1.4 Cauchy Sequence in $\mathbb{R}$

Definition. (1.4.1)

A sequence $x_n \in \mathbb{R}$ is said to converge to a limit x if

- $\forall \epsilon > 0, \quad \exists N \quad s.t. \quad n > N \Rightarrow |x_n - x| < \epsilon.$

A sequence $x_n \in \mathbb{R}$ is called Cauchy sequence if

- $\forall \epsilon, \quad \exists N \quad s.t. \quad n > N \,\&\, m > N \quad \Rightarrow \quad |x_n - x_m| < \epsilon.$

Proposition. (1.4.2)

Every convergent sequence is a Cauchy sequence.

Proof. Assume $x_k \to x$. Let $\epsilon > 0$ be given.

- $\exists N$ $s.t.$ $n > N \Rightarrow |x_n - x| < \frac{\epsilon}{2}$.
- $n, m \geq N \quad \Rightarrow$
  $|x_n - x_m| \leq \quad |x - x_n| + |x - x_m| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$

Theorem. (1.4.3; Bolzano-Weierstrass Property)

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

Proof. Suppose x_n is a bounded sequence in $\mathbb{R}$. $\exists M$ such that $-M \leq x_n \leq M, \; n = 1, 2, \cdots$. Select $x_{n_0} = x_1$.

- Bisect $I_0 := [-M, M]$ into $[-M, 0]$ and $[0, M]$.
- At least one of these (either $[-M, 0]$ or $[0, M]$) must contain x_n for **infinitely many indices** n.
- Call it I_1 and select $n_1 > n_0$ with $x_{n_1} \in I_0$.
- Continue in this way to get a subsequence x_{n_k} such that
    - $I_0 \supset I_1 \supset I_2 \supset I_3 \cdots$
    - $I_k = [a_k, b_k]$ with $|I_k| = 2^{-k} M$.
    - Choose $n_0 < n_1 < n_2 < \cdots$ with $x_{n_k} \in I_k$.
- Since $a_k \leq a_{k+1} \leq M$ (monotone and bounded), $a_k \rightarrow^{\exists} x$.
- Since $x_{n_k} \in I_k$ and $|I_k| = 2^{-k} M$, we have

$$|x_{n_k} - x| < |x_{n_k} - a_k| + |a_k - x| \leq 2^{-k-1} M + |a_k - x| \to 0 \text{ as } k \to \infty.$$

Corollary. (1.4.5; Compactness)

Every ***sequence in the closed interval*** $[a, b]$ *has a* ***subsequence*** *in* $\mathbb{R}$ *that converges to some point in* $\mathbb{R}$.

Proof. Assume $a \le x_n \le b$ for $n = 1, 2, \cdots$. By Theorem 1.4.3, $\exists$ a subsequence x_{n_k} and $a \le {}^{\exists}x \le b$ such that $x_{n_k} \to x$.

Lemma. (1.4.6; Boundedness of Cauchy sequence)

If x_n *is a* **Cauchy** *sequence,* x_n *is* **bounded**.

Proof. $\exists N$ s.t. $n \ge N \Rightarrow |x_n - x| < 1$. Then $\sup_n |x_n| \le 1 + \max\{|x_1|, \cdots, |\mathbf{x}_N|\}$ (Why?)

Theorem. (1.4.3; Completeness)

Every **Cauchy** *sequence in* $\mathbb{R}$ *converges to an element in* $[a, b]$.

Proof. Cauchy seq. $\Rightarrow$ bounded seq. $\Rightarrow$ convergent subseq.

1.5. Cluster Points of the sequence x_n

Definition. (1.5.1; cluster points)

A point x is called a cluster point of the sequence x_n if

- $\forall \epsilon > 0, \quad \exists$**infinitely many values of** n **with** $|x_n - x| < \epsilon$

In other words, a point x is a cluster point of the sequence x_n iff
$\forall \epsilon > 0$ & $\forall N, \quad \exists n > N \quad s.t. \quad |x_n - x| < \epsilon$

Example

- Both 1 and -1 are cluster points of the sequence $1, -1, 1, -1, \cdots$.
- The sequence $x_n = \frac{1}{n}$ has the only cluster point 0.
- The sequence $x_n = n$ does not have any cluster point.

Proposition.

1. *x is a cluster point of the sequence x_n iff $\exists$ a subsequence x_{n_k} $s.t.$ $x_{n_k} \longrightarrow x$.*
2. *$x_n \longrightarrow x$ iff every subsequence of x_n converges to x*
3. *$x_n \longrightarrow x$ iff the sequence $\{x_n\}$ is bounded and x is its only cluster points.*

Proof.

1. ($\Rightarrow$) Assume x is a cluster point. Then, we can choose $n_1 < n_2 < n_3 \cdots$ $s.t.$ $|x_{n_k} - x| < \frac{1}{k}$. (Why?) This gives a subsequence $x_{n_k} \longrightarrow x$.
2. Trivial
3. ($\Leftarrow$)If not, $\exists \epsilon$ and $\exists$ a subseq x_{n_k} so that $|x_{n_k} - x| > \epsilon$. Since x_{n_k} is bounded, $\exists$ a convergent subseq. The limit of that subseq would be a cluster pt of the seq x_n different from x, but there are no such pt. Contradiction.

Definition. (1.5.3; limit superior & limit inferior of seq x_n)

Define the limit superior $\overline{\lim} x_n$ *in the following way:*

- *If* x_n *is bounded above, then*

  $\limsup_{n\to\infty} x_n = \overline{\lim} x_n =$ *the largest cluster point*

  $\overline{\lim} x_n = -\infty$ *if the set cluster point is empty*

- *If* x_n *is NOT bounded above, then* $\overline{\lim} x_n = \infty$

Similarly, we can define the limit inferior $\underline{\lim} x_n$.

Examples

- For the seq $1, 0, -1, 1, 0, -1, \cdots$, $\overline{\lim} x_n = 1$ and $\underline{\lim} x_n = -1$.
- If $x_n = n$, then $\overline{\lim} x_n = \infty = \underline{\lim} x_n$
- Let $x_n = (-1)^n \frac{1+n}{n}$. Then $\overline{\lim} x_n = 1$ and $\underline{\lim} x_n = -1$.

Definition. (1.6.2; Vector space)

A real vector space $\mathcal{V}$ is a set of elements called **vectors**, *with given operations of vector addition $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ and scalar multiplication $\cdot : \mathbb{R} \times \mathcal{V} \to \mathcal{V}$ such that the followings hold for all $v, u, w \in \mathcal{V}$ and all $\lambda, \mu \in \mathbb{R}$:*

1. $v + w = w + v, \quad (v + u) + w = v + (u + w), \quad \lambda(v + w) = \lambda v + \lambda w, \quad \lambda(\mu v) = (\lambda\mu)v, \quad (\lambda + \mu)v = \lambda v + \mu v, \quad 1v = v.$
2. $\exists 0 \in \mathcal{V}$ *s.t.* $v + 0 = v. \qquad \exists - v \in \mathcal{V}$ *s.t.* $v - v = 0.$

- A subset of $\mathcal{V}$ is called a subspace if it is itself a vector space with the same operations.
- $\mathcal{W}$ is a vector subspace of $\mathcal{V}$ iff $\lambda v + \mu u \in \mathcal{W}$ whenever $u, v \in \mathcal{W}$ and $\lambda, \mu \in \mathbb{R}$.
- The straight line $\mathcal{W} = \{(x_1, x_2) \ : \ x_1 = 2x_2\}$ is a subspace of $\mathbb{R}^2$.

Euclidean space $\mathbb{R}^n$ & Definitions & Properties

The Euclidean n-space $\mathbb{R}^n$ with the operations $(x_1, \cdots, x_n) + (y_1, \cdots, y_n) = (x_1 + y_1, \cdots, x_n + y_n)$ & $\lambda(x_1, \cdots, x_n) = (\lambda x_1, \cdots, \lambda x_n)$ is a vector space of dimension n.

- **The standard basis of $\mathbb{R}^n$;** $\mathbf{e}_1 = (1, 0, \cdots, 0), \cdots, \mathbf{e}_n = (0, \cdots, 0, 1)$.
- **Unique representation:** $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ can be expressed **uniquely** as $x = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$.
- **Inner product of x and y:** $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$
- **Norm of x:** $\|x\| = \sqrt{\langle x, y\rangle}$.
- **Distance between x and y:** $\mathrm{dist}(x, y) = \|x - y\|$
- **Triangle inequality:** $\|x + y\| \leq \|x\| + \|y\|$.
- **Cauchy-Schwartz inequality:** $\langle x, y\rangle \leq \|x\| \, \|y\|$
- **Pythagorean theorem:** If $\langle x, y\rangle = 0$, then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

Definition. (1.7.1; Metric Space (M, d) equipped with d =distance)

A metric space (M, d) *is a set* M *and a function* $d : M \times M \to R$ *such that*

1. $d(x, y) \geq 0$ *for all* $x, y \in M$.
2. $d(x, y) = 0$ *iff* $x = y$.
3. $d(x, y) = d(y, x)$ *for all* $x, y \in M$.
4. $d(x, y) \leq d(x, z) + d(z, y)$ *for all* $x, y \in M$.

Example [Fingerprint Recognition] Let M be a data set of fingerprints in Seoul city police department.

