsilon)$. Then, choose $y_3 \in A \backslash \big(D(y_1, \varepsilon) \cup D(y_2, \varepsilon)\big)$. This process can go forever as A cannot be covered by finite number of balls of radius ε. So, we get sequence

$$y_n \in A \backslash \big(D(y_1, \varepsilon) \cup \cdots \cup D(y_{n-1}, \varepsilon)\big).$$

We have a sequence $\{y_n\}$ with the property that

$$d(x_n, x_m) > \varepsilon \quad \forall\, n \neq m.$$

So, $\{y_n\}$ does not have a convergent subsequence.

Everything convergent must be Cauchy. $d(x_n, x_m) > \varepsilon$ implies not Cauchy, so it must be non-convergent. ⨳ This contradicts with the assumption that A is sequentially compact (has a subsequence converges to some point in A). Hence, this claim must be true. □

Now, let $r > 0$ be as in Claim (1). By Claim (2), $\exists\, y_1, y_2, \ldots, y_N \in A\ s.t.$

$$A \subset \bigcup_{j=1}^{N} D(y_j, r).$$

Then, further by Claim (1), we get $D(y_j, r) \subset U_{i_j}$. So,

$$A \subset \bigcup_{j=1}^{N} D(y_j, r) \subset \bigcup_{j=1}^{N} U_{i_j}.$$

Therefore, A can be covered by a finite subcover. Hence, A is compact.

Q.E.D. ■

Theorem 3.1.9
$A \subset M$ is compact $\iff$ A is complete and totally bounded.

Remark 3.2 *So, if a set is not bounded/totally bounded, it cannot be compact.*

Proof 3. ($\Rightarrow$): Done when proving B-W Thm. □

($\Leftarrow$): Assume A is complete and totally bounded. [WTS: A is compact/sequentially compact]

Let $\{y_n\}$ be a sequence in A. [WTS: $\exists$ subsequence y_{n_k} converges in A]

WLOG, we may assume $\{y_n\}$ is formed by distinct terms. If we don't get distinct terms, we will have a constant sequence when n gets sufficiently large. Hence, it converges in A and is trivial to discuss.

Since A is totally bounded, for $\varepsilon_1 = 1$, A is covered by finite number of balls:

$$D\left(x_1^{(1)}, \varepsilon_1\right), \ldots, D\left(x_{L_1}^{(1)}, \varepsilon_1\right).$$

We can choose a subsequence $\{y_{1n}\}_{n=1}^{\infty}$ of $\{y_n\}$ that is contained one of the balls.

Repeat that for $\varepsilon_2 = \dfrac{1}{2}$, we have

$$A \subset D\left(x_1^{(2)}, \varepsilon_2\right) \cup \cdots \cup \left(x_{L_2}^{(2)}, \varepsilon_2\right).$$

We can choose a subsequence $\{y_{2n}\}_{n=1}^{\infty}$ of $\{y_n\}$ that is contained in one of the balls.

Continuing this process with $\varepsilon_m = \dfrac{1}{m}$, $m = 1, 2, \ldots$. We obtain a subsequence $\{y_{m_n}\}_{n=1}^{\infty}$ that is contained in a ball of radius $\varepsilon_m = \dfrac{1}{m}$. Then, we have the following subsequence:

$$\begin{array}{cccccc}
y_{11}, & y_{12}, & y_{13}, & \cdots, & y_{1n}, & \cdots \\
y_{21}, & y_{22}, & y_{23}, & \cdots, & y_{2n}, & \cdots \\
\vdots & & & & & \\
y_{m1}, & y_{m2}, & y_{m3}, & \cdots, & y_{mn}, & \cdots \\
\vdots & & & & &
\end{array}$$

Each subsequence is a subsubsequence of the proceeding subsequence.
Select $y_{11}, y_{22}, y_{33}, \ldots, y_{nn}, \ldots$ to form a subsequence of $\{y_n\}$.
Denote this subsequence as $\{z_n\} = \{y_{nn}\}$.
A is complete. To show z_n converge in A, we only need to show z_n is Cauchy.

Claim $\{z_n\}$ *is Cauchy.*

Proof. Assume $n > m$:

$$d(z_n, z_m) < \frac{2}{m}.$$

When $m \to \infty$, $d(z_n, z_m) \to 0$. So, $\{z_n\}$ is Cauchy. □

Since A is complete, $\{z_n\}$ is Cauchy, we have $z_n \to z \in A$. Hence, A is sequentially compact. By B-W Theorem, A is compact.

Q.E.D. ■

3.2 Compactness in $\mathbb{R}^n$

Theorem 3.2.1 Heine-Borel Theorem
A set $A \subset \mathbb{R}^n$ is compact $\iff$ A is bounded and closed.

Proof 1. ($\Rightarrow$): True in general metric space. □
($\Leftarrow$): Assume $A \subset \mathbb{R}^n$ is closed and bounded. [WTS: A is sequentially compact]
Given sequence $\{x_k\}$ in A, write

$$x_k = \left(x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(n)}\right) \in A \subset \mathbb{R}^n.$$

A is bounded $\implies$ $\{x_k\}$ is bounded $\implies$ $\left\{x_k^{(1)}\right\}$ is bounded in $\mathbb{R}$.
$\implies$ $\exists$ converging subsequence $\left\{x_{f_1(k)}^{(1)}\right\}_{k=1}^{\infty}$.
Similarly, $\left\{x_{f_1(k)}^{(2)}\right\}_{k=1}^{\infty}$ is bounded in $\mathbb{R}$. $\implies$ $\exists$ converging subsequence $\left\{x_{f_2(k)}^{(2)}\right\}_{k=1}^{\infty}$.
In this way, we obtain subsequence

$$x_{f_n(k)} = \left(x_{f_n(k)}^{(1)}, x_{f_n(k)}^{(2)}, \ldots, x_{f_n(k)}^{(n)}\right)$$

with $x_{f_n(k)}^{(i)} \xrightarrow{k \to \infty} x^{(i)}$ for $i = 1, 2, \ldots, n$. Hence,

$$x_{f_n(k)} \to \left(x^{(1)}, x^{(2)}, \ldots, x^{(n)}\right) \in A.$$

Therefore, A is sequentially compact.

Q.E.D. ■

Remark 3.3 *In Heine-Borel Theorem, ($\Leftarrow$) does not hold in general metric space. That is, A metric space that is closed and bounded does not imply compactness. For example, let M = infinite set with discrete* *metric*

$$d(x,y) = \begin{cases} 0, & x = y \\ 1, & x \neq y. \end{cases}$$

M is closed and bounded, but M is not compact.

Example 3.2.2

- $A \subset \mathbb{R}^n$ is bounded $\implies \mathrm{cl}(A)$ is compact.
- $A = [0,1] \subset \mathbb{R}^1$ is compact.
- $A = (0,1] \subset \mathbb{R}$ is not compact.
- $\mathbb{R}$ is not compact because it is not totally bounded.

3.3 Nested Set Property

Theorem 3.3.1 Nested Set Property
Let F_k be a set of non-empty compact sets in M $s.t.$

$$F_{k+1} \subset F_k \quad \forall k = 1, 2, \ldots.$$

Then,

$$\bigcap_{k=1}^{\infty} F_k \neq \varnothing.$$

Proof 1. For each $k = 1, 2, \ldots$, choose $x_k \in F_k$. Then, $\{x_k\} \subset F_1$. Since F_1 is compact, $\exists$ subsequence

$$x_{f(k)} \xrightarrow{k\to\infty} x \in F_1.$$

Claim $x \in F_n \quad \forall n$.

Proof. Fix $n > 1$. Then, for large k ($\exists N$ $s.t.$ $k \geq N$), we have $f(k) \geq n$. Then, $F_{f(k)} \subset F_n$. Recall that $x_{f(k)} \in F_{f(k)}$ and $x_{f(k)} \xrightarrow{k\to\infty} x$, then

$$x \in F_n$$

as F_n is closed. □

Hence, $x \in \bigcap_{k=1}^{\infty} F_k \neq \varnothing$.

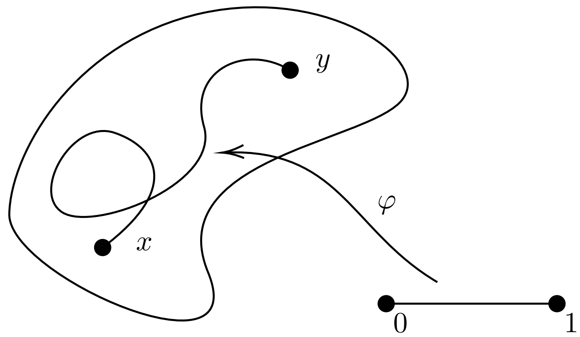
Q.E.D. ■

Remark 3.4 *"Compact" cannot be replaced by "open," "closed," or "bounded open."*

3.4 Connectedness

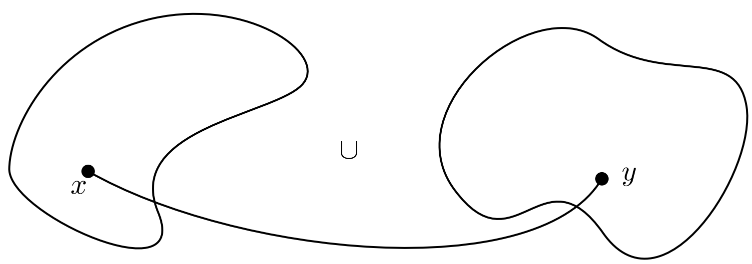
Definition 3.4.1 (Path-Connected, Geometric Point of View). A set $A \subset M$ is *path-connected* if each pair of points $x, y \in A$ can be joined by a continuous path given by a continuous map

$$\varphi : [0,1] \to A \quad s.t. \quad \varphi(0) = x \quad \text{and} \quad \varphi(1) = y$$



Example 3.4.2

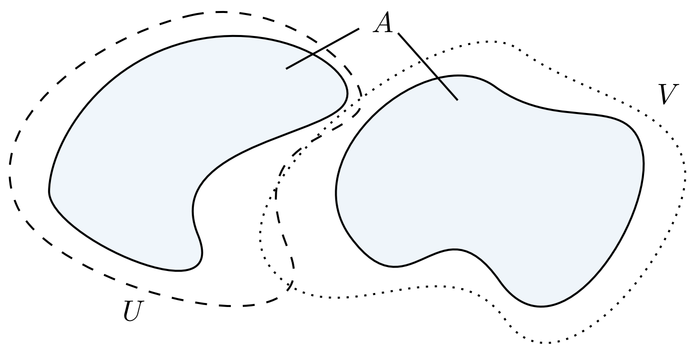
- This is not path-connected:



- Let $\varphi : [0,1] \to M$ be a continuous map. Then, $C = \varphi([0,1])] \subset M$ (the image of this maaping) is path-connected.

Definition 3.4.3 (Disconnected Set, Topological Point of View). A set $A \subset M$ is said to be *disconnected* if $\exists$ open sets $U, V \subset M$ that separate A:

- $U \cap V \cap A = \varnothing$
- $U \cap A \neq \varnothing$ and $V \cap A \neq = \varnothing$
- $A \subset U \cup V$



Definition 3.4.4 (Connected Set). If a set is not disconnected, then it is *connected*.

Remark 3.5 *It is easy to prove disconnectedness since we only need to find one pair of open sets satisfying the* 4 *conditions. To prove connectedness, we need to show* $\forall$ *open sets* $U, V \subset M$*, they cannot satisfy the* 4 *conditions at the same time.*

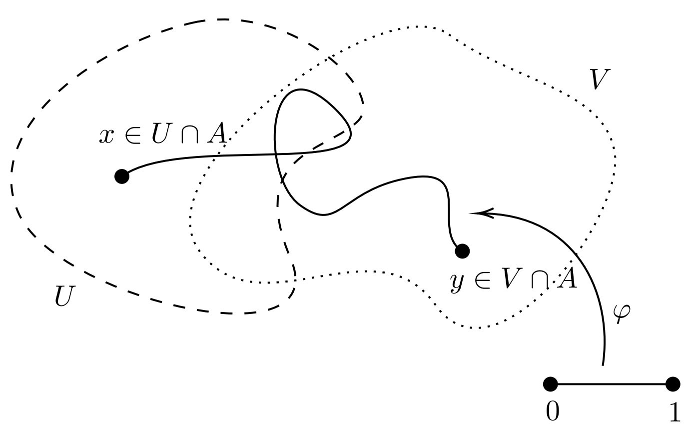
> **Theorem 3.4.5**
> Path-connectedness $\implies$ connectedness

Proof 1. We will start the proof with the following claim (The proof is trivial, and so we omit the proof):

Claim 3.4.6 The interval $[a, b] \subset \mathbb{R}^1$ is connected.

Suppose, for the sake of contradiction, that $A \subset M$ is path-connected but not connected. Then, $\exists$ open sets U, V that separates A as defined in Definition 3.4.3.

Fix $x \in U \cap A$ and $y \in V \cap A$.



Since A is path-connected, $\exists$ a continuous map $\varphi : [0, 1] \to A$ with $\varphi(0) = x$ and $\varphi(1) = y$. Let

$$\begin{aligned} C &= \varphi^{-1}(A \cap U) \subset [0, 1] \\ &:= \{t \in [0, 1] \mid \varphi(t) \in A \cap U\}. \end{aligned}$$

Similarly, we can define $D = \varphi^{-1}(A \cap V)$. Then, $0 \in C$ and $1 \in D$.

Claim 3.4.7 C is closed.

Proof. Let $t_k \in C$ $s.t.$ $t_k \to t$. Then, by continuity of φ, $\varphi(t_k) \to \varphi(t) \in A$. Suppose, for the sake of contradiction, $\varphi(t) \notin U$. Then, $\varphi(t) \in V$. Since V is open, $\varphi(t_k) \in V$ for large k. Hence,

$$\varphi(t_k) \in A \cap U \cap V = \varnothing.$$

※ We reach a contradiction. So, $\varphi(t) \in U$, which implies $t \in C$. As $t_k \to t \in C$, we have shown that C is closed. □

Corollary 3.4.8 : By symmetry of C and D, D is also closed.

To derive a contradiction with Claim 3.4.6, note that

$$A \cap U \cap V = \varnothing,$$

which implies $C \cap D = \varnothing$. Therefore, the two open sets $(\mathbb{R} \backslash C)$ and $(\mathbb{R} \backslash D)$ separates $[0,1]$.※ This contradicts with Claim 3.4.6 that $[0,1]$ is connected. Hence, our assumption was wrong, and A must be path-connected and connected. In other words, path-connectedness $\implies$ connectedness.

Q.E.D. ■

Remark 3.6 *The converse is not true.*

> ***Example 3.4.9***
>
> *Suppose* $A = \underbrace{\left\{\left(x, \sin\frac{1}{x}\right) \mid x > 0\right\}}_{\text{graph of } f(x) = \sin\left(\frac{1}{x}\right)} \cup \underbrace{\{(0,y) \mid -1 \le y \le 1\}}_{\text{segment of } y\text{-axis}} \subset \mathbb{R}^2.$
>
> *Then, A is connected but not path-connected.*

Proposition 3.4.10 : $A \subset \mathbb{R}^n$ open and connected $\implies$ path-connected.

Proof 2. (Sketch) Fix a point $x_0 \in A$ $s.t.$

$$B = \{y \in A \mid x_0 \text{ and } y \text{ can be joined by a continuous path } \in A\}.$$

Show:

- $B \neq \varnothing$ $[x_0 \in B]$
- B is open.
- B is closed in A.

Then, $B = A$. [If $B \neq A$, then $U = B$ and $V = A \backslash B$ separates A $\implies$ A is disconnected $\implies$ contradiction, it must be $A = B$.]

Q.E.D. ■

Theorem 3.4.11 Equivalent Ways to Describe Connectedness

- In Definition 3.4.3, one can replace "open sets" by "closed sets."

  $A \subset M$ is disconnected $\iff \exists$ closed sets E, F $s.t.$

  - $E \cap F \cap A = \varnothing$
  - $E \cap A \neq \varnothing$ and $F \cap A \neq \varnothing$
  - $A \subset E \cup F$

  [Take complement of open sets, we get closed sets]

- In Definition 3.4.3, one can replace "U, V" by "disjoint open sets."

  $A \subset M$ is disconnected $\iff \exists$ disjoint open sets U_1 and V_1 $s.t.$

  - $U_1 \cap V_1 \cap A = \varnothing$
  - $U_1 \cap A \neq \varnothing$ and $V_1 \cap A \neq \varnothing$
  - $A \subset U_1 \cup V_1$

Proof 3. (Hint of ②): Consider the distance function $d(x, A \cap V)$ given fixed $x \in U \cap A$.

Claim $\forall\, x \in A \cap U$, *define* $d(x) = d(x, A \cap V) = \inf\{d(x,a) \mid a \in A \cap V\}$. *Then,* $d(x) > 0$. *Similarly,* $\forall\, y \in A \cap V$, *define* $d(x) = d(y, A \cap U) = \inf\{d(y,a) \mid a \in A \cap U\}$. *Then,* $d(y) > 0$.

Define open sets U_1, V_1 as follows:

$$U_1 = \left\{D\left(x, \frac{1}{2}d(x)\right) \mid x \in A \cap U\right\} \quad \text{and} \quad V_1 = \left\{D\left(y, \frac{1}{2}d(y)\right) \mid y \in A \cap V\right\}$$

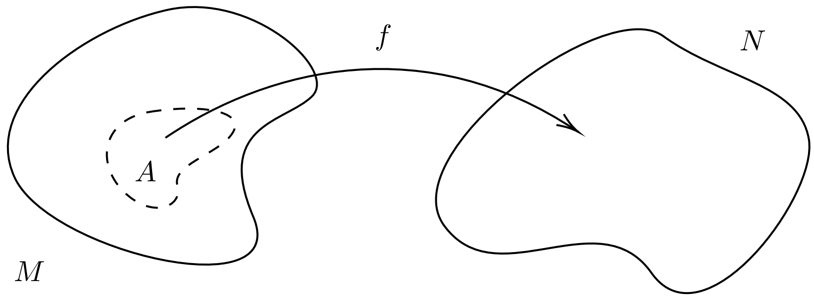
We have the desired disjoint U_1 and V_1.

Q.E.D. ■

4 Continuous Mappings

4.1 Continuity

Definition 4.1.1 (Maps). Suppose (M,d) and (N,ρ) are metric spaces. Let $A \subset M$. Then, $f : A \to N$ is a *map* (or a function)



Definition 4.1.2 (Continuous Maps). f is *continuous at a point* $x_0 \in A$ if

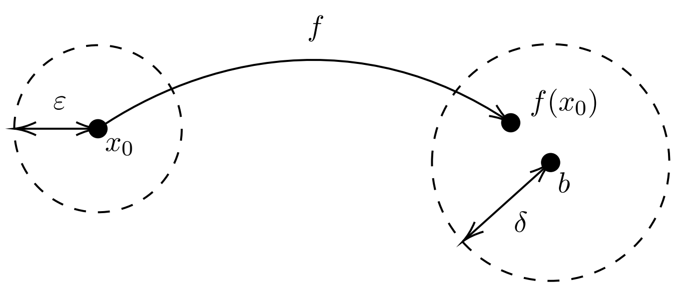
$$\lim_{\substack{x \to x_0 \\ x \in A}} f(x) = f(x_0).$$

f is *continuous* in A if it is continuous at each point in A.

Definition 4.1.3 (Limit of a Function). $b \in N$ is the limit of $f(x)$ at x_0, written as

$$\lim_{x \to x_0} f(x) = b,$$

if $\forall\, \varepsilon > 0$, $\exists\, \delta > 0\ s.t.\ x \in A$ and $d(x,x_0) < \delta \implies \rho(f(x),b) < \varepsilon$.



Definition 4.1.4 (Isolated Points). $x_0 \in A$ is an *isolated point* in A if $\exists\, \delta > 0\ s.t.\ D(x_0,\delta) \cap A = \{x_0\}$.

Remark 4.1

- *The continuous definition implies three things: the function is defined, the limit exists, and the limit value equals the function value.*
- *A point is either an isolated point or an accumulation point.*
- *For the limit definition, x_0 is not required to be in A. For example,*

$$f(x) = \frac{\sin(x)}{x},\ x \in (0,1)\ \lim_{x \to 0} f(x) = 0 \notin (0,1).$$

- *If x_0 is an* isolated point *in A, then $\lim_{x \to x_0} f(x) = f(x_0)$ is always true. Therefore, any function $f(x)$ is continuous at isolated points.*

Example 4.1.5

- $f(x) = x : \mathbb{R}^n \to \mathbb{R}^n$ (identity function) is continuous
- $g(x) = \begin{cases} x, & 0 \le x \le 1 \\ 2x, & 1 < x \le 3 \end{cases} : [0,3] \to \mathbb{R}^1$ is continuous at every point except for $x = 1$.
- $h(x) = \begin{cases} x, & x \neq 1 \\ 3, & x = 1 \end{cases} : \mathbb{R} \to \mathbb{R}$ is continuous at every point except $x = 1$.

Theorem 4.1.6 Equivalent Conditions for Continuity
Let $f : A \subset M \to N$. The following are equivalent:

- f is continuous on A.
- For each converging sequence $x_k \to x \in A$, $f(x_k) \to f(x)$.

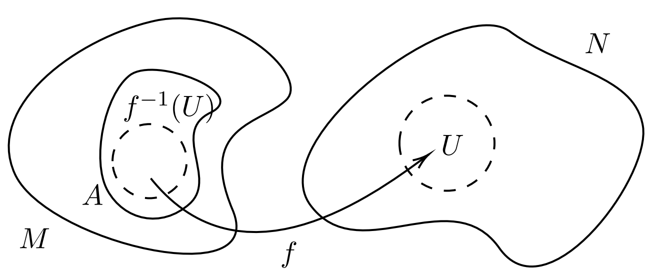
  Remark 4.2 *Continuous map preserves the convergence of sequences*

- For each open set $U \subset N$, the pre-image $f^{-1}(U) \subset A$ is open relative to A. That is

$$f^{-1}(U) = \{x \in A \mid f(x) \in U\} = A \cap V, \quad \text{where } V \subset M \text{ is open.}$$

- For each close set $F \subset N$, the pre-image $f^{-1}(F) \subset A$ is closed relative to A. That is,

$$f^{-1}(F) = A \cap E, \quad \text{where } E \subset M \text{ is closed.}$$



Proof 1. We will prove equivalence by the following cycle: ① $\Longrightarrow$ ② $\Longrightarrow$ ④ $\Longrightarrow$ ③ $\Longrightarrow$ ①.
(① $\Longrightarrow$ ②): Given sequence $x_k \in A$ with $x_k \to x \in A$. [WTS: $\lim_{k \to \infty} f(x_k) = f(x)$]

(② $\implies$ ④): Fix closed set $F \in N$. [WTS: $f^{-1}(F) = A \cap \mathrm{cl}\left(f^{-1}(F)\right)$] It is trivial that $f^{-1}(F) \subset A \cap \mathrm{cl}\left(f^{-1}(F)\right)$. So, we only need to prove the "$\supset$" direction. Given $x \in A \cap \mathrm{cl}\left(f^{-1}(F)\right)$, $\exists$ sequence $x_n \in f^{-1}(F) \subset A$ $s.t.$ $x_n \to x$. Then, $y_n = f(x_n) \to f(x) \in F$ by ② and closedness. So, $x \in f^{-1}(F)$. That is, $A \cap \mathrm{cl}\left(f^{-1}(F)\right) \supset f^{-1}(F)$. Hence, $f^{-1}(F) = A \cap \mathrm{cl}\left(f^{-1}(F)\right)$, implying $f^{-1}(F)$ is closed in A.

(④ $\implies$ ③): [Use complement: $U \subset N$ is open $\iff$ $F = N \backslash U$ is closed]

(③ $\implies$ ①): Given $x_0 \in A$. [WTS: $\lim\limits_{x \to x_0} f(x) = f(x_0)$] Fix any $\varepsilon > 0$. [WTS: $\exists\, \delta > 0$ $s.t.$ $d(x, x_0) < \delta \implies \rho(f(x), f(x_0)) < \varepsilon$] Let $U = D(f(x_0), \varepsilon) < r$ is open. By ③, $f^{-1}(U)$ is open in A. i.e.,

$$f^{-1}(U) = A \cap V, \quad V \subset M \text{ is open.}$$

Note that $x_0 \in f^{-1}(U) \implies x_0 \in V$. Since V is open, $\exists\, \delta > 0$ $s.t.$ $D(x, \delta) \subset V$. [WTS: $x \in A,\ d(x, x_0) < \delta \implies \rho(f(x), f(x_0)) < \varepsilon$] Suppose $x \in A$ with $d(x, x_0) < \delta$. Then, $x \in A$ and $x \in V$. That is, $x \in A \cap V = f^{-1}(U)$. Hence, $f(x) \in U$. By definition of U, we get $\rho(f(x), f(x_0)) < \varepsilon$ as desired.

Q.E.D. ■

4.2 Properties of Continuous Mappings

Theorem 4.2.1 Images of Compact and Connected Sets

Suppose $f : M \to N$ is continuous. Then,

- If $K \subset M$ is compact, then $f(K)$ is also compact.
- If $B \subset M$ is connected, then $f(B)$ is also connected.

Proof 1.

- Let x_k be a sequence in K. Then, $y_k = f(x_k)$ is a sequence in $f(K)$. [WTS: $f(K)$ is sequentially compact.] Suppose K is compact, $\exists\, x_{k_j} \to x_0 \in K$ when $j \to \infty$. By continuity of f, $f(x_{k_j}) \to f(x_0) \in f(K)$ when $k \to \infty$. So, for sequence $y_k = f(x_k)$, we find a subsequence $f(x_{k_j}) \to f(x_0) \in f(K)$. So, $f(K)$ is sequentially compact. □
- Given connected set $B \subset M$. Assume, for the sake of contradiction, that $f(B)$ is disconnected. Then, $\exists$ open sets U, V $s.t.$ $f(B) \cap U \cap V = \varnothing$ and $f(B) \cap U \neq \varnothing$, $f(B) \cap V \neq \varnothing$, $f(B) \subset U \cup V$. [We can derive that B is also disconnected, which is a contradiction.] So, it must be that $f(B)$ is also connected.

Q.E.D. ■

Theorem 4.2.2 Operations on Continuous Mapping

Addition, multiplication, divisions, and compositions of continuous functions (if they are well-defined) are also continuous.

Example 4.2.3

If $f(x) = \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}$ are continuous, then, $f(x)g(x)$ is also continuous.

Proof 2. Denote $F(x) = f(x)g(x)$. Then,

$$\begin{aligned}
|F(x) - F(x_0)| &= |f(x)g(x) - f(x_0)g(x_0)| \\
&\leq |f(x)g(x) - f(x)g(x_0)| + |f(x)g(x_0) - f(x_0)g(x_0)| \\
&= |f(x)||g(x) - g(x_0)| + |g(x_0)||f(x) - f(x_0)| \\
&\vdots \\
&< \varepsilon
\end{aligned}$$

Q.E.D. ■

Theorem 4.2.4 Maximum/Minimum Property

Let $K \subset M$ be compact and $f : K \to \mathbb{R}$ be continuous. Then,

- f is bounded on K (i.e., $f(K)$ is a bounded set)
- $\exists x_0, x_1 \in K \; s.t.$

$$f(x_1) = \max_{x \in K} f(x) \quad \text{and} \quad f(x_0) = \min_{x \in K} f(x).$$

That is, $f(x_0) \leq f(x) \leq f(x_1) \quad \forall x \in K$.

Proof 3.

- Since K is compact and f is continuous, $f(K)$ is compact. Since $f(K) \subset \mathbb{R}$ is compact, $f(K)$ is closed and bounded.
- Since $f(K)$ is bounded, we know $\inf(f(K))$ and $\sup(f(K))$ exist and are finite. Further since $f(K)$ is closed, $\inf(f(K), \sup(f(K) \in f(K)$. Hence, $\exists\, x_0 = \inf(f(K))$ and $x_1 = \sup(f(K)) \; s.t.$

$$f(x_0) \leq f(x) \leq f(x_1) \quad \forall x \in K.$$

Q.E.D. ■

Remark 4.3

- *The condition "compact" cannot be removed.*

Example 4.2.5

$$f(x) = \frac{1}{x} : (0, 1) \to \mathbb{R} \textit{ is continuous but not bounded}$$

$f(x) = x : (0, 1) \to \mathbb{R}$ *is bounded, but does not have max/min values*

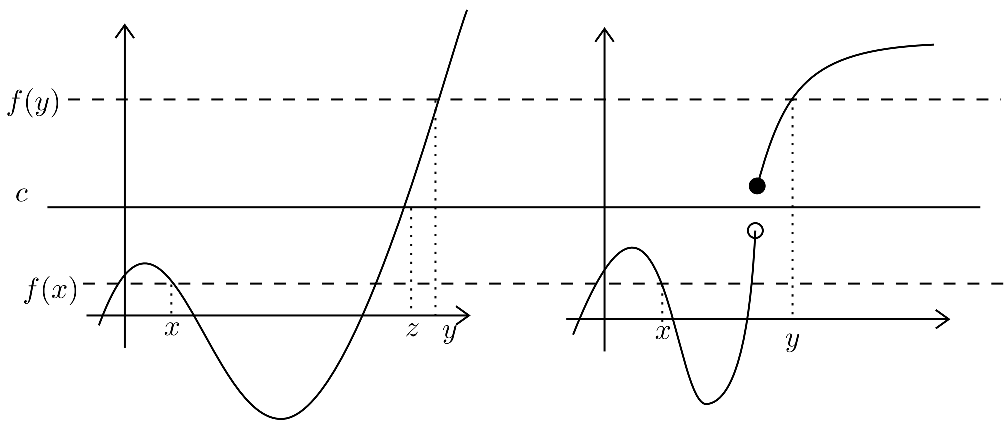
- *The condition "continuity" cannot be removed.*

Example 4.2.6

Consider function $f : [0, 1] \to \mathbb{R}$ *by* $f(x) = \begin{cases} \frac{1}{x}, & x > 0 \\ 2, & x = 0. \end{cases}$ *Although* $[0, 1]$ *is compact,* $f(x)$ *is nto continuous, and* f *is not bounded and does not have max/min values on* $[0, 1]$.

- *We don't need differentiability here.*

Theorem 4.2.7 Intermediate Value Theorem (IVT)
Let $K \subset M$ be connected and $f : K \to \mathbb{R}$ be continuous. Suppose $x, y \in K$ with $f(x) < f(y)$. Then, for any intermediate value c $s.t.$ $f(x) < c < f(y)$, $\exists\, z \in K$ with $x < z < y$ $s.t.$ $f(z) = c$.



Proof 4. Let $K \subset M$ be connected and $f : K \to \mathbb{R}$ be continuous. Suppose $x, y \in K$ with $f(x) < f(y)$. Assume, for the sake of contradiction, $\exists\, c$ with $f(x) < c < f(y)$ $s.t.$ $c \notin f(K)$.

Since K is connected and f is continuous, $f(K)$ is also connected. However, $U = (-\infty, c)$ and $V = (c, +\infty)$ separate $f(K)$, implying $f(K)$ is not connected. ※ We reach a contradiction. So, such a c does not exist.

Q.E.D. ■

Example 4.2.8 Application of IVT I

Let $f(x)$ be a polynomial of odd degree. Then, f has at least one real root.

Proof 5. Suppose $f(x) : \mathbb{R} \to \mathbb{R}$ is continuous. Write $f(x)$ as

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

where $a_n \neq 0$ and $n = 2k+1$ is odd.

WLOG, suppose $a_n > 0$. Then,

$$\lim_{x\to-\infty} f(x) = -\infty \quad \text{and} \quad \lim_{x\to\infty} f(x) = \infty.$$

So, $\exists x, y \in \mathbb{R}\ s.t.\ f(x) < 0$ and $f(y) > 0$. Therefore, by IVT, $\exists\, x_0 \in \mathbb{R}\ s.t.\ f(x_0) = c = 0$.

Q.E.D. ■

Definition 4.2.9 (Fixed Point). x is a *fixed point* of f if $f(x) = x$.

Example 4.2.10 Application of IVT II

Let $f : [1,2] \to [0,3]$ be continuous with $f(1) = 0$, $f(2) = 3$. Then, f has a fixed point.

Proof 6. Apply IVT to a new function: $F(x) = f(x) - x$. Take $c = 0$ as the intermediate value.

Q.E.D. ■

4.3 Uniform Continuity (UC)

Definition 4.3.1 (Uniform Continuity (UC)). A function $f : A \subset M \to N$ is *uniformly continuous* on A if $\forall\, \varepsilon > 0, \exists\, \delta > 0\ s.t.\ x, y \in A$ and $d(x,y) < \delta \implies \rho(f(x), f(y)) < \varepsilon$.

Remark 4.4

- *For uniform continuity, the δ depends only on ε not on points.*
- *For continuity (at x_0), the δ may depend on ε and the point x_0.*

Example 4.3.2

Consider $f(x) = \dfrac{1}{x} : (0,1) \to \mathbb{R}$. f is continuous at any point $x_0 \in (0,1)$. But to satisfy

$$|x - x_0| < \delta \implies |f(x) - f(x_0)| = \left|\frac{1}{x} - \frac{1}{x_0}\right| = \frac{|x - x_0|}{|x \cdot x_0|} < \varepsilon$$

we need to pick

$$|x - x_0| = \delta = \min\left\{\frac{1}{2}x_0^2\varepsilon,\ \frac{1}{2}x_0\right\}.$$

Theorem 4.3.3 Uniform Continuity on Compact Set

Let $f : K \subset M \to N$ be continuous and K be compact. Then, f is uniformly continuous on K.

Proof 1. Fix $\varepsilon > 0$. For each $x \in K$, since f is continuous at x, $\exists\, \delta_x\ s.t.$ for $y \in K$ with $d(x,y) < \delta_x$, we have $\rho(f(x), f(y)) < \dfrac{\varepsilon}{2}$.

Consider the open cover of K: $\left\{ D\left(x, \frac{\delta_x}{2}\right) \middle| x \in K \right\}$. Since K is compact, $\exists$ subcover:

$$D\left(x_i, \frac{\delta_{x_i}}{2}\right), \quad i = 1, 2, \ldots, L$$

Finally, let

$$\delta = \min_{1 \le i \le L} \left\{ \frac{\delta_{x_i}}{2} \right\}.$$

Claim $x, y \in K$ *with* $d(x,y) < \delta \implies \rho(f(x), f(y)) < \varepsilon$.

Proof. Note that

$$\begin{aligned} d(y, x_i) &\le d(y,x) + d(x, x_i) \\ &< \delta + \frac{\delta_{x_i}}{2} \\ &< \delta_{x_i}. \end{aligned}$$

One can continue to show that $\rho(f(x), f(y)) < \varepsilon$. □

Q.E.D. ■

Definition 4.3.4 (Lipschitz Continuity). A function $f : A \subset M \to N$ is called *Lipschitz* if $\exists$ constant L *s.t.*

$$\rho(f(x), f(y)) \le L \cdot d(x,y) \quad \forall\, x, y \in A.$$

Theorem 4.3.5 Lipschitz and Uniform Continuity
If $f : A \subset M \to N$ is Lipschitz, then f is uniformly continuous in A.

Corollary 4.3.6 : Suppose $f : (a,b) \to \mathbb{R}$ is differentiable and $\exists\, M > 0$ *s.t.* $|f'(x)| \le M \quad \forall x \in (a,b)$. Then, f is Lipschitz.

Proof 2. Given $x, y \in (a,b)$. Then,

$$\begin{aligned} |f(y) - f(x)| &= \left| f'(z)(y-x) \right| && \text{[Mean Value Theorem]} \\ &\le M|x-y|. \end{aligned}$$

Q.E.D. ■

Example 4.3.7 Lipschitz Functions
$f(x) = x$ and $f(x) = \sin x$ are Lipschitz functions.