- Motivation: Design an efficient access system to find a target.
- We need to define a **dissimilarity** function stating the distance between the data. The distance $d(x, y)$ between two data x and y must satisfy the above four rules.
- **Similarity queries.** For a given target $x^* \in M$ and $\epsilon > 0$, arrest all having finger print $y \in M$ such that $d(y, x^*) < \epsilon$.

Definition. (1.7.3. Normed Space $(\mathcal{V}, \|\cdot\|)$)

A **normed space** $(\mathcal{V}, \|\cdot\|)$ *is a vector space* $\mathcal{V}$ *and a function* $\|\cdot\| : \mathcal{V} \to \mathbb{R}$ *called a* **norm** *such that*

1. $\|\mathbf{v}\| \geq 0, \qquad \forall \mathbf{v} \in \mathcal{V}$
2. $\|\mathbf{v}\| = 0$ *iff* $\mathbf{v} = 0$.
3. $\|\lambda \mathbf{v}\| = |\lambda| \|\mathbf{v}\|, \qquad \forall \mathbf{v} \in \mathcal{V}$ *and every scaler* λ.
4. $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|, \qquad \forall \mathbf{v}, \mathbf{w} \in \mathcal{V}$

Examples

- $\mathcal{V} = \mathbb{R}$ and $\|x\| = |x|$ for all $x \in \mathbb{R}$.
- $\mathcal{V} = \mathbb{R}^2$ and $\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2}$ for all $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$.
- Let $\mathcal{V} = C([0,1])$=all continuous functions on the interval $[a, b]$. Define
  $\|f\| = \sup\{|f(x)| \ : \ x \in [0,1]\} \qquad$ (called supremum norm).

Proposition.

If $(\mathcal{V}, \|\cdot\|)$ *is a normed vector space and*

$$d(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|$$

, then d *is a metric in* $\mathcal{V}$.

Proof. EASY.

Examples

- For $\mathcal{V} = C([0,1])$, the metric is

$$d(f,g) = \|f - g\| = \sup\{|f(x) - g(x)| \;:\; x \in [0,1]\}.$$

The sup distance between functions is the largest vertical distance between their graphs.

Definition.

A vector space $\mathcal{V}$ *with a function* $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ *is called an* **inner product space** *if*

1. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ *for all* $\mathbf{v} \in \mathcal{V}$.
2. $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ *iff* $\mathbf{v} = 0$.
3. $\langle \lambda \mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{v}, \mathbf{w} \rangle, \qquad \forall \mathbf{v} \in \mathcal{V}$ *and every scaler* λ.
4. $\langle \mathbf{v} + \mathbf{w}, \mathbf{h} \rangle = \langle \mathbf{v}, \mathbf{h} \rangle + \langle \mathbf{w}, \mathbf{h} \rangle$.
5. $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$

Examples

1. $\mathcal{V} = \mathbb{R}^2$ and $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + v_2 w_2$.
   Two vectors $\mathbf{v}$ and $\mathbf{w}$ are orthogonal if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.
2. $\mathcal{V} = C[0, 1]$ and $\langle f, g \rangle = \int_0^1 f(x) g(x) dx$
3. $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ is a norm on $\mathcal{V}$.

Theorem. (Cauchy-Schwarz inequality)

If $\langle \cdot, \cdot \rangle$ is an inner product in a real vector space $\mathcal{V}$, then $|\langle f, g \rangle| \leq \|f\|\|g\|$

Proof:

- Suppose $g \neq 0$. Let $h = \frac{g}{\|g\|}$. It suffices to prove that $|\langle f, h \rangle| \leq \|f\|$. (Why? $|\langle f, g \rangle| \leq \|f\|\|g\|$ iff $|\langle f, h \rangle| \leq \|f\|$.)
- Denote $\alpha = \langle f, h \rangle$. Then

$$
\begin{aligned}
0 &\leq \|f - \alpha h\|^2 = \langle f - \alpha h, f - \alpha h \rangle \\
&= \|f\|^2 - \alpha \langle h, f \rangle - \alpha \langle f, h \rangle + |\alpha|^2 \\
&= \|f\|^2 - |\alpha|^2
\end{aligned}
$$

Hence, $|\alpha| = |\langle f, h \rangle| \leq \|f\|$. This completes the proof.

Chapter 2: Topology of $M = \mathbb{R}^n$

Throughout this chapter, assume $M = \mathbb{R}^n$ (the Euclidean space) with the metric $d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} = \|\mathbf{x} - \mathbf{y}\|$

Definition. ($D(\mathbf{x}, \epsilon)$, open, neighborhood)

- $D(\mathbf{x}, \epsilon) := \{\mathbf{y} \in M \ : \ d(\mathbf{y}, \mathbf{x}) < \epsilon\}$ *is called ϵ-ball (or ϵ-disk) about* $\mathbf{x}$.
- *$A \subset M$ is open if $\forall\ \mathbf{x} \in M$, $\quad \exists \epsilon > 0$ s.t. $D(\mathbf{x}, \epsilon) \subset A$.*
- *A neighborhood of* $\mathbf{x}$ *is an open set A containing* $\mathbf{x}$.

- Open sets: (a, b), $\quad D(\mathbf{x}, \epsilon)$, $\quad \{(x, y) \in \mathbb{R}^2 \ : \ 0 < x < 1\}$.
- The union of an **arbitrary collection** of open subsets of M is open. (Why?)
- The intersection of **a finite number** of open subsets of M is open. (Note that $\cap_{n=1}^{\infty}(-1/n, 1/n) = \{0\}$ is closed.)

2.2 Interior of a set A: $int(A)$

Definition. (2.2.1; Interior point & interior of A)

Let (M,d) is a metric space and $A \subset M$. **$\mathbf{x}$** *is called an* **interior point of A if $\exists D(\mathbf{x},\epsilon)$ s.t. $D(\mathbf{x},\epsilon) \subset A$.** *Denote*

$$int(A) := \text{ the collection of all interior points of } A.$$

Examples. Proofs are very easy.

- If $A = [0,1]$, then $int(A) = (0,1)$.
- $int\{(x,y) \in \mathbb{R}^2 : 0 < x \leq 1\} = \{(x,y) \in \mathbb{R}^2 : 0 < x < 1\}$.
- If A is open, then $int(A) = A$.
- Let (M,d) be a metric space and $\mathbf{x}_0 \in M$.
  $int\{\mathbf{y} \in M \ : \ d(\mathbf{y},\mathbf{x}_0) \leq 1\} = \{\mathbf{y} \in M \ : \ d(\mathbf{y},\mathbf{x}_0) < 1\}$

Definition. (2.3-4: Closed sets & Accumulation Points)

• *A set B in a metric space M is said to be closed if $M \setminus B$ is open.*

• $x \in M$ *is* **accumulation point (or cluster point) of a set** $A \subset M$ *if* $\forall \epsilon > 0, D(\mathbf{x}, \epsilon)$ *contains* $\mathbf{y} \in A$ *with* $\mathbf{y} \neq \mathbf{x}$.

Prove the followings:

- Closed sets: $[a, b]$, $\{\mathbf{y} \in \mathbb{R}^2 \ : \ d(\mathbf{y}, \mathbf{x}_0) \leq 1\}$.
- The union of an **a finite number** of closed subsets of M is closed. (Note that $\cup_{n=1}^{\infty}[1/n, 2 - 1/n] = (0, 2)$ is open.)
- The intersection of **an arbitrary family** of closed subsets of M is closed. Why?
- Every finite set in $\mathbb{R}^n$ is closed.
- A set $A \subset M$ is closed iff the accumulation points of A belongs to A.
- $A = \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdots\} \cup \{0\}$ is closed.

Definition. (Closure of A & Boundary of A)

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

- $cl(A)$:=*the intersection of all closed set containing A.*
- $\partial A = bd(A) = cl(A) \cap cl(M \setminus A)$ *is called the boundary of A*

Examples

- **Closure:** $cl((0,1)) = [0,1],\ cl\{(x,y) \in \mathbb{R}^2 : x > y\} = \{(x,y) \in \mathbb{R}^2 : x \geq y\}$.
- **Boundary** $bd((0,1)) = \{0,1\},\ bd\{(x,y) \in \mathbb{R}^2 : x > y\} = \{(x,y) \in \mathbb{R}^2 : x = y\}$.

Let (M, d) is a metric space and $A \subset M$. Prove that

- $cl(A) = A \cup \{$ accumnulation points of $A\}$.
- $\mathbf{x} \in cl(A)$ iff $\inf\{d(\mathbf{x}, \mathbf{y})\ :\ \mathbf{y} \in A\} = 0$.
- $\mathbf{x} \in bd(A)$ iff $\forall \epsilon > 0, D(\mathbf{x}, \epsilon) \cap A \neq \emptyset$ & $D(\mathbf{x}, \epsilon) \cap (M \setminus A) \neq \emptyset$.

Definition. (Sequences & Completes)

Let (M, d) is a metric space and x_k a sequence of points in M.

- $\lim_{k\to\infty} x_k = x$ *iff* $\forall \epsilon > 0, \exists N \ s.t. \ k \geq N \Rightarrow d(x, x_k) < \epsilon.$
- x_k *is* **Cauchy seq.**
  iff $\forall \epsilon > 0, \exists N \ s.t. \ k, l \geq N \Rightarrow d(x_k, x_l) < \epsilon.$
- x_k *is* **bounded**
  iff $\exists B > 0 \ \& \ x_0 \in M \ \ s.t. \ d(x_k, x_0) < B$ *for all* k.
- x *is a* **cluster point** *of the seq.* x_k
  iff $\forall \epsilon, \exists$ *infinitely many* k *with* $d(x_k, x) < \epsilon.$

The space M is called **complete** *if every Cauchy seq. in M converges to a point in M.*

In a metric space, it is easy to prove the followings:

- Every convergent seq. is a Cauchy seq.
- A Cauchy seq. is bounded.
- If a subseq. of a Cauchy seq. converges to x, then the sequence itself converges to x.

Chapter 3: Compact & Connected sets

Throughout this chapter, we assume that (M, d) is a metric space.

Definition. (3.1.1: Sequentially compact & Compact)

Let $A \subset M$.

- *A is called* **sequentially compact** *if* **EVERY** *sequence in A has a subsequence that converges to* **a point in A.**
- *A is* **compact** *if* **EVERY open cover** *of A has* **FINITE subcover.**
  - *An* **open cover** *of A is a collection $\{U_i\}$ of open sets such that $A \subset \cup_i U_i$.*
  - *An open cover $\{U_i\}$ of A is said to have* **finite subcover** *if a finite subcollection of $\{U_i\}$ covers A.*

- In chapter 1, we proved that every sequence $\mathbf{x}_n$ in the closed interval $[a, b]$ has a subsequence that converges to a point in $[a, b]$. Hence, $[a, b]$ is sequentially compact.

Examples of compact set

1. Prove that the entire line $\mathbb{R}$ is NOT compact.
   Proof. Clearly, $\{D(n,1) \ : \ n = 0, \pm 1, \pm 2, \cdots\}$ is open cover of $\mathbb{R}$ but does not have a finite subcover (why?).
2. Prove that $A = (0,1]$ is not compact. *Proof.* Clearly, $(0,1] = \cup_{n=1}^{\infty}(1/n, 2)$. Hence, $\{(1/n,2) : n = 1,2,\cdots\}$ is an open cover of $(0,1]$ but does not have a finite subcover.
3. **Heine-Borel thm.** Let $A \subset M = \mathbb{R}^n$. A is compact iff A is closed and bounded. *Proof.* later.
4. Give an example of a bounded and closed set that is not compact.
   Sol'n. Let $M = \{\mathbf{e}_n \ : \ n = 1,2,\cdots\}$ where $\mathbf{e}_1 = (1,0,0,\cdots)$, $\mathbf{e}_2 = (0,1,0,\cdots)$, $\cdots$. Let $d(\mathbf{e}_i, \mathbf{e}_j) = \sqrt{2}$ if $i \neq j$. Then (M,d) is a metric space.
   - The entire metric space M is closed and bounded (why?).
   - $\{D(\mathbf{e}_n, 1) \ ; \ n = 1, 2\cdots\}$ is open cover of M but but does not have a finite subcover (why?). Hence, M is not compact.

Theorem. (3.1.3; Bolzano-Weirstrass theorem)

$A \subset M$ is compact iff A is sequentially compact.

- **Lemma 1: Let $A \subset M$. If A is compact, then A is closed.**
  Proof . We will show $M \setminus A$ is open. Let $\mathbf{x} \in M \setminus A$.
  1. $A \subset \cup_{n=1}^{\infty} U_n$ where $U_n = M \setminus \overline{D(\mathbf{x}, 1/n)}$ open set.
  2. Since A is compact and $\{U_n\}$ covers A, $\exists$ a finite subcover, that is, $\exists N$ $s.t.$ $A \subset \cup_{n=1}^{N} U_n. = U_N$
  3. Hence, $D(\mathbf{x}, 1/N) \subset U_N^c \subset A^c = M \setminus A$ and therefore $M \setminus A$ is open.
- **Lemma 2: Let $A \subset B \subset M$. If B is compact and A is closed, then A is compact.**
  Proof. Let U_i be an open covering of A.
  1. Set $V = M \setminus A$. Note that V is open.
  2. Thus $\{U_i, V\}$ is an open cover of B.
  3. Since B is compact, B has a finite cover, say, $\{U_1, \cdots, U_N, V\}$. Hence, $A \subset U_1 \cup \cdots \cup U_N$.

- **Lemma 4: If A is sequentially compact, then A is totally bounded.**
  1. Definition of totally bounded: $A \subset M$ is totally bounded if $\forall \epsilon$, $\exists$ finite set $\{x_1, \cdots, x_N\} \subset M$ $s.t.$ $A \subset \cup_{i=1}^{N} D(x_i, \epsilon)$.
  2. Proof. If not, then for some $\epsilon > 0$ we cannot cover A with finitely many disks.
     (i) Choose $x_1 \in A$ and $x_2 \in A \setminus D(x_1, \epsilon)$.
     (ii) By assumption, we can repeat; choose $x_n \in A \setminus \cup_{i=1}^{n-1} D(x_i, \epsilon)$ for $n = 1, 2, \cdots$.
     (iii) This seq $\{x_n\}$ satisfies $d(x_n, x_m) > \epsilon$ for all $n \neq m$.
     (iv) Hence, x_n has no convergent subseq., a contradiction.

- **Summery.** Let $A \subset M$.
  - **A is compact $\Rightarrow$ A is closed**
  - **A is a closed subset of a compact set $\Rightarrow$ A is compact.**
  - **A is sequentially compact $\Rightarrow$ A is totally bounded.**

Proof of B-W thm (⇒): If A is compact, then A is sequentially compact.

Let A be compact. Let $\{x_n\}$ be a seq. in A.

1. To derive a contradiction, assume that $\{x_n\}$ has no convergent subseq.
2. Then, $\{x_n\}$ has infinitely many distinct points $\{y_k\}$ which has no accumulation points. (Why? If not, ∃ convergent subseq.)
3. Hence, ∃ some neighborhood U_k of y_k containing no other y_i.
4. $\{y_n\}$ is closed because it has no accumulation points. Hence, $\{y_n\}$ is compact by Lemma 2. **Lemma2: Any closed subset of the compact set A is compact.**
5. But $\{U_k\}$ is an open cover that has no finite subcover, **a contradiction.**
6. Hence, x_n has a convergence subsequence. The limit lies in A, since A is closed by Lemma 1.

Hence, x_n has a subsequence that converges to a point in A.

Proof of B-W thm ($\Leftarrow$): If A is sequentially compact, than A is compact.

Suppose $\{U_i\}$ is an open cover of A. We need to prove that $\{U_i\}$ has finite subcover.

- $\exists r > 0$ s.t. $\forall y \in A,\ D(y,r) \subset U_i$ for some U_i.
  Why?
  1. **If not**, $\exists y_n \in A$ s.t. $D(y_n, 1/n)$ is not contained in any U_i.
  2. By assumption, $\{y_n\}$ has a subseq., say, $y_{n_k} \to z \in A$. Since $z \in A \subset \cup_i U_i$, $z \in U_{i_0}$ for some U_{i_0}.
  3. Since U_{i_0} is open, $\exists \epsilon > 0$ s.t. $D(z,\epsilon) \subset U_{i_0}$.
  4. Since $y_{n_k} \to z$, $\exists N = n_{k_0} \geq 2/\epsilon$ s.t. $y_N \in D(z, \epsilon/2)$.
  5. But $D(y_N, 1/N) \subset D(z,\epsilon) \subset U_{i_0}$ (why?), **a contradiction**.
- Since A **is totally bounded** (see Lemma 4), we can write $A \subset D(y_1, r) \cup \cdots \cup D(y_n, r)$ for finitely many y_i.
- Since $D(y_k, r) \subset U_{i_k}$ for some U_{i_k}, $\quad A \subset U_{i_1} \cup \cdots \cup U_{i_n}$, **finite subcover**. Hence, A is compact.

Theorem. (3.15; Compact $\Leftrightarrow$ Closed and Totally Bounded)

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

(Proof of $\Rightarrow$) Assume A is compact.

1. A is compact $\Rightarrow$ totally bounded & sequentially compact.
2. A is sequentially compact $\Rightarrow$ A is complete.

(Proof of $\Leftarrow$) Assume A is is complete and totally bounded. It suffices to prove that A is sequentially compact. Assume that $\{y_n\}$ is a sequence in A.

1. We may assume that the y_k are all distinct. (Why? If not, ...)
2. Since A is totally bounded, for each $k = 1, 2, \cdots$

   $$\exists x_{k1}, \cdots, x_{kL_k} \in M \text{ s.t. } A \subset D(x_{k1}, 1/k) \cup \cdots \cup D(x_{kL_k}, 1/k)$$

3. Nest page...

Theorem. (Continue...)

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

(Proof of $\Leftarrow$) Assume A is is complete and totally bounded. It suffices to prove that A is sequentially compact.

Assume that $\{y_n\}$ is a sequence in A.

1. **We may assume that the y_k are all distinct. (Why? If not, ...)**
2. **Since A is totally bounded, for each $k = 1, 2, \cdots$**
   $\exists x_{k1}, \cdots, x_{kL_k} \in M$ s.t. $A \subset D(x_{k1}, 1/k) \cup \cdots \cup D(x_{kL_k}, 1/k)$
3. For $k = 1$, an infinitely many y_n lie in one of these disks $D(x_{1,j}, 1)$. Hence, we can select a subseq. $\{y_{11}, y_{12}, \cdots\}$ lying entirely in one of these disks.
4. Repeat the previous step for $k = 2$ and obtain the subseq. $\{y_{21}, y_{22}, \cdots\}$ of $\{y_{11}, y_{12}, \cdots\}$ lying entirely in one of these disks $D(x_{2j}, 1/2)$.
5. Now choose the diagonal subsequence $y_{11}, y_{22}, y_{33}, \cdots$. This sequence is Cauchy seq. because $d(y_{ii}, y_{jj}) \leq \max\{1/i, 1/j\}$.
6. Since A is complete, y_{ii} converges to a point in A.