Remark 4.5

- *If f has bounded derivative (or slope), then f is uniformly continuous.*

- *But if f is differentiable and uniformly continuous, f may not have bounded derivative.*
- *Open End-ed Questions:*
  - *$f : \mathbb{R} \to \mathbb{R}$ is bounded and continuous, f may not be uniformly continuous.*
  - *$f, g : \mathbb{R} \to \mathbb{R}$ are uniformly continuous, $f \cdot g$ is not uniformly continuous in general.*
  - *But if f, or g, or both are bounded and uniformly continuous, is $f \cdot g$ uniformly continuous?*

4.4 Differentiability

Remark 4.6 *Starting from this section, we will only consider functions $f :$ an interval $\to \mathbb{R}$.*

Definition 4.4.1 (Differentiability). A function f is *differentiable* at a point x_0 if it is defined in an open interval that contains x_0 and its derivative exists:

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}, \tag{D}$$

or equivalently, set $h = x - x_0$,

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

Remark 4.7 (Interpretation)

- *Rewrite* (D) *as*
$$\lim_{x \to x_0} \left[\frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} \right] = 0.$$
This implies the function $y = f(x)$ can be approximated by the linear function
$$y = f(x_0) + f'(x_0)(x - x_0)$$
in a neighborhood of x_0.
- *Rewrite* (D) *as*
$$\lim_{\Delta x \to 0} \left[\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} - f'(x_0) \right] = 0.$$
this implies the slope of tangent line is the limit of the slope of secant lines.

Theorem 4.4.2 Continuity of Differentiable Functions
Suppose $f : A \subset M \to N$ is differentiable at x_0. Then, it is continuous at x_0.

Proof 1. Given $\varepsilon > 0$. Since
$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0),$$
$\exists\, \delta_1 > 0$ *s.t.* $|x - x_0| < \delta_1 \implies \left| \dfrac{f(x) - f(x_0)}{x - x_0} \right| < |f'(x)| + 1$. Choose
$$\delta = \min \left\{ \frac{\varepsilon}{|f'(x)| + 1}, \delta_1 \right\}.$$

So, when $|x - x_0| < \delta$, we have

$$\begin{aligned}
|f(x) - f(x_0)| &= \frac{|f(x) - f(x_0)|}{|x - x_0|} \cdot |x - x_0| \\
&< \left(\left|f'(x)\right| + 1\right) \cdot \frac{\varepsilon}{|f'(x)| + 1} \\
&= \varepsilon.
\end{aligned}$$

Q.E.D. ■

Remark 4.8 *The converse if not true: continuity $\not\Rightarrow$ differentiability. Counterexample: $f(x) = |x|$.*

Proof 2. (Another Approach) Note that

$$\begin{aligned}
\lim_{x\to x_0} (f(x) - f(x_0)) &= \lim_{x\to x_0} \left[\frac{f(x) - f(x_0)}{x - x_0}\right] (x - x_0) \\
{}' &= \lim_{x\to x_0} \left[\frac{f(x) - f(x_0)}{x - x_0}\right] \lim_{x\to x_0} (x - x_0) && \text{[Product Rule of Limit]} \\
&= f'(x) \cdot 0 \\
&= 0.
\end{aligned}$$

So, the function is continuous.

Q.E.D. ■

Theorem 4.4.3 Rules of Differentiation

- Constant multiple rule:
$$(kf)'(x_0) = k \cdot f'(x_0).$$
- Sum rule:
$$(f + g)'(x_0) = f'(x_0) + g'(x_0)$$
- Product rule:
$$(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0).$$
- Quotient rule:
$$\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{(g(x_0))^2}$$
- Chain rule:
$$(g \circ f)'(x_0) = g'(f(x_0)) \cdot f'(x_0).$$

Lemma 4.4.4 : If $f : (a, b) \to \mathbb{R}$ is differentiable and f has a max (or min) at $c \in (a, b)$, then $f'(c) = 0$.

Proof 3. Assume f has a max at $c \in (a, b)$. Then,

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

[WTS: $f'(c) \geq 0$ and $f'(c) \leq 0$.]

As f has a max at c, $f(h+c) \leq f(c)$, and so

$$f(h+c) - f(c) \leq 0.$$

$\boxed{\text{Case I}}$ $h > 0$:

$$f'(c) = \lim_{h^+ \to 0} \frac{f(h+c) - f(c)}{h} \leq 0.$$

$\boxed{\text{Case II}}$ $h < 0$:

$$f'(c) = \lim_{h^- \to 0} \frac{f(h+c) - f(c)}{h} \geq 0.$$

As $f'(c) \geq 0$ and $f'(c) \leq 0$, it must be that $f'(c) = 0$.

Q.E.D. ■

Theorem 4.4.5 Rolle's Theorem

Let $f : [a,b] \to \mathbb{R}$ be continuous and f be differentiable on (a,b). If $f(a) = f(b) = 0$, then $\exists\, c \in (a,b)\ s.t.\ f'(c) = 0$.

Proof 4. f has max and min on $[a,b]$ as $[a,b]$ is compact. [WTS: This max/min occur in (a,b).]

Since $f(a) = f(b) = 0$, then max and min cannot both occur at the endpoint (i.e., either max or min occur in (a,b)) unless f is the constant function $f(x) = 0$.

Now, by Lemma 4.4.4, $\exists\, c \in (a,b)\ s.t.\ f'(c) = 0$, where c is either the max or min.

Q.E.D. ■

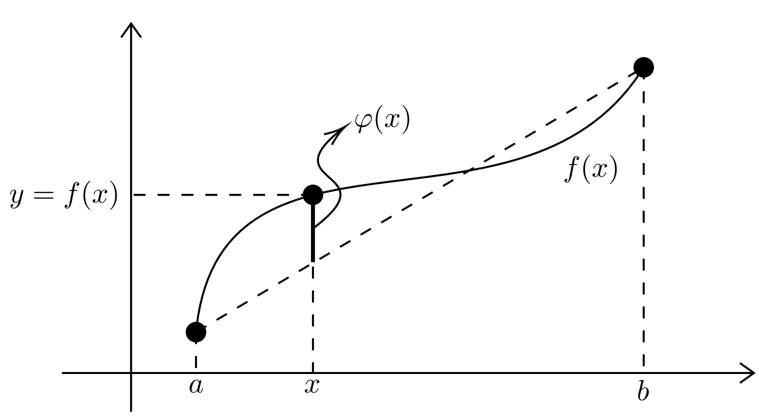
Theorem 4.4.6 Mean Value Theorem (MVT)

Suppose f is continuous on $[a,b]$ and differentiable on (a,b). Then, $\exists\, c \in (a,b)\ s.t.$

$$f(b) - f(a) = f'(c)(b-a) \quad \text{or} \quad f'(c) = \frac{f(b) - f(a)}{b-a}.$$

Remark 4.9 *Rolle's Theorem is a special case of MVT. We will use the special case to prove the general case.*

Proof 5.



Construct $\varphi(x)$:

$$\varphi(x) = f(x) - \left[f(a) + \frac{f(b) - f(a)}{b - a}(x - a)\right].$$

One can verify the following:

- $\varphi(a) = 0$;
- $\varphi(b) = 0$; and
- φ is continuous and differentiable.

Then, apply Rolle's Theorem to $\varphi(x)$: $\exists\, c \in (a, b)\ s.t.$

$$\varphi'(c) = 0.$$

Note that $\varphi'(c) = f'(c) - \dfrac{f(b) - f(a)}{b - a} = 0$, we have

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

Q.E.D. ■

Remark 4.10 (Geometric Interpretation) *There is at least one point where the instant change of rate is the same as the average change of rate.*

Definition 4.4.7 (Monotonecity).

- We say $f(x)$ is *increasing* (or *strictly increasing*) at a point x_0 if $\exists$ open interval (a, b) containing x_0 with:
  - $a < x < x_0 \implies f(x) \leq f(x_0)$ (or $f(x) < f(x_0)$);
  - $x_0 < x < b \implies f(x) \geq f(x_0)$ (or $f(x) > f(x_0)$).
- Similar definition for decreasing (or strictly decreasing) at a point x_0.
- $f(x)$ is increasing (or strictly increasing) on an interval I if for $x_1, x_2 \in I$

$$x_1 < x_2 \implies f(x_1) \leq f(x_2) \quad (\text{or } f(x_1) < f(x_2)).$$

- Similar definition for decreasing (or strictly decreasing) on an interval.

Theorem 4.4.8 Local Monotonecity and Derivative
Let f be differentiable at x_0. Then,

- f increasing at $x_0 \implies f'(x_0) \geq 0$; f decreasing at $x_0 \implies f'(x_0) \leq 0$.
- $f'(x_0) > 0 \implies f$ strictly increasing at x_0; $f'(x_0) < 0 \implies f$ strictly decreasing at x_0.

Proof 6. (of ①): Suppose f is increasing at x_0. Then

$$\begin{aligned} f(x_0+h)-f(x_0) &\geq 0 \quad \text{when } h>0 \\ &\leq 0 \quad \text{when } h<0. \end{aligned}$$

Then,

$$f'(x_0)=\lim_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h}\geq 0.$$

(of ②): Suppose $f'(x_0)>0$. Then, for $\varepsilon=\dfrac{1}{2}f'(x_0)>0$, $\exists\,\delta>0\ s.t.$

$$0<|h|<\delta \implies \left|\frac{f(x_0+h)-f(x_0)}{h}-f'(x_0)\right|<\varepsilon=\frac{1}{2}f'(x_0).$$

$$-\frac{1}{2}f'(x_0)<\frac{f(x_0+h)-f(x_0)}{h}-f'(x_0)<\frac{1}{2}f'(x_0) \implies 0<\frac{1}{2}f'(x_0)<\frac{f(x_0+h)-f(x_0)}{h}<\frac{3}{2}f'(x_0).$$

When $x<x_0$, $h=x-x_0<0$. As $\dfrac{f(x_0+h)-f(x_0)}{h}>0$,

$$f(x_0+h)-f(x_0)=f(x)-f(x_0)<0 \implies f(x)<f(x_0)$$

When $x>x_0$, $h=x-x_0>0$,

$$f(x_0+h)-f(x_0)=f(x)-f(x_0)>0 \implies f(x)>f(x_0).$$

Hence, f is strictly increasing.

Q.E.D. ■

Theorem 4.4.9 Global Monotonecity and Derivative
Let f be continuous on $[a,b]$ and differentiable on (a,b). Then,

- $f'(x)\geq 0 \quad \forall\, x\in(a,b) \implies f$ increasing on $[a,b]$.
- $f'(x)\leq 0 \quad \forall\, x\in(a,b) \implies f$ decreasing on $[a,b]$.
- $f'(x)>0 \quad \forall\, x\in(a,b) \implies f$ strictly increasing on $[a,b]$.
- $f'(x)<0 \quad \forall\, x\in(a,b) \implies f$ strictly decreasing on $[a,b]$.

Theorem 4.4.10 Local Max/Min and Derivatrive
Suppose f is continuous on $[a,b]$ and twice differentiable on (a,b). Let $x_0\in(a,b)$.

- $f'(x_0)=0$ and $f''(x_0)>0 \implies x_0$ is a strict local min of f.
- $f'(x_0)=0$ and $f''(x_0)<0 \implies x_0$ is a strict local max of f.

Proof 7. (of ①) By Theorem 3.3.8(2), $f''(x_0)>0 \implies f'(x)$ is strictly increasing at x_0. Then,

- $f'(x) < f'(x_0) = 0 \quad \forall x \in (x_0 - \delta, x_0) \implies f(x)$ strictly decreasing on $(x_0 - \delta, x_0)$
- $f'(x) > f'(x_0) = 0 \quad \forall x \in (x_0, x_0 + \delta) \implies f(x)$ strictly increasing on $(x_0, x_0 + \delta)$.

Q.E.D. ■

Theorem 4.4.11 Inverse Function Theorem (IFT)
Suppose $f'(x) > 0 \quad \forall x \in (a, b)$ (or, $f'(x) < 0 \quad \forall x \in (a, b)$). Then,

- $f : (a, b) \to \mathbb{R}$ is a bijection onto its image
- Inverse f^{-1} is differentiable on its domain.
- $(f^{-1})'(y) = \dfrac{1}{f'(x)}$, where $y = f(x)$.

Proof 8. Assume $f'(x) > 0 \quad \forall x \in (a, b)$. Then, f is strictly increasing. Then, f is 1-to-1 function $\implies f$ is a bijection $\implies f^{-1}$ exists. [WTS: f^{-1} is continuous.]

Let U be an open set in (a, b). [WTS: $(f^{-1})^{-1}(U) = f(U)$ is open.]

Finally, write $y = f(x)$. Then, $x = f^{-1}(y)$. Let $y_0 = f(x_0)$. Then,

$$\begin{aligned}
(f^{-1})'(y_0) &= \lim_{y \to y_0} \frac{f^{-1}(y) - f^{-1}(y_0)}{y - y_0} \\
&= \lim_{y \to y_0} \frac{x - x_0}{f(x) - f(x_0)} \\
&= \lim_{x \to x_0} \frac{1}{\dfrac{f(x) - f(x_0)}{x - x_0}} \\
&= \frac{1}{\lim\limits_{x \to x_0} \dfrac{f(x) - f(x_0)}{x - x_0}} \\
&= \frac{1}{f'(x_0)}.
\end{aligned}$$

Q.E.D. ■

4.5 Integration

Definition 4.5.1 (Riemann Integrable). Let $A \subset \mathbb{R}$ be bounded and $f : A \to \mathbb{R}$ be a bounded function. [We want to make sense $\displaystyle\int_A f(x)\mathrm{d}x$.]

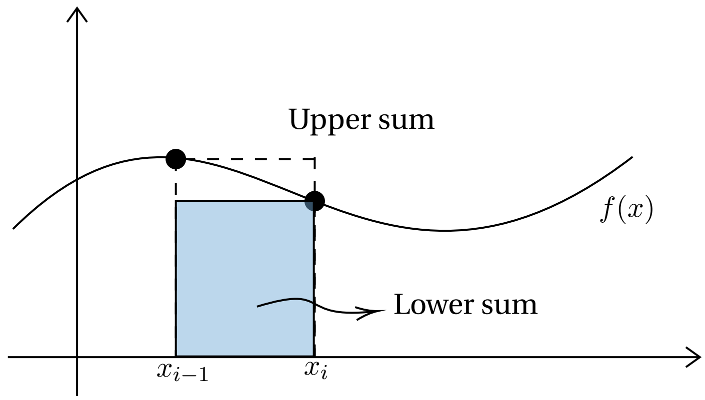
- Partition the interval:

  If interval $[a, b] \supset A$ and extend function $f(x)$ to $[a, b]$ by letting $f(x) = 0 \quad \forall x \notin A$.

  Partition the interval $[a, b]$ by points: $a = x_0 < x_1 < \cdots < x_n = b$. Denote P by

$$P = \{x_0, x_1, x_2, \cdots, x_n\}.$$

- Form Upper and Lower Sum of f.



For any fixed partition, let

$$U(f,P) = \sum_{i=1}^{n} \sup\left\{f(x) \mid x \in [x_{i-1}, x_i]\right\}(x_i - x_{i-1})$$

is the upper sum, and

$$L(f,P) = \sum_{i=1}^{n} \inf\left\{f(x) \mid x \in [x_{i-1}, x_i]\right\}(x_i - x_{i-1})$$

is the lower sum.

Claim *Suppose* $m \leq f(x) \leq M$*. Then,*

$$m(b-a) \leq L(f,P) \leq U(f,P) \leq M(b-a).$$

- Upper integral and Lower integral are defined as

$$\overline{\int_A} f = \inf\left\{U(f,P) : P \text{ is a partition}\right\} \qquad \text{(Upper Integral)}$$

$$\underline{\int_A} f = \sup\left\{L(f,P) : P \text{ os a partition}\right\} \qquad \text{(Lower Integral)}$$

- We say a function f is *Riemann integrable* if

$$\overline{\int_A} f = \underline{\int_A} f,$$

and we write

$$\int_A f = \overline{\int_A} f = \underline{\int_A} f.$$

Example 4.5.2 Riemann Integrable

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

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

Then, for any partition P,

$$U(f,P) = \sum_{i=1}^{n} 1 \cdot (x_i - x_{i-1}) = 1$$

and

$$L(f,P) = \sum_{i=1}^{n} 0 \cdot (x_i - x_{i-1}) = 0.$$

So,

$$\overline{\int_A} f \neq \underline{\int_A} f \implies f \text{ is not integrable}$$

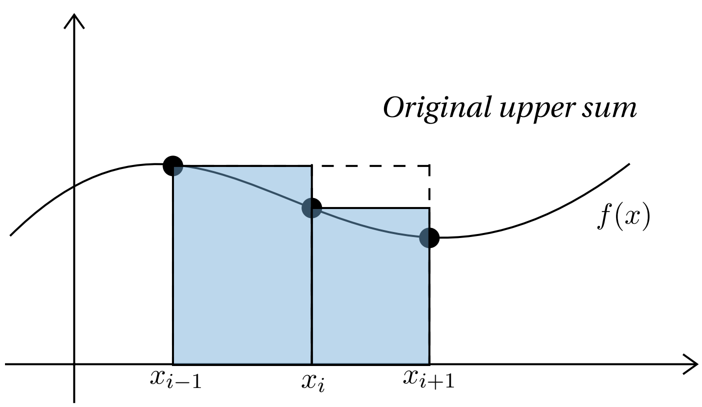
- Compute $\overline{\int_0^1} x \, \mathrm{d}x$ and $\underline{\int_0^1} x \, \mathrm{d}x$.

Hint: Consider partition $P_n = \left\{0 < \frac{1}{n} < \frac{2}{n} < \cdots < \frac{n}{n}\right\}$.

Lemma 4.5.3 : Let $f : [a,b] \to \mathbb{R}$ be bounded. If P, P' are partitions of $[a,b]$ with $P \subset P'$ (P' is a refinement of P), then

$$L(f.P) \leq L(f,P') \leq U(f,P') \leq U(f,P).$$

Remark 4.11 *In words, when the partition gets finer, lower sum increases but upper sum decreases.*



Proposition 4.5.4 :

$$\underline{\int_a^b} f \leq \overline{\int_a^b} f$$

Proof 1. For any fixed partition P and Q. As $P \subset P \cup Q$ and $Q \subset P \cup Q$, by Lemma 4.5.4, we have

$$L(f,P) \leq L(f,P \cup Q) \leq U(f,P \cup Q) \leq U(f,Q).$$

Then,

$$\underline{\int_a^b} f = \sup_P L(f,P) \leq U(f,Q) \quad \text{for any } Q$$

So,

$$\underline{\int_a^b} f \leq \inf_Q U(f,Q) = \overline{\int_a^b} f.$$

Q.E.D. ■

Theorem 4.5.5

- If $f : [a,b] \to \mathbb{R}$ is bounded and is continuous at all but finite many points, then f is integrable.
- If f is increasing or decreasing on $[a,b]$, then f is integrable.

Proof 2.

- (Proof of ①): Observe that $\forall$ partition P, $L(f,P) \leq \underline{\int_a^b} f \leq \overline{\int_a^b} \leq U(f,P)$. [To prove a function is integrable, it's sufficient to show that $\forall\, \varepsilon > 0$, $\exists$ partition P $s.t.$ $U(f,P) - L(f,P) < \varepsilon$.]

  - Suppose f is continuous on $[a,b]$ except at $a_1, a_2, \ldots, a_k$. Since f is bounded, $\exists\, m, M$ $s.t.$ $m \leq f(x) \leq M \quad \forall\, x \in [a,b]$. Choose partition P_1 $s.t.$ each subinterval containing some a_i has length $\leq \frac{\varepsilon}{2} \cdot \frac{1}{2k(M-m)}$.
    Let K be the union of the remaining subinterval in P_1. Then, K is compact and f is continuous on K. So, f is uniformly continuous on K. That is,

    $$\exists\, \delta > 0 \; s.t. \; x_1, x_2 \in K \; s.t. \; |x_1 - x_2| < \delta \implies |f(x_1 - f(x_2)| < \frac{\varepsilon}{2(b-a)}.$$

  - Construct the refinement P of P_1 so that each subinterval in P not containing some a_i has length $< \delta$. So,

    $$P = \{a = x_0 < x_1 < \cdots < x_n = b\} \quad \text{and} \quad I_j - [x_{j-1}, x_j].$$

    Denote

    $$M_j = \sup_{I_j} f(x) \quad \text{and} \quad m_j = \inf_{I_j} f(x).$$

    If I_j contains some a_i, then $m \leq m_j \leq M_j \leq M$.
    If I_j contains no discontinuous points, then $I_j \subset K$, and

    $$M_j - m_j = \max - \min < \frac{\varepsilon}{2(b-a)}.$$

– Finally, we have

$$\begin{aligned}
U(f,P)-L(f,P) &= \sum_{j=1}^{n}(M_j-m_j)(x_j-x_{j-1}) \\
&= \sum_{a_i\in I_j}(M_j-m_j)(x_j-x_{j-1}) + \sum_{a_i\notin I_j}(M_j-m_j)(x_j-x_{j-1}) \\
&< \overbrace{\cancel{2k}}^{\substack{\text{worse case:}\\ 2k \text{ such intervals}}} \quad \underbrace{\cancel{(M-m)}}_{\substack{\text{estimate of}\\ M_j-m_j}} \cdot \underbrace{\frac{\varepsilon}{2}\cdot\frac{1}{\cancel{2k}\cancel{(M-m)}}}_{\text{length of } I_j} + \underbrace{\frac{\varepsilon}{2\cancel{(b-a)}}}_{\substack{\text{estimate of}\\ M_j-m_j}} \overbrace{\cancel{(b-a)}}^{\text{total length}} \\
&= \frac{\varepsilon}{2}+\frac{\varepsilon}{2} \\
&= \varepsilon.
\end{aligned}$$

Therefore,

$$\underline{\int_a^b} f = \overline{\int_a^b} f \implies f \text{ is integrable.}$$

- (Proof of ②): Assume f is increasing. Given $\varepsilon > 0$. Consider an equal partition

$$P_n = \left\{a = x_0, x_1 = x_0 + \frac{b-a}{n}, x_2, \ldots, x_n = b\right\}.$$

Then, by equal partition and f is increasing, we have

$$U(f,P_n) = \sum_{j=1}^{n} f(x_j)(x_j - x_{j-1}) = \frac{b-a}{n}\sum_{j=1}^{n} f(x_j)$$

and

$$L(f,P_n) = \sum_{j=1}^{n} f(x_{j-1})(x_j - x_{j-1}) = \frac{b-a}{n}\sum_{j=1}^{n} f(x_{j-1}).$$

So,

$$\begin{aligned}
U(f,P_n)-L(f,P_n) &= \frac{b-a}{n}\sum_{j=1}^{n} f(x_j)-f(x_{j-1}) \\
&= \frac{b-a}{n}(f(x_n)-f(x_1)) && \text{[Intermediate terms cancel]} \\
&= \frac{b-a}{n}(f(b)-f(a)).
\end{aligned}$$

When $n\to\infty$, $U(f,P_n)-L(f,P_n) = \frac{b-a}{n}(f(b)-f(a)) \to 0$. Therefore,

$$U(f,P_n)-L(f,P_n) < \varepsilon \quad \text{for large } n \implies f \text{ is integrable.}$$

Q.E.D. ■

Remark 4.12 *To prove a function f is integrable, it is sufficient to show that* $\forall\, \varepsilon > 0$, $\exists$ *partition* P *s.t.*

$$U(f,P) = L(f,P) < \varepsilon.$$

Theorem 4.5.6 Rules of Integration

- $k\displaystyle\int_a^b f(x)\,\mathrm{d}x = \int_a^b kf(x)\,\mathrm{d}x$, k is a constant.
- $\displaystyle\int_a^b (f(x) \pm g(x))\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x \pm \int_a^b g(x)\,\mathrm{d}x$
- $\displaystyle\int_a^c f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_b^c f(x)\,\mathrm{d}x$, for $a \le b \le c$.
- If $f \le g$, then $\displaystyle\int_a^b f(x)\,\mathrm{d}x \le \int_a^b g(x)\,\mathrm{d}x$.
  In particular, $-|f| \le f \le |f|$, so
  $$-\int_a^b |f| \le \int_a^b f \le \int_a^b |f|.$$
  That is,
  $$\left|\int_a^b f\right| \le \int_a^b |f|.$$

Definition 4.5.7 (Antiderivative). Let $f(x) : [a,b] \to \mathbb{R}$. An *antiderivative* of f is a continuous function $F(x) : [a,b] \to \mathbb{R}$ *s.t.* $F'(x) = f(x) \quad \forall\, x \in (a,b)$.

Remark 4.13 (Antiderivative is not Unique) *Suppose $F(x)$ is an antiderivative of $f(x)$. If G is another antiderivative, then*

$$\frac{\mathrm{d}}{\mathrm{d}x}[G(x) - F(x)] = G'(x) - F'(x) = f(x) - f(x) = 0 \quad \forall\, x \in (a,b).$$

So, by MVT, $G(x) - F(x) = C$, where C is some constant, or

$$G(x) = F(x) + C.$$

Theorem 4.5.8 Fundamental Theorem of Calculus (FTC)
Let $f(x) : [a,b] \to \mathbb{R}$ be continuous. Then, f has an antiderivative F, and

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a) \left[= F(x)\Big|_a^b\right].$$

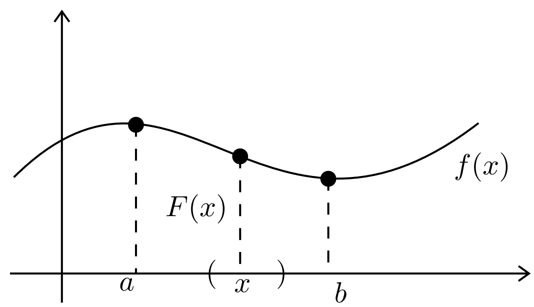
Proof 3. Define $F(x)$ by

$$F(x) = \int_a^x f(t)\,\mathrm{d}t$$

for $x \in [a,b]$.

Claim *$F(x)$ is an antiderivative of $f(x)$.*

Proof.



Fix $x \in (a,b)$. Let $h > 0$ $s.t.$ $(x-h, x+h) \subset (a,b)$. Then,

$$\begin{aligned}
\frac{F(x+h)-F(x)}{h} &= \frac{1}{h}\left(\int_a^{x+h} f(t)\,\mathrm{d}t - \int_a^x f(t)\,\mathrm{d}t\right) \\
&= \frac{1}{h}\left(\cancel{\int_a^x f(t)\,\mathrm{d}t} + \int_x^{x+h} f(t)\,\mathrm{d}t - \cancel{\int_a^x f(t)\,\mathrm{d}t}\right) = \frac{1}{h}\int_x^{x+h} f(t)\,\mathrm{d}t.
\end{aligned}$$

Note that

$$f(x) = \frac{1}{h}\int_x^{x+h} \underbrace{f(x)}_{\text{constant } w.r.t.\ t}\,\mathrm{d}t$$

So,

$$\begin{aligned}
\frac{F(x+h)-F(x)}{h} - f(x) &= \frac{1}{h}\int_x^{x+h} f(t)\,\mathrm{d}t - \frac{1}{h}\int_x^{x+h} f(x)\,\mathrm{d}t \\
&= \frac{1}{h}\int_x^{x+h} f(t) - f(x)\,\mathrm{d}t
\end{aligned}$$

Given $\varepsilon > 0$, f is continuous at x. So, $\exists\, \delta > 0$ $s.t.$

$$|t-x| < \delta \implies |f(t) - f(x)| < \varepsilon.$$

Then, when $|h| < \delta$, we have

$$\begin{aligned}
\left|\frac{F(x+h)-F(x)}{h}\right| \le \left|\frac{1}{h}\int_x^{x+h} f(t) - f(x)\,\mathrm{d}t\right| & \\
&\le \frac{1}{|h|}\int_x^{x+h} |f(t) - f(x)|\,\mathrm{d}t \\
&< \frac{1}{|h|}\int_x^{x+h} \varepsilon\,\mathrm{d}t \\
&= \frac{1}{\cancel{|h|}}\cdot\varepsilon\cdot\cancel{|h|} \\
&= \varepsilon.
\end{aligned}$$

So,

$$\lim_{h\to 0}\frac{F(x+h)-F(x)}{h}=f(x)\quad \text{i.e., } F'(x)=f(x).$$

Furthermore, one can show that $F(x)$ is continuous on $[a,b]$. [As $F(x)$ is differentiable on (a,b), it is continuous on (a,b). We only need to check for the endpoints.] □

Finally, by definition,

$$F(b)=\int_a^b f(t)\,\mathrm{d}t \quad \text{and} \quad F(a)=\int_a^b f(t)\,\mathrm{d}t=0.$$

So,

$$\int_a^b f(t)\,\mathrm{d}t=F(b)-F(a).$$

Q.E.D. ■

Remark 4.14 *In FTC, the continuity assumption of $f(x)$ cannot be removed. More specifically, it cannot be replaced by integrability. For example,*

$$f(x)=\begin{cases}0, & 0\le x\le 1\\ 1, & 1<x\le 2.\end{cases}$$

f is integrable, and its antiderivative

$$F(x)=\int_0^x f(t)\,\mathrm{d}t \quad \textit{is well-defined.}$$

However, $F'(x)=f(x)$ for $1<x\le x$. When $x=1$, $F'(x)$ does not even exist.

5 Uniform Convergence

5.1 Definition of Convergence

Definition 5.1.1 (Pointwise Convergence). Given a sequence of functions $f_k(x) : A \subset M \to N$ for $k = 1, 2, \dots$. We say $f_k(x) \to f(x)$ *converges pointwise* on A if $\forall\, x \in A$, the sequence of points $\{f_k(x)\}$ converges to $f(x)$. That is, $\forall\, x, \forall\, \varepsilon > 0, \exists\, K\ s.t.\ k \geq K \implies \rho(f_k(x), f(x)) < \varepsilon$.
Definition 5.1.2 (Uniform Convergence). $f_k(x) \to f(x)$ *converges uniformly on* A if $\forall\, \varepsilon > 0, \exists\, K\ s.t.\ k \geq K \implies \rho(f_k(x), f(x)) < \varepsilon \quad \forall\, x \in A$. We write $f_k \to f$ UC on A.

Remark 5.1 *For pointwise convergence, the choice of K depends both on ε and the point x. However, for uniform convergence, K only depends on ε but not specific point x.*

Definition 5.1.3 (Convergence of Series). Assume N is a normed space. Suppose $g_k : A \subset M \to N$. Then, $\sum_{k=1}^{\infty} g_k(x)$ converges to $g(x)$ *pointwise* or *uniformly*. Using sequence of partial sums, we have

$$f_n(x) = \sum_{k=1}^{n} g_k(x).$$

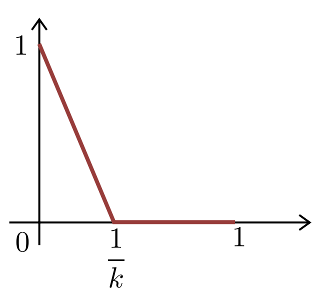
Remark 5.2 *UC is stronger: UC $\implies$ pointwise convergence.*
However, pointwise convergence $\;\not\Longrightarrow\;$ UC in general.

Example 5.1.4
Consider $A = [0, 1]$ and

$$f_k(x) = \begin{cases} 0 & \text{if } \frac{1}{k} \leq x \leq 1 \\ 1 & \text{if } 0 \leq x \leq \frac{1}{k} \end{cases}$$

Note that $f_k(x) \to f(x)$ pointwise, where

$$f(x) = \begin{cases} 0 & x > 0 \\ 1 & x = 0. \end{cases}$$



However, this convergence is not uniform: $\exists\, \varepsilon_0 > 0\ s.t.\ \forall\, K,\ \exists\, k \geq K\ s.t.\ \rho(f_k(x), f(x)) > \varepsilon_0$ for some $x \in A$. For example, take $\varepsilon_0 =$ and $0 < x < \frac{1}{k}$.

Theorem 5.1.5 Continuity of Uniform Limit
Let $f_k : A \subset M \to N$ be a sequence of continuous functions and $f_k \to f$ uniformly converges on A. Then, f is also continuous.

Proof 1. Fix $x_0 \in A$. Given $\varepsilon > 0$. By UC, $\exists\, K$ $s.t.$ $\rho(f_K(x), f(x)) < \dfrac{\varepsilon}{3}$ $\forall\, x \in A$. Since f_K is continuous, $\exists\, \delta > 0$ $s.t.$

$$x \in A,\ d(x, x_0) < \delta \implies \rho(f_K(x), f(x_0)) < \frac{\varepsilon}{3}.$$

Therefore, by triangle inequality, we have

$$\begin{aligned}\rho(f(x), f(x_0)) &\leq \rho(f(x), f_K(x)) + \rho(f_K(x), f_K(x_0)) + \rho(f_K(x_0), f_(x_0)) \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\ &= \varepsilon.\end{aligned}$$

So, f is continuous at x_0.

Q.E.D. ■

Remark 5.3 *This result can be used to show that a convergence is not uniform.*

Example 5.1.6

- $f_n(x) = \dfrac{x^n}{1+x^n}$, with $A = [0, 2]$.
  - Find pointwise limit
    $$f_n(x) \to f(x) = \begin{cases} 0, & 0 \leq x \leq 1 \\ \frac{1}{2}, & x = 0 \\ 1, & 1 < x \leq 2. \end{cases}$$
  - Determine uniform convergence:
    The convergence is not uniform because f is not continuous.
- Geometric Series: *Counterexample to the converse of Theorem 5.1.5*
  $$\sum_{k=0}^{\infty} x^k \quad \text{with } A = (-1, 1).$$
  - Converge pointwise to $g(x) = \dfrac{1}{1-x}$.
    Find partial sum:
    $$S_n(x) = \sum_{k=0}^{n} x^k = \frac{1 - x^{n+1}}{1-x}.$$

Since $x \in (-1,1)$, as $n \to \infty$, $x^{n-1} \to 0$. So,

$$S_n(x) = \frac{1-x^{n+1}}{1-x} \quad \xrightarrow{n\to\infty} \quad \frac{1}{1-x} \quad \text{for } x \in (-1,1).$$

– Uniform convergence on subinterval $[-a,a]$ for any $0 < a < 1$.

Estimate the error term:

$$|S_n(x) - g(x)| = \frac{|x|^{n+1}}{|1-x|}.$$

When $x \to 1$, $|S_n(x) - g(x)| \to \infty$ as $|1-x| \to 0$. However, if we restrict $x \in [-a,a]$ for some $0 < a < 1$, then $|1-x| \geq 1-a$, and we have

$$\forall \varepsilon > 0, \quad \exists N \; s.t. \; n \geq N \implies \frac{a^{n+1}}{1-a} < \varepsilon.$$

$$\implies |S_n(x) - g(x)| \leq \frac{a^{n+1}}{1-a} < \varepsilon \quad \forall x \in [-a,a].$$

– Convergence is not uniform on $(-1,1)$.

Observe that for any fixed N, we have $\frac{|x|^{N+1}}{|1-x|} \xrightarrow{x\to 1^-} \infty$. Therefore,

$$\exists x_0 < 1 \; s.t. \; \frac{|x_0|^{N+1}}{|1-x_0|} = |S_N(x_0) - g(x_0)| \geq 1 = \varepsilon_0.$$

Definition 5.1.7 (Uniformly Cauchy Sequence). A sequence of functions $f_k : A \subset M \to N$ is *uniformly Cauchy sequence* if $\forall \varepsilon > 0, \; \exists L > 0 \; s.t. \; j,k \geq L \implies \rho(f_k(x), f_j(x)) < \varepsilon \quad \forall x \in A$.

Theorem 5.1.8 Cauchy Criterion

Let (N,ρ) be a *complete* metric space and $f_k : A \subset M \to N$ be a sequence of functions. Then, f_k converges uniformly on $A \iff \forall \varepsilon > 0, \; \exists L > 0 \; s.t.$

$$j,k \geq L \implies \rho(f_k(x), f_j(x)) < \varepsilon \quad \forall x \in A.$$

Proof 2. $(\Rightarrow)$ Assume $f_k \to f$ uniformly. [WTS: f_k is uniformly Cauchy.]

$$\rho(f_k(x), f_j(x)) \leq \rho(f_k(x), f(x)) + \rho(f(x) + f_j(x)). \qquad \square$$

$(\Leftarrow)$ Assume $\{f_k\}$ is uniformly Cauchy.