Theorem. (3.2.1, Heine-Borel thm.)

Let $A \subset M = \mathbb{R}^n$. A is compact iff A is closed and bounded.

Proof.

- Recall Thm 3.1.5: A is compact iff A is closed and totally bounded.
- Since $M = \mathbb{R}^n$ is Euclidean space,

$$A \text{ is bounded} \quad \Leftrightarrow \quad A \text{ is totally bounded}$$

Caution: If M is not Euclidean space, the above statement is not true. See Example 3.1.8 where there is an example that A is bounded but not totally bounded.

Theorem. (3.3.1: Nested Set Property)

Let F_k be a sequence of **compact non-empty set** *in a metric space M such that $F_1 \supseteq F_2 \supseteq F_3 \supseteq \cdots$. Then,*

$$\cap_{k=1}^{\infty} F_k \neq \emptyset.$$

1. For each n, choose $x_n \in F_n$.
2. Since $\{x_n\} \subset F_1$ and F_1 is compact, $\exists$ a subseq $\{x_{n_k}\}$ that converges to some point z in F_1, that is,
$$x_{n_k} \longrightarrow z \in F_1$$
3. With a rearrangement, we may assume that $x_n \to z$. (why?)
4. $n > N \implies x_n \in F_n \subset F_N \implies x_n \in F_N$
5. Since $\lim_{j\to\infty} x_{N+j} = z$ & $x_{N+j} \in F_N$ & F_N is compact, it must be
$$z \in F_N, \qquad N = 1, 2, 3, \cdots$$
This completes the proof.

Definition. (Path-Connected Sets)

- $\phi : [a,b] \to M$ *is said to be* **continuous** *if*

$$t_k \in [a,b] \to t \implies \phi(t_k) \to \phi(t)$$

- *A* **continuous path joining** $x, y \in M$ *is a continuous mapping* $\phi : [a,b] \to M$ *such that* $\phi(a) = x, \phi(b) = y$.
- $A \subset M$ *is said to be* **path-connected** *if for any* $x, y \in A$, *there exists a continuous path* $\phi : [a,b] \to M$ *joining* x *and* y *such that*

$$\phi([a,b]) \subset A.$$

Definition. (3.5.1: Separate, Connected Sets)

Let A be a subset of a metric space M.

- *Two open set U, V are said to be* **separate** *A if*
  1. $U \cap V \cap A = \emptyset$
  2. $U \cap A \neq \emptyset$ & $V \cap A \neq \emptyset$
  3. $A \subset U \cup V$.
- *A is* **disconnected** *if such sets U, V exist.*
- *A is* **connected** *if such sets U, V do not exist.*

Theorem. (3.3.1)

Path-connected sets are connected.

1. Clearly, $[a, b]$ is connected.
2. To derive a **contradiction**, suppose A is path-connected but not connected. Then $\exists$ open sets U, V such that
   (i) $U \cap V \cap A = \emptyset \quad \& \quad A \subset U \cup V$
   (ii) ${}^{\exists}x \in U \cap A \quad \& \quad {}^{\exists}y \in V \cap A$
3. Since A is path-connected, $\exists$ a continuous path $\phi : [a, b] \to M$ s.t. $\phi(a) = x, \ \phi(b) = y, \ \phi([a, b]) \subset A$.
4. From Theorem 4.2.1 which we will learn soon, $\phi([a, b])$ **is connected.** This is **a contradiction** since U, V separate $\phi([a, b])$.

Example 3.1

- **Show that** $A := \{x \in \mathbb{R}^n \;:\; \|x\| \leq 1\}$ **is compact and connected.**

  Proof.

  1. Since A is closed and bounded, A is compact by Heine-Borel thm.
  2. To prove connectedness, let $x, y \in A$.
  3. Define $\phi : [0, 1] \to \mathbb{R}^n$ by $\phi(t) = tx + (1 - t)y$. Clearly, ϕ is continuous path joining $\phi(0) = x$ and $\phi(1) = y$.
  4. $\|\phi(t)\| \leq t\|x\| + (1 - t)\|y\| \leq t + (1 - t) = 1$ for $t \in [0, 1]$. Hence, $\phi([0, 1]) \subset A$.
  5. Hence, A is path-connected.

Example 3.2

- **Let $A \subset \mathbb{R}^n, x \in A$ and $y \in \mathbb{R}^n \setminus A$. Let $\phi : [0,1] \to \mathbb{R}^n$ be a continuous path joining x and y.**
  Show that $\exists\ t_0$ s.t. $\phi(t_0) \in bd(A)$.
  1. Let $t_0 = \sup\{t \ : \ \phi([0,t]) \subset A\}$. This is well-defined because $\phi(0) = x \in A$.
  2. If $t_0 = 1$, clearly $y = \phi(t_0) \in bd(A)$.
  3. Assume $0 \le t_0 < 1$. From the definition of t_0, for $n = 1, 2, \cdots,$ $\exists t_n \quad s.t. \quad t_0 \le t_k \le t_0 + \frac{1}{n} \quad \& \quad \phi(t_n) \in A^c$
  4. Since $\phi(t_n) \in A^c \to \phi(t_0), \quad \phi(t_0) \in bd(A)$.

Chapter 4. Continuous Mappings

Throughout this chapter, we assume that $M = \mathbb{R}^n$ and $N = \mathbb{R}^m$ are Euclidean space with the standard metric

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{j=1}^{n} (x_i - y_j)^2}, \quad \mathbf{x}, \mathbf{y} \in M \\ \rho(\mathbf{v}, \mathbf{w}) &= \|\mathbf{v} - \mathbf{w}\| = \sqrt{\sum_{j=1}^{m} (v_i - w_j)^2}, \quad \mathbf{v}, \mathbf{w} \in N \end{aligned}$$

Please note that the same symbol $\| \cdot \|$ may have different norm depending on its context.

Throughout this chapter, we assume that $A \subset M = \mathbb{R}^n$ and

$$f : A \rightarrow N = \mathbb{R}^m \qquad \text{is a mapping.}$$

Definition. (4.1.1: Continuity of $f : A \to N$)

- *Suppose that* $\mathbf{x}_0 \in \{$*accumulation points of A*$\}$. *We write* $\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = \mathbf{b}$ *if* $\forall \epsilon > 0, \ \exists\, \delta > 0$ *s.t.*

$$0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta \ \ \& \ \ \mathbf{x} \in A \quad \Rightarrow \quad \|f(\mathbf{x}) - \mathbf{b}\| < \epsilon$$

- *Let* $\mathbf{x}_0 \in A$. *We say that* f *is continuous at* $\mathbf{x}_0$ *if either* $\mathbf{x}_0 \neq \{$*accumulation points of A*$\}$ *or* $\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$.
- *Let* $B \subset A$. f *is called continuous on* B *if* f *is continuous at each point on* B. *If* $A = B$, *we just say that* f *is continuous.*

Theorem. (4.1.4: Continuity of $f : A \to N$)

The following assertions are equivalent.

1. f *is continuous on* A
2. *For every convergent seq* $\mathbf{x}_k \to \mathbf{x}_0$ *in* A*, we have* $f(\mathbf{x}_k) \to f(\mathbf{x}_0)$.
3. *For each open set* U *in* N, $f^{-1}(U)$ *is* **open relative to** A*; that is,* $f^{-1}(U) = A \cap V$ *for some open* V
4. *For each closed set* F *in* N, $f^{-1}(F)$ *is* **closed relative to** A*; that is,* $f^{-1}(F) = A \cap G$ *for some close* G

Proof. $1 \overset{easy}{\Longrightarrow} 2 \overset{?}{\Longrightarrow} 4 \overset{easy}{\Longrightarrow} 3 \overset{?}{\Longrightarrow} 1$

Proof of $(2 \Longrightarrow 4)$

Let $F \subset N$ be closed. We want to prove that $f^{-1}(F)$ is closed **relative to** A. We begin with reviewing the definition of closed.

1. B is closed iff $B = B \cup \{\text{accumulation points of } B\}$.
2. B is closed iff for every sequence $\{\mathbf{x}_k\} \subset B$ that $\mathbf{x}_k \to \mathbf{x}_0$, we necessary have $\mathbf{x}_0 \in B$.
3. $B \subset A$ is closed **relative to** A iff $B = (B \cup \{\text{accumulation points of } B\}) \cap A$
4. $B \subset A$ is closed **relative to** A iff for every sequence $\{\mathbf{x}_k\} \subset B$ that $\mathbf{x}_k \to \mathbf{x}_0 \in A$, we necessary have $\mathbf{x}_0 \in B$.
5. **Proof of $(2 \Longrightarrow 4)$.** Let $\mathbf{x}_k \in f^{-1}(F)$ and let $\mathbf{x}_k \to \mathbf{x}_0 \in A$. By 2, $f(\mathbf{x}_k) \to f(\mathbf{x}_0)$. Since F is closed, $f(\mathbf{x}_0) \in F$. $\therefore\ \mathbf{x}_0 \in f^{-1}(F)$. $\therefore f^{-1}(F)$ is closed **relative to** A.

Proof of $(3 \Longrightarrow 1)$

For given $\mathbf{x}_0 \in A$ and $\epsilon > 0$, we must find $\delta > 0$ such that

$$\underbrace{\|\mathbf{x} - \mathbf{x}_0\| < \delta \ \ \& \ \ \mathbf{x} \in A}_{\mathbf{x} \ \in \ D(\mathbf{x}_0,\delta)\cap A} \quad \Rightarrow \quad \underbrace{\|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \epsilon}_{f(\mathbf{x}) \ \in \ D(f(\mathbf{x}_0),\epsilon)}$$

1. Since $D(f(\mathbf{x}_0), \epsilon)$ is open, by 3

$$f^{-1}\left(D(f(\mathbf{x}_0), \epsilon)\right) \quad \text{is open relative to } A.$$

$$\therefore \quad f^{-1}\left(D(f(\mathbf{x}_0), \epsilon)\right) = A \cap V \quad \text{for some open set } V.$$

2. Since $\mathbf{x}_0 \in V$ and V is open,

$$\exists \ \ \delta > 0 \ \ s.t. \ \ D(\mathbf{x}_0, \delta) \subset V.$$

3. Hence, $D(\mathbf{x}_0, \delta) \cap A \subset f^{-1}\left(D(f(\mathbf{x}_0), \epsilon)\right)$ and this completes the proof.

Theorem. (4.2.1: $f(connected)$ is connected if $f \in C(M)$)

Suppose that $f : M \to N$ is **continuous** *and let $K \subset M$.*

(i) *If K is connected, so is $f(K)$.*

(ii) *If K is path-connected, so is $f(K)$.*

Proof of (i). Suppose $f(K)$ is not connected.

1. From the definition of disconnectedness, $\exists$ *open* U, V $s.t.$

   $$f(K) \subset U \cup V, \quad U \cap V \cap f(K) = \emptyset, \quad U \cap f(K) \neq \emptyset, \quad V \cap f(K) \neq \emptyset$$

2. Since f is continuous, $f^{-1}(U)$ and $f^{-1}(V)$ are open. Moreover, $K \subset f^{-1}(U) \cup f^{-1}(V)$, $f^{-1}(U) \cap f^{-1}(V) \cap K = \emptyset$, $f^{-1}(U) \cap K \neq \emptyset$, $f^{-1}(V) \cap K \neq \emptyset$.

3. Hence, K is disconnected, **a contradiction.**

Proof of (ii). If K is path-connected, so is $f(K)$.

1. Let $\mathbf{v}, \mathbf{w} \in f(K)$ and let $\mathbf{x}, \mathbf{y} \in K$ s.t. $f(\mathbf{x}) = \mathbf{v}, f(\mathbf{y}) = \mathbf{w}$.
2. Since K is path-connected, $\exists$ a continuous curve $c : [0,1] \to M$ s.t.

$$c(t) \in K \;\; (0 \leq t \leq 1), \quad c(0) = \mathbf{x}, \; c(1) = \mathbf{y}$$

3. Since f is continuous, it is easy to show that $\tilde{c}(t) = f(c(t)) \in f(K)$ for $0 \leq t \leq 1$ and and $\tilde{c} : [0,1] \to N$ is continuous path joining $\mathbf{v}$ and $\mathbf{w}$.
4. Hence, $f(K)$ is path-connected

Theorem. (4.2.2: $f(compact)$ is compact if $f \in C(M)$)

Suppose that $f : M \to N$ is **continuous** *and $K \subset M$ is* **compact**. *Then $f(K)$ is* **compact**.

It suffices to prove that $f(K)$ is **sequentially compact**.

1. Let $\mathbf{v}_n \in f(K)$. Let $\mathbf{x}_n \in K$ s.t. $f(\mathbf{x}_n) = \mathbf{v}_n$.
2. Since K is compact, $\exists$ a convergent subsequence, say, $\mathbf{x}_{n_k} \to \mathbf{x}_0 \in K$.
3. Since f is continuous, $\mathbf{v}_{n_k} = f(\mathbf{x}_{n_k}) \to f(\mathbf{x}_0) \in f(K)$. This proves that $f(K)$ is **sequentially compact**.

Examples

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a continuous map. Denote $\mathbf{x} = (x_1, x_2)$.

- Let $f(\mathbf{x}) = x_1$ for $x \in \mathbb{R}^2$. If $K \subset \mathbb{R}^2$ be compact, so is $f(K) = \{x_1 \ : \ \mathbf{x} = (x_1, x_2) \in K\}$. (Why? Since f is continuous and K is compact, $f(K)$ is compact.)
- Let $f(\mathbf{x}) = 7$ for $x \in \mathbb{R}^2$. The set $\{7\}$ is compact, while $\mathbb{R}^2 = f^{-1}(\{7\})$ is **not** compact.
- The set $A = \{f(\mathbf{x}) \ : \ \|\mathbf{x}\| = 1\}$ is a closed interval. (Why? $K = \{\mathbf{x} \in \mathbb{R}^2 \ : \ \|\mathbf{x}\| = 1\}$ is compact and connected. Hence, $A = f(K)$ is compact and connected.)

Theorem.

(1) Let $f : A \subset N \to M$ and $g : A \subset N \to M$ be continuous at $\mathbf{x}_0$. Then

- *$f \pm \alpha g$ is continuous at $\mathbf{x}_0$ for any $\alpha \in \mathbb{R}$.*
- *fg is continuous at $\mathbf{x}_0$*
- *f/g is continuous at $\mathbf{x}_0$ if $g(\mathbf{x}_0) \neq 0$.*

(2) Suppose $f : A \subset N \to M$ and $h : B \subset N \to \mathbb{R}^p$ are continuous and $f(A) \subset B$. Then $h \circ f : A \subset N \to \mathbb{R}^p$ is also continuous.

Proof. EASY

Theorem. (4.4.1: Maximum-Minimum Principle)

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

- $f(K)$ *is bounded.*
- $\exists x_0, y_0 \in K$ *such that*

$$f(x_0) = \inf f(K) = \inf_{x \in K} f(x) \;\;\&\;\; f(y_0) = \sup f(K) = \sup_{x \in K} f(x).$$

Proof. Since K is compact and f is continuous on $K \subset A$, $f(K)$ is compact. Hence, $f(K)$ is closed and bounded in $\mathbb{R}$ by Heine-Borel thm. This completes the proof.

Theorem. (4.5.1: Intermediate Value Theorem)

Let $f : A \subset M \to \mathbb{R}$ be continuous. Assume K is a connected subset in A and $x, y \in K$ and $f(x) < f(y)$. Then,

- *For every number $c \in \mathbb{R}$ such that $f(x) < c < f(y)$,*

$$\exists\, z \in K \;\; s.t. \;\; f(z) = c$$

Proof. Since K is connected and f is continuous on $K \subset A$, $f(K)$ is connected. Hence, $[f(x), f(y)] \subset f(K)$.
$\therefore \exists\, z \in K \;\; s.t. \;\; f(z) = c$. This completes the proof.

4.6 Uniform Continuity

Throughout this section, we assume that $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ is continuous.

- **Definition.** Let $B \subset A$. f is **uniformly continuous** on B if for every $\epsilon > 0$, there is $\delta > 0$ s.t.

$$\|x - y\| < \delta \;\&\; x, y \in B \;\Rightarrow\; \|f(x) - f(y)\| < \epsilon.$$

- **Example.** Consider $f : \mathbb{R} \to \mathbb{R},\; f(x) = x^2$. Then f is continuous on $\mathbb{R}$, but it is not uniformly continuous. Why? Let $x_n = n + 1/n$ and $y_n = n$. Then $|x_n - y_n| = 1/n \to 0$, while $|f(x_n) - f(y_n)| \geq 1$.
- **Example.** Consider $f : (0, 1) \to \mathbb{R},\; f(x) = 1/x$. Then f is continuous on $(0, 1)$, but it is not uniformly continuous. Why? Let $x_n = 1/n$. Then $|x_{n+1} - x_n| < 1/n \to 0$, while $|f(x_{n+1}) - f(x_n)| = 1$.

Theorem. (Uniform Continuity Theorem)

Let $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ be continuous and let $K \subset A$ be compact. Then f **uniformly continuous** *on K.*

1. Let ϵ be given. Since f is continuous on K, for each $x \in K$, $\exists \delta_x > 0$ s.t. $f(D(x, \delta_x) \cap K) \subset D\left(f(x), \frac{\epsilon}{2}\right)$
2. Since $K \subset \cup_x D(x, \delta_x/2)$ and K is compact, $\exists \{x_1, \cdots, x_N\} \subset K$ s.t. $K \subset \cup_{j=1}^N D(x_j, \delta)$ where $\delta = \frac{1}{2} \min\{\delta_{x_1}, \cdots, \delta_{x_N}\}$.
3. If $|x - y| < \delta$, $x, y \in K$, then $\exists x_j$ s.t. $|x - x_j| < \delta$. Since $|y - x_j| \le |y - x| + |x - x_j| < 2\delta \le \delta_{x_j}$, $\|f(x) - f(y)\| \le \|f(x) - f(x_j)\| + \|f(x_j) - f(y)\| \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$.