- Find the limit function (pointwise)

  For each fixed $x \in A$, the sequence of points $\{f_k(x)\}$ is Cauchy in N. By completeness of N, $f_k(x)$ converges to some point in N. Denoted by $f(x)$.

- Show $f_k(x) \to f(x)$ UC

  Given $\varepsilon > 0$, $\exists\, L_1\ s.t.\ j,k \geq L_1 \implies \rho(f_k(x), f_j(x)) < \frac{\varepsilon}{2} \quad \forall\, x \in A$. Furthermore, as $f_k(x) \to f(x)$ pointwise, for any $x \in A$, $\exists L_x \geq L_1\ s.t.\ j \geq L_x \implies \rho(f_j(x), f(x)) < \frac{\varepsilon}{2}$.

  Now, let $K = L_1$. Then, when $k \geq K$ we have

$$\begin{aligned}\rho(f_k(x), f(x)) &\leq \rho(f_k(x), f_{L_x}(x)) + \rho(f_{L_x}(x), f(x)) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon \quad \forall\, x \in A.\end{aligned}$$

  Just pick $j = L_x$, we have different intermediate term for different x's.

Q.E.D. ■

Corollary 5.1.9 Weierstrass M Test: Let N be a complete normed space and $g_k : A \to N$ be a sequence of functions $s.t.$ $\exists$ constants M_k with

- $\|g_k(x)\| \leq M_k$ for all $x \in A$, and
- $\sum_{k=1}^{\infty} M_k$ converges.

Then, the series $\sum_{k=1}^{\infty} g_k(x)$ converge uniformly.

Proof 3. The sequence of partial sums $\{f_n(x)\}$ is uniformly Cauchy.

$$f_n(x) = \sum_{k=1}^{n} g_k(x).$$

Then, apply Cauchy criterion.

Q.E.D. ■

Example 5.1.10

- $\sum_{n=1}^{\infty} \frac{(\sin nx)^2}{n^2}, \quad A = \mathbb{R}.$

  Set $g_n(x) = \frac{(\sin nx)^2}{n^2}$. Then, $|g_n(x)| \leq \frac{1}{n^2}$.

  As $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges, by M test, $\sum_{n=1}^{\infty} \frac{(\sin nx)^2}{n^2}$ converges uniformly.

- $\sum_{n=0}^{\infty} \left(\frac{x^n}{n!}\right)^2 \to f(x)$ on $\mathbb{R}$ pointwise

  If we limit $A = [-a, a]$, then $\sum_{n=0}^{\infty} \left(\frac{x^n}{n!}\right)^2$ uniformly converges.

5.2 Integration and Differentiation of Series

Theorem 5.2.1

Suppose $f_n : [a,b] \to \mathbb{R}$ and integrable and $f_n \to f$ uniformly on $[a,b]$. Then, f is integrable, and

$$\lim_{n\to\infty} \int_a^b f_n(x)\,\mathrm{d}x = \int_a^b \lim_{n\to\infty} f_n(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x.$$

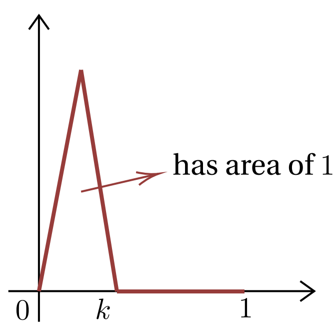
Proof 1. Assume f is integrable. Then,

$$\begin{aligned}\left|\int_a^b f_n(x)\,\mathrm{d}x - \int_a^b f(x)\,\mathrm{d}x\right| &\le \int_a^b \underbrace{|f_n(x) - f(x)|}_{<\varepsilon \quad \forall x, \text{ by UC}}\,\mathrm{d}x \\ &< \int_a^b \varepsilon\,\mathrm{d}x = \varepsilon(b-a).\end{aligned}$$

Q.E.D. ■

Remark 5.4 *One cannot replace uniform convergence by pointwise convergence.*

Example 5.2.2



Define $f_k(x) : [0,1] \to \mathbb{R}$ *s.t.*

$$\int_0^1 f_k(x)\,\mathrm{d}x = 1.$$

Observe that $f_k(x) \xrightarrow{\text{pointwise}} f(x) \equiv 0 \quad \forall x \in [0,1]$. So,

$$\int_0^1 f_k(x)\,\mathrm{d}x \quad \not\to \int_0^1 f(x)\,\mathrm{d}x$$

Remark 5.5 *The same result is not true for differentiation. One cannot simply replace integrable with differentiable. For example, consider*

$$f_n(x) = \frac{x^{n+1}}{n+1} \quad \textit{on } [0,1] \implies f_n'(x) = x^n.$$

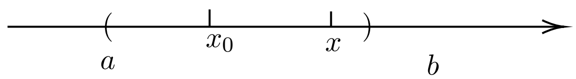
We have $f_n(x) \xrightarrow{UC} f(x) \equiv 0$. *However,*

$$\lim_{n\to\infty} f_n'(x) \neq \lim_{n\to\infty} f'(x).$$

Theorem 5.2.3
Let $f_n : (a,b) \to \mathbb{R}$ be differentiable, converging pointwise to $f(x) : (a,b) \to \mathbb{R}$. If $f_n'(x)$ are continuous and converges uniformly to a function g, then $f'(x) = g(x)$. i.e.,

$$\lim_{n\to\infty} \frac{\mathrm{d}}{\mathrm{d}x}(f_n(x)) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\lim_{n\to\infty} f_n(x)\right) = \frac{\mathrm{d}}{\mathrm{d}x} f(x) = g(x).$$

Proof 2.



Use Fundamental Theorem of Calculus,

$$\begin{aligned} f_n(x) &= f_n(x_0) + f_n(x) - f_n(x_0) \\ &= f_n(x_0) + \int_{x_0}^{x} f_n'(t)\,\mathrm{d}t. \end{aligned}$$

When $n \to \infty$, for fixed $x \in A$,

$$f_n(x) \to f(x), \quad f_n(x_0) \to f(x_0), \quad \int_{x_0}^{x} f_n'(t)\,\mathrm{d}t \to \int_{x_0}^{x} g(t)\,\mathrm{d}t.$$

So,

$$\begin{aligned} f(x) &= f(x_0) + \int_{x_0}^{x} g(t)\,\mathrm{d}t \\ \frac{\mathrm{d}}{\mathrm{d}x}(f(x)) &= \frac{\mathrm{d}}{\mathrm{d}X}(f(x_0)) + \frac{\mathrm{d}}{\mathrm{d}x}\int_{x_0}^{x} g(t)\,\mathrm{d}t \\ \lim_{n\to\infty} f_n'(x) &= f'(x) = 0 + g(x) = g(x). \end{aligned}$$

Q.E.D. ■

Example 5.2.4 One cannot replace UC with pointwise convergence

$$f_n = \frac{nx^2}{1+nx^2}, \quad -1 \leq x \leq 1 \implies f_n'(x) \xrightarrow{\text{pointwise}} g(x)$$

However, $f_n'(x) \neq g(x)$.

5.3 The Space of Continuous Functions

Notation 5.1. Let $A \subset M$ be a metric space and N is a normal vector space. Then

- $\mathcal{C} = \mathcal{C}(A, N) = \{f \mid f\ A \to N \text{ continuous}\}$: the collection of all continuous functions $f : A \to N$
- $\mathcal{C}_b = \mathcal{C}(A, N) = \{f \in \mathcal{C} \mid f \text{ is bounded}\}$: the collection of all bounded continuous functions ($\exists\, M\ s.t.\ \|f(x)\|_N \leq M \quad \forall\, x \in A$)

Example 5.3.2

$A = [0, 1] \subset \mathbb{R}$, $N = \mathbb{R}$. Then,

$$\mathcal{C}_b = \mathcal{C}, \quad \text{the set of all continuous functions on } [0, 1].$$

Remark 5.6

- *C_b and $\mathcal{C}$ are vector spaces;*
- *Goal: Study $\mathcal{C}_b$ as a normed vector spaces as $\mathbb{R}^n$.*

Definition 5.3.3 (Norm on $\mathcal{C}_b$). Given $f \in \mathcal{C}_b$. Define $\|f\|$ as follows:

$$\|f\| = \sup\{\|f(x)\|_N \mid x \in A\}.$$

This is called the *maximum absolute value norm.*

Theorem 5.3.4

$\|\cdot\|$ defined in Definition 5.3.3 is a norm in $\mathcal{C}_b$. i.e.,

- Positive definiteness: $\|f\| \geq 0$ and $\|f\| = 0 \iff f = 0$;
- Scalar multiplicity: $\|\alpha f\| = |\alpha| \|f\| \quad \forall\, \alpha \in \mathbb{R}$
- Triangle inequality: $\|f + g\| \leq \|f\| + \|g\|$

Proof 1. (of ③) By definition, $\|f + g\| = \sup\{\|f(x) + g(x)\|_N \mid x \in A\}$. [WTS: $\|f\| + \|g\|$ is an upper bound.] Note that

$$\begin{aligned} \|f(x) + g(x)\|_N &\leq \|f(x)\|_N + \|g(x)\|_N && \text{[triangle inequality in } N\text{]} \\ &\leq \|f\| + \|g\| && \text{[definition]} \end{aligned}$$

So, $\|f + g\| \leq \|f\| + \|g\|$.

Q.E.D. ■

Definition 5.3.5 (Convergence in $\mathcal{C}_b$). $f_k \to f$ in $\mathcal{C}_b$ means that $\|f_k - f\| \to 0$ as $k \to \infty$.

Theorem 5.3.6

$f_k \to f$ in $\mathcal{C}_b$ (convergence in norm as vectors) $\iff$ $f_k \to f$ uniformly on A (convergence in function)

Proof 2. ($\Rightarrow$): Assume $\|f_k - f\| \to 0$. Then, $\forall\,\varepsilon > 0,\ \exists\,K\ s.t.\ k \geq K \implies \|f_k - f\| \leq \varepsilon$. Thus, $\forall x \in A$, by definition of norm, for $k \geq K$,

$$\|f_k(x) - f(x)\|_N \leq \|f_k - f\| < \varepsilon.$$

So, $f_k(x) \to f(x)$ uniformly on A. $\square$

($\Leftarrow$): Assume $f_k(x) \to f(x)$ uniformly on A. Then, $\forall\,\varepsilon > 0, \exists\,K\ s.t.\ k \geq K \implies \|f_k(x) - f(x)\|_N < \varepsilon$. Then, ε is an upper bound. Note that

$$\|f_k - f\| = \sup\left\{\|f_k(x) - f(x)\|_N \mid x \in\right\}$$

is a least upper bound. So,

$$\|f_k - f\| = \sup\left\{\|f_k(x) - f(x)\|_N \mid x \in A\right\} < \varepsilon$$

So, $\|f_k f\| \to 0$ as $k \to \infty$.

Q.E.D. ■

> **Theorem 5.3.7 Completeness of $\mathcal{C}_b$**
> If N is complete, so is $\mathcal{C}_b(A, N)$.

Proof 3. Let $\{f_k\}$ be a Cauchy sequence in $\mathcal{C}_b$. Then, $\forall\,\varepsilon > 0, \exists\,K\ s.t.\ j, k \geq K \implies \|f_j - f_k\| < \varepsilon$. By definition, we have

$$\|f_j(x) - f_k(x)\|_N \leq \|f_j - f_k\| < \varepsilon \quad \forall x \in A.$$

So, $\{f_k(x)\}$ is a uniform Cauchy sequence on A. By Cauchy criterion,

$$f_k(x) \to f(x) \quad \text{uniformly on } A.$$

f is also continuous since UC preserves continuity. By Theorem 5.3.6, we have $f_k \to f$ in $\mathcal{C}_b$. So, $\mathcal{C}_b$ is complete.

Q.E.D. ■

Remark 5.7 (Comparison Between $\mathcal{C}_b$ and $\mathbb{R}^n$) *Let $A \subset M$ be compact and $N = \mathbb{R}^n$.*

Properties	$\mathbb{R}^n$	$\mathcal{C}_b(A, N = \mathbb{R}^n)$
Normed Space	✓	✓
Completeness	✓	✓
Finite Dimension	✓	✗
Compact Subset	*<u>Heine-Borel</u>:* *$B \subset \mathbb{R}^n$ is compact* *$\iff$ B is closed and bounded*	*<u>Arzela-Ascoli</u>: $A \subset M$ compact.* *Then, $\mathcal{B} \subset \mathcal{C}_b$ is compact* *$\iff$ $\mathcal{B}$ is closed, bounded, and equicontinuous in A*

Definition 5.3.8 (Equicontinuous). A family of function $\mathcal{B}$ is equicontinuous at a point $x \in A$ if $\forall \varepsilon > 0,\ \exists \delta > 0\ s.t.\ y \in D(x, \delta) \cap A \implies \|f(x) - f(y)\|_N < \varepsilon \quad \forall f \in \mathcal{B}$.

Remark 5.8 *δ is independent of $f \in \mathcal{B}$.*

Example 5.3.9

- $\mathcal{B} = \{f \in \mathcal{C}_b(\mathbb{R}, \mathbb{R}) \mid f(x) > 0 \quad \forall x \in \mathbb{R}\}$.
  - Is $\mathcal{B}$ open? No.

    Suppose $f \to 0$ as $x \in \infty$. Then, no matter how small we take the δ, some part of $D(f, \delta)$ will not be contained in $\mathcal{B}$.
  - What is $\mathrm{cl}(\mathcal{B})$?
    $$\mathrm{cl}(\mathcal{B}) = \{f \in \mathcal{C}_b(\mathbb{R}, \mathbb{R}) \mid f(x) \geq 0 \quad \forall x \in \mathbb{R}\}.$$
  - What is $\mathrm{int}(\mathcal{B})$?
    $$\mathrm{int}(\mathcal{B}) = \{f \in \mathcal{C}_b(\mathbb{R}, \mathbb{R}) \mid \inf(f(x)) > 0 \quad \forall x \in \mathbb{R}\}.$$

    Think of $\inf(f(x)) > 0$ in this way: we need a buffer zone.
- $\mathcal{B} = \{f \in \mathcal{C}_b([0, 1], \mathbb{R}) \mid f(x) > 0 \quad \forall x \in [0, 1]\}$.

5.4 The Contraction Mapping Principle (CMP)

Theorem 5.4.1 CMP
Let (M, d) be a complete metric space, and $\Phi : M \to M$ be a map. Suppose $\exists$ constant $k\ s.t.\ 0 < k < 1\ s.t.$
$$d(\Phi(x), \Phi(y)) \leq k \cdot d(x, y) \quad \forall x, y \in M.$$

Then,

- Φ has a unique fixed point in M. That is, $\exists! x^* \in M\ s.t.\ \Phi(x^*) = x^*$.
- The fixed point can be constructed (or approximated) as follows:

  Fix any point $x_0 \in M$. Let $x_1 = \Phi(x_0), x_2 = \Phi(x_1), \ldots, x_{n+1} = \Phi(x_n), \ldots$. Then,
  $$\lim_{n \to \infty} x_n = x^*.$$

Remark 5.9 *Φ is continuous. Further, Φ is Lipschitz $\implies$ Φ is uniform continuous.*

Proof 1. Fix $x_0 \in M$. Let $x_{n+1} = \Phi(x_n)$ for $n = 0, 1, 2, \ldots$.

Claim *$\{x_n\}$ is Cauchy.*

Note that $\forall\, n \geq 1$,

$$\begin{aligned} d(x_n, x_{n+1}) = d(\Phi(x_{n-1}), \Phi(x_n)) &\leq kd(x_{n-1}, x_n) \\ &\leq k^2 d(x_{n-1}, x_{n-1}) \\ &\vdots \\ &\leq k^n d(x_0, x_1). \end{aligned}$$

Thus, $\forall\, p \geq 1$,

$$\begin{aligned} d(x_n, x_{n+p}) &\leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \cdots + d(x_{n+p-1}, x_{n+p}) \\ &\leq k^n d(x_0, x_1) + k^{n+1} d(x_0, x_1) + \cdots + k^{n+p-1} d(x_0, x_1) \\ &= \underbrace{\left(k^n + k^{n+1} + \cdots + k^{n+p-1}\right)}_{\text{geometric series}} d(x_0, x_1) \to 0 \quad \text{as} \quad n \to \infty. \end{aligned}$$

As the geometric series converges, $\{x_n\}$ is Cauchy.

Since M is complete, $x_n \to x^* \in M$.

Claim x^* *is a fixed point.*

Since Φ is continuous,

$$\lim_{n\to\infty} \Phi(x_n) = \Phi\left(\lim_{n\to\infty} x_n\right) = \Phi(x^*).$$

Meanwhile, $\Phi(x_n) = x_{n+1}$, so

$$\lim_{n\to\infty} \Phi(x_n) = \lim_{n\to\infty} x_{n+1} = x^*.$$

Hence, $x^* = \Phi(x^*)$, implying x^* is a fixed point.

Claim *The fixed point is unique.*

Let $y^* \in M$ be another fixed point. One can show

$$\begin{aligned} d(x^*, y^*) &\leq d(\Phi(x^*), \Phi(y^*)) && [x^*, y^* \text{ are fixed points}] \\ &\leq kd(x^*, y^*) && [\Phi \text{ is a contraction mapping}] \end{aligned}$$

$\implies d(x^*, y^*) = 0.$

Q.E.D. ■

Example 5.4.2 Application in ODE

Consider the following initial value problem (IVP):

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(t, x) \quad x(t_0) = x_0 \tag{IVP}$$

- Basic Assumptions:

1. $f(t,x)$ is continuous in a neighborhood U of $(t_0, x_0) \in \mathbb{R}^2$
2. $f(t,x)$ is Lipschitz in x: $\exists$ constant K $s.t.$

$$|f(t,x_1) - f(t,x_2)| \leq K|x_1 - x_2| \quad \forall (t_1, x_1), (t_1, x_2) \in U$$

- Apply CMP:

Theorem 5.4.3
If $f(t,x)$ is continuous in U an Lipschitz in x, then (IVP) has a unique solution $x = \varphi(t)$ in the neighborhood of t_0: $(t_0 - \delta, t_0 + \delta)$. i.e.,

$$\varphi'(t) = f(t, \varphi(t)), \quad \varphi(t_0) = x_0.$$

- Solving (IVP) is equivalent to finding a function $\varphi(t)$ $s.t.$

$$\varphi'(t) = f(t, \varphi(t)).$$

Or, by integration:

$$\varphi'(t) = x_0 + \int_{t_0}^{t} f(s, \varphi(s))\,\mathrm{d}s \quad [x_0 \text{ comes from plugging in the initial condition}]$$

This is just a fixed point for the following map (an integral operator):

$$\Phi : g(t) \longmapsto \Phi(g) = x_0 + \int_{t_0}^{t} f(s, g(s))\,\mathrm{d}s$$

Theorem 5.4.4
We need to construct an appropriate metric space $M \subset \mathcal{C}_b$ $s.t.$ $\Phi : M \to M$ is a contraction mapping.

Algorithm 1: Iterative Method to Approximate the Solution to (IVP)

begin
  $\varphi_0 \equiv x_0$;
  for $n = 0, 1, 2, \ldots$ **do**
    $\varphi_{n+1}(t) = \Phi(\varphi_n(t)) = x_0 + \int_{t_0}^{t} f(s, \varphi_n(s))\,\mathrm{d}s$;

Example 5.4.5

Consider the IVP: $f(t,x) = tx^2 + x^3, \quad x(0) = 1$.

Let $\varphi_0(t) = 1$. Then,

$$\begin{aligned}
\varphi_1(t) &= 1 + \int_0^t s\varphi_0(s)^2 + \varphi_0(s)^3 \,\mathrm{d}s \\
&= 1 + \int_0^t s + 1 \,\mathrm{d}s \\
&= 1 + \left[\frac{1}{2}s^2 + s\right]_0^t \\
&= 1 + \frac{1}{2}t^2 + t \\
\varphi_2(t) &= 1 + \int_0^t s\varphi_1(s)^2 + \varphi_1(s)^3 \,\mathrm{d}s \\
&= 1 + \int_0^t s\left(1 + \frac{1}{2}s^2 + s\right)^2 + \left(1 + \frac{1}{2}s^2 + s\right)^3 \mathrm{d}s \\
&\vdots
\end{aligned}$$

6 Differential Mappings

6.1 Definition and Matrix Representation of a Differential

Definition 6.1.1 (Linear Transformation). A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a *linear transformation* if $\forall\, x, y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, we have

- $T(x + y) = T(x) + T(y)$
- $T(\lambda x) = \lambda T(x)$

These two properties can be combined and written equivalently as $T(ax+by) = aT(x)+bT(y) \quad \forall\, x, y \in \mathbb{R}^n$ and $\forall\, a, b \in \mathbb{R}$.

6.1.2 Matrix Representation of T.

Observation: Given $m \times n$ matrix A, define function $T : \mathbb{R}^n \to \mathbb{R}^m$ by $T(x) = A \cdot x$. Then, T is a linear transformation.

Proof 1.

$$T(ax + by) = A(ax + by) = A(ax) + A(by) = aAx + bAy = aT(x) + bT(y).$$

Q.E.D. ∎

Example 6.1.3

Suppose $A = \begin{bmatrix} 2 & 3 \\ 1 & -1 \\ 0 & 4 \end{bmatrix}$. Then,

$$T(x) = A \cdot x = \begin{bmatrix} 2 & 3 \\ 1 & -1 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2 \\ x_1 - x_2 \\ 4x_2 \end{bmatrix} \in \mathbb{R}^3.$$

Theorem 6.1.4 Fact
Every linear transformation T is determined by a matrix in such a way as above (via matrix multiplication).

Proof 2. Given $T : \mathbb{R}^n \to \mathbb{R}^m$ linear, we need to find a matrix A $(m \times n)$ such that

$$T(x) = A \cdot x \quad \forall\, x \in \mathbb{R}^n.$$

To construct A, consider the standard basis for $\mathbb{R}^n : \{e_1, e_2, \ldots, e_n\}$ and for $\mathbb{R}^m : \{e'_1, e'_2 \ldots, e'_m\}$. Then,

$$T(e_j) = \sum_{i=1}^{m} a_{ij} e'_i, \quad \forall\, j = 1, 2, \ldots, n.$$

Let $A = \left(a_{ij}\right)_{m\times n}$.

Claim $T(x) = Ax \quad \forall x \in \mathbb{R}^n$.

In fact, let $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$. Then, we can rewrite x as a linear combination of standard basis:

$$x = \sum_{j=1}^{n} x_j e_j.$$

So,

$$T(x) = x_1T(e_1) + x_2T(e_2) + \cdots + x_nT(e_n) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = Ax. \qquad [T \text{ is Linear}]$$

Q.E.D. ■

Remark 6.1 *The collection of* {*linear transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$} *forms a* 1*-to-*1 *correspondence with the collection of* {$m \times n$ *matrices* A}.

Theorem 6.1.5 Continuity of T

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then it is Lipschitz, and hence continuous.

Proof 3. Recall the definition of Lipschitz: $|f(x) - f(y)| \le L \cdot |x - y|$.
Since $T(x) - T(y) = T(x - y)$, we only need to show that

$$\|T(x)\| \le M \cdot \|x\| \text{ for some } M \in \mathbb{R}.$$

Let $x = \sum x_j e_j$. Then, $T(x) = \sum x_j T(e_j)$. So, $\|T(x)\| \le \sum_j |x_j| \cdot \|T(e_j)\|$.

Recall that $\|x\| = \sqrt{\sum_j x_j^2}$. So, $|x_j| \le \|x\|$. Hence,

$$\|T(x)\| \le \sum_j \|x\| \cdot \|T(e_j)\| = \underbrace{\left(\sum_{j=1}^{n} \|T(e_j)\|\right)}_{M,\text{ independent of } x} \cdot \|x\| = M \cdot \|x\|$$

Q.E.D. ■

6.1.6 Derivative (Differential) as a Linear Transformation.

- Recall one variable case: Let $f:(a,b)\to\mathbb{R}$. Then, we can rewrite $f'(x_0)=\lim\limits_{x\to x_0}\dfrac{f(x)-f(x_0)}{x-x_0}$ as

$$\lim_{x\to x_0}\left[\frac{f(x)-f(x_0)-f'(x_0)(x-x_0)}{x-x_0}\right]=0.$$

- **Definition 6.1.7 (Generalization to Higher Dimensions).**A map $f:A\subset\mathbb{R}^n\to\mathbb{R}^m$ is said to be *differentiable* at $x_0\in A$ if there is a linear map, denoted by $\mathbb{D}f(x_0):\mathbb{R}^n\to\mathbb{R}^m$ with

$$\lim_{x\to x_0}\frac{\|f(x)-f(x_0)-\mathbb{D}f(x_0)(x-x_0)\|}{\|x-x_0\|} \tag{$\star$}$$

Remark 6.2 *Interpretations of* ($\star$)*:*

1. *Rewrite* ($\star$)*:* $\forall\,\varepsilon>0,\ \exists\,\delta>0\ s.t.\ \forall\,x\in A$,

$$\|x-x_0\|>\delta\implies\|f(x)-f(x_0)-\mathbb{D}f(x_0)(x-x_0)\|<\varepsilon\|x-x_0\|.$$

2. $f(x)\approx f(x_0)+\underbrace{\mathbb{D}f(x_0)\cdot(x-x_0)}_{\textit{linear map}}$ *is called the* affine map.

3. *Geometric Interpretation:* $z=f(x):\mathbb{R}^n\to\mathbb{R}^1$*. Then,* $z-f(x_0)=\mathbb{D}f(x_0)(x-x_0)$ *represents the tangent plane of the surface* $z=f(x)$.

4. *For* $f:\mathbb{R}^1\to\mathbb{R}^1$, $\mathbb{D}f(x)$ *is the differential, representing a linear transformation, whereas* $f'(x)$ *or* $\dfrac{\mathrm{d}f}{\mathrm{d}x}$ *is the derivative, which is just a number.*
   For example, $f(x)=x^2$*. Then,* $f'(x)=2x$*. However,* $\mathbb{D}f(x)$ *is a linear transformation* $\mathbb{R}^1\to\mathbb{R}^1$*, defined as*

$$\mathbb{D}f(x)(h)=2xh,\quad\forall\,h\in\mathbb{R}^1.$$

- Uniqueness of Differential

Theorem 6.1.8
Let $A\in\mathbb{R}^n$ be open and $f:A\to\mathbb{R}^m$ be differentiable at $x_0\in A$. Then, the differential $\mathbb{D}f(x_0)$ is uniquely determined by f.

Proof 4. Let L_1 and L_2 be two linear transformations such that

$$\lim_{x\to x_0}\frac{\|f(x)-f(x_0)-L_1(x-x_0)\|}{\|x-x_0\|}=0=\lim_{x\to x_0}\frac{\|f(x)-f(x_0)-L_2(x-x_0)\|}{\|x-x_0\|}.$$

We need to show that $L_1=L_2$. i.e., $L_1(h)=L_2(h)\quad\forall\,h\in\mathbb{R}^n$.

Fix any unit vector $e\in\mathbb{R}^n$. Let $x=x_0+te$, where $t\in\mathbb{R}$ and $t\neq 0$ (*This makes sense because* A *is* *open by assumption*). Then,

$$\begin{aligned}
\|L_1(e) - L_2(e)\| &= \frac{\|L_1(te) - L_2(te)\|}{|t|} \\
&= \frac{\|L_1(x-x_0) - L_2(x-x_0)\|}{\|x-x_0\|} \qquad [|t| = \|x-x_0\|] \\
&= \frac{\|L_1(x-x_0) - (f(x)-f(x_0)) + (f(x)-f(x_0)) - L_2(x-x_0)\|}{\|x-x_0\|} \\
&\le \frac{\|L_1(x-x_0) - (f(x)-f(x_0))\| + \|(f(x)-f(x_0)) - L_2(x-x_0)\|}{\|x-x_0\|} \\
&= \frac{\|L_1(x-x_0) - (f(x)-f(x_0))\|}{\|x-x_0\|} + \frac{\|(f(x)-f(x_0)) - L_2(x-x_0)\|}{\|x-x_0\|}.
\end{aligned}$$

Note that both parts $\to 0$ as $x \to x_0$. So, $\|L_1(e) - L_2(e)\| = 0$, and thus $L_1(e) = L_2(e) \quad \forall$ unit vector e. Using linear transformation, $L_1(h) = L_2(h) \quad \forall\, h \in \mathbb{R}^n$.

Q.E.D. ■

Remark 6.3 *Theorem 6.1.8 is not true if A is not open. A trivial example would be when $A = \{x_0\}$, the set of just one point. Then, any linear map satisfies the differential definition. That is,*

$$\lim_{\substack{x \to x_0 \\ x \in A}} \frac{\|f(x) - f(x_0) - T(x-x_0)\|}{\|x-x_0\|} = 0 \quad \forall \textit{ linear map } T.$$

Or, equivalently, $\|f(x) - f(x_0) - T(x-x_0)\| < \varepsilon\|x-x_0\|$.

6.1.9 Matrix Representation of the Differential $\mathbb{D}f(x)$.

Question: Given f, how do we find the linear transformation $\mathbb{D}f(x)$?

Definition 6.1.10 (Partial Derivative). Write $f(x) = \Big(f_1(x_1,\dots,x_n), f_2(x_1,\dots,x_n),\dots,f_m(x_1,\dots,x_n)\Big) \in \mathbb{R}^m$. Then,

$$\frac{\partial f_j}{\partial x_i} = \lim_{h\to 0} \frac{f_j(x_1,\dots,x_{i-1},x_i+h,x_{i+1},\dots,x_n) - f_j(x_1,\dots,x_i,\dots,x_n)}{h}.$$

Theorem 6.1.11 Relation Between Differential $\mathbb{D}f(x)$ and Partial Derivatives
Suppose $A \subset \mathbb{R}^n$ is open and $f : A \to \mathbb{R}^m$ is differentiable at $x \in A$. Then, $\dfrac{\partial f_j}{\partial x_i}$ exists and the matrix of the linear map $\mathbb{D}f(x)$ is given by

$$\begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m\times n},$$

and we denotes this matrix as $J_f(x)$, the *Jacobian matrix* of f at x.

Proof 5. Denote the matrix of $\mathbb{D}f(x)$ by $B = \Big(b_{ji}\Big)_{m\times n}$. We need to show $b_{ji} = \dfrac{\partial f_j}{\partial x_i}$.

Recall: $b_{ji} = j-$th component of $\mathbb{D}f(x)(e_i) = \sum_{j=1}^{m} b_{ji}e'_j$. Fix i, j and let $y = x + he_i, \quad h \in \mathbb{R}$. Then, by definition of differential,

$$\frac{\|f(y) - f(x) - \mathbb{D}f(x)(y-x)\|}{\|y-x\|} \to 0 \quad \text{as } y - x \to 0.$$

Taking the j-th component,

$$\frac{|f_j(x_1, \ldots, x_i + h, \ldots, x_n) - f_j(x_1, \ldots, x_n) - b_{ji} \cdot h|}{|h|} \to 0 \quad \text{as } h \to 0.$$

So,

$$\lim_{h \to 0} \frac{f_j(x_1, \ldots, x_i + h, \ldots, x_n) - f_j(x_1, \ldots, x_n)}{h} = b_{ji}.$$

Hence,

$$\frac{\partial f_j}{\partial x_i} = b_{ji} \quad \forall\, i, j.$$

So, $\mathbb{D}f(x)$ is determined by the Jacobian matrix $J_f(x)$.

Q.E.D. ■

Example 6.1.12

- $f(x, y, z) = (x^4 y, xe^z) : \mathbb{R}^3 \to \mathbb{R}^2$.

$$J_f(x,y,z) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x} & \dfrac{\partial f_1}{\partial y} & \dfrac{\partial f_1}{\partial z} \\ \\ \dfrac{\partial f_2}{\partial x} & \dfrac{\partial f_2}{\partial y} & \dfrac{\partial f_2}{\partial z} \end{bmatrix} = \begin{bmatrix} 4x^3y & x^4 & 0 \\ e^z & 0 & xe^z \end{bmatrix}.$$

- Special Case: $m = 1$: $f : \mathbb{R}^n \to \mathbb{R}$. Then,

$$J_f(x) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} & \cdots & \dfrac{\partial f}{\partial x_n} \end{bmatrix} \quad \text{is a } 1 \times n \text{ matrix.}$$

  Definition 6.1.13 (Gradient). The *gradient*, $\operatorname{grad} f$ or ∇f, is defined by the following vector:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right).$$

  Gradient points towards the direction of fastest growth.

- $f(x,y,z) = \dfrac{x \sin y}{z}$. Computing $\mathbb{D}f$ and ∇f.

  Solution 6.

$$\mathbb{D}f(x) = J_f(x) = \begin{bmatrix} \dfrac{\sin y}{z} & \dfrac{x \cos y}{z} & -\dfrac{x \sin y}{z^2} \end{bmatrix}.$$

$$\nabla f(x) = \left(\frac{\sin y}{z}, \frac{x\cos y}{z}, -\frac{x \sin y}{z^2}\right).$$

$\square$

Remark 6.4 (Relation Between $\mathbb{D}f(x)$ and ∇f) *For any $h \in \mathbb{R}^n$, we have*

$$\textit{matrix multiplication} \leftarrow \quad \mathbb{D}f(x)h = \langle \nabla f, h\rangle \quad \rightarrow \textit{inner product/dot product}$$

- Special Case: $n = 1$. Consider $x = c(t) : [a, b] \subset \mathbb{R} \to \mathbb{R}^m$. Then,

$$\mathbb{D}x(t) = c'(t) = \Big(c_1'(t), c_2'(t), \ldots, c_m'(t)\Big)$$

is the tangent vector.

6.2 Necessary and Sufficient Conditions for Differentiability

Definition 6.2.1 (Locally Lipschitz). f is *locally Lipschitz* at x_0 if $\forall\, x_0 \in A$, $\exists\, \delta > 0$ and M $s.t.$

$$\|x - x_0\| < \delta \implies \|f(x) - f(x_0)\| < M \cdot \|x - x_0\|.$$

Theorem 6.2.2 Necessary Condition for Differentiability I
Suppose $A \subset \mathbb{R}^n$ is open and $f : A \to \mathbb{R}^m$ is differentiable. Then, f is locally Lipschitz.

Remark 6.5 (Ideas to Prove this Theorem)

- *Linear map $\mathbb{D}f(x)$ is Lipschitz;*
- *$f(x)$ can be approximated by $\mathbb{D}f(x_0)$ locally.*

Proof 1. Fix $x_0 \in A$. By definition,

$$\lim_{x \to x_0} \frac{\big\|f(x) - f(x_0) - \mathbb{D}f(x_0)(x - x_)\big\|}{\|x - x_0\|} = 0.$$

For $\varepsilon = 1$, $\exists\, \delta > 0$ $s.t.$

$$\|x - x_0\| < \delta \implies \|f(x) - f(x_0) - \mathbb{D}f(x_0)(x - x_0)\| \le \varepsilon \cdot \|x - x_0\| = \|x - x_0\|.$$

By triangle inequality,

$$\|f(x) - f(x_0)\| \le \|\mathbb{D}f(x_0)(x - x_0)\| + \|x - x_0\|.$$

Since $\mathbb{D}f(x_0)$ is Lipschitz, $\exists\, L\ s.t.$

$$\|\mathbb{D}f(x_0)(x-x_0)\| \leq L \cdot \|x-x_0\|.$$

So, $\|x-x_0\| < \delta \implies$

$$\begin{aligned}\|f(x)-f(x_0)\| &\leq L\cdot\|x-x_0\| + \|x-x_0\| \\ &= \underbrace{(L+1)}_{M}\cdot\|x-x_0\| \\ &= M\cdot\|x-x_0\|.\end{aligned}$$

Q.E.D. ■

Remark 6.6

- *Continuity is not sufficient to guarantee differentiability. For instance,* $f(x)=|x|$.

  However, differentiability $\implies$ *continuity.*

- *Derivative of a differentiable function may not be continuous. For example, consider the function* $f(x)=\begin{cases} x^2\sin\dfrac{1}{x}, & x\neq 0 \\ 0, & x=0.\end{cases}$ *;* $f:\mathbb{R}^1\to\mathbb{R}^1$*. Then, we have*

$$f'(0)=\lim_{x\to 0}\frac{f(x)-f(0)}{x}=\lim_{x\to 0}\frac{x^2\sin\frac{1}{x}}{x}=\lim_{x\to 0}x\sin\frac{1}{x}=0.$$

  When $x\neq 0$,

$$f'(x)=2x\sin\frac{1}{x}+x^2\cos\frac{1}{x}(-\frac{1}{x^2})=1x\sin\frac{1}{x}-\cos\frac{1}{x}.$$

  Conclusion: f *is differentiable in* $\mathbb{R}^1$*. However,*

$$f'(x)=\begin{cases} 2x\sin\dfrac{1}{x}-\cos\dfrac{1}{x}, & x\neq 0 \\ 0, & x=0\end{cases}$$

  is not continuous at $x=0$.

Theorem 6.2.3 Necessary Condition for Differentiability II

Suppose $f: A\subset\mathbb{R}^n\to\mathbb{R}^m$ is differentiable. Then, the partial derivatives, $\dfrac{\partial f_j}{\partial x_i}$, exists $\forall\, i,j$.

Example 6.2.4 The Converse is not True

The converse of Theorem 6.2.3 is, in general, not true. Here we will consider a counterexample.

Consider function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x,y) = \begin{cases} \dfrac{xy}{\sqrt{x^2+y^2}}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0) \end{cases}$$

Claim *f is continuous at $(0,0)$.*

In fact, we have $(a-b)^2 \geq 0 \implies a^2 - 2ab + b^2 \geq 0$. So,

$$ab \leq \frac{a^2+b^2}{2} \quad a, b \in \mathbb{R}.$$

Then,

$$|xy| \leq \frac{1}{2}(a^2+b^2) \implies \frac{xy}{\sqrt{x^2+y^2}} \to 0 \quad \text{as } (x,y) \to (0,0).$$

Claim $\dfrac{\partial f(0,0)}{\partial x} = 0$ *and* $\dfrac{\partial f(0,0)}{\partial y} = 0$.

$$\frac{\partial f(0,0)}{\partial x} = \lim_{x\to 0} \frac{f(x,0) - f(0,0)}{x - 0} = \lim_{x \to 0} \frac{0-0}{x} = 0.$$

Claim *f is not differentiable at $(0,0)$.*

If f were differentiable, the matrix of $\mathbb{D}f(0,0)$ is given by

$$J_f(0,0) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = \Big(0, 0\Big).$$

However, note that

$$\frac{\|f(x,y) - f(0,0) - \mathbb{D}f(x,y)\|}{\|(x,y) - (0,0)\|} = \frac{\dfrac{|xy|}{\sqrt{x^2+y^2}}}{\sqrt{x^2+y^2}} = \frac{|xy|}{x^2+y^2}.$$

Since $\dfrac{|xy|}{x^2+y^2}$ does not $\to 0$ as $(x,y) \to (0,0)$, f is not differentiable at $(0,0)$.

Conclusion: Continuity + Existence of Partial Derivative $\dfrac{\partial f_j}{\partial x_i} \quad \not\Longrightarrow \quad$ Differentiability.

Theorem 6.2.5 Sufficient Condition for Differentiability
Let $A \subset \mathbb{R}^n$ be open and $f = (f_1, \ldots, f_m) : A \to \mathbb{R}^m$. If all the partials $\dfrac{\partial f_j}{\partial x_i}$ exist and continuous on A, then f is differentiable on A.

Proof 2. WTS: $\forall\, x \in A$,

$$\lim_{y \to x} \frac{\|f(y) - f(x) - J_f(x)(y-x)\|}{\|y-x\|} = 0.$$

It is sufficient to show that this is true for each component f_j of $f = (f_1, f_2, \ldots, f_m)$. Thus, we may assume $m = 1$: $f : A \subset \mathbb{R}^n \to \mathbb{R}^1$. Then,

$$J_f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} & \cdots & \dfrac{\partial f}{\partial x_n} \end{pmatrix}.$$

So,

$$J_f(y-x) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(y_i - x_i),$$

and

$$\begin{aligned} f(y) - f(x) &= f(y_1, y_2, \ldots, y_n) - f(x_1, x_2, \ldots, x_n) \\ &= f(y_1, y_2, \ldots, y_n) - f(x_1, y_2, \ldots, y_n) \\ &\quad + f(x_1, y_2, \ldots, y_n) - f(x_1, x_2, \ldots, y_n) \\ &\quad + f(x_1, x_2, \ldots, y_n) - \cdots \qquad \text{each time, we change one component} \\ &\quad + f(x_1, x_2, \ldots, y_n) - f(x_1, x_2, \ldots, x_n) \end{aligned}$$

By MVT,

$$f(y_1, y_2, \ldots, y_m) - f(x_1, y_2, \ldots, y_n) = \frac{\partial f}{\partial x_1}(u_1, y_2, \ldots, y_n) \cdot (y_1 - x_1).$$

Applying MVT to other terms, we obtain

$$f(y) - f(x) = \frac{\partial f(z^{(1)})}{\partial x_1}(y_1 - x_1) + \frac{\partial f(z^{(2)})}{\partial x_2}(y_2 - x_2) + \cdots + \frac{\partial f(z^{(n)})}{\partial x_n}(y_n - x_n).$$

Thus,

$$\begin{aligned} \|f(y) - f(x) - J_f(x)(y-x)\| &= \left\| \sum_{i=1}^{n} \frac{\partial f(z^{(i)})}{\partial x_i}(y_i - x_i) - \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i}(y_i - x_i) \right\| \\ &\le \sum_{i=1}^{n} \left| \frac{\partial f(z^{(i)})}{\partial x_i} - \frac{\partial f(x)}{\partial x_i} \right| \cdot \|y - x\| \qquad \text{Triangle Inequality: } |y_i - x_i| \le \|y - x\| \end{aligned}$$

By continuity of partial derivative, $\forall\, \varepsilon > 0$, $\exists\, \delta > 0$ $s.t.$

$$\|y - x\| < \delta \implies \sum_{i=1}^{n} \left| \frac{\partial f(z^{(i)})}{\partial x_i} - \frac{\partial f(x)}{\partial x_i} \right| < \varepsilon$$

Hence,

$$\|f(y) - f(x) - J_f(x)(y-x)\| < \varepsilon\|y-x\|.$$

Q.E.D. ■

Definition 6.2.6 (Directional Derivative). Let $f : \mathbb{R}^n \to \mathbb{R}$ and $e \in \mathbb{R}^n$ be a unit vector. The directional derivative of f at x_0 in the direction e is given by

$$D_e f(x_0) = \frac{\mathrm{d}}{\mathrm{d}t} f(x_0 + te)\bigg|_{t=0} = \lim_{t\to 0} \frac{f(x_0+te) - f(x_0)}{t}.$$

Claim 6.2.7 If f is differentiable at x_0, then $D_e f(x_0) = \mathbb{D}f(x_0)\cdot e$

Proof 3.

$$\begin{aligned} \lim_{t\to 0} \frac{\|f(x_0+te) - f(x_0) - \mathbb{D}f(x_0)(te)\|}{\|te\|} &= 0 \\ \lim_{t\to 0} \frac{f(x_0+te) - f(x_0)}{t} &= \mathbb{D}f(x_0)(e) \\ D_e f(x_0) &= \mathbb{D}f(x_0)(e). \end{aligned}$$

Q.E.D. ■

Remark 6.7 *Exitence of directional derivatives* $\quad\not\Longrightarrow\quad$ *differentiability*

Example 6.2.8

Continuity of f + Existence of directional derivative $\quad\not\Longrightarrow\quad$ differentiability.

Consider function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x,y) = \begin{cases} \dfrac{xy}{\sqrt{x^2+y^2}}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0). \end{cases}$$

Claim $D_e f(0,0)$ *exists for any direction* $e \in \mathbb{R}^2$.

$$\lim_{t\to 0} \frac{f((0,0)+te) - f(0,0)}{t} \quad \text{exists} \quad \forall\, e \in \mathbb{R}^2.$$

Definition 6.2.9 (Tangent Line/Plane).

- The *tangent line* to the curve $y = f(x)$ at x_0 is given by

$$y = f(x_0) + f'(x_0)(x - x_0).$$

- The *tangent plane* to the surface $z = f(x)$ at x_0 is given by

$$z = f(x_0) + \mathbb{D}f(x_0)(x - x_0).$$

Example 6.2.10

Find the tangent plane at $(1,2)$ to the surface $z = x^2 + y^2$.

Solution 4.

$$J_f(x) = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 2x & 2y \end{pmatrix}.$$

The tangent plane is given by

$$\begin{aligned} z &= f(1,2) + \mathbb{D}f(1,2)((x,y) - (1,2)) \\ &= 1^2 + 2^2 + \begin{bmatrix} 2x & 2y \end{bmatrix}\Big|_{(x,y)=(1,2)} \begin{bmatrix} x-1 \\ y-2 \end{bmatrix} \\ &= 5 + \begin{bmatrix} 2 & 4 \end{bmatrix} \begin{bmatrix} x-1 \\ y-2 \end{bmatrix} \\ z &= 5 + 2(x-1) + 4(y-2). \end{aligned}$$

$\square$

Summary III: Relations among Properties of $f : \mathbb{R}^n \to \mathbb{R}^m$

Differentiability

$\implies$ Existence of Directional Derivative

$\implies$ Existence of Partial Derivative (moving in direction of the basis)

$+$ **Theorem 6.2.5** $\implies$ Differentiability

$\not\Longrightarrow$ Continuity

$\implies$ Continuity

$\not\Longrightarrow$ Existence of Partial Derivative

6.3 Differentiation Rules

6.3.1 Chain Rule

Recall the one variable case: $h = g(u)$, $u = f(x)$. Then,

$$h = f \circ f(x) = g(f(x)),$$

and

$$\frac{\mathrm{d}h}{\mathrm{d}x} = \frac{\mathrm{d}h}{\mathrm{d}u} \cdot \frac{\mathrm{d}u}{\mathrm{d}x} = g'(f(x)) \cdot f'(x).$$

Theorem 6.3.1 General Case Chain Rule
Let $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ and $g : B \to \mathbb{R}^p$ be differentiable with $f(A) \subset B$. Then, the composite $g \circ f : A \to \mathbb{R}^p$ is differentiable, and

$$\mathbb{D}(g \circ f)(x) = \mathbb{D}g(f(x)) \circ \mathbb{D}f(x),$$

a composition of linear mappings.
In matrix notation, define $h = g(u)$ and $u = f(x)$. Then, $h = g \circ f(x) = g(f(x))$, and

$$\begin{aligned} J_h(x) &= J_g(f(x)) \cdot J_f(x) && \text{product of matrices} \\ &= \begin{bmatrix} \frac{\partial g_1}{\partial u_1} & \cdots & \frac{\partial g_1}{\partial u_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_p}{\partial u_1} & \cdots & \frac{\partial g_p}{\partial u_m} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \end{aligned}$$

Proof 1. (Sketch). We need to show: for fixed $x \in A \subset \mathbb{R}^n$,

$$\lim_{y \to x} \frac{\|h(y) - h(x) - \mathbb{D}h(x)(y-x)\|}{\|y-x\|} = 0,$$

or

$$\lim_{y \to x} \frac{\|g(f(y)) - g(f(x)) - \mathbb{D}g(f(x))[\mathbb{D}f(x)(y-x)]\|}{\|y-x\|} = 0.$$

Work with the numerator:

$$\begin{aligned} \text{numerator} &= \|g(f(y)) - g(f(x)) - \mathbb{D}g(f(x))(f(y) - f(x)) \\ &\quad + \mathbb{D}g(f(x))(f(y) - f(x)) - \mathbb{D}g(f(x))[\mathbb{D}f(x)(y-x)]\| \\ &\le \|g(f(y)) - g(f(x)) - \mathbb{D}g(f(x))[\mathbb{D}f(x)(y-x)]\| && \text{triangle inequality} \\ &\quad + \|\mathbb{D}g(f(x))(f(y) - f(x)) - \mathbb{D}g(f(x))[\mathbb{D}f(x)(y-x)]\| \\ &\le \varepsilon_1 \|f(y) - f(x)\| + \|\mathbb{D}g(f(x))\| \cdot \|f(y) - f(x) - \mathbb{D}f(x)(y-x)\| \\ &\quad (\varepsilon_1 : g \text{ is differentiable; } \mathrm{d}g(f(x)) : \text{common factor}) \\ &\le \varepsilon_1 \cdot L\|y-x\| + M \cdot \varepsilon_2 \|y-x\| \\ &\quad (L : \text{local Lipschitz; } M : \text{differential bounded; } \varepsilon_2 : f \text{ is differentiable}) \\ &= (L\varepsilon_1 + M\varepsilon_2) \cdot \|y-x\|. \end{aligned}$$

Therefore,

$$\lim_{y \to x} \frac{\text{numerator}}{\|y-x\|} = \lim_{y \to x} \frac{(L\varepsilon_1 + M\varepsilon_2)\|y-x\|}{\|y-x\|} = \lim_{y \to x} L\varepsilon_1 + M\varepsilon_2 = 0.$$

Q.E.D. ■

Example 6.3.2

- Change of Variable

$$(x,y,z) \longleftrightarrow (r,\theta,z) : \begin{cases} x = r\cos\theta \\ y = r\sin\theta \\ z = z \end{cases} \qquad \text{(cylindrical coordinate)}$$

  Let $h(r,\theta,z) = f(x,y,z) = f(x(r,\theta,z), y(r,\theta,z), z(r,\theta,z))$. Then,

$$\mathbb{D}h = \frac{\partial h}{\partial(r,\theta,z)} = \frac{\partial f}{\partial(x,y,z)} \cdot \frac{\partial(x,y,z)}{\partial(r,\theta,z)} = J_f \cdot \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Consider composition of the maps $[0,1] \xrightarrow{\gamma} \mathbb{R}^n \xrightarrow{f} \mathbb{R}$. Then, $h(t) = f(\gamma(t))$. By chain rule,

$$\begin{aligned} h'(t) = \mathbb{D}f \circ \mathbb{D}\gamma &= \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{pmatrix} \begin{pmatrix} x_1'(t) \\ x_2'(t) \\ \vdots \\ x_n'(t) \end{pmatrix} \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} x_i'(t) = \left\langle \boldsymbol{\nabla} f, \gamma'(t) \right\rangle. \end{aligned}$$

6.3.2 Other Differentiation Rules

Theorem 6.3.3 Product Rule
Let $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ and $g : A \to \mathbb{R}$ be differentiable. Then, the product $gf : A \to \mathbb{R}^m$ is differentiable, and

$$\mathbb{D}(gf) = g(\mathbb{D}f) + (\mathbb{D}g)f.$$

More precisely, for each $x \in A$ and $h \in \mathbb{R}^n$,

$$\mathbb{D}(gf) \cdot h = \underbrace{g(x)}_{\text{scalar}} \cdot \underbrace{\mathbb{D}f(x)(h)}_{\in\mathbb{R}^m} + \underbrace{\mathbb{D}g(x)(h)}_{\text{scalar}} \cdot \underbrace{f(x)}_{\in\mathbb{R}^m}.$$

In particular,

$$\frac{\partial g(f_j)}{\partial x_i} = g \cdot \frac{\partial f_j}{\partial x_i} + \frac{\partial g}{\partial x_i} \cdot f_j.$$

Theorem 6.3.4 Other Differentiation Rules

$$\mathbb{D}(f+g) = \mathbb{D}f + \mathbb{D}g$$

$$\mathbb{D}(\lambda f) = \lambda \mathbb{D}f$$

$$\mathbb{D}\left(\frac{f}{g}\right) = \frac{g\mathbb{D}f - (\mathbb{D}g)f}{g^2} \quad \left(\text{derived from product rule: } \frac{f}{g} = f \cdot \frac{1}{g}\right)$$

6.4 Geometric Interpretation of Gradient

Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ be differentiable.

Definition 6.4.1 ($\mathbb{D}f(x), \nabla f(x), D_e f(x)$).

- Differential of f: a matrix/linear transformation

$$\mathbb{D}f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

- Gradient of f: a vector

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right).$$

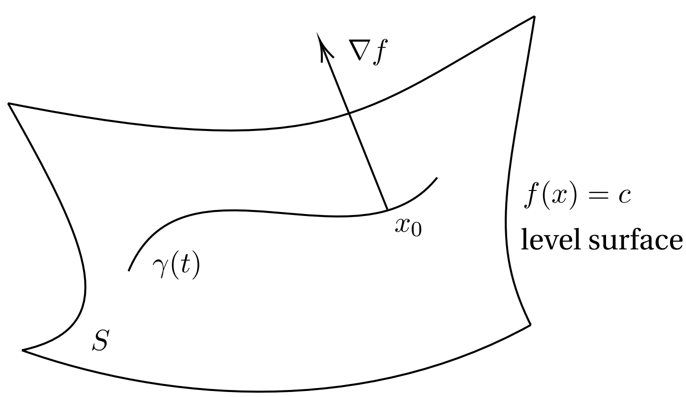
- Directional derivative of f in the direction e:

$$D_e f(x) = \mathbb{D}f(x)e = \langle \nabla f(x), e \rangle .$$

  Geometric meaning of $D_e f(x)$: Rate of change in the direction of e.

6.4.2 Geometric Meaning of Gradient.

Claim 6.4.3 ∇f is perpendicular to the level surface S defined by $f(x) =$ constant.



Proof 1. Fix any curve $\gamma(t)$ on S: $\gamma : [a, b] \to S$. Then, $f(\gamma(t)) = c$. By chain rule,

$$\mathbb{D}f(\gamma(t)) \cdot \gamma'(t) = 0 \implies \left\langle \nabla f(x_0), \gamma'(x_0) \right\rangle = 0.$$

So, $\nabla f(x_0) \perp \gamma'(x_0)$. That is, $\nabla f \perp$ curve γ on $S \implies \nabla f \perp S$.

Q.E.D. ■

Corollary 6.4.4 Tangent Plane: The tangent plane at x_0 of the level surface is given by

$$\langle \nabla f(x_0), x - x_0 \rangle = 0.$$

Example 6.4.5

Find the tangent plane at $(1, 0, 1)$ to the surface $x^2 - y^2 + xyz = 1$.

Claim 6.4.6 The direction of ∇f is the direction in which f has the greatest rate of change, which is given by $\|\nabla f\|$.

Proof 2. Fix a direction $e \in \mathbb{R}^n$. Then, the rate of change in direction e is given by

$$D_e f(x_0) = \langle \nabla f, e \rangle = \|\nabla f\| \|e\| \cos\theta,$$

where θ is the angle between $\nabla f(x_0)$ and e. Then, the rate of change is maximized when $\cos\theta = 1$. So, $\theta = 0$. That is, e is in the direction of ∇f.

Q.E.D. ■

6.5 Mean Value Theorem (MVT)

Theorem 6.5.1 MVT in 1-D

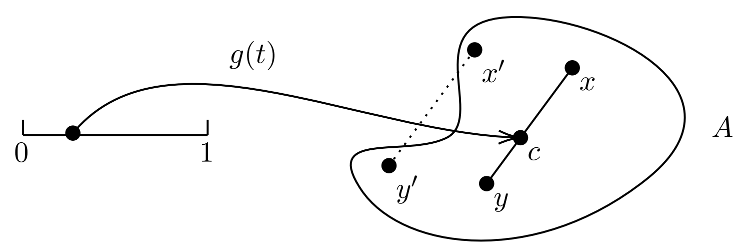
Let $f : [a, b] \to \mathbb{R}^1$ be continuous and differentiable on (a, b). Then, $\exists\, c \in (a, b)$ *s.t.*

$$f'(c) = \frac{f(b) - f(a)}{b - a} \quad \text{or} \quad f(b) - f(a) = f'(c)(b - a).$$

Theorem 6.5.2 MVT in Higher Dimension

Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ be differentiable on an open set A. Then, for any pair of points $x, y \in A$ *s.t.* the line segment $[x, y]$ joining x and y is contained in A, $\exists$ a point $c \in [x, y]$ *s.t.*

$$f(y) - f(x) = \mathbb{D}f(c)(y - x).$$



Proof 1. Let $g(t) = (1-t)x + ty$ for $0 \le t \le 1$ and

$$h(t) = f \circ g(t) = f((1-t)x + ty) : [0,1] \to \mathbb{R}.$$

Apply Theorem 6.5.1 to h, we know $\exists t_0 \in (0,1)$ $s.t.$

$$\begin{aligned} h(1) - h(0) &= h'(t_0)(1-0) \\ f(y) - f(x) &= \mathbb{D}f(g(t_0)) \cdot g'(t_0) && \text{[Chain Rule]} \\ &= \mathbb{D}f(g(t_0)) \cdot (y-x). \end{aligned}$$

Denote $g(t_0) = c \in [x,y]$. Then,

$$f(y) - f(x) = \mathbb{D}f(c)(y-x).$$

Q.E.D. ■

Definition 6.5.3 (Convex Set). A set $A \subset \mathbb{R}^n$ is *convex* if $\forall\, x, y \in A$, $[x,y] \subset A$.

Corollary 6.5.4 : Let $A \subset \mathbb{R}^n$ be open and convex, and $f : A \to \mathbb{R}^m$ differentiable. If $\mathbb{D}f \equiv 0$, then f is constant in A.

Proof 2. (Sketch)

Apply MVT to each component of $f = (f_1, f_2, \ldots, f_m)$.

Q.E.D. ■

6.6 Taylor's Theorem & Higher Order Differentials

6.6.1 One Dimensional Case

Theorem 6.6.1 Taylor's Formula

Let $f : (a,b) \to \mathbb{R}$ be one of class $\mathcal{C}^r$ (i.e., $f'(x), f''(x), \ldots, f^{(r)}(x)$ are continuous). Then, for any $x_0, x \in (a,b)$, we have

$$f(x) = \underbrace{\boxed{f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \cdots + \frac{f^{(r-1)}(x_0)}{(r-1)!}(x-x_0)^{r-1}}}_{\text{Taylor's polynomial of degree } r-1} + \overbrace{R_{r-1}(x_0)}^{\text{Remainder}},$$

where R_{r-1} is the remainder at x_0 and can be written as

$$R_{r-1}(x_0) = \frac{f^{(r)}(c)}{r!}(x-x_0)^r \quad \text{for some } c \text{ between } x \text{ and } x_0.$$

Remark 6.8 (Key Idea to Prove) *Use integration by parts in a reversed way multiple times.*

Proof 1. Write $h = x - x_0$. Then, by Fundamental Theorem of Calculus,

$$f(x) - f(x_0) = f(x_0 + h) - f(x_0) = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} f(x_0 + th)\,\mathrm{d}t.$$

Now, apply integration by parts. Taking $u = \dfrac{\mathrm{d}}{\mathrm{d}t} f(x_0 + th) = f'(x_0+th)h$ and $\mathrm{d}v = \mathrm{d}t \implies v = t-1$, we have

$$\begin{aligned}
f(x) - f(x_0) &= \int_0^1 u \,\mathrm{d}v \\
&= uv\Big|_0^1 - \int_0^1 v\,\mathrm{d}u \\
&= -(-1)f'(x_0)h - \int_0^1 (t-1)f''(x_0+th)h^2)\,\mathrm{d}t \\
&= f'(x_0)h - \int_0^1 f''(x_0+th)h^2(t-1)\,\mathrm{d}t.
\end{aligned}$$

Apply integration by parts again with

$$u = f''(x_0+th)h^2 \quad \text{and} \quad \mathrm{d}v = (t-1)\,\mathrm{d}t \implies v = \frac{1}{2}(t-1)^2.$$

Then, we obtain

$$\begin{aligned}
\int_0^1 f''(x_0+th)h^2(t-1)\,\mathrm{d}t &= f''(x_0+th)h^2\frac{1}{2}(t-1)^2\Big|_0^1 - \int_0^1 \frac{1}{2}(t-1)^2 f''(x_0+th)h^3\,\mathrm{d}t \\
&= \frac{f''(x_0)}{2}h^2 + \int_0^1 f^{(3)}(x_0+th)h^3 \cdot \frac{1}{2}(t-1)^2\,\mathrm{d}t.
\end{aligned}$$

So,

$$f(x) - f(x_0) = f'(x_0)h + \frac{f''(x_0)}{2}h^2 + \int_0^1 f^{(3)}(x_0+th)h^3 \cdot \frac{1}{2}(t-1)^2\,\mathrm{d}t.$$

By induction, we obtain that

$$\begin{aligned}
f(x) - f(x_0) = &\overbrace{\boxed{f'(x_0)h + \frac{f''(x_0)}{2}h^2 + \frac{f'''(x_0)}{3!}h^3 + \cdots + \frac{f^{(r-1)}(x_0)}{(r-1)!}h^{r-1}}}^{\text{Taylor's polynomial}} \\
&+ \underbrace{\boxed{(-1)^{r-1}\int_0^1 f^{(r)}(x_0+th)h^r\,\frac{(t-1)^{r-1}}{(r-1)!}\,\mathrm{d}t}}_{\text{Remainder}}
\end{aligned}$$

Lemma 6.6.2 2^{nd} MVT for Integral: If $g \geq 0$, then $\displaystyle\int_a^b f(x)g(x)\,\mathrm{d}x = f(\lambda)\int_a^b g(x)\,\mathrm{d}x$.

Apply 2^{nd} MVT to the remainder, we have

$$\begin{aligned}
R_{r-1} &= (-1)^{r-1}f^{(r)}(x_0+t_0h)h^r\int_0^1 \frac{(t-1)^{r-1}}{(r-1)!}\,\mathrm{d}t \\
&= f^{(r)}(x_0+t_0h)h^r \cdot \frac{1}{r} && \left[\begin{array}{c}(-1)^{r-1} \text{ is absorbed when} \\ \text{evaluating the integral}\end{array}\right] \\
&= \frac{f^{(r)}(c)}{r!}h^r && \left[\begin{array}{c}\text{Denote } c{=}x_0{+}t_0h\text{, a point} \\ \text{between } x_0 \text{ and } x\end{array}\right]
\end{aligned}$$

Combining everything, we get exactly what we have claimed.

Q.E.D. ■

Summary IV: Taylor's Formula & Taylor's Approximation

- Taylor's Formula:

$$f(x) = P_{r-1}(x) + R_{r-1}.$$

- Taylor's Approximation:

$$f(x) \approx P_{r-1}(x).$$

6.6.2 Taylor Series

Definition 6.6.3 (Taylor Series). Let $f \in \mathcal{C}^\infty$. Then, the *Taylor series* is defined as

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2}(x-x_0)^2 + \cdots .$$

Definition 6.6.4 (Real Analytic). f is *(real) analytic* at x_0 if its Taylor series converges to $f(x)$ in a neighborhood of x_0. i.e.,

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n, \quad |x-x_0| < R.$$

Corollary 6.6.5 : If $f \in \mathcal{C}^\infty(\mathbb{R})$ and for each interval $[a,b]$, $\exists$ constant M $s.t.$

$$\left| f^{(n)}(x) \right| \leq M^n \quad \forall\, n \text{ and } x \in [a,b],$$

then, f is real analytic at each point x_0 and it has Taylor series representation. Namely,

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n, \quad |x-x_0| < \infty.$$

Proof 2. Fix $x_0 \in \mathbb{R}$. For any $x \in \mathbb{R}$, choose $b > 0$ $s.t.$ $x_0, x \in [-b,b]$. By Taylor's Formula,

$$f(x) = \underbrace{P_{n-1}(x)}_{\substack{\text{partial sum}\\ \text{of the series}}} + R_{n-1}.$$

Recall:

$$R_{n-1} = \frac{f^{(n)}(c)}{n!}(x-x_0)^n \quad \text{for some } c.$$

Then,

$$|R_{n-1}| \leq \frac{M^n}{n!}|x-x_0|^n \quad \forall\, x \in [-b,b].$$

Since the series $\sum_{n=0}^{\infty} \frac{M^n}{n!}|x-x_0|^n$ converges by ratio test, its n-th term,

$$\frac{M^n}{n!}|x-x_0|^n \to 0 \quad \text{as } n \to \infty.$$

Hence, $R_{n-1} \to 0$ as $n \to \infty$. Then, $P_{n-1}(x) \to f(x)$ as $n \to \infty$.

Q.E.D. ■

Example 6.6.6

- e^x and $\sin x$ are real analytic in $\mathbb{R}$. Find Taylor series at $x_0 = 0$:

  Solution 3.

  $$e^n = \sum_{n=0}^{\infty} \frac{1}{n!} x^n, \quad |x-x_0| < \infty.$$

  □

- Is every $\mathcal{C}^\infty$ real analytic? No.

  Counterexample 6.7. Consider the function $f(x)$:

  $$f(x) = \begin{cases} 0, & x = 0 \\ e^{-1/x^2}, & x \neq 0. \end{cases}$$

  Claim $f(x) \in \mathcal{C}^\infty$.

  Proof. At $x = 0$,

  $$f'(0) = \lim_{x\to 0} \frac{f(x)-f(0)}{x-0} = \lim_{x\to 0} \frac{e^{-1/x^2}}{x} = 0 \quad \text{(by L.H.)}$$

  At $x \neq 0$,

  $$f'(x) = e^{-1/x^2}\left(\frac{2}{x^3}\right) = \frac{2/x^3}{e^{1/x^2}} \to 0 \quad \text{as } x \to 0 \quad \text{(by L.H.)}$$

  So, $f'(x)$ is continuous at $x = 0$, and

  $$f'(x) = \begin{cases} 0, & x = 0 \\ e^{-1/x^2}\left(\frac{2}{x^3}\right) & x \neq 0. \end{cases}$$

  By induction, one can show that

  - $f^{(n)}(0) = 0 \quad \forall n$
  - $f^{(n)}(x) \to 0$ as $x \to 0$.

So, $f^{(n)}(x)$ is continuous at $x = 0$. So, $f \in \mathcal{C}^\infty$. $\square$

Claim *$f(x)$ is not real analytic at $x = 0$.*

Proof. Taylor series:

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}(x)^n = 0.$$

So, the Taylor series does not converge to $f(x)$ on any neighborhood of $x = 0$. $\square$

6.6.3 Higher Dimensional Case

Observation: Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$.

- Differential: $\mathbb{D}f(x)$ is a linear transformation $\mathbb{R}^n \to \mathbb{R}$.
- Let $g(x) = \mathbb{D}f(x)$. Then, $g : A \subset \mathbb{R}^n \to \mathbf{L}(\mathbb{R}^n, \mathbb{R}) \approx \mathbb{R}^n$, where $\mathbf{L}(M, N)$ is the space of linear transformation from M to N.
- $\mathbb{D}g(x)$ is a linear transformation $\mathbb{R}^n \to \mathbb{R}^n$ or $\mathbf{L}(\mathbb{R}^n, \mathbb{R})$.

Notation 6.8. Higher Order Differential The second order differential of f at x is denoted as

$$\mathbb{D}^2 f(x) = \mathbb{D}g(x) = \mathbb{D}(\mathbb{D}f(x)).$$

Definition 6.6.9 (Bilinear Maps). Given f and $x \in A$. Define a *bilinear map*, $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

$$\mathbb{D}^2 f(x)(u, v) = \left[\mathbb{D}^2 f(x)(u)\right](v),$$

where $u, v \in \mathbb{R}^n$ and $\mathbb{D}^2 f(x)(u) \in \mathbf{L}(\mathbb{R}^n, \mathbb{R})$. In matrix notation,

$$uBv^\top,$$

where u is $1 \times n$, B is $n \times n$, and $v^\top$ is $n \times 1$.

Definition 6.6.10 (Matrix Representation of the Bilinear Map). $\mathbb{D}^2 f(x) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is given by

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}_{n \times n}$$

This matrix is denoted as $H_x(f)$, the Hessian matrix of f at x. Then, in matrix form, we have that for

$u = (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n$ and $v = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$, and

$$\mathbb{D}^2 f(x)(u, v) = u \cdot H_x(f) \cdot v^\top \in \mathbb{R}.$$

Proof 4. Note that

$$g(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right) : \mathbb{R}^n \to \mathbb{R}^n.$$

Then,

$$\begin{aligned}\mathbb{D}^2 f(x) &= \mathbb{D} g(x) \\ &= \begin{bmatrix} \frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_1} & \frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_1} & \cdots & \frac{\partial}{\partial x_n}\frac{\partial f}{\partial x_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1}\frac{\partial f}{\partial x_n} & \frac{\partial}{\partial x_2}\frac{\partial f}{\partial x_n} & \cdots & \frac{\partial}{\partial x_n}\frac{\partial f}{\partial x_n} \end{bmatrix}.\end{aligned}$$

Q.E.D. ■

Lemma 6.6.11 Symmetry of the Partials and Differentials: Let $f(x, y) : A \subset \mathbb{R}^2 \to \mathbb{R}$ be of class $\mathcal{C}^2$. Then,

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}.$$

In general, for $f : A \subset \mathbb{R}^n \to \mathbb{R}$ in class $\mathcal{C}^2$,

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \quad \forall\, i, j.$$

Extension 6.1 *If $f \in \mathcal{C}^{(n)}$, the order of taking n-th derivative does not matter.*

Corollary 6.6.12 : If f is of class $\mathcal{C}^2$, then $\mathbb{D}^2 f(x) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is symmetric. That is,

$$\mathbb{D}^2 f(x)(u, v) = \mathbb{D}^2 f(x)(v, u).$$

Proof 5.

$$\mathbb{D}^2 f(x)(u, v) = u \cdot H_x(f) \cdot v^\top$$

Since $\mathbb{D}^2 f(x)(u, v) \in \mathbb{R}$, we have

$$\begin{aligned}\mathbb{D}^2 f(x)(u, v) = \left[\mathbb{D}^2 f(x)(u, v)\right]^\top &= (u \cdot H_x(f) \cdot v^\top)^\top \\ &= v \cdot H_x(f)^\top \cdot u^\top \\ &= v \cdot H_x(f) \cdot u^\top && \text{[by symmetry of } H_x(f)\text{]} \\ &= \mathbb{D}^2 f(x)(u, v).\end{aligned}$$

Q.E.D. ■

Example 6.6.13 Symmetry of Partials

Let $f(x, y, z) = e^{x,y} + xyz : \mathbb{R}^3 \to \mathbb{R}$. Verify the symmetry of the partials.