Chapter 5. Uniform Convergence

This chapter deals with very important results in physical science:

- **a basic iteration technique called the contraction mapping principle** (5.7.1)
- **some applications to differential and integral equations and some problems in control theory.** (5.7.2, 5.7.3, 5.7.10)

To study such results, we need

- **compactness in a complete metric space** (5.5.3)
- **uniform convergence, equi-continuity** (5.6.2)

Definition. (Pointwise convergence & Uniform Convergence)

Let N be a metric space with the metric ρ, A a set, and $f_k : A \to N, k = 1, 2, \cdots$

- $f_k \to f$ **pointwise** *if for each* $x \in A$, $\lim_{k\to\infty} f_k(x) = f(x)$, *i.e.*

$$\forall x \in A, \quad \lim_{k\to\infty} \rho(f_k(x), f(x)) = 0$$

- $f_k \to f$ **uniformly** *if* $\lim_{k\to\infty} \sup_{x\in A} \rho(f_k(x), f(x)) = 0$, *i.e.*

$$\forall \epsilon > 0, \ \ \exists\ N \ \ s.t. \ \ k > N \ \Rightarrow \ \ \sup_{x\in A} \rho(f_k(x), f(x)) < \epsilon$$

Examples:

- $f_k(x) = x^k \to 0$ pointwise in $(0, 1)$. (Why?)
- $f_k(x) = x^k$ **does NOT** converge to 0 uniformly in $(0, 1)$.
- Show that $f_n(x) = \frac{x^n}{1+x^n}$ converges pointwise on $[0, 2]$ but that the convergence is not uniform.

Definition. (5.1.3: Does $\sum_k g_k$ makes sense ?)

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

- $\sum_k g_k = f$ *(pointwise) if* $f_n \to f$ *pointwise.*
- $\sum_k g_k = f$ *uniformly if* $f_n \to f$ *uniformly.*

Examples.

- $\sum_{k=0}^\infty \frac{(-1)^k x^{2k+1}}{(2k+1)!} = \sin x$ uniformly in the interval $[-100, 100]$.
- $\sum_k x^k = \frac{1}{1-x}$ converges uniformly in $[-0.9, 0.9]$
- $\sum_k x^k = \frac{1}{1-x}$ converges pointwise (NOT uniformly) in $(-1, 1)$
- $\sum_k x^k$ does not converge in $\mathbb{R} \setminus (-1, 1)$

The Weierstrass M-test

Theorem. (5.2.1: Cauchy Criterion)

Let $\mathcal{V}$ be a complete normed vector space with norm $\|\cdot\|$, and let A be a set. Let $f_k : A \to \mathcal{V}$ is a sequence of functions. Then f_k converges uniformly on A iff

$$\forall\ \epsilon > 0, \quad \exists\ N \ \ s.t.\ \ l, k > N \ \Rightarrow\ \sup_{x \in A} \|f_k(x) - f_l(x)\| < \epsilon$$

Proof of $\Rightarrow$.

1. Assume $f_k \longrightarrow f$ uniformly. Let $\epsilon > 0$ be given.
2. Then
   $\exists\ N \ \ s.t.\ \ k \geq N \ \Rightarrow\ \|f_k - f\| = \sup_{x \in A} |f_k(x) - f(x)| \ < \epsilon/2.$
3. Hence,
   $l, k \geq N \ \Rightarrow\ \|f_k - f_l\| \leq \|f_k - f\| + \|f_l - f\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$

Theorem. (5.2.1: Continue...)

♣♢♡ *Then* f_k *converges uniformly on* A *iff*
$\forall\ \epsilon > 0, \quad \exists\ N \ \ s.t.\ \ l, k > N \ \Rightarrow\ \sup_{x\in A} \|f_k(x) - f_l(x)\| < \epsilon$

Proof of $\Leftarrow$.

1. From the assumption, $f_k(x)$ is Cauchy sequence for all $x \in A$.
2. Hence, for all $x \in A$, $\exists \lim_k f_k(x)$ and we can define $f(x) = \lim_k f_k(x)$.
3. Let $\epsilon > 0$ be given. From the assumption, $\exists\ N \ \ s.t.\ \ l, k > N \ \Rightarrow\ \sup_{x\in A} \|f_l(x) - f_l(x)\| < \epsilon/2$.
4. From 2, $\forall x \in A, \ \ \exists\ N_x \ \ s.t.\ \ l > N_x \ \Rightarrow\ \|f(x) - f_l(x)\| < \epsilon/2$.
5. From 3 and 4, if $k \geq N$ and $x \in A$, then $\|f_k(x) - f(x)\| \leq \|f_k(x) - f_l(x)\| + \|f_l(x) - f(x)\| \ < \epsilon/2 + \epsilon/2$ for any $l \geq N_x$.
6. From 5, $k \geq N \ \Rightarrow\ \sup_{x\in A} \|f_k(x) - f(x)\| < \epsilon$.

Theorem. (5.2.2: Weierstrass M test)

Let $\mathcal{V}$ be a complete normed vector space with norm $\|\cdot\|$, and let A be a set. Suppose that $g_k : A \to \mathcal{V}$ are functions such that $\sup_{x\in A} \|g_k(x)\| < M_k$ and $\sum_{k=1}^{\infty} M_k < \infty$. Then $\sum_{k=1}^{\infty} g_k$ converges uniformly.

Proof.

1. Denote $f_n(x) = \int_{k=1}^{n} g_k(x)$.
2. Then $\|f_n(x) - f_{n+\ell}(x)\| = \|\sum_{k=n}^{n+\ell} g_k(x)\| \leq \sum_{k=n}^{n+\ell} M_k$.
3. Since $\lim_{n\to\infty} \sum_{k=n}^{\infty} M_k = 0$, it follows from 2 and Theorem 5.2.1 that f_n converges uniformly.

5.5 The space of continuous functions

Throughout this section, we assume $M = \mathbb{R}^n$, $A \subset M$, and $N = \mathbb{R}^n$. (N,M: complete normed space)

- Denote $\mathcal{C}(A, N) = \{f \ : \ f : A \to N \text{ is continuous } \}$. Then $\mathcal{C}$ is a vector space.
- For $f \in \mathcal{C}(A, N)$, f is said to be bounded if there is a constant C such that $\|f(x)\| < C$ for all $x \in A$.
- Denote $\mathcal{C}_b(A, N) = \{f \in \mathcal{C} \ : \ f \text{ is bounded } \}$.
- Define

$$\|f\| = \sup_{x \in A} \|f(x)\|$$

- $\|f\|$ is a measure of the size of f and is called the **norm of f**.

Theorem. (5.5.1-3: $\mathcal{C}_b(A, N)$ is a complete normed space)

Let $A \subset M = \mathbb{R}^m, N = \mathbb{R}^n$. The set $\mathcal{C}_b(A, N)$ is a complete normed space equipped with the norm $\|f\| = \sup_{x \in A} \|f(x)\|$; that is,

1. $\mathcal{C}_b(A, N)$ *is a* **normed space**.
   - $\|f\| \geq 0$ *and* $\|f\| = 0$ *iff* $f = 0$.
   - $\|\alpha f\| = |\alpha| \|f\|$ *for* $\alpha \in \mathbb{R}$, $f \in \mathcal{C}_b$.
   - $\|f + g\| \leq \|f\| + \|g\|$.
2. **Completeness:** *Every Cauchy sequence* $\{f_k\}$ *in* $\mathcal{C}_b(A, N)$ *converges to a function* $f \in \mathcal{C}_b(A, N)$, *that is,*

$$\lim_{k \to \infty} \|f_k - f\| = \lim_{k \to \infty} \sup_{x \in A} \|f_k(x) - f(x)\| = 0.$$

- Clearly, $\mathcal{C}_b(A, N)$ is a normed space. (EASY!)
- From the definition, $f_k \to f$ uniformly iff $f_k \to f$ in $\mathcal{C}_b$.
- From Cauchy criterion (Theorem 5.2.1), $\mathcal{C}_b(A, N)$ is complete.

Examples

- Let $B = \{f \in C([0,1], \mathbb{R}) \ : \ f(x) > 0 \ \text{ for all } x \in [0,1]\}$.
  Show that B is open in $C([0,1], \mathbb{R})$.
  Proof.
    1. In order to prove that B is open, we must show that $\forall\ f \in B, \quad \exists \epsilon > 0 \ \ s.t. \ \ D(f, \epsilon) \subset B$.
    2. Let $f \in B$. Since $[0,1]$ is compact, f has a minimum value-say, m- at some point in $[0,1]$. Hence, $\inf_{x \in [0,1]} f(x) = m$.
    3. Let $\epsilon = \frac{m}{2}$. We will show $D(f, \epsilon) \subset B$.
       Proof. If $g \in D(f, \epsilon)$, then $\|g - f\| < \epsilon$, and $\therefore \ \ g(x) \geq f(x) - |g(x) - f(x)| \geq f(x) - \|f - g\| \geq m - \epsilon = \frac{m}{2}$ for all $x \in [0,1]$. Hence, $g \in B$. $\therefore D(f, \epsilon) \subset B$
- Prove that $\overline{B}$ is $D = \{f \in \mathcal{C}_b \ : \ \inf_{x \in [0,1]} f(x) \ \geq 0\}$.
  Proof.
    1. $\overline{D} = D$ because if $f_n \in D \ \to f$ uniformly, then $f_n(x) \to f(x)$ pointwise and $\therefore \ \inf_{x \in [0,1]} f(x) \geq 0$.
    2. If $f \in D$, then $f_n(x) := f(x) + \frac{1}{n} \in B$ and $\|f_n - f\| = \frac{1}{n} \to 0$. $\therefore B \subset D \subset \overline{B}$.

Examples

- Consider a sequence $f_n \in \mathcal{C}_b$ such that $\|f_{n+1} - f_n\| \leq r_n$, where $\sum r_n$ is convergent. Prove that f_n converges.

Proof.