Solution 6.

$$\frac{\partial f}{\partial x} = ye^{xy} + yz; \qquad \frac{\partial f}{\partial y} = xe^{xy} + yz; \qquad \frac{\partial f}{\partial z} = xy.$$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = e^{xy} + xye^{xy} + z;$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = e^{xy} + xye^{xy} + z.$$

□

Summary V: Higher Order Differentials

- 1-st Order Differential: $\mathbb{D}f(x_0) : \mathbb{R}^n \to \mathbb{R}$: 1-linear form

$$\mathbb{D}f(x_0)(v) = J_f(x_0) \cdot v = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} v_i.$$

- 2-nd Order Differential: $\mathbb{D}^2 f(x_0) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$: bilinear form

$$\mathbb{D}^2 f(x_0)(v, w) = v \cdot H_f(x_0) \cdot w^\top = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j} v_i \cdot w_j.$$

- k-th Order Differential: $\mathbb{D}^k f(x_0) : \mathbb{R}^\times\mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$: k-linear form

$$\mathbb{D}^k f(x_0)(v^{(1)}, v^{(2)}, \ldots, v^{(k)}) = \sum_{i_1, i_2, \ldots, i_k = 1}^{n} \frac{\partial^2 f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_k}} v_{i_1}^{(1)} v_{i_2}^{(2)} \cdots v_{ik}^{(k)}$$

  In particular, denote $h = x - x_0 \in \mathbb{R}^n$, then

$$\mathbb{D}^k f(x_0)(h, h, \ldots, h) = \sum \frac{\partial^2 f}{\partial x_{i_1} \partial x_{i_2} \partial x_{i_k}} h_{i_1} h_{i_2} \cdots h_{i_k}.$$

- Speical case: $n = 2$: Write $\mathbb{D}^k f(x_0)(h, h) = \left(\frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2}\right)^k f(x_0) \cdot (h, h)$. Then,

$$\mathbb{D}^1 f = \left(\frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2}\right)^1 f = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2}; \quad \mathbb{D}^2 f = \left(\frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2}\right)^2 f = \frac{\partial^2 f}{\partial x_1^2} + 2\frac{\partial^2 f}{\partial x_1 \partial x_2} + \frac{\partial^2 f}{\partial x_2^2},$$

$$\mathbb{D}^3 f(h, h, h) = \frac{\partial^3 f}{\partial x_1^3} h_1^3 + 3\frac{\partial^3 f}{\partial x_1^2 \partial x_2} h_1^2 h_2 + 3\frac{\partial^3 f}{\partial x_1 \partial x_2^2} h_1 h_2^2 + \frac{\partial^3 f}{\partial x_2^3} h_2^3$$

Theorem 6.6.14 Taylor's Theorem

Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ be of class $\mathcal{C}^r$. Suppose $x, x_0 \in A$ $s.t.$ the line segment joining x and x_0, $[x, x_0] \subset A$. Then, $\exists\, c \in [x, x_0]$ $s.t.$

$$\begin{aligned} f(x) = f(x_0) + \mathbb{D}f(x_0)(x - x_0) + \frac{1}{2!}\mathbb{D}^2 f(x_0)(x - x_0, x - x_0) + \cdots \\ + \frac{1}{(r-1)!}\mathbb{D}^{r-1} f(x_0)(x - x_0, x - x_0, \ldots, x - x_0) + R_{r-1}, \end{aligned}$$

where R_{r-1} is the remainder given by

$$R_{r-1} = \frac{1}{r!}\mathbb{D}^r f(c)(x - x_0, \ldots, x - x_0)$$

and satisfies

$$\frac{R_{r-1}(x_0)}{\|x - x_0\|^{r-1}} \to 0 \quad \text{as} \quad x \to x_0.$$

Proof 7. Consider 1-variable form, $\varphi(t) = x_0 + t(x - x_0)$. Define

$$g(t) = f(x_0 + t(x - x_0))$$

for $t \in (a, b)$ with $[0, 1] \subset (a, b)$.

Apply Taylor's Theorem in 1-D to $g(t)$, we get

$$\begin{aligned} g(1) &= g(0) + g'(0)(1 - 0) + \frac{g''(0)}{2!}(1 - 0)^2 + \cdots + \frac{g^{(r-1)}(0)}{(r-1)!}(1 - 0)^{r-1} + R_{r-1} \\ f(x) &= f(x_0) + \sum_{k=1}^{r-1} \frac{g^{(k)}(0)}{k!} + \frac{1}{r!}g^{(r)}(\tilde{c}), \quad \tilde{c} \in [0, 1]. \end{aligned}$$

By chain rule, one can get

$$\begin{aligned} g'(t) &= \mathbb{D}f(\varphi(t))\varphi'(t) \\ g'(0) &= \mathbb{D}f(x_0)(x - x_0) \\ g''(t) &= \mathbb{D}^2 f(\varphi(t))\varphi'(t) \cdot \varphi'(t) \\ g''(0) &= \mathbb{D}^2 f(x_0)(x - x_0)^2 = \mathbb{D}^2 f(x_0)(x - x_0, x - x_0). \end{aligned}$$

So,

$$g^{(k)}(0) = \mathbb{D}^k f(x_0)(x - x_0, x - x_0, \ldots, x - x_0).$$

Q.E.D. ■

Example 6.6.15 Polynomial Approximation using Taylor's Theorem

Determine the 2-nd order Taylor's formula for $f(x, y) = e^{(x-1)^2}\cos y$ at $(1, 0)$.

Solution 8.

- Compute partials:

$$\frac{\partial f}{\partial x} = 2(x-1)e^{(x-1)^2}\cos y; \quad \frac{\partial f}{\partial y} = -e^{(x-1)^2}\sin y.$$

$$\frac{\partial^2 f}{\partial x^2} = 2e^{(x-1)^2}\cos y + 4(x-1)^2 e^{(x-1)^2}\cos y; \quad \frac{\partial^2 f}{\partial y^2} = -e^{(x-1)^2}\cos y.$$

$$\frac{\partial^2 f}{\partial x \partial y} = -2(x-1)e^{(x-1)^2}\sin y$$

- Evaluate at base point $(1,0)$:

$$\left.\frac{\partial f}{\partial x}\right|_{(1,0)} = 0, \quad \left.\frac{\partial f}{\partial y}\right|_{(1,0)} = 0, \quad \left.\frac{\partial^2 f}{\partial x^2}\right|_{(1,0)} = 2, \quad \left.\frac{\partial^2 f}{\partial x \partial y}\right|_{(1,0)} = 0, \quad \left.\frac{\partial^2 f}{\partial y^2}\right|_{(1,0)} = 1.$$

- Taylor's Formula: $h = x - x_0 = (x,y) - (1,0)$.

$$f(x,y) = f(1,0) + \mathbb{D}f(1,0)(h) + \mathbb{D}^2 f(1,0)(h,h) + R_2,$$

where $f(1,0) = 1$, $\mathbb{D}f(1,0) = \begin{bmatrix} 0 & 0 \end{bmatrix}$, and $\mathbb{D}^2 f(1,0) = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$. So,

$$\begin{aligned} \mathbb{D}f(1,0)(h) &= 0 \\ \mathbb{D}^2 f(1,0(h,h) &= \begin{pmatrix} x-1, y \end{pmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix} \begin{pmatrix} x-1 \\ y \end{pmatrix} = 2(x-1)^2 - y^2. \end{aligned}$$

Then,

$$f(x,y) = 1 + \frac{1}{2}\left[2(x-1)^2 - y^2\right] + R_2,$$

where

$$\frac{R_2}{\|(x-1,y)\|^2} \to 0 \quad \text{as} \quad (x-1,y) \to (1,0).$$

$\square$

6.7 Minima & Maxima in $\mathbb{R}^n$

Question: Given function $f : A \subset \mathbb{R}^n \to \mathbb{R}$, how do we find (local) maximum or minimum points for f in A?

6.7.1 Optimization in 1-D. Suppose $f : (a,b) \to \mathbb{R}$

- A local max/min point (or extreme point) x_0 must be a critical point:

$$f'(x_0) = 0 \quad \text{or} \quad f'(x_0) \text{ D.N.E.}$$

- 2-nd Order Derivative Test (for critical points):

$$f''(x_0) > 0 : \text{ local min;} \quad f''(x_0) < 0 : \text{ local max.}$$

Definition 6.7.2 (Extrema). Suppose $f : A \subset \mathbb{R}^n \to \mathbb{R}$.

- Then, $x_0 \in A$ is a *local minimum* if $\exists\, \delta > 0\ s.t.\ x \in A$ and

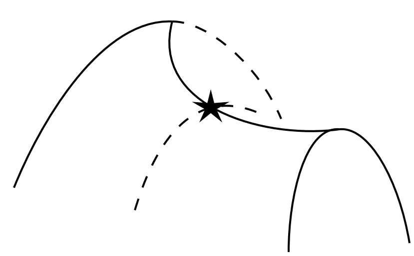
$$|x - x_0| < \delta \implies f(x) \geq f(x_0).$$

- Similarly, $x_0 \in A$ is a *local maximum* if $\exists\, \delta > 0\ s.t.\ x \in A$ and

$$|x - x_0| < \delta \implies f(x) \leq f(x_0).$$

Theorem 6.7.3 Necessary Condition for Extreme Points
If $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable and $x_0 \in A$ is an extreme point for f, then x_0 is a *critical point*, i.e., $\mathbb{D}f(x_0) = 0$.

Remark 6.9 *This is only a necessary condition but not sufficient. For example, in* $\mathbb{R}^1$, $f(x) = x^2$ *at* $(0,0)$ *or in* $\mathbb{R}^2$, $f(x,y) = x^2 - y^2$ *at* $(0,0)$.



For a critical point that is not an extreme point, we call it a saddle point.

Proof 1. (Sketch).

Assume $\mathbb{D}f(x_0) \neq 0$. Then, WLOG, $\exists\, v \in \mathbb{R}^n\ s.t.\ \mathbb{D}f(x_0)(v) = c > 0$. By definition of differential, choose $\delta > 0\ s.t.$

$$\|f(x_0 + h) - f(x_0) - \mathbb{D}f(x_0)(h)\| < \underbrace{\frac{c}{2\|v\|}}_{=\varepsilon} \cdot \|h\| \quad \forall\, \|h\| < \delta.$$

Choose $h = \lambda v$ with $\lambda > 0$ and $\|h\| < \delta$. Then, by triangle inequality,

$$f(x_0 + \lambda v) - f(x_0) > 0 \quad \text{but} \quad f(x_0 - \lambda v) - f(x_0) < 0.$$

Contradiction!

Q.E.D. ■

Definition 6.7.4 (Positive/Negative (Semi)definite). A bilinear form $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is call *positive definite* (or *negative definite*) if $B(x,x) > 0$ (or < 0) $\forall\, x \in \mathbb{R}^n,\ x \neq 0$. We say B is *positive (or negative) semidefinite* if $B(x,x) \geq 0$ (or ≤ 0) $\forall\, x \in \mathbb{R}^n$.

Definition 6.7.5 (Major Diagonal Factors). Recall B is determined by a matrix H as follows:

$$B(x,x) = xHx^\top, \quad \text{where } H = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}_{n\times n}$$

The *major diagonal factors* are given by

$$\begin{aligned} \Delta_1 &= \det\begin{pmatrix} a_{11} \end{pmatrix} = a_{11} \\ \Delta_2 &= \det\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} \\ &\vdots \\ \Delta_n &= \det(H). \end{aligned}$$

Lemma 6.7.6 :

- H is positive definitie $\iff \Delta_k > 0 \quad \forall\, k = 1, \ldots . n$
- H is positive semi-definite $\implies \Delta_k \geq 0 \quad \forall\, k = 1, \ldots, n$.
- H is negative definite $\iff (-H)$ is positive definite.

Example 6.7.7

$$H = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix} \implies \Delta_1 = 2,\, \delta_2 = 5 \implies H \text{ is positive definite.}$$

Theorem 6.7.8 Second Order Sufficient Condition

Suppose $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is of class $\mathcal{C}^2$ and $x_0 \in A$ is a critical point (i.e., $\mathbb{D}f(x) = 0$).

- If $H_f(x_0)$ is negative (or positive) definite, then x_0 is a local maximum (or minimum).
- If x_0 is a local maximum (or minimum), then $H_f(x_0)$ is negative (or positive) semidefinite.

Remark 6.10

- *Max of f* $\iff$ *Min of* $(-f)$
- *About minimum point:*
  - $\Delta_k > 0 \quad \forall k$, $H_f(x_0)$ *is positive definite* $\implies x_0$ *is local minimum.*

- x_0 *is a local minimum* $\implies$ $H_f(x_0)$ *is positive semidefinite* $\implies$ $\Delta_k \geq 0 \quad \forall\, k$.
- $\Delta_k < 0$ *for some* k $\implies$ x_0 *is not a local minimum.*

- *About maximum point:*
  - $\Delta_k < 0$ *for odd* k *and* $\Delta_k > 0$ *for even* k $\implies$ $(-H_f(x_0))$ *is negative definite* $\implies$ $H_f(x_0)$ *is negative definite* $\implies$ x_0 *is local maximum.*
  - x_0 *is local maximum* $\implies$ $H_f(x_0)$ *is negative semidefinite* $\implies$ $\Delta_k \leq 0$ *for odd* k *and* $\Delta k \geq 0$ *for even* k.
  - $\Delta_k < 0$ *for some even* k $\implies$ x_0 *is not a local maximum* $\implies$ x_0 *is a saddle point.*

Proof 2. (of ①)

- Set-up: Suppose H_f is negative definite. Need to show:

$$\exists \delta > 0 \; s.t. \; \|y - x\| < \delta \implies f(y) \leq f(x_0). \tag{$\star$}$$

Scartch:

By Taylor's Theorem

$$f(y) = f(x_0) + \underbrace{\mathbb{D}f(x_0)}_{=0,\text{critical point}} (y - x) + \frac{1}{2}\mathbb{D}^2 f(c)(y - x_0, y - x_0)$$

$$f(y) - f(x_0) = \frac{1}{2}\mathbb{D}^2 f(c)(y - x_0, y - x_0).$$

If $\mathbb{D}^2 f(c)$ is negative semidefinite, we are done with the proof. However, we only know definiteness at x_0. Let's add and subtract $\mathbb{D}^2 f(x_0)$:

$$f(y) - f(x_0) = \frac{1}{2}\underbrace{\mathbb{D}^2 f(x_0)(y - x_0, y - x_0)}_{\text{negative}} + \frac{1}{2}\underbrace{\left[\mathbb{D}^2 f(c) - \mathbb{D}^2 f(x_0)\right]}_{\text{make it small}}(y - x_0, y - x_0)$$

- Consider the function

$$g(x) = \mathbb{D}^2 f(x_0)(x, x) : \mathbb{R}^n \to \mathbb{R}.$$

Denote $\mathbb{D}^2 f(x_0) = H$, then $g(x) = H(x, x)$. g is continuous. Then, $\exists\, \overline{x} \in S = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ $s.t.$

$$H(x, x) \leq H(\overline{x}, \overline{x}).$$

Extreme Value Theorem: Continuous function on closed and bounded set attains its maximum and minimum. Since H is negative definite, $H(\overline{x}, \overline{x}) < 0$. Let $\varepsilon = -H(\overline{x}, \overline{x}) > 0$. Then, for any

$h \in \mathbb{R}^n$ with $h \neq 0$, we have

$$H(h,h) = \left\|h^2\right\| \cdot H\left(\frac{h}{\|h\|}, \frac{h}{\|h\|}\right) \leq \left\|h^2\right\| \cdot H(\delta x, \overline{x}).$$

So,

$$H(h,h) \leq -\varepsilon \left\|h^2\right\| \tag{I}$$

- Prove ($\star$) is true in a neighborhood.

  By continuity of $\mathbb{D}^2 f$ at x_0, $\exists\, \delta > 0$ $s.t.$

$$\|y - x_0\| < \delta \implies y \in A,\ \underbrace{\left\|\mathbb{D}^2 f(y) - \mathbb{D}^2 f(x_0)\right\|}_{\text{operator norm}} < \frac{\varepsilon}{2} \tag{II}$$

  Operator norm satisfies: $\|T(x,y)\| \leq \|T\| \cdot \|x\| \cdot \|y\|$.

  By Taylor's Formula, because $\mathbb{D} f(x_0) = 0$, we have

$$f(y) - f(x) = \frac{1}{2}\mathbb{D}^2 f(c)(h,h),$$

  where $y \in B(x_0, \delta)$, $h = y - x_0$, and $c \in [x_0, y]$. Note that

$$\begin{aligned}
\mathbb{D}^2 f(c)(h,h) &= \left[\mathbb{D}^2 f(c) - \mathbb{D}^2 f(x_0)\right](h,h) + \mathbb{D}^2 f(x_0)(h,h) \\
&\leq \left\|\mathbb{D}^2 f(c) - \mathbb{D}^2 f(x_0)\right\| \cdot \|h\|^2 + (-\varepsilon)\|h\|^2 && \text{By (I)} \\
&\leq \frac{1}{2}\varepsilon\|h\|^2 + (-\varepsilon)\|h\|^2 && \text{By (II)} \\
&= -\frac{\varepsilon}{2}\|h\|^2 \leq 0.
\end{aligned}$$

  Then, $f(y) \leq f(x) \quad \forall\, y \in B(x_0, \delta)$. So, x_0 is the local maximum.

Q.E.D. ■

Example 6.7.9

Find and classify the critical points for $f(x, yz) = \cos 2x \sin y + z^2$.

Solution 3.

- Find the critical point:

$$\frac{\partial f}{\partial x} = -2\sin 2x \sin y; \quad \frac{\partial f}{\partial y} = \cos 2x \cos y; \quad \frac{\partial f}{\partial z} = 2z.$$

  Set

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = \frac{\partial f}{\partial z} = 0.$$

Then,

$$\begin{cases} -2\sin 2x \sin y = 0 \\ \cos 2x \cos y = 0 \\ 2z = 0 \end{cases} \implies \begin{cases} x = \dfrac{k\pi}{2} \\ y = \dfrac{2j+1}{2}\pi \\ z = 0 \end{cases} \quad \text{or} \quad \begin{cases} x = \dfrac{2k+1}{4}\pi \\ y = j\pi \\ z = 0. \end{cases}$$

- Classify critical points:

Compute the Hessian

$$\frac{\partial^2 f}{\partial x^2} = -4\cos 2x \sin y; \quad \frac{\partial^2 f}{\partial y \partial x} = -2\sin 2x \cos y; \quad \frac{\partial^2 f}{\partial z \partial x} = 0$$

$$\frac{\partial^2 f}{\partial y^2} = -\cos 2x \sin y; \quad \frac{\partial^2 f}{\partial z \partial y} = 0; \quad \frac{\partial^2 f}{\partial z^2} = 2.$$

So,

$$H_f(x) = \begin{bmatrix} -4\cos 2x \sin y & -2\sin 2x \cos y & 0 \\ -2\sin 2x \cos y & -\cos 2x \sin y & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

$\boxed{\text{Case I}}$ $x = \dfrac{k\pi}{2}$, $y = \dfrac{2j+1}{2}\pi$, $z = 0$. Then,

$$H_f\left(\frac{k\pi}{2}, \frac{2j+1}{2}\pi, 0\right) = \begin{bmatrix} -4(-1)^k(-1)^j & 0 & 0 \\ 0 & -1(-1)^k(-1)^j & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Then, $\Delta_1 = -4(-1)^{j+k}$, $\Delta_2 = 4(-1)^{2k}(-1)^{2j} = 4 > 0$, and $\Delta_3 = 2 \cdot \Delta_2 = 8 > 0$.

- If $j + k$ is odd, then $\Delta_1 > 0$. Then, H_f is positive definite, and the critical point is a local minimum.
- If $j + k$ is even, then $\Delta_1 < 0$. Then, the critical point is not a local minimum. But $\Delta_3 = 0 > 0$, so it cannot be a local maximum. Hence, it must be a saddle point.

$\boxed{\text{Case II}}$ $x = \dfrac{2k+1}{4}\pi$, $y = j\pi$, $z = 0$. Then,

$$H_f\left(\frac{2k+1}{4}\pi, j\pi, 0\right) = \begin{bmatrix} 0 & (-2)(-1)^k(-1)^j & 0 \\ (-2)(-1)^k(-1)^j & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Then, $\Delta_1 = 0$, $\Delta_2 = -(-2)(-1)^{k+j} \cdot (-2)(-1)^{k+j} = -4(-1)^{2(k+j)} = -4 < 0$, and $\Delta_3 = 0$. As $\Delta_2 < 0$, they are saddle points.

- Conclusion:

$$\left(\frac{k\pi}{2}, \frac{2j+1}{2}\pi, 0\right) \begin{cases} \text{local minimum when } k+j \text{ is odd} \\ \text{saddle point when } k+j \text{ is even.} \end{cases}$$

$$\left(\frac{2k+1}{4}\pi, j\pi, 0\right) : \text{saddle point.}$$

$\square$

7 Inverse and Implicit Function Theorem

7.1 Inverse Function Theorem

7.1.1 Linear Case.

- Consider a linear map: $y = f(x) : \mathbb{R}^n \to \mathbb{R}^n$ given by

$$\begin{cases} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ &\vdots \\ y_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n. \end{cases}$$

  Or, in matrix notation:

$$Ax = y \tag{$\star$}$$

- Given $y \in \mathbb{R}^n$, ($\star$) is a linear system of equations.
- **Fact**: ($\star$) has unique solution $x \iff A$ is invertible. i.e., $\det(A) \neq 0$. In this case, the solution is given by $x = A^{-1}y$.
- $x = A^{-1}y$ is the inverse function of $y = f(x)$.

7.1.2 When can we solve a nonlinear system?.

- Nonlinear System:

$$\begin{cases} f_1(x_1, x_2, \dots, x_n) &= y_1 \\ &\vdots \\ f_n(x_1, x_2, \dots, x_n) &= y_n \end{cases}, \quad \text{or } f(x) = y.$$

  In order to have inverse, dimension must match.

- **Notation 7.3.**

  1. $y = f(x) : A \subset \mathbb{R}^n \to \mathbb{R}^n$, where A is open and f is differentiable on A. Suppose $y = (y_1, y_2, \dots, y_n)$, $x = (x_1, x_2, \dots, x_n)$, and $f = (f_1, f_2, \dots, f_n)$.
  2. $\mathbb{D}f(x) = \left(\dfrac{\partial f_j}{\partial x_i}\right)_{ij}$ and $J_f(x) = \det(\mathbb{D}f(x))$ is the *Jacobian determinant* of f at x.

Theorem 7.1.4 Inverse Function Theorem
Let $y = f(x) : A \subset \mathbb{R}^n \to \mathbb{R}^n$ be of class $\mathcal{C}^1$. Suppose $x_0 \in A$ and $J_f(x_0) \neq 0$. Then, $\exists$ neighborhoods U of x_0 and W of $y_0 = f(x_0)$ $s.t.$

1. $f(U) = W$ and $f : U \to W$ has an inverse $f^{-1} : W \to U$
2. $f^{-1} : W \to U$ is of class $\mathcal{C}^1$. *Additionally, if $f \in \mathcal{C}^r$, then $f^{-1} \in \mathcal{C}^r$.*
3. $\mathbb{D}f^{-1}(y) = [\mathbb{D}f(x)]^{-1} \quad \forall\, y \in W$ and $y = f(x)$.

▶ **Proof 1 of Inverse Function Theorem**

Theorem (Contraction Mapping Principle / CMP) *Let $\mathcal{X}$ be a complete metric space and $\varphi : \mathcal{X} \to \mathcal{X}$. Suppose $\exists 0 < k < 1$ s.t.*

$$\mathrm{d}(\varphi(x), \varphi(y)) \leq k \cdot d(x, y) \quad \forall\, x, y \in \mathcal{X}.$$

Then, $\exists$ unique fixed point x^* s.t. $\varphi(x^*) = x^*$.

$\boxed{\textbf{Step } 1}$ **Reductions**

- We may assume that $\mathbb{D}f(x_0) = I$.

  In fact, let $T = \mathbb{D}f(x_0)$. Then, $J_f(x_0) \neq 0 \implies T^{-1}$ exists. Consider a new map: $T^{-1} \circ f : A \to \mathbb{R}$. Then,

$$\begin{aligned} \mathbb{D}(T^{-1} \circ f) &= \mathbb{D}T^{-1}(f(x_0)) \circ \mathbb{D}f(x_0) \\ &= T^{-1} \circ T \\ &= I. \end{aligned}$$

  If the inverse of $T^{-1} \circ f$ exists, then the inverse of f also exists. *So, once the identity case is true, we just multiply $T{-}1$ to f and we can get the general case is true.*

- We may assume that $x_0 = 0$ and $f(x_0) = 0$.

  To see this, let $h(x) = f(x + x_0) - f(x_0)$. Then, $h(0) = 0$ and $\mathbb{D}h(0) = \mathbb{D}f(x_0)$. If the inverse of $h(x)$ exists, the n the equation $f(x) = y$ can be solved:

$$\begin{aligned} f(x) = h(x - x_0) + f(x_0) &= y \\ h(x - x_0) &= y - f(x_0) \\ x - x_0 &= h^{-1}(y - f(x_0)) \\ x &= h^{-1}(y - f(x_0)) + x_0. \end{aligned}$$

$\boxed{\textbf{Step } 2}$ **Existence of Inverse**

- By reduction above, we have $x_0 = 0$, $y_0 = f(x_0) = 0$, $\mathbb{D}f(x_0) = \mathbb{D}f(0) = I$.

  WTS: $\exists$ neighborhoods U, W of 0 s.t. the map $y = f(x) : U \to W$ has an inverse in W. i.e., $\forall\, y \in W$, $\exists$ unique $x \in U$ s.t. $y = f(x)$.

  For a fixed $y \in \mathbb{R}^n$, define $g_y(x) := y + x - f(x) : A \to \mathbb{R}^n$.

  If $g_y(x)$ has a fixed point: $g_y(x^*) = x^* = y + x^* - f(x^*) \implies y - f(x^*) = 0$. So, we want to show $g_y(x)$ has a unique fixed point.

- Construction of neighborhoods U and W.

  Let $g(x) = x - f(x)$. Then,

$$\mathbb{D}g(0) = I - \mathbb{D}f(0) = I - I = 0.$$

Since $f \in \mathcal{C}^1$, $g \in \mathcal{C}^1$. Then, $\mathbb{D}g(x)$ is continuous at 0. Then, $\forall\, \varepsilon = \dfrac{1}{2n}$, $\exists\, \delta > 0\ s.t.$

$$\|x - 0\| < \delta \implies \|\mathbb{D}g_i(x) - \mathbb{D}g_i(0)\| = \|\mathbb{D}g_i(x) - 0\| = \|\mathbb{D}g_i(x)\| < \frac{1}{2n},$$

where $g = (g_1, g_2, \ldots, g_n)$.

Apply MVT to each of g_i, we obtain $\forall x \in \overline{B}(x_0, \delta),\ \exists\, c_i \in [0, x]\ s.t.$

$$g_i(x) = g_i(x) - g_i(0) = \mathbb{D}g_i(c_i)(x - 0).$$

So,

$$\begin{aligned}
\|g(x)\| \le \sum_{i=1}^{n} \|g_i(x)\| &= \sum_{i=1}^{n} |\mathbb{D}g_i(c_i) \cdot x| \\
&\le \sum_{i=1}^{n} \|\mathbb{D}g_i(c_i)\| \cdot \|x\| && \text{[operator norm]} \\
&\le \sum_{i=1}^{n} \frac{1}{2n}\|x\| && \text{[continuity of } \mathbb{D}g\text{]} \\
&= \frac{1}{2}\|x\|.
\end{aligned}$$

i.e., $\|g(x)\| \le \frac{1}{2}\|x\|$. Thus, $g : \overline{B}(0, \delta) \to \overline{B}(0, \frac{1}{2}\delta) \subset \overline{B}(0, \delta)$ is a contraction map. Let $W = B\left(0, \dfrac{\delta}{2}\right)$ and $U = \{x \in B(0, \delta) : f(x) \in W\}$. WTS: U and W are the desired neighborhoods.

- Show existence of $f^{-1} : W \to U$.

  Fix $y \in W$. Then, $\forall\, x \in \overline{B}(0, \delta)$,

$$\begin{aligned}
\|g_y(x)\| = \|y + g(x)\| &\le \|y\| + \|g(x)\| \\
&< \frac{\delta}{2} + \frac{1}{2}\delta = \delta \qquad \left[y \in W = B\left(0, \frac{\delta}{2}\right),\ \|g(x)\| \le \frac{1}{2}\|x\|,\ x \in U = B(0, \delta)\right]
\end{aligned}$$

  Then, $g_y(x) : \overline{B}(0, \delta) \to \overline{B}(0, \delta)$ and g_y is also a contraction map with $k = \dfrac{1}{2}$. Then, by CMP, $\exists$ unique $x\ s.t.\ g_y(x) = x$. Then,

$$\begin{aligned}
g_y(x) = y + x - f(x) &= x \\
y - f(x) = 0 &\implies y = f(x).
\end{aligned}$$

  So, for fixed y, $\exists$ unique $x\ s.t.\ y = f(x)$. Then, f is a bijection, and thus the inverse exists.

$\boxed{\textbf{Step } 3}$ **Continuity of f^{-1}.**

WTS: f^{-1} is Lipschitz continuous.

Fix $y_1, y_2 \in W$. Let $x_i = f^{-1}(y_i)$ for $i = 1, 2$. Then,

$$\begin{aligned}\left\|f^{-1}(y_1) - f^{-1}(y_2)\right\| = \|x_1 - x_2\| &= \|g(x_1) + f(x_1) - g(x_2) - f(x_2)\| \\ &\le \|g(x_1) - g(x_2)\| + \|f(x_1) - f(x_2)\| \\ &= \|g(x_1) - g(x_2)\| + \|y_1 - y_2\|.\end{aligned}$$

Since $\|\mathbb{D}g(x)\| \le \dfrac{1}{2}$ for $x \in \overline{B}(0, \delta)$, by Mean Value Inequality,

$$\|g(x_1) - g(x_2)\| \le \frac{1}{2}\|x_1 - x_2\|.$$

Then,

$$\|x_1 - x_2\| \le \frac{1}{2}\|x_1 - x_2\| + \|y_1 - y_2\|.$$

So,

$$\frac{1}{2}\|x_1 - x_2\| \le \|y_1 - y_2\| \implies \|x_1 - x_2\| \le 2\|y_1 - y_2\|.$$

That is,

$$\left\|f^{-1}(y_1) - f^{-1}(y_2)\right\| \le 2\|y_1 - y_2\| \tag{$\star$}$$

Thus, f^{-1} is Lipschitz and thus continuous.

$\boxed{\textbf{Step } 4}$ **Differentiability of f^{-1}**

- **Proposition** $[\mathbb{D}f(0)]^{-1}$ *exists and* $\mathbb{D}f(x)$ *is continuous at* $0 \implies \exists\, \delta' > 0$ *s.t.* $[\mathbb{D}f(x)]^{-1}$ *exists and bounded by* M*:*

$$\underbrace{\|\mathbb{D}f(x) \cdot (v)\|}_{\text{operator norm}} \le \|M\| \cdot \|v\| \quad \forall\, \|x\| < \delta' \text{ and } v \in \mathbb{R}^n.$$

- WTS: $f^{-1}(y)$ is differentiable at any fixed point $y_0 \in W$ and

$$\mathbb{D}f^{-1}(y_0) = [\mathbb{D}f(x_0)]^{-1} \quad \text{with } y_0 = f(x_0).$$

Fix $y_0 \in W$. Then,

$$\begin{aligned}&\frac{\left\|f^{1}(y) - f^{-1}(y_0) - \mathbb{D}f^{-1}(y_0) \cdot (y - y_0)\right\|}{\|y - y_0\|} \\ =&\frac{\left\|[\mathbb{D}f(x_0)]^{-1} \cdot [\mathbb{D}f(x_0) \cdot f^{-1}(y) - \mathbb{D}f(x_0) \cdot f^{-1}(y_0) - (y - y_0)]\right\|}{\|y - y_0\|} && [\text{factor out } \mathbb{D}f^{-1}(y_0) = [\mathbb{D}f(x_0)]^{-1}] \\ =&\frac{\left\|[\mathbb{D}f(x_0)]^{-1} \cdot [\mathbb{D}f(x_0)(x - x_0) - (f(x) - f(x_0))]\right\|}{\left\|f(x) - f(x_)\right\|} && [y = f(x)] \\ =&\frac{\left\|[\mathbb{D}f(x_0)]^{-1} \cdot [\mathbb{D}f(x_0)(x - x_0) - (f(x) - f(x_0))]\right\| \cdot \|x - x_0\|}{\|f(x) - f(x_0)\| \cdot \|x - x_0\|} && [\text{Multiply by magic } 1] \\ \le&\frac{2\left\|[\mathbb{D}f(x_0)]^{-1}[\mathbb{D}f(x_0)(x - x_0) - (f(x) - f(x_0))]\right\|}{\|x - x_0\|} && [\text{Lipschitz continuity, Eq } (\star)] \\ &\to 0 \quad \text{as } x \to x_0.\end{aligned}$$

So, f^{-1} is differentiable, and

$$\left[\mathbb{D}f^{-1}(y)\right] = [\mathbb{D}f(x)]^{-1}.$$

Q.E.D. ■

Example 7.1.5

Investigate the invertibility (both local and global) for the map $W = (u, v) = f(x, y) : \mathbb{R}^2 \to \mathbb{R}^2$ given by $u = e^x \cos y$ and $v = e^x \sin y$.

Solution 2.

Firstly, we know $f \in \mathcal{C}^\infty$. Compute the Jacobian determinant:

$$\begin{aligned} J_f(x, y) = \det(\mathbb{D}f(x)) &= \det\left(\begin{bmatrix} \partial u/\partial x & \partial u/\partial y \\ \partial v/\partial x & \partial v/\partial y \end{bmatrix}\right) \\ &= \det\left(\begin{bmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{bmatrix}\right) \\ &= e^{2x} \cos^2 y + e^{2x} \sin y^2 \\ &= e^{2x} > 1. \end{aligned}$$

So, by the Inverse Function Theorem, f is invertible locally at any point, and the differentiable of the inverse is given by

$$\mathbb{D}f^{-1}(u, v) = [\mathbb{D}f(x, y)]^{-1} = \begin{bmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{bmatrix}^{-1}.$$

Now, let's examine if f is globally invertible (i.e., if f is a one-to-one function on $\mathbb{R}^2$). Note that

$$f(x_0, y_0) = e^{x_0} \cos y_0$$

and

$$f(x_0, y_0 + 2\pi) = e^{x_0} \cos(y + 2\pi) = e^{x_0} \cos(y_0) \text{ and } f(x_0, y_0 - 2\pi) = e^{x_0} \cos(y_0 - 2\pi) = e^{x_0} \cos(y_0).$$

So, f is not globally invertible since f is not an injection.

□

Remark 7.1 *f can be written in complex notation: $f(z) = e^z$, where $z = x + \mathrm{i}y \in \mathbb{C}$. Then,*

$$f(z) = e^z = e^{x+\mathrm{i}y} = e^x(\cos x + \mathrm{i} \sin y).$$

Meanwhile, $f^{-1}(z) = \ln(z)$.

7.2 Implicit Function Thm and Applications

Motivation

- Given a function $f : \mathbb{R} \to \mathbb{R}$. Consider an equation $f(y) = x$. If it can be solved for y (uniquely in terms of x), then the solution $y = g(x)$ is the inverse of f. That is, $(f \circ g)(x) = x$.
- Reinterpretation of Inverse:
  Rewrite $f(y) = x$ as $x - f(y) = 0 \quad ①$.
  Then, f is invertible $\iff$ Equation ① is solvable for y.
- **Question**: When can we solve a general equation for y, $F(x, y) = 0$ ($F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$)?
  The solution of $f(x, y) = 0$, denoted by $y = g(x)$, is called the *implicit function* determined by $F(x, y) = 0$.

Example 7.2.1

Consider equation $x^2 + y^2 - 1 = 0$ to be $F(x, y) : \mathbb{R}^1 \times \mathbb{R}^1 \to \mathbb{R}^1$.
Given (x_0, y_0) $s.t.$ $F(x_0, y_0) = 0$ with $y_0 \neq 0$. Then, $\exists$ a unique solution

$$y = \begin{cases} \sqrt{1 - x^2} & \text{if } y_0 > 0 \\ -\sqrt{1 - x^2} & \text{if } y_0 < 0. \end{cases}$$

in the neighborhood of x_0.
Note that $\left.\dfrac{\partial F}{\partial y}\right|_{y=y_0} = 2y_0 \neq 0$ when $y_0 \neq 0$.

Theorem 7.2.2 Implicit Function Theorem
Let $A \subset \mathbb{R}^n \times \mathbb{R}^m$ and $F(x, y) : A \to \mathbb{R}^m$ be of class $\mathcal{C}^1$. Suppose $(x_0, y_0) \in A$ with $F(x_0, y_0) = 0$. If

$$\begin{aligned}\Delta = \det\left(\frac{\partial F}{\partial y}\right) &= \det\left(\frac{\partial(F_1, \ldots, F_m)}{\partial y_1, \ldots, y_m}\right) \\ &= \det\begin{bmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{bmatrix} \neq 0 \quad \text{at } (x_0, y_0),\end{aligned}$$

then $\exists$ neighborhoods U of x_0, V of y_0, and a unique function $y = f(x) : U \to V$ such that $F(x, f(x)) = 0 \quad \forall\, x \in U$. i.e., $y = f(x)$ is the solution of $F(x, y) = 0$.
Furthermore, if $F \in \mathcal{C}^r$, then $f \in \mathcal{C}^r$.