1. Let $\epsilon > 0$ be given.
2. Since $\sum r_n$ is convergent,

$$\exists\ N \ \ s.t. \ \ n > N \ \Rightarrow \ \sum_{k=n}^{\infty} r_k < \epsilon$$

3. Hence, if $n \geq N$, then

$$\|f_{n+k} - f_n\| = \|\sum_{j=n}^{n+k-1} (f_{j+1} - f_j)\| \leq \sum_{j=n}^{n+k-1} \|f_{j+1} - f_j\| \leq \sum_{j=n}^{\infty} r_j < \epsilon$$

4. From 3, f_n is a Cauchy sequence, so it converges.

Arzela-Ascoli Theorem

Throughout this section, we assume that
$M = \mathbb{R}^m, \; A \subset M, \; N = \mathbb{R}^n$ (N,M: complete normed space).

Definition. (5.6.1: Equi-continuous)

Assume $\mathcal{B} \subset \mathcal{C}(A, N)$.

- *We say that $\mathcal{B}$ is* **equi-continuous** *if*

$$\forall\, \epsilon > 0, \quad \exists\, \delta > 0 \;\; s.t. \qquad \|x - y\| < \delta \;\&\; x, y \in A$$
$$\Rightarrow \quad \sup_{f \in \mathcal{B}} \|f(x) - f(y)\| < \epsilon$$

- *We say $\mathcal{B}$ is* **pointwise compact** *iff* $\mathcal{B}_x = \{f(x) \;:\; f \in \mathcal{B}\}$ *is compact in N for each $x \in A$.*

Example 5.6.4 (Compact sequence)

Let $f_n \subset \mathcal{C}_b([0,1],\mathbb{R})$ and be such that f_n' exist and

$$\sup_n \|f_n\| \le C \quad \& \quad \sup_n \left(\sup_{x\in(0,1)} |f_n'(x)| \right) \le C$$

for a positive constant C. Prove that $\mathcal{B} := \{f_n\}$ is equicontinuous.

Proof.

- By the mean value theorem,

$$|f_n(x) - f_n(y)| \le \sup_{z\in(0,1)} |f_n'(z)|\, |x-y| \le C|x-y|, \qquad \text{for all } n$$

- Hence, for given $\epsilon > 0$, we can choose $\delta = \frac{\epsilon}{C}$ and

$$|x-y| < \delta \,\&\, x,y \in [0,1] \;\Rightarrow\; \sup_n |f_n(x)-f_n(y)| < C|x-y| < \epsilon.$$

Hence, $\mathcal{B} := \{f_n\}$ is equi-continuous. (So, $\{f_n\}$ has a convergent subsequence. Why? See Arzela-Ascoli theorem.)

Theorem. (5.6.2:Arzela-Ascoli theorem)

Let A be compact and $\mathcal{B} \subset \mathcal{C}(A, N)$. If $\mathcal{B}$ is closed, equi-continuous, and pointwise compact, then $\mathcal{B}$ is compact, that is, any sequence f_n in $\mathcal{B}$ has a uniformly convergent subsequence.

The proof strategy is based on Bolzano-Wierstrass properties.

Theorem. (Special case of Arzela-Ascoli theorem)

Let $\mathcal{B} \subset \mathcal{C}([0,1], \mathbb{R})$. If $\mathcal{B}$ is closed, equi-continuous, and bounded, then $\mathcal{B}$ is compact.

Proof.

1. Assume f_n is a sequence in $\mathcal{B}$.
2. Denote $C_{1/n} = \{\frac{1}{n}, \frac{2}{n}, \cdots, \frac{n-1}{n}, 1\}$. Let $C = \cup_n C_{1/n}$.
3. Since C is countable, we can write $C = \{x_1, x_2, \cdots\}$.

Theorem. (Special case of Arzela-Ascoli theorem)

Let $\mathcal{B} \subset \mathcal{C}([0,1],\mathbb{R})$. If $\mathcal{B}$ is closed, equi-continuous, and bounded, then $\mathcal{B}$ is compact.

Proof.

1. Assume f_n is a sequence in $\mathcal{B}$.
2. Denote $C_{1/n} = \{\frac{1}{n}, \frac{2}{n}, \cdots, \frac{n-1}{n}, 1\}$. Let $C = \cup_n C_{1/n}$.
3. Since C is countable, we can write $C = \{x_1, x_2, \cdots\}$.
4. Since $\mathcal{B}_{x_1}$ is compact, $\exists$ a convergent subseq of $f_n(x_1)$. Let us denote this subsequence by

$$f_{11}(x_1),\ f_{12}(x_1), \cdots, f_{1k}(x_1), \cdots$$

5. Similarly, the sequence $f_{1k}(x_2)$ has a subsequence

$$f_{21}(x_2),\ f_{22}(x_2), \cdots, f_{2k}(x_2), \cdots \qquad \text{which is converegnt.}$$

6. We proceed in this way and then set $g_n = f_{nn}$.

Proof of Arzela-Ascoli theorem

7. $g_n = f_{nn}$ is obtained by picking out the diagonal

$$\begin{array}{cccccccc} f_{11} & f_{12} & f_{13} & \cdots & f_{1n} & \cdots & & \text{(1st subseq.)} \\ f_{21} & f_{22} & f_{23} & \cdots & f_{11} & \cdots & & \text{(2nd sub seq.)} \\ \vdots & \vdots & \vdots & & \vdots & & & \\ f_{n1} & f_{n2} & f_{n3} & \cdots & f_{nn} & \cdots & & \text{(n-th subseq.)} \end{array}$$

8. From the construction from the diagonal process,

$$\lim_{n\to\infty} g_n(x_i) \quad \text{exists for all } x_i \in C.$$

9. Now, we are ready to prove

$$\|g_n - g_m\| = \sup_{x\in[0,1]} |g_n(x) - g_m(x)| \to 0 \qquad \text{as} \quad m, n \to \infty.$$

Continue...

9. Proof of $\lim_{n,m\to\infty} \sup_{x\in A} |g_n(x) - g_m(x)| = 0$.
   a. Let $\epsilon > 0$ be given.
   b. From equi-continuity of $\{g_n\} \subset \mathcal{B}$, we can choose δ s.t.

   $$|x-y| < \delta \ \ \& \ \ x, y \in A = [0,1] \quad \Rightarrow \quad \sup_n |g_n(x) - g_n(y)| < \epsilon/3$$

   c. Choose $L \geq \frac{1}{\delta}$. From 8,

   $$\exists\ N \ \ s.t. \ \ n, m > N \quad \Rightarrow \quad \sup_{x_i \in \mathcal{C}_{1/L}} |g_n(x_i) - g_m(x_i)| \ < \frac{\epsilon}{3}.$$

   d. For each $x \in A$, there exist $y_j \in \mathcal{C}_{1/L}$ s.t. $|x - y_j| < \delta$.
   Therefore, if $n, m > N$, then

   $$\begin{aligned} |g_n(x) - g_m(x)| \quad &\leq \quad |g_n(x) - g_n(y_j)| + |g_n(y_j) - g_m(y_j)| \\ &\quad + |g_m(x) - g_m(y_j)| \quad \leq \quad \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \end{aligned}$$

   This proves $\lim_{n,m\to\infty} \sup_{x\in A} |g_n(x) - g_m(x)| = 0$.

Continue...

10. From 9, g_n is a Cauchy sequence in $\mathcal{C}([0,1], N)$.
11. Since $\mathcal{C}([0,1], N)$ is the complete normed space, g_n converges to some $g \in \mathcal{C}([0,1], N)$.
12. Since $\mathcal{B}$ is closed, it must be $g \in \mathcal{B}$.
13. From 1, 11, and 12, $\mathcal{B}$ is sequentially compact, so it is compact.

♣ ♣ ♣ ♣ ♣ ♣ ♣

The proof of Arzela-Ascoli theorem is exactly the same as the special case discussed above except the step 2.
For the replacement of the step 2, we use the fact that the compact set A is **totally bounded** . The compactness of A provides that, for each $\delta > 0$, there exist a finite set $C_\delta = \{y_1, \cdots, y_k\}$ such that $A \subset \cup_{j=1}^{k} D(y_j, \delta)$.

5.7 The contraction mapping principle

Theorem. (5.7.1: Contraction mapping principle)

Let M be a complete normed space and $\Phi : M \to M$ a given mapping. Assume

$$\exists \ k \in [0,1) \ \ s.t. \ \ \|\Phi(f) - \Phi(g)\| \le k \, \|f - g\| \quad \textit{for all} \ \ f, g \in M$$

Then there exists a unique fixed point $f_* \in M$ s.t. $\Phi(f_*) = f_*$. In fact, if $f_0 \in M$ and $f_{n+1} = \Phi(f_n), n = 0, 1, 2, \cdots$, then

$$\lim_{n \to \infty} \|f_n - f_*\| \ = \ 0$$

Key idea: Φ is shrinking distances:

$$\|f_{n+1} - f_n\| = \|\Phi(f_n) - \Phi(f_n)\| \le k\|f_n - f_{n-1}\| \le \cdots \le k^n\|f_1 - f_0\|$$

The proof of contraction mapping principle: $\exists\ f_* \in M \quad s.t. \quad \Phi(f_*) = f_*$

1. Let $f_0 \in M$ and $f_{n+1} = \Phi(f_n), n = 0, 1, 2, \cdots$.
2. $\|f_2 - f_1\| = \|\Phi(f_1) - \Phi(f_0)\| \leq k\|f_1 - f_0\|$.
3. $\|f_3 - f_2\| = \|\Phi(f_2) - \Phi(f_1)\| \leq k\|f_2 - f_1\| \leq k^2\|f_1 - f_0\|$.
4. Inductively, $\|f_{n+1} - f_n\| \leq k^n\|f_1 - f_0\|$.
5. Hence, $\sum_{n=0}^{\infty} \|f_{n+1} - f_n\| \leq \|f_1 - f_0\| \sum_{n=0}^{\infty} k^n = \|f_1 - f_0\| \frac{1}{1-k} < \infty$.
6. From the proof in Example 5.5.6 and 4, f_n converges.
7. Since M is complete, $\lim_{n\to\infty} f_n = f_*$ for some $f_* \in M$.
8. Φ is uniformly continuous because $\|\Phi(f) - \Phi(g)\| \leq k\ \|f - g\|$.
9. From 8, $\lim_{n\to\infty} \Phi(f_n) = \Phi(f_*)$.
10. Hence, $f_* = \lim_{n\to\infty} f_{n+1} = \lim_{n\to\infty} \Phi(f_n) = \Phi(f_*)$.