Remark 7.2

- $y = f(x)$ *is called the implicit function determined by the equation* $F(x, y) = 0$ *based at the point* (x_0, y_0).

- *Differential of implicit function:*

  Suppose $n = m = 1$ *and* $F(x, y) = 0$. *Then, by chain rule,*

$$
\begin{aligned}
\frac{\partial F}{\partial x} \cdot \frac{\mathrm{d}x}{\mathrm{d}x} + \frac{\partial F}{\partial y} \cdot \frac{\mathrm{d}y}{\mathrm{d}x} &= 0 \\
\frac{\mathrm{d}y}{\mathrm{d}x} &= -\frac{\partial F/\partial x}{\partial F/\partial y}.
\end{aligned}
$$

  In the general case, let $y = f(x) = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$. *Let* f *be the implicit function determined by* $F(x, y) = 0$. *Then,*

$$
\mathbb{D}f = -\left(\frac{\partial F}{\partial y}\right)^{-1} \cdot \left(\frac{\partial F}{\partial x}\right).
$$

▶ **Proof 1 of Implicit Function Theorem**

Given $F(x, y) = A \subset \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$. Consider the map $G : A \subset \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ given by

$$
G(x, y) = (x, F(x, y)).
$$

We want to use Inverse Function Theorem. So, we need a map that maps to the same dimension.

Suppose G^{-1} exists in a neighborhood of (x_0, y_0). Write

$$
G^{-1}(x, 0) = (x, f(x)).
$$

Then, $y = f(x)$ is the solution of $F(x, y) = 0$ because

$$
\begin{aligned}
G(x, f(x)) &= (x, 0) \\
&= (x, F(x, f(x))).
\end{aligned}
$$

So, $F(x, f(x)) = 0$.

It remains to show that G is invertible. This follows from the inverse function theorem. Consider

$$
\mathbb{D}G\Big|_{(x,y)=(x_0,y_0)} = \left[\begin{array}{cccc|cccc}
1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\
\hline
\frac{\partial F_1}{\partial x_1} & \cdots & & \frac{\partial F_1}{\partial x_n} & \frac{\partial F_1}{\partial y_1} & \cdots & & \frac{\partial F_1}{\partial y_m} \\
\vdots & \ddots & & \vdots & \vdots & \ddots & & \vdots \\
\frac{\partial F_m}{\partial x_1} & \cdots & & \frac{\partial F_m}{\partial x_n} & \frac{\partial F_m}{\partial y_1} & \cdots & & \frac{\partial F_m}{\partial y_m}
\end{array}\right].
$$

So,

$$J_G(x_0, y_0) = \det \begin{bmatrix} \partial F_1/\partial y_1 & \cdots & \partial F_1/\partial y_m \\ \vdots & \ddots & \vdots \\ \partial F_m/\partial y_1 & \cdots & \partial F_m/\partial y_m \end{bmatrix} = \Delta \neq 0,$$

as assued in implicit function theorem. Therefore, by the inverse function theorem, G is invertible.

Q.E.D. ■

Example 7.2.3

Discuss the solvability of $\begin{cases} y + x + uv = 0 \\ uxy + v = 0 \end{cases}$ for u, v in terms of x, y near the point $(0,0,0,0)$ and the point $(1, 1, \sqrt{2}, -\sqrt{2})$. If impossible, compute $\dfrac{\partial u}{\partial x}$ and $\dfrac{\partial v}{\partial x}$ if exists.

Solution 2.

$F(x, y, u, v) = 0$ and $\begin{cases} F_1 = y + x + uv \\ F_2 = uxy + v. \end{cases}$ Let's compute Δ:

$$\begin{aligned} \Delta = \det\left(\frac{\partial(F_1, F_2)}{\partial(u, v)}\right) &= \det \begin{bmatrix} \partial F_1/\partial u & \partial F_1/\partial v \\ \partial F_2/\partial u & \partial F_2/\partial v \end{bmatrix} \\ &= \det \begin{bmatrix} v & u \\ xy & 1 \end{bmatrix} \\ &= v - uxy. \end{aligned}$$

Then, $\Delta(0,0,0,0) = 0$. So, Implicit Function Theorem does not apply. On the other hand,

$$\Delta(1, 1, \sqrt{2}, -\sqrt{2}) = -\sqrt{2} - \sqrt{2} = -2\sqrt{2} \neq 0.$$

So, by Implicit Function Theorem, $\exists$ unique solution $u = u(x, y)$ and $v = v(x, y)$ in a neighborhood.

Furthermore, the differentiable is given by

$$\begin{aligned} \frac{\partial(u, v)}{\partial(x, y)} &= -\left(\frac{\partial F}{\partial(u, v)}\right)^{-1} \left(\frac{\partial F}{\partial(x, y)}\right) \\ &= -\begin{bmatrix} v & u \\ xy & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 \\ uy & ux \end{bmatrix}. \end{aligned}$$

□

Theorem 7.2.4 Application: Domain-Straightening Theorem
Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$. Suppose $\mathbb{D}f(x_0) \neq 0$ and $f(x_0) = 0$. Then, $\exists$ open sets U and V (with $x_0 \in V$) and invertible map $h : U \to V$ s.t. $f(h(x_1, \ldots, x_n)) = x_0$.

Remark 7.3 *Under change of variables h, one can flatten the level curves of function $f(x)$.*

Theorem 7.2.5 Application: Range-Straightening Theorem
Suppose $f : A \subset \mathbb{R}^p \to \mathbb{R}^n$ with $p < n$ and rank of $\mathbb{D}f(x_0) = p$. Then, $\exists$ neighborhoods U, V, and invertible map $g : U \to V\ s.t.\ g \circ f(x_1, \dots, x_p) = (x_1, \dots, x_p, 0, \dots, 0)$.

7.3 Constrained Extrema

7.3.1 Morse Theory: Local Behavior Near a Critical Point

Let $f(x) : A \subset \mathbb{R}^n \to \mathbb{R}$ be of class $\mathcal{C}^2$ and x_0 is a critical point. Then, one can use $H_f(x_0)$ to classify critical point x_0.

- Morse Theory makes this classification more prcise.
- **Lemma 7.3.1 Morse Lemma:** Let $f(x) : A \subset \mathbb{R}^n \to \mathbb{R}$ be of class $\mathcal{C}^2$ with critical point $x_0 \in A$. If $H_f(x_0)$ is nondegenerate (i.e., $\det(H_f(x_0)) \neq 0$), then $\exists$ neighborhoods U for x_0 and V for 0, and invertible map $g : V \to U\ s.t.$ the function $h = f \circ g$ has the form

$$h(y) = f(x_0) - \left[y_1 62 + y_2^2 + \dots + y_\lambda^2\right] + [y_\lambda^2 + \dots + y_n^2],$$

  where λ is an integer called the *index* of f at x_0.
- Interpretation/Application:
  1. $\lambda = 0$: x_0 is a local minimum. Paraboloid open up.
  2. $\lambda = n$: x_0 is a local maximum. Paraboloid open down.
  3. $0 < \lambda < n$: x_0 is a saddle point. Hyperboloid.
- What is λ?

  λ (the index of f at x_0) is the number of negative eigenvalues of $H_f(x_0)$.

Example 7.3.2
Determine the shape of the surface given by $z = x^2 + 3xy - y^2$ near critical point $(0, 0)$.
Solution 1.
$\mathbb{D}f = \begin{pmatrix} 2x + 3y & 3x - 2y \end{pmatrix}$. Therefore,

$$H_f(x, y) = \begin{bmatrix} 2 & 3 \\ 3 & -2 \end{bmatrix}$$

The eigenvalues are $t = \pm\sqrt{13}$. So, index $\lambda = 1$. As $0 < \lambda < n$, $(0, 0)$ is a saddle point. The shape is thus a hyperboloid.

□

7.3.2 Constrained Extremal Problem

Goal: To maximum (or minimize) a function $f(x) : \mathbb{R}^n \to \mathbb{R}$ under the constraint $g(x) = c$.

Tool: Lagrange Multiplier Method.

Theorem 7.3.3 Necessary Condition
Let $f, g : U \subset \mathbb{R}^n \to \mathbb{R}$ be of class $\mathcal{C}^1$. Assume $g(x_0) = c_0$ with $\nabla g(x_0) \neq 0$. If f restricted to the surface $S : g(x) = c_0$ has maximum or minimum at x_0, then $\exists\, \lambda \in \mathbb{R}\ s.t.$

$$\nabla f(x_0) = \lambda \nabla g(x_0).$$

Remark 7.4 (Geometric Meaning) $\nabla f(X_0)$ *is parallel to* $\nabla g(x_0)$.

Proof 2.

- Geometric proof: WTS: $\nabla f(x_0) \perp S$.

  Fix curve $c(t)$ at t_0. So, $c(t_0) = x_0$. WTS: $\nabla f(x) \perp c'(t)$.

  Since f restricted to S has a maximum at x_0, $h(t) = f(c(t))$ has a maximum at t_0. Then,

  $$0 = h'(t_0) = \nabla f(x_0) \cdot c'(t_0) = \left\langle \nabla f(x_0), c'(t_0) \right\rangle .$$

  So, $\nabla f(x_0) \perp c'(t_0)$, and thus $\nabla f(x_0) \perp S$.

- Analytical proof: Substitute the condition $g(x) = c_0$ into $f(x)$

  Since
  $$\nabla g(x_0) = \left(\frac{\partial g}{\partial x_1}, \ldots, \frac{\partial g}{\partial x_n} \right) \neq \vec{\mathbf{0}},$$

  then $\exists \dfrac{\partial g}{\partial x_i} \neq 0$ for some $i = 1, \ldots, n$. WLOG, assume $\dfrac{\partial g}{\partial x_n}(x_0) \neq 0$. By Implicit Function Theorem, the equation

  $$g(x_1, \ldots, x_n) = c_0$$

  can be uniquely solve for x_0:

  $$x_n = h(x_1, \ldots, x_{n-1}).$$

  Let $k(x_1, \ldots, x_{n-1}) = f(x_1, \ldots, x_{n-1}, h(x_1, \ldots, x_{n-1}))$. Then, the maximum of f correspond to maximum of k. Then,

  $$0 = \frac{\partial k}{\partial x_i} = \frac{\partial f}{\partial x_i} + \frac{\partial f}{\partial x_n} \cdot \frac{\partial h}{\partial x_i} \quad \text{for } i = 1, \ldots, n-1. \tag{1}$$

  Furthermore, $g(x) = c_0$. So, $g(x_1, \ldots, x_{n-1}, h(x_1, \ldots, x_{n-1})) = c_0$. Then,

  $$\frac{\partial g}{\partial x_i} + \frac{\partial g}{\partial x_n} \cdot \frac{\partial h}{\partial x_i} = 0 \quad \text{for } i = 1, \ldots, n-1.$$

Then,

$$\frac{\partial h}{\partial x_i} = -\frac{\partial g/\partial x_i}{\partial g/\partial x_n} \tag{2}$$

Substitute (2) into (1):

$$\begin{aligned}\frac{\partial f}{\partial x_i} = -\frac{\partial f}{\partial x_n}\cdot\frac{\partial h}{\partial x_i} &= -\frac{\partial f}{\partial x_n}\cdot\frac{-\partial g/\partial x_i}{\partial g/\partial x_n}\\ &= \underbrace{\frac{\partial f/\partial x_n}{\partial g/\partial x_n}}_{\lambda}\cdot\frac{\partial g}{\partial x_i}\\ &= \lambda\frac{\partial g}{\partial x_i}.\end{aligned}$$

So,

$$\frac{\partial f}{\partial x_i} = \lambda\frac{\partial g}{\partial x_i} \quad \forall\, i = 1,\ldots,n.$$

That is,

$$\nabla f(x) = \lambda\nabla g(x).$$

Q.E.D. ■

Theorem 7.3.4 General Procedure to Solve an Extremal Problem

- Solve the equations for $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$:

$$\begin{cases} g(x) = c_0 \\ \nabla f(x) = \lambda\nabla g(x) \end{cases}$$

- Compare values of f at these points.

Example 7.3.5

Find extrema for the function $f(x,y) = x^2 - y^2$ subject to the constraint $x^2+y^2=1$.

Solution 3.

Solve the equations:

$$\begin{cases} g(x) = c_0 \\ \nabla f(x) = \lambda\nabla g(x) \end{cases} \implies \begin{cases} x^2+y^2=1 \\ \begin{bmatrix} 2x \\ -2y \end{bmatrix} = \lambda\begin{bmatrix} 2x \\ 2y \end{bmatrix}. \end{cases} \implies \begin{cases} x^2+y^2=1 \\ 2x = \lambda 2x \\ -2y = \lambda 2y. \end{cases}$$

- If $x = 0,\ y = \pm 1$, and $\lambda = -1$.
- If $y = 0,\ x = \pm 1$, and $\lambda = 1$.

Possible candidates: $(0,1)$, $(0,-1)$, $(1,0)$, and $(-1,0)$.

- At $(0, 1)$, $f(0, 1) = 0^2 - 1^2 = -1$.
- At $(0, -1)$, $f(0, -1) = 0^2 - (-1)^2 = -1$.
- At $(1, 0)$, $f(1, 0) = 1^2 - 0^2 = 1$.
- At $(-1, 0)$, $f(-1, 0) = (-1)^2 - 0 = 1$.

Then, $(0, 1)$ and $(0, -1)$ are local minimum, and $(1, 0)$ and $(-1, 0)$ are local maximum.

$\square$

Theorem 7.3.6 Extremal Problem with Multiple Constraints
Maximize/Minimize $f(x)$ with constraints $g_1(x) = c_1, \ldots, g_m(x) = c_m$. Then, we solve

$$\begin{cases} g_1(x) = c_1 \\ \qquad \vdots \\ g_m(x) = c_m \\ \nabla f(x) = \lambda_1 \nabla g_1(x) + \cdots + \lambda_m \nabla g_m(x). \end{cases}$$

8 Integration

8.1 Definition of Integration

8.1.1 Geometric Motivation. To compute the area of region under the curve $y = f(x)$.

- Form the upper and lower approximation:

$$U(f, \mathcal{P}) = \sum_{i=1}^{n} \sup_{I_i} f(x)\ell(I_i)$$

$$L(f, \mathcal{P}) = \sum_{i=1}^{n} \inf_{I_i} f(x)\ell(I_i).$$

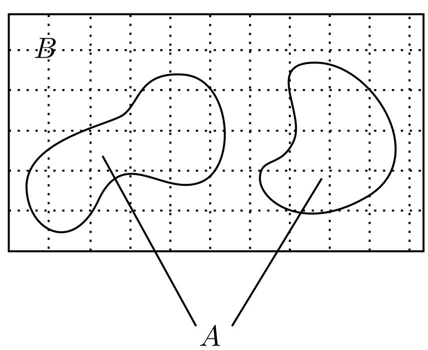
- Form the upper and lower integral:

$$\overline{\int_A} f = \inf_{\mathcal{P}} U(f, \mathcal{P})$$

$$\underline{\int_A} f = \sup_{\mathcal{P}} L(f, \mathcal{P}).$$

8.1.2 General Formulation of Integral.

- Set-up: Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ be a bounded function on a bounded set A.
- Goal: define the volume of the region under the surface $y = f(x)$ (or the integral $\displaystyle\int_A f \,\mathrm{d}x$).
- Step 1: choose a rectangle $B = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ that contains A. Extend f $s.t.$ $f(x) = 0$ when $x \notin A$.



Then, the volume over A is the same as the volume over B. That is,

$$\int_A f(x) \,\mathrm{d}x = \int_B f(x) \,\mathrm{d}x.$$

- Step 2: partition B: divide slides of B into sub-intervals to obtain a partition P, collection of smaller rectangles.

- Step 3: Form upper and lower sums:

$$U(f,\mathcal{P}) = \sum_{R\in\mathcal{P}} \underbrace{\sup_R f(x)}_{\text{height}} \cdot \underbrace{v(R)}_{\text{base}} \qquad \text{(Upper Sum of } f\,w.r.t.\mathcal{P}\text{)}$$

$$L(f,\mathcal{P}) = \sum_{R\in\mathcal{P}} \inf_R f(x)\cdot v(R) \qquad \text{(Lower Sum of } f\,w.r.t.\mathcal{P}\text{)}$$

- Step 4: Form upper and lower integrals:

$$\overline{\int_A} f = \inf_{\mathcal{P}}(U,\mathcal{P}) \quad \text{and} \quad \underline{\int_A} f = \sup_{\mathcal{P}} L(f,\mathcal{P}).$$

- Observation:

$$L(f,\mathcal{P}) \le \text{real volume} \le U(f,\mathcal{P}) \implies \underline{\int_A} f \le \text{real volume} \le \overline{\int_A} f.$$

- **Definition 8.1.3 (Integrable).**We say f is *Riemann integrable* if

$$\underline{\int_A} f = \overline{\int_A} f.$$

The integral of f on the set A is defined as $\int_A f(x)\,\mathrm{d}x = \underline{\int_A} f = \overline{\int_A} f$. Sometimes, the integral is also written as $\int_A f$ or $\int_A f(x)\,\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n$.

Theorem 8.1.4 Equivalent Conditions for Integrability
Suppose $f : A\subset\mathbb{R}^n \to \mathbb{R}$ is bounded and A and B are bounded. Let B be a rectangle in $\mathbb{R}^n$. Assume $f(x)=0$ for $x\notin A$. Then, the following are equivalent conditions for f to be integrable:

- (Riemann's Condition): $\forall\,\varepsilon>0$, $\exists$ partition $\mathcal{P}_\varepsilon$ (of B) $s.t.$

$$0 \le U(f,\mathcal{P}_\varepsilon) - L(f,\mathcal{P}_\varepsilon) < \varepsilon.$$

- (Darboux's Condition): $\exists$ a number I $s.t.\ \forall\,\varepsilon>0,\ \exists\,\delta>0\ s.t.$

  1. $\mathcal{P}$ is any partition of B into rectangles $B_1, B_2, \ldots, B_N$ with side length less than δ, and
  2. If $x_1\in B_1,\ x_2\in B_2,\ldots,x_N\in B_N$, then we have

$$\left|\sum_{i=1}^N f(x_i)v(B_i) - I\right| < \varepsilon.$$

Remark 8.1 • *The number I is the value of the integral*

- $\displaystyle\sum_{i=1}^{N} f(x_i)v(B_i)$ *is called the Riemann sum of* $f w.r.t.\mathcal{P}$*.*
- *Interpretation: Darboux's condition says that when the partition is fine enough (side length* $< \delta$*), then the Riemann sum is a good approximation of the integral.*

▶ **Proof 1 of Equivalent Conditions for Integrability**

$\boxed{\textbf{Step } 1}$ **f integrable $\implies$ Riemann's Condition**

Given $\varepsilon > 0$, need to find a partition $\mathcal{P}_\varepsilon$ $s.t.$ $U(f, \mathcal{P}_\varepsilon) - L(f, \mathcal{P}_\varepsilon) < \varepsilon$.

Since

$$\overline{\int_A} f = \inf_{\mathcal{P}} U(f, \mathcal{P}),$$

by definition of infimum,

$$\exists\, \mathcal{P}_1 \ s.t.\ U(f, \mathcal{P}_1) < \overline{\int_A} f + \frac{\varepsilon}{2}.$$

Similarly,

$$\exists\, \mathcal{P}_2 \ s.t.\ L(f, \mathcal{P}_2) > \underline{\int_A} f - \frac{\varepsilon}{2}.$$

Let $\mathcal{P}_\varepsilon = \mathcal{P}_1 \cup P_2$ (partition refinement). Then, $\mathcal{P}_\varepsilon$ is a refinement of $\mathcal{P}_1$ and $\mathcal{P}_2$. Therefore,

$$U(f, \mathcal{P}_\varepsilon) \le U(f, \mathcal{P}_1) < \overline{\int_A} f + \frac{\varepsilon}{2}, \quad \text{and} \quad L(f, \mathcal{P}_\varepsilon) \ge L(f, \mathcal{P}_2) > \underline{\int_A} f - \frac{\varepsilon}{2}.$$

Hence,

$$\begin{aligned}
U(f, \mathcal{P}_\varepsilon) - L(f, \mathcal{P}_\varepsilon) &\le U(f, \mathcal{P}_1) - L(f, \mathcal{P}_2) \\
&< \overline{\int_A} f + \frac{\varepsilon}{2} - \int_A f + \frac{\varepsilon}{2} \\
&= \overline{\int_A} f - \underline{\int_A} f + \varepsilon \\
&= 0 + \varepsilon && [f \text{ integrable}] \\
&= \varepsilon. \qquad \square
\end{aligned}$$

$\boxed{\textbf{Step } 2}$ **Riemann's Condition $\implies$ f integrable**

By Assumption, $\forall\, \varepsilon > 0$, $\exists$ partition $\mathcal{P}_\varepsilon$ $s.t.$

$$U(f, \mathcal{P}_\varepsilon) - L(f, \mathcal{P}_\varepsilon) < \varepsilon.$$

Since $\overline{\int_A} f = \inf_{\mathcal{P}} U(f, \mathcal{P})$, we have

$$\overline{\int_A} f \le U(f, \mathcal{P}_\varepsilon).$$

Similarly, we have $\underline{\int}_A f \geq L(f, \mathcal{P}_\varepsilon)$. Then,

$$0 \leq \overline{\int}_A f - \underline{\int}_A f \leq U(f, \mathcal{P}_e) - L(f, \mathcal{P}_\varepsilon) < \varepsilon.$$

Thus,

$$\overline{\int}_A f = \underline{\int}_A f \implies f \text{ is integrable.} \quad \square$$

$\boxed{\textbf{Step } 3}$ **Darboux's Condition $\implies$ Integrability**

Let I be the number in Darboux's condition.

$$\text{WTS: } \overline{\int}_A f = I = \underline{\int}_A f.$$

Claim 8.1.5 $\forall\, \varepsilon > 0$, $\exists$ partition $\mathcal{P}$ $s.t.$

$$|L(f, \mathcal{P}) - I| < \varepsilon \tag{$\star$}$$

Scratch:

$$|L(f, \mathcal{P}) - I| < \underbrace{\left| L(f, \mathcal{P}) - \sum_{i=1}^{N} f(x_i) v(B_i) \right|}_{= \sum_{i=1}^{N} \left| \inf_{B_i} f(x_i) - f(x_i) \right| v(B_i)} + \underbrace{\left| \sum_{i=1}^{N} f(x_i) v(B_i) - I \right|}_{< \frac{\varepsilon}{2},\ \text{by Darboux}}$$

So, we will make

$$\left| \inf_{B_i} f(x_i) - f(x_i) \right| < \frac{\varepsilon}{2 v(B_i) N}$$

since we want $\frac{\varepsilon}{2}$ eventually. Then,

$$|L(f, \mathcal{P}) - I| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Given $\varepsilon > 0$. By Darboux's condition, $\exists\, \delta > 0$ $s.t.$ $\forall\, \mathcal{P} = \{B_1, B_2, \ldots, B_N\}$ with sides $< \delta$, we have

$$\left| \sum_{i=1}^{N} f(x_i) v(B_i) - I \right| < \frac{\varepsilon}{2}.$$

for any $x_i \in B_i$, where $i = 1, \ldots, N$.

To prove ($\star$), we can choose $x_i \in B_i$ $s.t.$

$$0 \leq f(x_i) - \inf_{B_i} f(x_i) < \frac{\varepsilon}{2 v(B_i) N}.$$

Then, it follows that

$$
\begin{aligned}
|L(f,\mathcal{P}) - I| &< \left|L(f,\mathcal{P}) - \sum_i f(x_i)v(B_i)\right| + \left|\sum_i f(x_i)v(x_i) - I\right| \\
&< \sum_{i=1}^{N} \left|\inf_{B_i} f(x_i) - f(x_i)\right| v(B_i) + \frac{\varepsilon}{2} \\
&< \sum_{i=1}^{N} \frac{\varepsilon}{2N \cdot \cancel{v(B_i)}} \cdot \cancel{v(B_i)} + \frac{\varepsilon}{2} \\
&= \cancel{N} \cdot \frac{\varepsilon}{2\cancel{N}} + \frac{\varepsilon}{2} \\
&= \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \implies (\star)
\end{aligned}
$$

Furthermore, $(\star) \implies L(f,\mathcal{P}) > I - \varepsilon \quad \forall \varepsilon > 0$. So,

$$\underline{\int_A} f = \sup_{\mathcal{P}} L(f,\mathcal{P}) \geq I.$$

Similarly, $\forall \varepsilon > 0, \exists \mathcal{P}\ s.t.\ |U(f,\mathcal{P}) - I| < \varepsilon \implies U(f,\mathcal{P}) < I + \varepsilon$. Then,

$$\overline{\int_A} f = \inf_{\mathcal{P}} U(f,\mathcal{P}) \leq I.$$

So, it must be

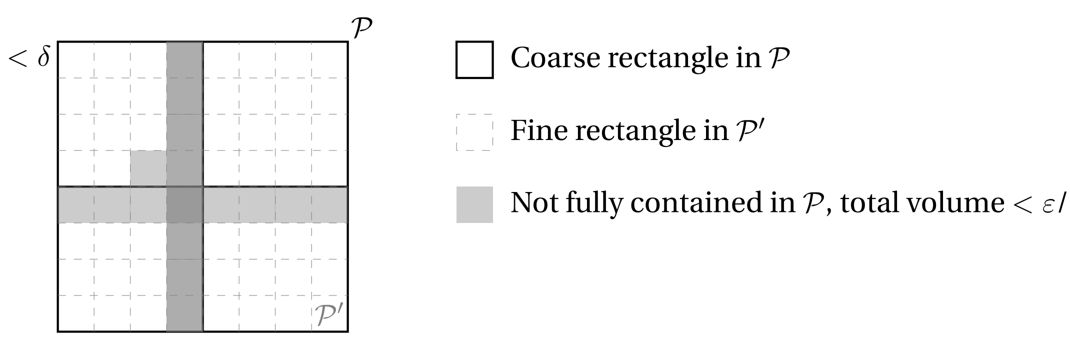
$$\overline{\int_A} f = \underline{\int_A} f = I.$$

$\boxed{\textbf{Step } 4}$ **Integrability $\implies$ Darboux's Condition (Scratch)**

- Given $\varepsilon > 0, \exists \mathcal{P}\ s.t.$

$$I - \frac{\varepsilon}{2} < L(f,\mathcal{P}) \leq \sum_i f(x_i)v(B_i) \leq U(f,\mathcal{P}) < I + \frac{\varepsilon}{2}.$$

- Given partition $\mathcal{P}$, $\exists \delta > 0\ s.t.$ for any partition $\mathcal{P}'$ with side length $< \delta$, the sum of volumes of sub-rectangles in $\mathcal{P}'$ that are not completely/entirely contained in a sub-rectangle in $\mathcal{P}$ is less than ε.



Q.E.D. ■

Example 8.1.6 An Exercise

Compute the upper and lower sums for $\int_0^1 x \,\mathrm{d}x$ over special partition $\mathcal{P}$:

$$\mathcal{P} = \left\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}\right\}.$$

8.2 Criterion for Integrability

Question: When is f integrable? How can we tell from other properties?

Short Answer: f is integrable when the set of discontinuity is "small."

8.2.1 Measure Zero: How to Measure the Size of a Set

Definition 8.2.1 (Volume of A). Given a bounded set $A \subset \mathbb{R}^n$, define *characteristic function* of A by

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}.$$

We say that A has *volume* (or *Jordan measurable*) if $\mathbb{1}_A(x)$ is integrable on A. We write

$$v(A) = \int_A \mathbb{1}_A(x) \,\mathrm{d}x.$$

Remark 8.2 *When $n = 1$, $v(A)$ is the length of A. When $n = 2$, $v(A)$ is the area of A.*

Fact: A set has volume 0 (i.e., $v(A) = 0$) $\iff$ $\forall\, \varepsilon > 0$, $\exists$ finite cover of A by rectangles $S_1, S_2, \dots, S_N$ *s.t.*

$$\sum_{i=1}^{N} v(S_i) < \varepsilon.$$

Proof 1. Suppose $v(A) = \displaystyle\int_A \mathbb{1}_A(x) \,\mathrm{d}x = 0$. Then, $\forall\, \varepsilon > 0$, $\exists$ partition $\mathcal{P} = \{\mathcal{P}_1, \dots, \mathcal{P}_N\}$ of B *s.t.*

$$U(\mathbb{1}_A(x), \mathcal{P}) < I + \varepsilon = \varepsilon.$$

$$\implies \sum_{\mathcal{P}_j \cap A \neq} \underbrace{\sup_{\mathcal{P}} \mathbb{1}_A(x)}_{=1} \cdot v(\mathcal{P}_i) = \sum_{\mathcal{P}_j \cap A \neq \varnothing} v(\mathcal{P}_i) < \varepsilon.$$

Note that $\{\mathcal{P}_j \mid \mathcal{P}_j \cap A \neq \varnothing\}$ is a finite cover of A.

Q.E.D. ■

Definition 8.2.2 (Measure Zero Set). A set $A \subset \mathbb{R}^n$ (not necessarily bounded) is said to have measure zero, $m(A) = 0$, if $\forall\, \varepsilon > 0$, $\exists$ countable cover of A by rectangles $\{S_i\}$ $s.t.$

$$\sum_{i=1}^{\infty} v(S_i) < \varepsilon.$$

Remark 8.3

- $v(A) = 0 \implies m(A) = 0$
- *Any finite set has volume zero.*
- *Any countable set has measure zero.* *(use geometric sum: first point covered by $\frac{\varepsilon}{2}$, second point covered by $\frac{\varepsilon}{4}$,..., N-th point covered by $\frac{\varepsilon}{2^N}$)*

Example 8.2.3

Let A be the x-axis (real line).

- If A is considered as a subset of $\mathbb{R}^2$, then $m(A) = 0$.

  Proof 2. To cover the x-axis, we can cover it interval by interval. But the volumes of the rectangles need to get smaller and smaller:

  $$S_n = [n, n+1] \times \left[-\frac{\varepsilon}{2^{|n|+2}}, \frac{\varepsilon}{2^{|n|+2}}\right]$$

  for $n = 0, \pm 1, \pm 2, \ldots$.

  Q.E.D. ■

- However, if A is considered as a subset of $\mathbb{R}^1$, then $m(A) \neq 0$.

Theorem 8.2.4

Suppose $A_i \subset \mathbb{R}^n$ with $m(A_i) = 0 \quad \forall\, i = 1, 2, \ldots$. Then,

$$A = \bigcup_{i=1}^{\infty} A_i$$

has measure zero.

Proof 3. Given $\varepsilon > 0$ for each $i = 1, 2, \ldots, m(A_i) = 0$. So, $\exists$ rectangles $\left\{S_j^{(i)}\right\}_{j=0}^{\infty}$ $s.t.$ $A_i \subset \bigcup_{j=1}^{\infty} S_j^{(i)}$ with $\sum_{j=1}^{\infty} v\left(S_j^{(i)}\right) < \frac{\varepsilon}{2^i}$. Then, $\left\{S_j^{(i)}\right\}_{i,j=1}^{\infty}$ is a countable collection of rectangles with

- $A = \bigcup_{i=1}^{\infty} A_i \subset \bigcup_{i=1}^{\infty} \bigcup_{j=1}^{\infty} S_j^{(i)}$

- $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} v\left(S_j^{(i)}\right) < \sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon.$

So, by definition, $m(A) = 0$.

Q.E.D. ■

Remark 8.4

- *The above theorem is not true for volume zero sets. A counterexample if the rationals in* $[0, 1]$. *Each rational is volume zero, but their union is not volume zero as* $\mathbb{1}_A$ *is not integrable.*
- *In Definition 8.2.2, we can replace "closed rectangles* S_i*" by "open rectangles* S_i*."*

8.2.2 Lebesgue's Theorem

Theorem 8.2.5 Lebesgue's Theorem
Let A be a bounded set in $\mathbb{R}^n$ and f be a bounded function on A. Extend f to $\mathbb{R}^n$ by letting $f(x) = 0 \quad \forall x \notin A$. Then, f is integrable on $A \iff$ the points on which the *extended function* f is discontinuous form a set of measure zero. That is, extended f has support on A, and if D denotes the set of discontinuity of extended f, then $m(D) = 0$.

Example 8.2.6

- $A = [0, 1]$ and
$$f(x) = \begin{cases} 1 & x \text{ rational} \\ 0 & \text{o/w.} \end{cases}$$
Then, the set of discontinuity is $D = [0, 1]$, and $m(D) \neq 0$. By Lebesgue's Theorem, f is not integrable.
- $A = \{\text{rationals} \in [0, 1]\}$ and $f(x) : A \to \mathbb{R}$ by $f(x) \equiv 1$. Then, f is continuous on A, but it is not integrable on A. The extended f has $D = [0, 1]$, not measure zero. So, by Lebesgue's Theorem, f is not integrable.
- $A = \left\{(x, y) \mid x^2 + y^2 < 1\right\} \subset \mathbb{R}^2$ and $f(x) : A \to \mathbb{R}$ by
$$f(x, y) = \begin{cases} x^2 + \sin\left(\frac{1}{y}\right) & y \neq 0 \\ x^2 & y = 0. \end{cases}$$
Then, the set of discontinuity is $D = [-1, 0] \times [1, 0] + \partial A$. Then, $m(D) = 0$ in $\mathbb{R}^2$. So, by Lebesgue's Theorem, f is integrable on A.

Corollary 8.2.7 of Lebesgue's Theorem:

- A bounded set $A \subset \mathbb{R}^n$ has volume $\iff$ ∂A has measure 0.

  Proof 4. Assume $v(A)$ exists. Then, $\mathbb{1}_A(x)$ is integrable. So, the set of discontinuity of extended $\mathbb{1}_A(x)$ is $D = \partial A$. By Lebesgue's Theorem, $f = \mathbb{1}_A(x)$ is integrable $\iff$ $m(\partial A) = 0$.

  Q.E.D. ■

- Let $A \subset \mathbb{R}^n$ be a bounded set with volume. If $f : A \to \mathbb{R}$ is bounded and has only a (finite or) countable number of discontinuity, then f is integrable.

  Proof 5. Denote the set of discontinuity of f on A as M. The set of discontinuity of the extended f will be $D \subset \partial A \cup M$. Since A has volume, by the previous Corollary, we know $m(\partial A) = 0$. Since M is countable, $m(M) = 0$. Then, $m(\partial A \cup M) = 0 \implies D \subset \partial A \cup M$ has measure zero. By Lebesgue's Theorem, f is integrable.

  Q.E.D. ■

▶ Proof 6 of Lebesgue's Theorem

$\boxed{\textbf{Step } 1}$ **Preparation and Reduction**

- The set-up: Fix a rectangle $B \supset A$ (so $\mathrm{cl}(A) \subset \int(B)$) and define $g : B \to \mathbb{R}$ by

$$g(x) = \begin{cases} f(x), & x \in A \\ 0, & x \notin A. \end{cases}$$

  Let D denote the set of discontinuity of $g(x)$. That is,

$$D = \{x \in B \mid g(x) \text{ is not continuous at } x\}.$$

  Need to show: f integrable on A $\iff$ $m(D) = 0$.

- How to quantify discontinuity?

  1. **Definition 8.2.8 (Oscillation).** The *oscillation* of a function $h(x)$ at a point x_0 is

  $$\mathcal{O}(h, x_0) = \inf\Big\{\sup\big\{|h(x_2) - h(x_1)| : x_1, x_2 \in U\big\} : U \text{ is a neighborhood of } x_0\Big\},$$

     where $\mathcal{O}(f, U) = \sup\big\{|h(x_2) - h(x_1)| : x_1, x_2 \in U\big\}$ is the oscillation in a neighborhood U, and $\inf$ takes over all possible neighborhoods of x_0.

  2. **Claim 8.2.9** h is continuous at $x_0 \implies \mathcal{O}(h, x_0) = 0$.

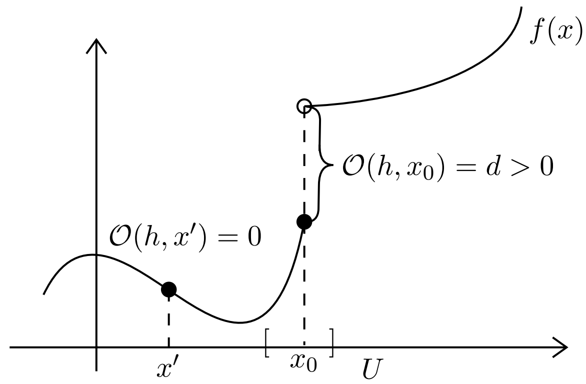
     Proof. h is continuous at $x_0 \implies \forall\, \varepsilon > 0,\ \exists\, \delta > 0\ s.t.$

  $$|x - x_0| < \delta \implies |h(x) - h(x_0)| < \frac{\varepsilon}{2}.$$

For $U = \{|x - x_0| < \delta\} \cap A$,

$$x_1, x_2 \in U \implies |h(x_2) - h(x_1)| \le |h(x_2) - h(x_0)| + |h(x_0) - h(x_1)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Then, $\mathcal{O}(h, U) < \varepsilon \implies \mathcal{O}(h, x_0) = 0.$ □



$\boxed{\textbf{Step } 2}$ **($\Leftarrow$) Assume $m(D) = 0$. Prove g is integrable.**

We will show: g satisfies Riemann condition.

- Set up:

  Fix $\varepsilon > 0$. Let $D_\varepsilon = \{x \in B \mid \mathcal{O}(g, x) > \varepsilon\}$. Then, $D_\varepsilon \subset D$. So, $m(D_\varepsilon) = 0$.

  By Definition, $\exists$ collection of open rectangles $|B_i|$ $s.t.$

$$D_\varepsilon \subset \bigcup_i B_i \quad \text{and} \quad \sum_i v(B_i) < \varepsilon.$$

  Claim 8.2.10 D_ε is closed (and hence compact).

  Proof. (Sketch) D_ε contains all its limits points. That is,

$$x_n \in D_\varepsilon,\ \{x_n\} \to x \implies x \in D_\varepsilon.$$

  Assume, for the sake of contradiction,

$$x \notin D_\varepsilon \implies \mathcal{O}(g, x) < \varepsilon.$$

  But $\mathcal{O}(g, x_n) \ge \varepsilon$, we can derive a contradiction from there. □

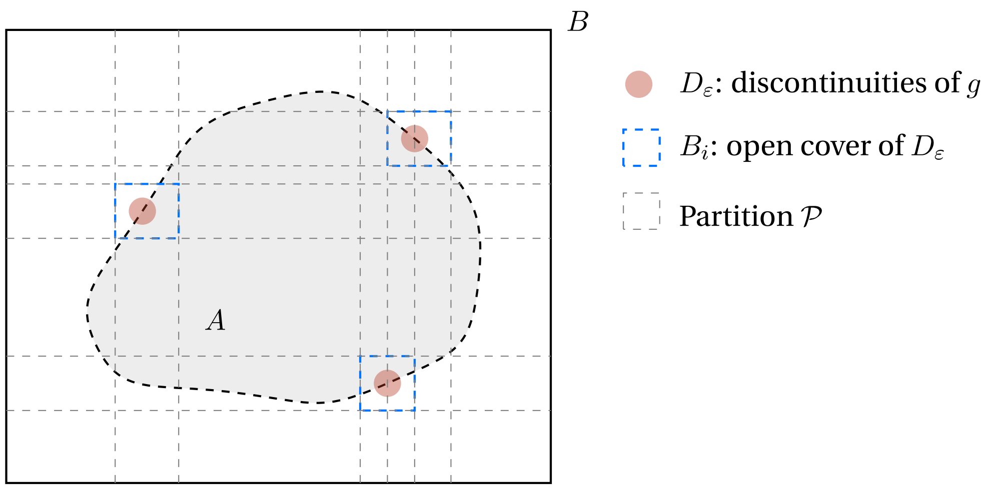
  Since D_ε is compact, it has a finite subcover:

$$\{B_1, B_2, \ldots, B_N\}\ s.t. \sum_{i=1}^{N} v(B_i) < \varepsilon.$$

- Initial Partition of B:

  Construct a partition $\mathcal{P}$ from $\{B_i\}_{i=1}^N$ $s.t.$ each rectangle $S \in \mathcal{P}$ is either:

  1. disjoint form D_ε, or
  2. its interior is contained in one of the B_i's.



  The way to construct $\mathcal{P}$ is to extend the sides of B_i to form a partition on B.

  Let $C_1 = \{S \in \mathcal{P} : \text{int}(S) \text{ is contained in one of } B_i\}$ and $C_2 = \{S \in \mathcal{P} : S \cap D_\varepsilon = \varnothing\}$.

- Refinement of $\mathcal{P}$

  Fix $S \in C_2$, $S \cap D_\varepsilon = \varnothing \implies \mathcal{O}(g, x) < \varepsilon \quad \forall\, x \in S$. Then, $\forall\, x \in S$, $\exists$ neighborhood U_x $s.t.$

$$\sup\{|g(x_1) - g(x_2)| : x_1, x_2 \in U_x\} < \mathcal{O}(g, x) + \delta,$$

  where $\delta = \dfrac{1}{2}(\varepsilon - \mathcal{O}(g, x))$. Then,

$$\sup_{U_x} g - \inf_{U_x} g < \mathcal{O}(g, x) + 2\delta = \varepsilon.$$

  Denote $M_{U_x}(g) = \sup\limits_{U_x} g$ and $m_{U_x}(g) = \inf\limits_{U_x} g$. Then,

$$\boxed{M_{U_x}(g) - m_{U_x}(g) < \varepsilon} \tag{$\star$}$$

  Since S is compact and $S \subset \bigcup\limits_{x \in S} U_x$.

  $\implies$ $\exists$ finite collection of neighborhoods $\{U_{x_i}\}$ that covers S. Partition S so that each rectangle is contained in some U_{x_i}. Do this partition for each $S \in C_2$, ad we obtain a refinement of $\mathcal{P}$, denoted by $\mathcal{P}'$.

- Verify Riemann's condition for $\mathcal{P}'$:

Note that

$$\begin{aligned} U(g,\mathcal{P}')-L(g,\mathcal{P}') &= \sum_{S'\in\mathcal{P}'}(M_S(g)-m_S(g))v(S) \\ &= \sum_{S'\subset S\in C_1}(M_{S'}(g)-m_{S'}(g))v(S')+\sum_{S'\subset S\in C_2}(M_{S'}(g)-m_{S'}(g))v(S') \\ &\le \sum_{S'\subset S\in C_1}2Mv(S')+\sum_{S'\subset S\in C_2}\varepsilon v(S') \qquad [|g(x)|\le M \quad \text{and} \quad (\star)] \\ &\le 2M\subset\sum_i v(B_i)+\varepsilon v(B) \qquad [C_1 \text{ is covered by } B_i's] \qquad < 2M\varepsilon+\varepsilon v(B) \qquad [I \\ &= \varepsilon(2M+v(B)). \end{aligned}$$

In summary, given $\varepsilon>0$, $\exists$ partition $\mathcal{P}'$ $s.t.$

$$U(g,\mathcal{P}')-L(g,\mathcal{P}')<\varepsilon(2M+v(B)).$$

So, we satisfy Riemann condition. $\square$

$\boxed{\textbf{Step } 3}$ **($\Rightarrow$) f is integrable $\implies m(D)=0$.**

For $n=1,2,\dots$, let

$$D_{1/n}=\left\{x\in D \mid \mathcal{O}(g,x)\ge\frac{1}{n}\right\}.$$

Then,

$$D=\bigcup_{i=1}^{\infty}D_{1/n}.$$

Need to show: $m(D_{1/n})=0 \quad \forall n.$

Fix $n\ge1$. For any partition $\mathcal{P}$, write

$$D_{1/n}=S_1\cup S_2,$$

where

$$S_1=\{x\in D_{1/n}\mid x \text{ is on the boundary of some } S\in\mathcal{P}\}$$

and

$$S_2=\{x\in D_{1/n}\mid x\in\operatorname{int}(S) \text{ for some } S\in\mathcal{P}\}.$$

Then, $m(S_1)=0$. We need to show $m(S_2)=0$.

Given $\varepsilon>0$. By Riemann's condition, $\exists$ partition $\mathcal{P}$ $s.t.$

$$\sum_{S\in\mathcal{P}}(M_S(g)-m_S(g))v(S)<\frac{\varepsilon}{n}.$$

Let C denote the collection of all $S\in\mathcal{P}$ $s.t.$ $D_{1/n}\cap\operatorname{int}(S)\ne\varnothing$. Then, C covers S_2 and for any $S\in C$,

$$M_S(g)-m_S(g)\ge\mathcal{O}(g,x)\ge\frac{1}{n}.$$

Thus,

$$\sum_{S\in C}(M_S(g)-m_S(g))v(S) \leq \sum_{S\in\mathcal{P}}(M_S(g)-m_S(g))v(S) < \frac{\varepsilon}{n}.$$

Since

$$\sum_{S\in C}(M_S(g)-m_S(g))v(S) \geq \sum_{S\in C}\frac{1}{n}v(S) = \frac{1}{n}\sum_{S\in C}v(S),$$

we have

$$\frac{1}{n}\sum_{S\in C}v(S) \leq \sum_{S\in C}(M_S(g)-m_S(g))v(S) < \frac{\varepsilon}{n}.$$

That is,

$$\frac{1}{n}\sum_{S\in C}v(S) < \frac{\varepsilon}{n} \implies \sum_{S\in C}v(S) < \varepsilon.$$

Therefore, $m(S_2)=0$ as well.

Since $m(S_1)=m(S_2)=0$ and $D_{1/n}=S_1\cup S_2$, $m(D_{1/n})=0 \quad \forall\, n$. Then,

$$m(D) = m\left(\bigcup_{i=1}^{\infty} D_{1/n}\right) = 0.$$

Q.E.D. ■

Theorem 8.2.11 Properties of Integration

Let $A, B \subset \mathbb{R}^n$ be bounded, $c\in\mathbb{R}$, and $f,g : A\to\mathbb{R}$ be integrable. Then,

- $f+g$ is integrable and $\displaystyle\int_a (f+g) = \int_A f + \int_A g$.
- cf is integrable and $\displaystyle\int_A (cf) = c\int_A f$.
- $|f|$ is integrable and $\displaystyle\left|\int_A f\right| \leq \int_A |f|$
- If $f\leq g$, then $\displaystyle\int_A f \leq \int_A g$.
- If A has volume and $|f|\leq M$, then $\displaystyle\left|\int_A f\right| \leq Mv(A)$.
- **(Mean Value Theorem for Integrals)**: If $f : A\to\mathbb{R}$ is continuous and A has volume and is compact and connected, then $\exists\, x_0 \in A\ s.t.\ \displaystyle\int_A f(x)\,\mathrm{d}x = f(x_0)v(A)$. The quantitive $\displaystyle\frac{1}{v(A)}\cdot\int_A f$ is called the *average* of f over A.
- Let $f : A\cup B\to\mathbb{R}$. If the sets A and B are such that $A\cap B$ has measure zero and $f \mid (A\cap B)$, $f\mid A$, and $f\mid B$ are all integrable, then f is integrable on $A\cup B$ and $\displaystyle\int_{A\cup B} = \int_A f + \int_B f$.

8.3 Improper Integrals

Goal: Study integral of the form $\int_A f(x)$, where $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is an arbitrary function and $A \subset \mathbb{R}^n$ is an arbitrary set.

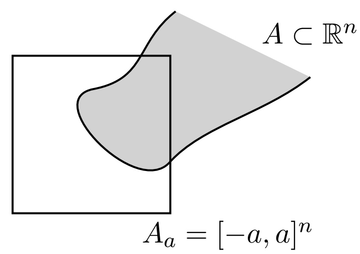
Definition 8.3.1 (Integral).

- If $A \subset \mathbb{R}^n$ is bounded and f is bounded, then

$$\int_A f(x) = \overline{\int_A} f(x) = \underline{\int_A} f(x) \qquad \text{(Riemann Condition)}$$

- $f(x) \geq 0$ bounded and A is arbitrary, then

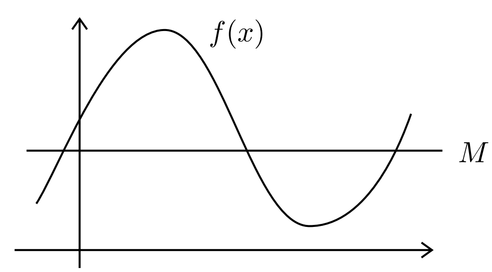
$$\int_A f(x) = \lim_{a\to\infty} \int_{A_a} f(x)$$



- $f(x) \geq 0$ unbounded and A is arbitrary.

  For $M > 0$, define

$$f_M(x) = \begin{cases} f(x) & \text{for } f(x) \leq M \\ 0 & \text{o/w.} \end{cases}$$



  Then,

$$\int_A f(x) = \lim_{M\to\infty} \int_A f_M(x).$$

- f is arbitrary and A is arbitrary.

  Let

$$f^+(x) = \begin{cases} f(x) & f(x) \geq 0 \\ 0 & f(x) < 0, \end{cases} \quad \text{and} \quad f^-(x) = \begin{cases} 0 & f(x) \geq 0 \\ -f(x) & f(x) < 0. \end{cases}$$

Remark 8.5 1. *$f^+(x)$ is the positive part of f and $f^-(x)$ is the negative part of f.*

2. *$f^+, f^- \geq 0$.*

3. *$f(x) = f^+(x) - f^-(x)$. We can write any function as the difference of two non-negative functions.*

4. $|f(x)| = f^+(x) + f^-(x)$.

So, f is integrable on A if both f^+ and f^- are integrable on A. We write

$$\int_A f(x) = \int_A f^+(x) - \int_A f^-(x).$$

Remark 8.6 *1. One can show this definition preserves linearity of integral from bounded case.*

2. ***Observation:*** *f integrable $\implies$ f^+ and f^- integrable $\implies$ $|f| = f^+ + f^-$ is also integrable. However, $|f|$ integrable $\not\Longrightarrow$ f integrable. For counterexample,*

$$f(x) = \begin{cases} 1 & x \text{ rational} \\ -1 & x \text{ irrational} \end{cases} \quad \text{on } [0,1].$$

$|f(x)| \equiv 1 \implies$ integrable. But $f^+,\ f^-$, or f are not integrable.

Theorem 8.3.2 Comparison Principle
Suppose

- $0 \leq g \leq f$ on A and $\displaystyle\int_A f$ converges (i.e., f is integrable on A)
- g is integrable on each finite rectangle $[-a,a]^n$.

Then, g is also integrable on A, and $\displaystyle\int_A g \leq \int_A f$.

Remark 8.7 *The second condition is crucial and cannot be removed.*

Proof 1. Since $g \geq 0$ and is integrable on $[-a,a]^n$, define

$$G(a) = \int_{[-a,a]^n} g(x).$$

Then, $G(a)$ is an increasing function of a. Furthermore,

$$g \leq f \implies G(a) = \int_{[-a,a]^n} g(x) \leq \int_{[-a,a]^n} f(x) \leq \int_A f(x).$$

So,

$$\int_A g(x) = \lim_{a\to\infty} G(a) \leq \int_A f(x).$$

Q.E.D. ■

Question: When does an integrable $\displaystyle\int_a^b f(x)$ (one-variable function) converge? If it converges, how to compute?

Theorem 8.3.3 Integral of Functions of One-Variable

- Suppose $f : [a, \infty] \to \mathbb{R}$ is continuous with $f(x) \geq 0$. Let $F'(x) = f(x)$. Then,

$$\int_a^\infty f(x)\,\mathrm{d}x = \lim_{b\to\infty} \int_a^b f(x)\,\mathrm{d}x = \lim_{b\to\infty} \Big[F(b) - F(a)\Big].$$

- Suppose $f : (a, b] \to \mathbb{R}$ is continuous with $f(x) \geq 0$. Then,

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{\varepsilon\to 0^+} \int_{a+\varepsilon}^b f(x)\,\mathrm{d}x.$$

Example 8.3.4

- Consider $\displaystyle\int_1^\infty x^p\,\mathrm{d}x$.

  Solution 2.

  For $b \geq 1$,

$$\int_1^b x^p\,\mathrm{d}x = \begin{cases} \ln b & p = -1 \\ \dfrac{1}{p+1}\big(b^{p+1} - 1\big) & p \neq -1. \end{cases}$$

  When $b \to \infty$, $\displaystyle\int_1^b x^p\,\mathrm{d}x$ diverges when $p \geq -1$ and converges when $p < -1$. So,

$$\int_1^\infty x^p\,\mathrm{d}x \quad \text{is divergent when } p \geq -1$$

  and

$$\int_1^\infty x^p\,\mathrm{d}x = -\frac{1}{p+1} \quad \text{is convergent when } p < -1.$$

  □

- Consider $\displaystyle\int_1^\infty e^{-x^2} x^3\,\mathrm{d}x$.

  Solution 3.

  Converges by comparison.

  □

Definition 8.3.5 (Conditional Convergence).

$$\int_a^\infty f(x)\,\mathrm{d}x \quad \text{(conditional)} = \lim_{b\to\infty} \int_a^b f(x)\,\mathrm{d}x.$$

Remark 8.8 (Types of Convergence) *For an improper integral* $\int_a^\infty f(x)\,\mathrm{d}x$*, there are three types of convergence:*

- ***Absolute Convergence****:* $\int_a^\infty |f(x)|\,\mathrm{d}x$ *exists.*
- ***Conditional Convergence****:* $\lim_{b\to\infty}\int_a^b f(x)\,\mathrm{d}x$ *exists.*
- ***Divergence****.*

For general function, absolute convergence $\not\Longrightarrow$ *conditional convergence. For continuous function, absolute convergence is stronger, and* $\Longrightarrow$ *conditional convergence.*

Example 8.3.6

Determine whether the integral $\int_1^\infty \frac{\cos x}{x}\,\mathrm{d}x$ is absolute convergence, conditional convergence, or neither (divergence).

Solution 4.

- First, consider absolute convergence.

  Observe that

$$\begin{aligned}\int_0^\infty \left|\frac{\cos x}{x}\right|\mathrm{d}x = \int_1^\infty \frac{|\cos x|}{x}\,\mathrm{d}x &\geq \int_{\pi/2}^{n\pi/2} \frac{|\cos x|}{x}\,\mathrm{d}x \\ &= \sum_{k=1}^{n-1}\int_{k\pi/2}^{(k+1)\pi/2}\frac{|\cos x|}{x}\,\mathrm{d}x \\ &\geq \sum_{k=1}^{n-1}\underbrace{\frac{1}{(k+1)\frac{\pi}{2}}}_{\text{harmonic}}\int_{k\pi/2}^{(k+1)\pi/2}|\cos x|\,\mathrm{d}x \\ &\to \infty \quad \text{as} \quad n\to\infty.\end{aligned}$$

  So, $\int_1^\infty \left|\frac{\cos x}{x}\right|\mathrm{d}x$ diverges, and thus $\int_1^\infty \frac{\cos x}{x}\,\mathrm{d}x$ is not absolutely convergent.

- Conditional convergence:

$$\int_1^b \frac{\cos x}{x}\,\mathrm{d}x = \frac{\sin x}{x}\bigg|_1^b + \int_1^b \frac{\sin x}{x^2}\,\mathrm{d}x \qquad \text{[Integration by Parts]}$$

  When $b\to\infty$,

$$\lim_{b\to\infty}\frac{\sin x}{x}\bigg|_1^b = \frac{\sin 1}{1} \quad \text{converges.}$$

  Further,

$$\left|\frac{\sin x}{x^2}\right| \leq \left|\frac{1}{x^2}\right| = \frac{1}{x^2} \implies \int_1^\infty \left|\frac{\sin x}{x^2}\right|\mathrm{d}x \leq \int_1^\infty \frac{1}{x^2}\,\mathrm{d}x.$$

So, $\displaystyle\int_1^\infty \frac{\sin x}{x^2}\,\mathrm{d}x$ absolutely converges by comparison.

Then, $\displaystyle\int_1^b \frac{\cos x}{x}\,\mathrm{d}x$ is conditional convergence.

□

8.4 Lebesgue Convergence Theorem

Goal: When do we have

$$\lim_{n\to\infty}\int_A f(x)\,\mathrm{d}x = \int_A \Big(\lim_{n\to\infty} f(x)\Big)\,\mathrm{d}x? \tag{$\star$}$$

Theorem 8.4.1 Lebesgue Monotone Convergence Theorem (LMCT)
Let $g_n : [0,1] \to \mathbb{R}$ be a sequence of non-negative integrable function such that

- $g_{n+1}(x) \leq g_n(x) \quad \forall\, x \in [0,1]$ (decreasing sequence)
- $\lim\limits_{n\to\infty} g_n(x) = 0 \quad \forall\, x \in [0,1]$.

Then,

$$\lim_{n\to\infty}\int_0^1 g_n(x)\,\mathrm{d}x = \int_0^1 0\,\mathrm{d}x = 0.$$

Corollary 8.4.2 : Suppose $f_n(x), f(x) : [0,1] \to \mathbb{R}$ with

- $f_n \leq f_{n+1}(x) \leq f(x) \quad \forall\, x \in [0,1]$
- $f_n(x) \to f(x) \quad \forall\, x$.

Then,

$$\lim_{n\to\infty}\int_0^1 f_n(X)\,\mathrm{d}x = \int_0^1 f(x)\,\mathrm{d}x.$$

Proof 1. Apply LMCT to the sequence $g_n(x) = f(x) - f_n(x) \geq 0$.

Q.E.D. ■

Remark 8.9

- *For ($\star$) to hold, we only need $f_n(x) \uparrow f(x)$ ($f_n(x)$ is monotone increasing and the limit of $f_n(x)$ is $f(x)$)*
- *The assumption that $A = [0,1] \subset \mathbb{R}$ is not essential. Result is true for any set $A \subset \mathbb{R}^n$.*
- *The monotonicity assumption cannot be removed. For example:*

$$g_n(x) = \begin{cases} n, & 0 < x < \dfrac{1}{n} \\ 0, & \textit{o/w} \end{cases}$$

Then, we have $g_n(x) \to 0 \quad \forall\, x \in [0,1]$. *However,*

$$\int_0^1 g_n(x)\,\mathrm{d}x = 1 \quad \forall\, n \quad \textit{and} \quad \int_0^1 0\,\mathrm{d}x = 0.$$

So,

$$\int_0^1 g_n\,\mathrm{d}x \neq \int_0^1 0\,\mathrm{d}x,$$

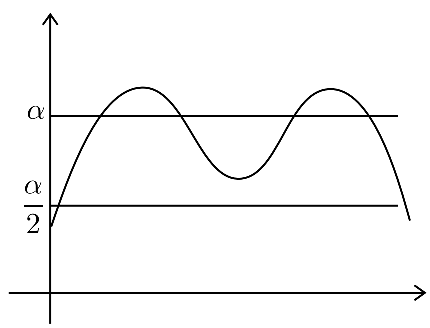
and LMCT does not hold anymore.

▶ **Proof 2 of Lebesgue Monotone Convergence Theorem**

Lemma 8.4.3 : Suppose $f : [0,1] \to \mathbb{R}$ be integrable with $|f| \le M$ and $\displaystyle\int_0^1 f \ge \alpha > 0$. Then, the set

$$E = \left\{ x \in [0,1] \mid f(x) \ge \frac{\alpha}{2} \right\}$$

contains a finite union of disjoint open intervals of total length $\ge \dfrac{\alpha}{4M}$.



Proof. By definition of integral, $\exists$ partition $\mathcal{P}$ $s.t.$

$$0 \le \int_0^1 f - L(f,\mathcal{P}) < \frac{\alpha}{4}.$$

Then,

$$L(f,\mathcal{P}) > \int_0^1 f - \frac{\alpha}{4} \ge \alpha - \frac{\alpha}{4} = \frac{3\alpha}{4}.$$

Let ℓ denote the total length of the intervals I in $\mathcal{P}$ with $I \subset E$. Then,

$$\begin{aligned}
\frac{3\alpha}{4} < L(f,\mathcal{P}) &= \sum_{I\in\mathcal{P}} \left(\inf_I f(x)\right)\ell(I) \\
&= \sum_{I\in\mathcal{P}\cap E} \left(\inf_I f(x)\right)\ell(x) + \sum_{I\in\mathcal{P}\setminus E} \left(\inf_I f(x)\right)\ell(I) \\
&\le \sum_{I\in\mathcal{P}\cap E} M\cdot\ell(I) + \sum_{I\in\mathcal{P}\setminus E} \frac{\alpha}{2}\ell(I) \\
&\le \ell M + \frac{\alpha}{2}\cdot 1 && \left[\text{If } I \notin E,\ f(x) \le \frac{\alpha}{2}\right]
\end{aligned}$$

So, $\ell \cdot M \geq \dfrac{\alpha}{4} \implies \ell \geq \dfrac{\alpha}{4M}$. Remove endpoints from I, we get open intervals. $\square$

- $\boxed{\text{Step 1}}$ Set up and Reduction:

$$0 \leq g_{n+1} \leq g_n \implies \int_0^1 g_{n+1}(x)\,\mathrm{d}x \leq \int_0^1 g_n(x)\,\mathrm{d}x.$$

  Then, the limit exists:

$$\lim_{n\to\infty} \int_0^1 g_n(x)\,\mathrm{d}x =: \lambda \geq 0.$$

  Need to show: $\lambda = 0$.

  Assume $\lambda > 0$, and we will derive a contradiction (with the assumption $g_n(x) \to 0 \quad \forall\, x \in [0,1]$).

- $\boxed{\text{Step 2}}$ Apply the above Lemma 8.4.3 to the cut-off function $(g_n)_M$, where $M > 0$.

$$(g_n(x))_M \coloneqq \begin{cases} g_n(x), & g_n(x) \leq M \\ M, & g_n(x) > M. \end{cases}$$

  Then,

$$\int_0^1 g_n(x)\,\mathrm{d}x = \lim_{M\to\infty} \int_0^1 (g_n)_M.$$

  Choose $M = \dfrac{2\lambda}{5}$ *s.t.*

$$0 \leq \int_0^1 (g_n - (g_n)_M) \leq \int_0^1 (g_1 - (g_1)_M) \leq \frac{\lambda}{5}.$$

  Let $E_n = \left\{ x \in [0,1] \mid g_n(x) \geq \dfrac{2\lambda}{5} \right\}$. Then,

  1. $E_{n+1} \subset E_n$ by monotonicity
  2. $\left\{ x \in [0,1] \mid (g_n)_M(x) \geq \dfrac{\alpha}{2} \right\} \subset E_n$. Choose α *s.t.* $\dfrac{2\lambda}{5} = \dfrac{\alpha}{2}$ to apply the Lemma. $\implies \alpha = \dfrac{4\lambda}{5}$.

  Apply Lemma 8.4.3 to $(g_n)_M$ and $\alpha = \dfrac{4\lambda}{5}$. Then, E_n contains a finite union of disjoint open intervals of total length

$$\ell \geq \frac{\alpha}{4M} = \frac{4\lambda}{5} \cdot \frac{1}{4M} = \frac{\lambda}{5M}$$

- $\boxed{\text{Step 3}}$ Show that $\bigcap_{n=1}^{\infty} E_n \neq \varnothing$.

  Let

$$D = \bigcup_{n=1}^{\infty} \{x \in [0,1] \mid g_n \text{ not continuous at } x\} = \bigcup_{n=1}^{\infty} D_n.$$

  Since g_n is integrable, we have $m(D_n) = 0$. So,

$$m(D) = m\left(\bigcup_{n=1}^{\infty} D_n \right) = 0.$$

That is, D is covered by U, a countable union of open intervals of total length $< \varepsilon = \dfrac{\lambda}{5M}$.

By Step 2, $E_n \not\subset U$.

Claim 8.4.4 $\mathrm{cl}(E_n) \subset E_n \cup U$.

Proof. In fact, if $x_0 \in \mathrm{cl}(E_n) \backslash E_n$, then [WTS: $x_0 \in U$]

$$g_n(x_0) < \frac{2\lambda}{5} \implies g_n \text{ is not continuous at } x_0.$$

Suppose $x_0 \in \mathrm{cl}(E_n) \implies \exists x_k \in E_n$ $s.t.$ $x_k \to x_0$ as $k \to \infty$. Also, $g_n(x_k) \geq \dfrac{2\lambda}{5}$, but $g_n(x_0) < \dfrac{2\lambda}{5} \implies g_n(x_k) \neq g_n(x_0) \implies$ discontinuous

So, $x_0 \in D_n$, and thus $x_0 \in U$. So, this Claim 8.4.4 is true. □

Note, let $F_n = \mathrm{cl}(E_n) \backslash U$. Then,

1. F_n is compact
2. $F_n \subset E_n$ (by Clam 8.4.4)

So, by the nested set property: $\displaystyle\bigcap_{n=1}^{\infty} F_n \neq \varnothing$. As $F_n \subset E_n$, we further have $\displaystyle\bigcap_{n=1}^{\infty} E_n \neq \varnothing$.

Let $x_0 \in \displaystyle\bigcap_{n=1}^{\infty} E_n$, then $g_n(x_0) \geq \dfrac{2\lambda}{5}$. Then, $\displaystyle\lim_{n\to\infty} g_n(x_0) \neq 0$. ※This derives a contradiction with the second assumption in LMCT (i.e., $g_n(x) \to 0$). So, $\lambda > 0$ is impossible, and it must be that $\lambda = 0$.

Q.E.D. ■

Corollary 8.4.5 : Let $g_n : A \to \mathbb{R}$ be integrable and non-negative. Assume

$$g(x) = \sum_{n=1}^{\infty} g_n(x)$$

is also integrable. Then,

$$\int_A g(x) = \int_A \sum_{n=1}^{\infty} g_n(x) = \sum_{n=1}^{\infty} \int_A g_n(x).$$

Proof 3. Let $f_n(x) = \displaystyle\sum_{k=1}^{n} g_k(x)$, the partial sum.

Then,

$$\int_A f_n(x) = \int_A \sum_{k=1}^{n} g_k(x) = \sum_{k=1}^{n} \int_A g_k(x) \quad \text{[property of integral]}$$

As $n \to \infty$, $f_n \to g(x)$, and $f_{n+1} \geq f_n$ (g_n is non-negative). Then, apply Corollary 8.4.2, we have

$$\int_A g(x) = \sum_{n=1}^{\infty} \int_A g_n(x).$$

Q.E.D. ■

9 Computing Integrals

Question: In practice, how do we compute the integral $\int_A f(x)\,\mathrm{d}x$?

- In $\mathbb{R}^1$: Fundamental Theorem of Calculus.

$$\int_a^b f(x)\,\mathrm{d}x = F(x)\Big|_a^b = F(b) - F(a).$$

- In $\mathbb{R}^n$: Reduce to $\mathbb{R}^1$ case by *Fubini's Theorem*. Or, use *change of variable* (substitution first), and then use Fubini's Theorem.

9.1 Fubini's Theorem

Theorem 9.1.1 Fubini's Theorem
Let $A = \{(x,y) \mid a \le x \le b,\ c \le y \le d\}$ be a rectangle in $\mathbb{R}^2$ and $f : A \to \mathbb{R}$ be integrable. Suppose for each fixed $x \in [a,b]$, the following integral exists:

$$g(x) = \int_c^d f(x,y)\,\mathrm{d}y.$$

Then, $g(x)$ is integrable on $[a,b]$, and

$$\int_A f(x,y) = \int_a^b g(x)\,\mathrm{d}x = \int_a^b \left(\int_c^d f(x,y)\,\mathrm{d}y \right) \mathrm{d}x.$$

Corollary 9.1.2 : If $f : A \to \mathbb{R}$ is continuous, then

$$\int_A f(x,y) = \int_a^b \left(\int_c^d f(x,y)\,\mathrm{d}y \right) \mathrm{d}x \overset{\text{symmetry}}{=\!=\!=} \int_c^d \left(\int_a^b f(x,y)\,\mathrm{d}x \right) \mathrm{d}y.$$

Corollary 9.1.3 Generalization: Let A be a region given by $A = \{(x,y) \mid a \le x \le b,\ \varphi(x) \le y \le \psi(x)\}$, where φ and ψ are continuous. If $f : A \to \mathbb{R}$ is continuous, then

$$\int_A f(x,y) = \int_a^b \left(\int_{\varphi(x)}^{\psi(x)} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x.$$

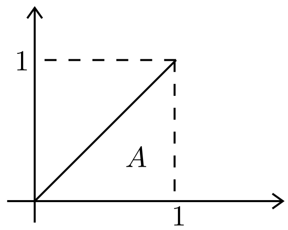
Remark 9.1

- *The roles of x and y can be interchanged.*
- *Results are true in higher dimensions. For example, let $C = A \times B \subset \mathbb{R}^{n+m}$, where $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$. Fix $x \in A$ and $y \in B$. Then,*

$$\int_{A\times B} f = \int_A \left(\int_B f(x,y)\,\mathrm{d}y \right) \mathrm{d}x.$$

Example 9.1.4 Computing Integral

Compute $\int_A (x+y)\,\mathrm{d}x\mathrm{d}y$, where A is the following region:



Solution 1.

$$
\begin{aligned}
\int_A (x+y)\,\mathrm{d}x\mathrm{d}y &= \int_0^1 \left(\int_0^x (x+y)\,\mathrm{d}y \right) \mathrm{d}x \\
&= \int_0^1 \left(xy + \frac{1}{2}y^2 \right) \Bigg|_0^1 \mathrm{d}x \\
&= \int_0^1 x^2 + \frac{1}{2}x^2\,\mathrm{d}x \\
&= \frac{3}{2} \cdot \frac{1}{3}\, x^3 \Big|_0^1 \\
&= \frac{1}{2}.
\end{aligned}
$$

□

► **Proof 2 of Fubini's Theorem**

- Let $g(x) = \int_c^d f(x,y)\,\mathrm{d}y$. WTS: (1) g is integrable on $[a,b]$, and (2) $\int_a^b g\,\mathrm{d}x = \int_A f$. We will compute the upper and lower sums of f and g.
- Fix any partition $\mathcal{P}_A$ of A, where $\mathcal{P}_A = \{S_{i,j}\}_{i,j}$, where $S_{i,j} = v_i \times w_j$. Then, $\mathcal{P}_A$ induces a partition of $[a,b]$, where $\mathcal{P}_{[a,b]} = \{v_i\}_i$ and a partition of $[c,d]$, $\mathcal{P}_{[c,d]} = \{w_j\}_j$.
- Next, estimate the lower sum $L(f, \mathcal{P}_A)$:

$$
\begin{aligned}
L(f,\mathcal{P}_A) &= \sum_{i,j} \underbrace{\inf_{x \in S_{i,j}} f(x)}_{\text{denote as } m_{i,j}(f)} v(S_{i,j}) \\
&= \sum_{i,j} m_{i,j}(f) v(v_i \times w_j) \\
&= \sum_{i,j} m_{i,j}(f) v(v_i) \cdot v(w_j).
\end{aligned}
$$

Key Observation:

$$\inf\{f(x,y) \mid (x,y) \in v_i \times w_j\} \leq \underbrace{\inf\{f(x,y) : y \in w_j\}}_{\text{fix } x \text{, allow } y \text{ to vary}} \quad \forall\, x \in v_i.$$

Denote $\inf\{f(x,y) \mid y \in w_j\} = m_j(f,x)$. Then, for any fixed $x \in [a,b]$,

$$\begin{aligned} m_{i,j}(f) &\leq m_j(f,x) \\ m_{i,j}(f)v(w_j) &\leq m_j(f,x)v(w_j) \\ \sum_j m_{i,j}(f)v(w_j) &\leq \underbrace{\sum_j m_j(f,x) \cdot v(w_j)}_{\substack{\text{lower sum of } f(x,y) \text{ in the} \\ \text{variable } y \text{ } w.r.t. \text{ partition } \mathcal{P}_{[c,d]}}} \\ &= L\big(f(x,y), \mathcal{P}_{[c,d]}\big) \\ &\leq \int_c^d f(x,y)\,\mathrm{d}y = g(x) \quad \forall x. \end{aligned}$$

Thus,

$$\begin{aligned} \sum_j m_{i,j}(f)v(w_j) &\leq \inf_{v_i} g(x) \\ \sum_j m_{i,j}(f)v(w_j)v(v_i) &\leq \inf_{v_i} g(x)v(v_i) \\ \sum_i \sum_j m_{i,j}(f)v(w_j)v(v_i) &\leq \sum_i \inf_{v_i} g(x)v(v_i) \\ \underbrace{\sum_{i,j} m_{i,j}(f)v(w_j)v(v_i)}_{L(f,\mathcal{P}_A)} &\leq \underbrace{\sum_i \inf_{v_i} g(x)v(v_i)}_{L(g,\mathcal{P}_{[a,b]})} \end{aligned}$$

So,

$$L(f,\mathcal{P}_A) \leq L\big(g,\mathcal{P}_{[a,b]}\big).$$

- Similarly, we have

$$U(f,\mathcal{P}_A) \geq U\big(g,\mathcal{P}_{[a,b]}\big).$$

- Therefore, we have

$$L(f,\mathcal{P}_A) \leq L\big(g,\mathcal{P}_{[a,b]}\big) \leq U\big(g,\mathcal{P}_{[a,b]}\big) \leq U(f,\mathcal{P}_A).$$

Since f is integrable, by Riemann's condition,

$$0 \leq U(f,\mathcal{P}_A) - L(f,\mathcal{P}_A) < \varepsilon.$$

Then,

$$0 \leq U\big(g,\mathcal{P}_{[a,b]}\big) - L\big(g,\mathcal{P}_{[a,b]}\big) < \varepsilon.$$

So, g is integrable as well. Moreover,

$$\int_a^b g(x)\,\mathrm{d}x = \int_A f.$$

Q.E.D. ■

Example 9.1.5

Compute the volume of the region

$$A = \{(x,y,z) \mid x \geq 0,\ y \geq 0,\ z \geq 0,\ x+y+z \leq 1\}$$

by integration.

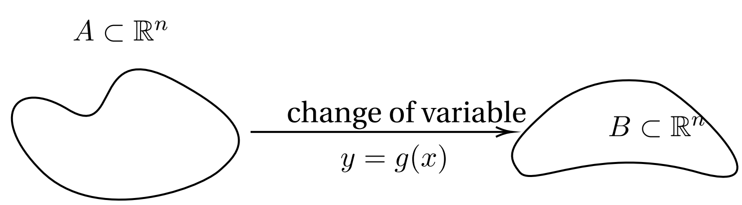
Solution 3.

$$v(A) \int_A \mathbb{1}_A = \int_0^1 \int_0^{1-x} \int_0^{1-x-y} 1\,\mathrm{d}z\mathrm{d}y\mathrm{d}x.$$

□

9.2 Change of Variable

General Setting: $f : B \to \mathbb{R}$ bounded is an integrable function



Goal: Transform integral $\int_B f(y)$ to an integral on A.

Example 9.2.1 1D Case

$$\int f(y)\,\mathrm{d}y = \int f(g(x))\,\underbrace{g'(x)\,\mathrm{d}x}_{\mathrm{d}y}.$$

Theorem 9.2.2 Change of Variable Formula in Higher Dimension

Assume $J_g(x) \neq 0 \quad \forall x \in A$. If $f : B \to \mathbb{R}$ is bounded and integrable on $B = g(A)$, then $f \circ g(x) \cdot (J_g(x))$ is integrable on A, and

$$\int_B f(y)\,\mathrm{d}y = \int_A f(g(x)) \cdot \underbrace{|J_g(x)|\,\mathrm{d}x}_{\mathrm{d}y}.$$

Proof 1. (Sketch)

- Change of volume under linear map:

  Let $\mathbf{L} : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map given by

$$\mathbf{L}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

  Denote $y = \mathbf{L}x$. Then,

$$v(L(s)) = \left|\det \begin{bmatrix} a & b \\ c & d \end{bmatrix}\right| \cdot v(S).$$

- Linear approximation of $g : A \to B$:

  Fix $x_0 \in A$. Then, in a neighborhood of x_0, g can be approximated by a linear map:

$$g(x) = g(x_0) + \mathbb{D}g(x_0)(x - x_0) + \text{error}.$$

- Conversion into integral formula:

  Fix small rectangles S in A. Then, $g(S)$ is "1nearly" parallelogram. So,

$$v(g(S)) \approx |J_g(x_0)|v(S).$$

  Do this for each rectangle S_{ij} in a partition:

$$v(g(S_{ij}) \approx |J_g(x_{ij})|v(S_{ij}).$$

  Then,

$$\begin{aligned} f(y_{ij})v(g(S_{ij})) &\approx f(g(x_{ij}))|J_g(x_{ij})|v(S_{ij}) \\ \sum f(y_{ij})v(g(S_{ij})) &\approx \sum f(g(x_{ij}))|J_g(x_{ij})|v(S_{ij}) \end{aligned}$$

  Through the summation and limit process:

$$\int_B f(y)\,\mathrm{d}y = \int_A f(g(x))|J_g(x)|\,\mathrm{d}x.$$

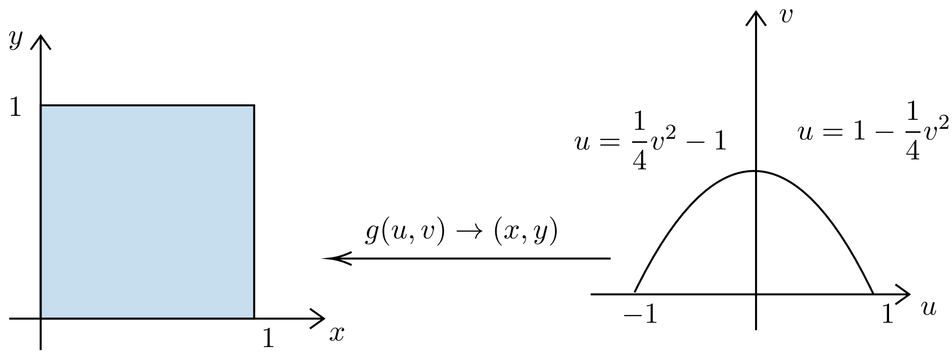
Q.E.D. ■

Example 9.2.3

Evaluate the integral using the change of variables $u = x^2 - y^2$ and $v = 2xy$.

$$\int_0^1 \int_0^1 (x^2 + y^2)\sin\left(x^2 - y^2\right)\,\mathrm{d}x\mathrm{d}y.$$

1. Sketch the regions in xy-plane and uv-plane:



$$\begin{cases} u = x^2 - y^2 \\ v = 2xy \end{cases} \quad \text{and} \quad g : (u, v) \to (x, y) \implies u = x^2 - \frac{v^2}{4x^2}.$$

2. Compute the determinant: $g^{-1} : (u, v) \to (x, y)$.

$$J_{g^{-1}}(x, y) = \begin{vmatrix} \partial u/\partial x & \partial u/\partial y \\ \partial v/\partial x & \partial v/\partial y \end{vmatrix} = \begin{vmatrix} 2x & -2y \\ 2y & 2x \end{vmatrix} = 4x^2 + 4y^2.$$

So,

$$J_g = \frac{1}{J_{g^{-1}}(x, y)} = \frac{1}{4(x^2 + y^2)}$$

3. Apply the change of variable formula:

$$\begin{aligned} \int_0^1 \int_0^1 (x^2 + y^2) \sin\big(x^2 - y^2\big)\, \mathrm{d}x\mathrm{d}y &= \int_0^2 \int_{(1/4)\cdot v^2 - 1}^{1-(1/4)\cdot v^2} (x^2 + y^2) \sin\big(x^2 - y^2\big) |J_g(x)|\, \mathrm{d}u\mathrm{d}v \\ &= \int_0^2 \int_{(1/4)\cdot v^2 - 1}^{1-(1/4)\cdot v^2} \cancel{(x^2 + y^2)} \sin(u) \frac{1}{4\cancel{(x^2 + y^2)}}\, \mathrm{d}u\mathrm{d}v \\ &= \frac{1}{4} \int_0^2 \int_{(1/4)\cdot v^2 - 1}^{1-(1/4)\cdot v^2} \sin(u)\, \mathrm{d}u\mathrm{d}v. \end{aligned}$$

Remark 9.2 (Special Coordinate Systems)

- *Polar Coordinate in* $\mathbb{R}^2$*:*

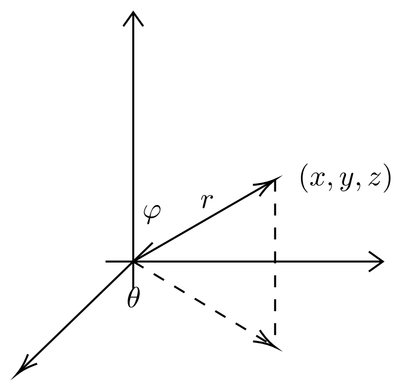
$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases} \qquad J_g(r, \theta) = r,$$

$$\implies \int_B f(x, y)\, \mathrm{d}x\mathrm{d}y = \int_A f(r\cos\theta, r\sin\theta) r\, \mathrm{d}r\mathrm{d}\theta.$$

- *Spherical Coordinate in* $\mathbb{R}^3$*:*

$$\begin{cases} x = r\sin\varphi\cos\theta \\ y = r\sin\varphi\sin\theta \\ z = r\cos\varphi \end{cases} \qquad J_g(r,\theta,\varphi) = r^2\sin\varphi$$

$$\Longrightarrow \int_B f(x,y,z)\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_A f(r\sin\varphi\cos\theta, r\sin\varphi\sin\theta, r\cos\varphi)r^2\sin\varphi\,\mathrm{d}r\mathrm{d}\theta\mathrm{d}\varphi.$$



- *Cylindrical Coordinate in* $\mathbb{R}^3$*:*

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \\ z = z \end{cases} \qquad J_g(r,\theta,z) = r$$

$$\Longrightarrow \int_B f(x,y,z)\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_A f(r\cos\theta, r\sin\theta, z)r\,\mathrm{d}r\mathrm{d}\theta\mathrm{d}z.$$

Example 9.2.4

- Evaluate $\displaystyle\int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x$

  Solution 2.

  $\boxed{\text{Step 1}}$ Evaluate integral

$$\int_{\mathbb{R}^2} e^{-x^2-y^2}\,\mathrm{d}x\mathrm{d}y$$

by polar coordinate ($x = r\cos\theta$ and $y = r\sin\theta$). Let D_R denote the circle centered at origin with radius R Then,

$$\int_{D_R} e^{-x^2-y^2}\,\mathrm{d}x\mathrm{d}y = \int_0^{2\pi}\int_0^R e^{-r^2}r\,\mathrm{d}r\mathrm{d}\theta = \int_0^{2\pi}\left(-\frac{1}{2}e^{-r^2}\right)\Big|_0^R\,\mathrm{d}\theta$$
$$= 2\pi\left(-\frac{1}{2}e^{-R^2}+\frac{1}{2}\right) = -\pi e^{-R^2}+\pi.$$

So,

$$\begin{aligned}\int_{\mathbb{R}^2} e^{-x^2-y^2}\,\mathrm{d}x\mathrm{d}y &= \lim_{\mathbb{R}\to\infty}\int_{D_R} e^{-x^2-y^2}\,\mathrm{d}x\mathrm{d}y\\ &= \lim_{R\to\infty}\left(-\pi e^{-R^2}+\pi\right)\\ &= \pi.\end{aligned}$$

$\boxed{\text{Step 2}}$ Evaluate

$$\int_{\mathbb{R}^2} e^{-x^2-y^2}\,\mathrm{d}x\mathrm{d}y$$

by Fubini's Theorem.

Let $S_b = [-b,b]\times[-b,b]\subset\mathbb{R}^2$. Then,

$$\begin{aligned}\int_{\mathbb{R}^2} e^{-x^2-y^2}\,\mathrm{d}x\mathrm{d}y &= \lim_{b\to\infty}\int_{S_b} e^{-x^2-y^2}\,\mathrm{d}x\mathrm{d}y\\ &= \lim_{b\to\infty}\int_{-b}^{b}\int_{-b}^{b} e^{-x^2}\cdot e^{-y^2}\,\mathrm{d}x\mathrm{d}y\\ &= \lim_{b\to\infty}\left(\int_{-b}^{b} e^{-x^2}\,\mathrm{d}x\right)\cdot\left(\int_{-b}^{b} e^{-y^2}\,\mathrm{d}y\right)\\ &= \left(\int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x\right)^2\end{aligned}$$

$\boxed{\text{Step 3}}$ Combine Steps 1 and 2:

$$\pi = \int_{\mathbb{R}^2} e^{-x^2-y^2}\,\mathrm{d}x\mathrm{d}y = \left(\int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x\right)^2$$

So,

$$\int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x = \sqrt{\pi}.$$

$\square$

- Evaluate $\displaystyle\int_{\mathbb{R}^3}\frac{1}{x^2+y^2+x^2}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z$
- Evaluate $\displaystyle\int_R 2e^{x^2-y^2}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z$, where $R=\left\{(x,y,z)\mid x^2+y^2\le 1, 1\le x\le 2\right\}$.

10 Fourier Analysis

10.1 Introduction

General Idea: Try to decompose certain objects into simpler components.

- Algebraic Model: $\mathbb{R}^n$
$$x = \sum_{i=1}^{n} x_i e_i,$$
where e_i's are the standard basis.
- Calculus Model: Taylor Series
$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n.$$
- Fourier Analysis: Theory of infinite dimensional inner product space of functions.

Goal: Decompose a function $f(x)$ into a "linear combination of basis:"

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \varphi_n(x).$$

Physics Motivation: Decompose complicated waves into harmonies.

10.2 Inner Product Space of Functions

10.2.1 Basic Concepts

Definition 10.2.1 (Inner Product). Let V be a complex vector space. Then, an *inner product* on V is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ *s.t.* $\forall\, f, g, h \in V$ and $a, b, \in \mathbb{C}$, we have

- Linearity:
$$\langle af + bg, h \rangle = a \langle f, h \rangle + b \langle g, h \rangle .$$
- Conjugate Symmetry:
$$\langle f, g \rangle = \overline{\langle g, f \rangle}.$$
- Positive Definiteness:
$$\langle f, f \rangle \geq 0 \quad \text{and} \quad \langle f, f \rangle = 0 \iff f = 0.$$

Example 10.2.2

$\mathbb{C}$ is an inner product space under the inner product:

$$\langle z_1, z_2 \rangle = z_1 \overline{z_2}.$$

Corollary 10.2.3 Conjugate Linearity in the Second Component:

$$\langle h, af + bg \rangle = \overline{a} \langle h, f \rangle + \overline{b} \langle h, g \rangle .$$

Proof 1.

$$\begin{aligned} \langle h, af + bg \rangle &= \overline{\langle af + bg, h \rangle} && \text{[Conjugate symmetry]} \\ &= \overline{a \langle f, g \rangle} + \overline{b \langle g, h \rangle} && \text{[Linearity]} \\ &= \overline{a} \langle h, f \rangle + \overline{b} \langle h, g \rangle . && \text{[Conjugate symmetry]} \end{aligned}$$

Q.E.D. ■

Definition 10.2.4 (Norm and Distance Induced by Inner Product).

- Norm:
$$\|f\| := \sqrt{\langle f, f \rangle}.$$
- Distance from f to g:
$$d(f, g) := \|f - g\|.$$

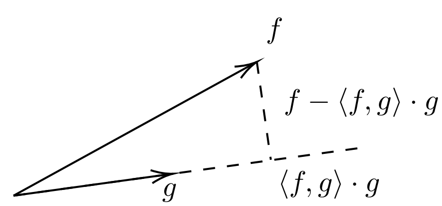
Corollary 10.2.5 Facts:

- $(V, \|\cdot\|)$ is a normed space.
- (V, d) is a metric space.

Lemma 10.2.6 Cauchy-Schwarz Inequality:

$$|\langle f, g \rangle| \leq \|f\| \cdot \|g\|$$

Proof 2.



The projection should have the smallest length:

$$\begin{aligned} 0 \leq \|f - \langle f, g \rangle g\|^2 &= \langle f, \langle f, g \rangle g, f - \langle f, g \rangle g \rangle \\ &= \langle f, f - \langle f, g \rangle g \rangle - \langle f, g \rangle \langle g, f - \langle f, g \rangle g \rangle \\ &= \langle f, f \rangle - \overline{\langle f, g \rangle} \langle f, g \rangle - \langle f, g \rangle \langle g, f \rangle + \langle f, g \rangle \overline{\langle f, g \rangle} \langle g, g \rangle \\ &= \|f\|^2 - |\langle f, g \rangle|^2 - |\langle f, g \rangle|^2 + |\langle f, g \rangle|^2 \|g\|^2. \end{aligned}$$

Normalize: let $\|g\| = 1$. Then,

$$\begin{aligned} 0 &\le \|f\|^2 - |\langle f, g\rangle|^2 \\ |\langle f, g\rangle|^2 &\le \|f\|^2 \\ |\langle f, g\rangle| &\le \|f\| = \|f\| \cdot \|g\|. \end{aligned}$$

Q.E.D. ■

Definition 10.2.7 (Convergence). Suppose $f_n, f \in V$. Then, $f_n \to f$ in V if $\|f_n - f\| \to 0$ as $n \to \infty$. We call this *convergence in norm.*

10.2.2 The Space $\mathcal{C}$ and L^2

Definition 10.2.8 (Integral of Complex Valued Functions). Suppose $f(x) = f_1(x) + \mathrm{i} f_2(x) : [a, b] \to \mathbb{C}$ be a complex-valued function, where $f_1, f_2 : [a, b] \to \mathbb{R}$. Then,

$$\int_a^b f(x)\,\mathrm{d}x := \int_a^b f_1(x)\,\mathrm{d}x + \mathrm{i}\int_a^b f_2(x)\,\mathrm{d}x.$$

Definition 10.2.9 (The Space $\mathcal{C}$ and L^2). Fix an interval $[a, b]$.

- $\mathcal{C} := \{f(x) \mid f : [a, b] \to \mathbb{C}$ is continuous$\}$.
- $L^2 := \left\{ f : [a, b] \to \mathbb{C} \mid \displaystyle\int_a^b |f(x)|^2\,\mathrm{d}x < \infty \right\}$.

  The condition $\displaystyle\int_a^b |f(x)|^2\,\mathrm{d}x < \infty$ is called L^2 *integrable.*

Corollary 10.2.10 Facts:

- $\mathcal{C}$ and L^2 are vectors spaces. $\mathcal{C}$ is a subspace of L^2.
- Zero vector in $\mathcal{C}$: $f(x) \equiv 0$.
- Zero vector in L^2: $f(x) = 0 \quad \text{a. e.}$ (almost everywhere).

  That is, $m(\{x \in [a, b] \mid f(x) \neq 0\}) = 0$.
- $\underbrace{f_1 = f_2}_{\text{vectors}}$ in $L^2 \iff \underbrace{f_1(x) = f_2(x)}_{\text{function}} \quad \text{a. e.}$
- Inner Product:

$$\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}\,\mathrm{d}x.$$

Claim 10.2.11 With the above definition of inner product, $\mathcal{C}$ and L^2 are inner product spaces.

10.3 Fourier Analysis on Inner Product Space

10.3.1 Geometry of an Inner Product Space

Definition 10.3.1 (Orthogonality). $f, g \in V$ are *orthogonal* (denoted as $f \perp g$) if $\langle f, g \rangle = 0$.

Definition 10.3.2 (Orthonormal Family). A family $\{\varphi_1, \varphi_2, \dots,\} \subset V$ is called an *orthonormal family* if

- $\langle \varphi_i, \varphi_j \rangle = 0 \quad \forall\, i \neq j$
- $norm \varphi_i = 1 \quad \forall\, i.$

Or equivalently,

$$\langle \varphi_i, \varphi_j \rangle = \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases}$$

Example 10.3.3

In $\mathbb{R}^n$: $\{e_1, e_2, \dots, e_n\}$, the standard basis, is an orthonormal basis.

Theorem 10.3.4 Gram-Schmidt Process: Generate Orthonormal Family from Linear Independent Family

$$\underbrace{\{g_0, g_1, \dots\}}_{\text{linear independent}} \to \underbrace{\{\varphi_0, \dots, \varphi_1, \dots\}}_{\text{orthonormal}}$$

1. Orthogonal projection:
$$x = \sum_i c_i e_i,$$
where $c_i = \langle x, e_i \rangle$. Then, we have
$$\langle x - \langle x, e_i \rangle e_i, e_i \rangle = 0.$$

2. Inductive Process:
$$\begin{aligned} \varphi_0 &= \frac{g_0}{\|g_0\|} \\ f_1 &= g_1 - \langle g_1, \varphi_0 \rangle \varphi_0, && \implies \varphi_1 = \frac{f_1}{\|f_1\|} \\ &\vdots \\ f_n &= g_n - \sum_{i=0}^{n-1} \langle g_n, \varphi_i \rangle \varphi_i, && \implies \varphi_n = \frac{f_n}{\|f_n\|}. \end{aligned}$$

10.3.2 Fourier Series and Complete Family

Definition 10.3.5 (Complete Orthonormal Family). An orthonormal family $\{\varphi_0, \varphi_1, \dots\}$ (countable) is called *complete* if each $f \in V$ can be written as

$$f = \sum_{k=0}^{\infty} c_k \varphi_k \tag{$\star$}$$

Remark 10.1

- *The meaning of* ($\star$):
$$\left\| f - \sum_{k=0}^{n} c_k \varphi_k \right\| \to 0 \quad \textit{as} \quad n \to \infty.$$
- ($\star$) *is called the* Fourier series *of* f $w.r.t.$ $\{\varphi_0, \varphi_1, \dots\}$.
- *If* $\{\varphi_0, \varphi_1, \dots\}$ *is complete, then it is an* orthonormal basis *of* V.

Objective: Find suitable complete orthonormal family and expand $f \in V$ into Fourier series.

Theorem 10.3.6
If f has Fourier series expansion:
$$f = \sum_{k=0}^{\infty} c_k \varphi_k,$$
then,
$$c_k = \langle f, \varphi_k \rangle \qquad \text{for } k = 0, 1, \dots.$$
c_k's are called the *Fourier coefficients* of f.

Proof 1. Let
$$S_n = \sum_{k=0}^{n} c_k \varphi_j.$$
Then,
$$\|f - S_n\| \to 0 \quad \text{ans} \quad n \to \infty.$$
Fix $m \geq 0$. Then, for any $n \geq m$,
$$\begin{aligned} c_m \overset{\text{want}}{=\!=} \langle f, \varphi_m \rangle &= \langle f - S_n + S_n, \varphi_m \rangle \\ &= \langle f - S_n, \varphi_m \rangle + \langle s_n, \varphi_m \rangle \\ &= \langle f - S_n, \varphi_m \rangle + c_m && \text{[orthogonality]} \\ &= 0 + c_m \quad \text{as } n \to \infty && \left[\substack{\text{Cauchy-Schwarz} \\ \|f - S_n\| \to 0}\right] \end{aligned}$$
So, $\langle f, \varphi_m \rangle = c_m$.

Q.E.D. ■

Question: Given f and $\{\varphi_1, \varphi_2, \dots\}$, does the series

$$\sum_{k=0}^{\infty} \langle f, \varphi_k \rangle \varphi_k$$

converge to f?

Theorem 10.3.7 Properties of Fourier Coefficients
Assume $\{\varphi_0, \varphi_1, \dots\}$ is an orthonormal family in V.

- Bessel's Inequality:
$$\sum_{k=0}^{\infty} |\langle f, \varphi_k \rangle|^2 \le \|f\|^2.$$
- Parseval's Equality (One can View this as the Pythagorean Theorem):
  If
$$f = \sum_{k=0}^{\infty} \langle f, \varphi_k \rangle \varphi_k,$$
  then
$$\sum_{k=0}^{\infty} |\langle f, \varphi_j \rangle|^2 = \|f\|^2.$$

Proof 2. Let $S_n = \sum_{k=0}^{n} \langle f, \varphi_k \rangle \varphi_k$. Denote $c_k = \langle f, \varphi_k \rangle$.

$$\begin{aligned}
\|f\|^2 &= \|f - S_n + S_n\|^2 \\
&= \langle f - S_n + S_n, f - S_n + S_n \rangle && \text{[definition]} \\
&= \|f - S_n\|^2 + \|S_n\|^2 && [\text{Linearity, } f - S_n \perp S_n] \\
\|S_n\| &= \langle S_n, S_n \rangle = \sum_{k=0}^{n} |c_k|^2.
\end{aligned}$$

Then,

$$\|f\|^2 = \underbrace{\|f - S_n\|^2}_{\ge 0} + \sum_{k=0}^{n} |c_k|^2 \implies \|f\|^2 \ge \sum_{k=0}^{n} |c_k|^2 = \sum_{k=0}^{n} |\langle f, \varphi_k \rangle|^2$$

true for any n. So, we get ① by letting $n \to \infty$.

Under the assumption of ②, when $n \to \infty$, we have $\|f - S_n\|^2 \to 0$. So,

$$\|f\|^2 = \sum_{k=0}^{\infty} |c_k|^2 = \sum_{k=0}^{\infty} |\langle f, \varphi_k \rangle|^2.$$

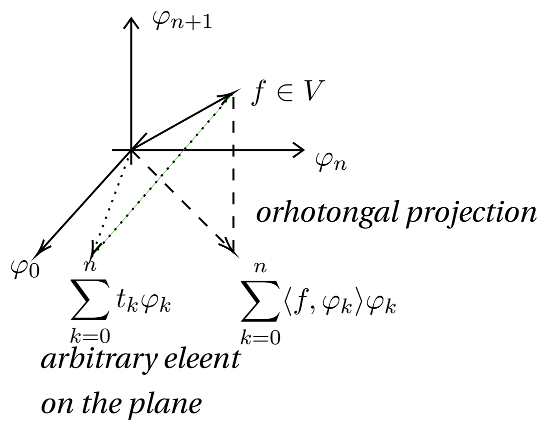
Q.E.D. ■

Theorem 10.3.8 Best mean Approximation Theorem (BMAT)
Assume $\{\varphi_0, \varphi_1, \dots\}$ is an orthonormal family in V. For any scalars $t_0, t_1, \dots, t_n \in \mathcal{C}$, we have

$$\left\| f - \sum_{k=0}^{n} t_k \varphi_k \right\| \geq \left\| f - \sum_{k=0}^{n} \langle f, \varphi_k \rangle \varphi_k \right\|.$$

- The first sum is an arbitrary element in the plane formed by $\{\varphi_0, \dots, \varphi_n\}$.
- The second sum is the orthogonal projection of f onto the plane.

Remark 10.2 (Geometric Inpterpretation)



LHS $\leq$ RHS: the shortest distance from a point f to the plane is achieved by the orthogonal projection (or, the perpendicular line).

Proof 3. Let $h_n = \sum_{k=0}^{n} t_k \varphi_k$. Then,

$$\begin{aligned}
\|f - h_n\|^2 &= \langle f - h_n, f - h_n \rangle \\
&= \langle f, f \rangle - \langle h_n, f \rangle - \langle f, h_n \rangle + \langle h_n, h_n \rangle \\
&= \|f\|^2 - \sum_{k=0}^{n} t_k \overline{c_k} - \sum_{k=0}^{n} \overline{t_k} c_k + \sum_{k=0}^{n} |t_k|^2 \\
&\ \vdots \quad \text{linearity} \\
&= \|f\|^2 - \sum_{k=0}^{n} |c_k|^2 + \sum_{k=0}^{n} |t_k - c_k|^2 \\
&= \|f - f_n\|^2 + \underbrace{\sum_{k=0}^{n} |t_k - c_k|^2}_{\geq 0}.
\end{aligned}$$

So, BMAT is proven.

Q.E.D. ■

10.4 Completeness and Convergence in L^2

Theorem 10.4.1 Orthogonal Functions in L^2
Let $V = L^2([a,b])$, where $[a,b] = [0, 2\pi]$.

- Exponential family:

$$\varphi_n(x) = \frac{e^{inx}}{\sqrt{2\pi}}, \qquad n = 0, \pm 1, \pm 2, \ldots$$

- Trig. family:

$$\frac{1}{\sqrt{2\pi}}, \ \frac{\cos mx}{\sqrt{2\pi}}, \ \frac{\sin nx}{\sqrt{2\pi}}, \qquad n, m = 1, 2, \ldots$$

Claim 10.4.2 Both families are orthogonal.

Proof 1. (of exponential family)

WTS:

$$\langle \varphi_n, \varphi_m \rangle = \delta_{n,m} = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases}$$

$$\begin{aligned} \langle \varphi_n, \varphi_m \rangle &= \int_0^{2\pi} \varphi_n(x) \overline{\varphi_m(x)} \, \mathrm{d}x \\ &= \frac{1}{2\pi} \int_0^{2\pi} e^{\mathrm{i}nx} \cdot e^{-\mathrm{i}mx} \, \mathrm{d}x \\ &= \frac{1}{2\pi} \int_0^{2\pi} e^{\mathrm{i}(n-m)x} \, \mathrm{d}x \\ &= \begin{cases} 1, & n = m \\ \dfrac{1}{2\pi} \cdot \dfrac{1}{\mathrm{i}(n-m)} e^{\mathrm{i}(n-m)x} \Big|_0^{2\pi} = 0, & n \neq m. \end{cases} \end{aligned}$$

Q.E.D. ■

Theorem 10.4.3 Mean Convergence Property/Completeness
The exponential family $\{\varphi_n\}_{n=-\infty}^{\infty}$ is complete in L^2

Remark 10.3 *To prove this Theorem, we aim to show: any function $f(x) \in L^2$ can be represented by its Fourier series:*

$$f(x) = \sum_{n=-\infty}^{\infty} \langle f, \varphi_n \rangle \varphi_n.$$

i.e.,.

$$\left\| f(x) - \sum_{k=-n}^{n} \langle f, \varphi_k \rangle \varphi_k \right\|_{L^2} \xrightarrow{(n \to \infty)} 0.$$

Lemma 10.4.4 Stone-Weierstrass Theorem: Continuous functions can be approximated by polynomials of $e^{\mathrm{i}x}$ and $e^{-\mathrm{i}x}$. More precisely, given $f : [0, 2\pi] \to \mathbb{C}$ continuous with $f(0) = f(2\pi)$. Then, $\forall\, \varepsilon > 0$,

$\exists\, n \geq 1$ and c_k, $k = 0, \pm 1, \ldots$ $s.t.$

$$|f(x) - p_n(x)| < \varepsilon \quad \forall\, x \in [0, 2\pi],$$

where

$$p_n(x) = \sum_{k=-n}^{n} c_k e^{\mathrm{i}kx},$$

a polynomial in $e^{\mathrm{i}x}$ and $e^{-\mathrm{i}x}$.

Lemma 10.4.5 : Integrable functions can be approximated by continuous functions. That is, let $f \in L^2$ and $\varepsilon > 0$ be given, $\exists$ continuous function $g : [0, 2\pi] \to \mathbb{C}$ with $g(0) = g(2\pi)$ $s.t.$

$$\|f - g\| < \varepsilon.$$

► Proof 2 of Mean Convergence Property

- $\boxed{\text{Step } 1}$ Special Case:

  Let f be continuous with $f(0) = f(2\pi)$. Write

  $$S_n = \sum_{k=-n}^{n} \langle f, \varphi_k \rangle \varphi_k, \qquad \text{where } \varphi_k(x) = \frac{e^{\mathrm{i}kx}}{\sqrt{2\pi}}$$

  WTS: $\|f - S_n\| \to 0$ as $n \to \infty$.

  Fix $\varepsilon > 0$. By Lemma 10.4.4, we can choose $p_N(x)$ $s.t.$

  $$|f(x) - p_N(x)| < \frac{\varepsilon}{\sqrt{2\pi}} \quad \forall\, x \in [0, 2\pi].$$

  Then,

  $$\|f - p_N\| \leq \left(\int_0^{2\pi} \left(\frac{\varepsilon}{\sqrt{2\pi}} \right)^2 \right)^{1/2} = \varepsilon.$$

  Thus, $\forall\, n \geq N$, we have

  $$\begin{aligned} \|f - S_n\| &\leq \|f - S_N\| && [\text{BMAT. } L_N \subset L_n \implies S_N \in L_n.] \\ &\leq \|f - p_N\| && [\text{BMAT. } p_N \in L_N] \\ &\leq \varepsilon. \end{aligned}$$

  So, $\|f - S_n\| \to 0$ as $n \to \infty$.

- $\boxed{\text{Step } 2}$ General Case:

  Fix $f \in L^2$. WTS: $f = \displaystyle\sum_{k=-\infty}^{\infty} \langle f, \varphi_k \rangle \varphi_k$.

By Lemma 10.4.5, $\exists$ sequence of continuous functions $g_n : [0, 2\pi] \to \mathbb{C}$ with $g(0) = g(2\pi)$ $s.t.$

$$\|f - g_n\| \to 0 \quad \text{as} \quad n \to \infty.$$

By Step 1, for each g_n, we have

$$g_n = \sum_{k=-\infty}^{\infty} \langle g_n, \varphi_k \rangle \varphi_k.$$

WTS: $\|f - S_n\| \to 0$.

Fix $\varepsilon > 0$. Choose N $s.t.$

$$\|f - g_N\| < \frac{\varepsilon}{3}.$$

Then, choose M $s.t.$

$$n \geq M \implies \|g_N - S_n(g_N)\| < \frac{\varepsilon}{3},$$

where $S_n(g_N)$ denotes the partial sum of Fourier series of g_N.

$$S_n(g_N) = \sum_{k=-n}^{n} \langle g_N, \varphi_k \rangle \varphi_k.$$

Thus, $\forall\, n \geq M$, we have

$$\begin{aligned}
\|f - S_n\| &= \|S_n - S_n(g_N) + S_n(g_N) - g_N + g_N - f\| \\
&\leq \|S_n - S_n(g_N)\| + \|S_n(g_N) - g_N\| + \|g_N - f\| \\
\|S_n - S_n(g_N)\| &= \left\| \sum_{k=-n}^{n} \langle f, \varphi_k \rangle \varphi_k - \sum_{k=-n}^{n} \langle g_N, \varphi_k \rangle \varphi_k \right\| \\
&= \left\| \sum_{k=-n}^{n} \langle f - g_N, \varphi_k \rangle \varphi_k \right\| \\
&= \left\langle \sum_{k=-n}^{n} \langle f - g_N, \varphi_k \rangle \varphi_k, \sum_{k=-n}^{n} \langle f - g_N, \varphi_k \rangle \varphi_k \right\rangle^{1/2} \\
&= \left(\sum_{k=-n}^{n} |\langle f - g_N, \varphi_k \rangle|^2 \right)^{1/2} && \text{[Pythagorean Theorem]} \\
&\leq \|f - g_N\| < \frac{\varepsilon}{3}.
\end{aligned}$$

So,

$$n \geq M \implies \|f - S_n\| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

Therefore,

$$\|f - S_n\| \to 0 \quad \text{as} \quad n \to \infty.$$

Q.E.D. ■

With that, these notes mark the end of a journey through the rigorous landscapes of Real Analysis. From the foundational structure of $\mathbb{R}$ *to the elegance of Fourier series in* L^2*, this document reflects not only the theorems and proofs, but also the quiet persistence of curiosity.*

I would like to express my sincere gratitude to:

Professor Shanshuang Yang
for his guidance, clarity, and intellectual generosity throughout MATH 411 & 412.

Jerrold E. Marsden and Michael J. Hoffman
whose text Elementary Classical Analysis *served as a source of inspiration, challenge, and reflection.*

My peers and friends
for the many shared ideas, problem-solving moments, and philosophical tangents.

Myself
for persistence, patience, and the willingness to wrestle with the abstract.

End of Notes

Jiuru Lyu *April 29, 2025*