The proof of contraction mapping principle: Uniqueness of the fixed point f_*

11. To prove the uniqueness, assume g_* is another fixed point, i.e., $\Phi(g_*) = g_*$
12. Then $f_* - g_* = \Phi(f_*) - \Phi(g_*)$ and

$$\|f_* - g_*\| \ = \ \|\Phi(f_*) - \Phi(g_*)\| \leq k\|f_* - g_*\|$$

    Hence, $(1-k)\|f_* - g_*\| \leq 0$.
13. Since $1 \leq k < 1$, it must be

$$\|f_* - g_*\| = 0$$

    Hence, $f_* = g_*$

Theorem. (5.7.2: Existence of sol'n of Differential equations)

Let $A \subset \mathbb{R}^2$ be an open neighborhood of (t_0, x_0). Assume $f : A \to \mathbb{R}$ is continuous function satisfying the following **Lipschitz condition**:

$$|f(t,x_1) - f(t,x_2)| \le K|x_1 - x_2| \quad \text{for all } (t,x_1),(t,x_2) \in A.$$

Then, there is a $\delta > 0$ s.t. the equation

$$\frac{dx(t)}{dt} = f(t,x), \quad x(t_0) = x_0$$

has a unique C^1-solution $x = \phi(t)$ with $\phi(t_0) = x_0$, for $t \in (t_0 - \delta, t_0 + \delta)$, i.e.,

$$\phi'(t) = f(t,\phi(t)) \quad \text{for all} \quad t \in (t_0 - \delta, t_0 + \delta) \quad \& \quad \phi(t_0) = x_0$$

C^1-solution = continuously differentiable solution

Get insight: Proof of Theorem 5.7.2

Before the proof, let us get some insight. Imagine that ϕ is the solution of $\frac{dx(t)}{dt} = f(t, x),\ x(t_0) = x_0$.
Since $\phi'(t) = f(t, \phi(t))$ with $\phi(t_0) = x_0$,

$$\phi(t) = \phi(t_0) + \int_{t_0}^{t} \phi'(s)ds = x_0 + \int_{t_0}^{t} f(s, \phi(s))ds$$

Hence, ϕ is a fixed point for the map $\Phi : M \to M$ defined by

$$\Phi(\phi) = x_0 + \int_{t_0}^{t} f(s, \phi(s))ds$$

In order to apply the contraction mapping principle, we need to choose **a suitable space** M. In practice, the solution ϕ can be achieved from the following iterative method:

$$\phi_{n+1}(t) = \Phi(\phi_n) = x_0 + \int_{t_0}^{t} f(s, \phi_n(s))ds \quad \& \quad \phi_0 = x_0$$

Proof of Theorem 5.7.2

1. Let $L = \sup_{(x,t)\in\tilde{A}} |f(x,t)|$ where $\tilde{A}$ is a closed subset of A.

   Since f is continuous in A, $L < \infty$.

2. Choose δ such that $K\delta < 1$ and

$$\{(t,x) \ : \ |t-t_0| < \delta, \ |x-x_0| < L\delta\} \ \subset \tilde{A}$$

3. Denote $\mathcal{C} = \mathcal{C}([t_0+\delta, t_0+\delta], \mathbb{R})$. From theorem 5.5.3, $\mathcal{C}$ is a complete normed space (or Banach space) with norm

$$\|\phi\| = \sup_{t\in[t_0+\delta, t_0+\delta]} |\phi(t)|$$

4. Let

$$M = \{\phi \in \mathcal{C} \ : \ \phi(t_0) = x_0 \ \ \& \ \ |\phi(t) - x_0| \leq L\delta\}$$

5. **Then, M is also a complete normed space.** (Why? M is closed subset of $\mathcal{C}$ w.r.t. the norm $\|\cdot\|$.)

Proof of Theorem 5.7.2

5. Define $\Phi : M \to \mathcal{C}$ by (Please find its motivation from the previous slide)

$$\Phi(\phi) = x_0 + \int_{t_0}^{t} f(s, \phi(s))ds$$

6. Claim: $\phi \in M \;\Rightarrow\; \Phi(\phi) \in M$.
   Proof. Let $\phi \in M$ and $\psi = \Phi(\phi)$.
   - $\psi(t_0) = x_0$ and $\psi \in \mathcal{C}$ because

$$\lim_{h\to 0} |\psi(t+h) - \psi(t)| = \lim_{h\to 0} \left| \int_{t}^{t+h} f(s, \phi(s))ds \right| \le \lim_{h\to 0} Lh = 0$$

   - From 1,

$$|t - t_0| \le \delta \;\Rightarrow\; |\psi(t) - x_0| = \left| \int_{t_0}^{t} f(s, \phi(s))ds \right| \le L|t - t_0| \le L\delta$$

   - Hence, $\psi \in M$.

7. From 6, Φ maps M to M. See the condition of Theorem 5.7.1.

Proof of Theorem 5.7.2

7. Using the Lipschitz condition,

$$\begin{aligned}\|\Phi(\phi_1)-\Phi(\phi_2)\| &= \sup_{t\in[t_0+\delta,t_0+\delta]}\left|\int_{t_0}^{t} f(s,\phi_1(s))-f(s,\phi_2(s))ds\right| \\ &\le \sup_{t\in[t_0+\delta,t_0+\delta]}\left|\int_{t_0}^{t} K|\phi_1(s)-\phi_2(s)|ds\right| \\ &\le \delta K\|\phi_1-\phi_2\|\end{aligned}$$

8. Since $\delta K < 1$,

$$\|\Phi(\phi_1)-\Phi(\phi_2)\| \le k\|\phi_1-\phi_2\|, \quad k=\delta K\in[0,1)$$

9. From 5.7.1, $\exists\ \phi_* \in M$ s.t. $\Phi(\phi_*)=\phi_*$.

Theorem. (5.7.3: Fredholm equation)

Assume that $K(x,y)$ *is continuous on* $[a,b]\times[a,b]$ *and*

$$M = \sup_{x,y\in[a,b]} |K(x,y)|$$

If $|\lambda| M |b-a| < 1$*, then the following Fredholm equation has a unique solution in* $\mathcal{C}([a,b],\mathbb{R})$*:*

$$f(x) = \lambda \int_a^b K(x,y)\ f(y)\ dy + \phi(x), \quad x \in [a,b]$$

where $\lambda \in \mathbb{R},\ \phi \in \mathcal{C}([a,b],\mathbb{R})$.

Proof. For $f \in \mathcal{C}([a,b],\mathbb{R})$, we define

$$(\Phi(f))(x) = \lambda \int_a^b K(x,y)\ f(y)\ dy + \phi(x)$$

Proof of 5.7.3

1. Claim: Φ maps from $\mathcal{C}([a,b],\mathbb{R})$ to $\mathcal{C}([a,b],\mathbb{R})$.
   Proof. Let $f \in \mathcal{C}([a,b],\mathbb{R})$. We need to show that $\Phi(f)$ is continuous. Let $\epsilon > 0$ be given.
   - Since $[a,b]\times[a,b]$ is compact, $K(x,y)$ is uniformly continuous.
   - Hence, $\exists\ \delta$
     s.t. $\|(x_1,y)-(x_2,y)\| < \delta$ & $(x_1,y),\ (x_2,y) \in [a,b]\times[a,b]$
     imply $|K(x_1,y)-K(x_2,y)| < \frac{\epsilon}{\|f\|\ |b-a|+1}$.
   - If $|x_1-x_2| < \delta$ and $x_1, x_2 \in [a,b]$, then
     $|(\Phi(f))(x_1)-(\Phi(f))(x_2)| =$
     $\int_a^b |K(x_1,y)-K(x_2,y)||f(y)|dy \le \delta\ \|f\||b-a| < \epsilon.$
2. Set $k = |\lambda|M|b-a|$. Then $k < 1$ and

$$\|\Phi(f)-\Phi(g)\| = \sup_{x\in[a,b]}\left|\int_a^b K(x,y)(f(y)-g(y)|dy\right| \le k\|f-g\|$$

3. From 5.7.1, $\exists$ unique $f_* \in \mathcal{C}([a,b],\mathbb{R})$ s.t. $\Phi(f_*) = f_*$.

Theorem. (5.7.4: Volterra integral equation)

Assuming $K(x, y)$ is continuous on $[a, b] \times [a, b]$, the Volterra integral equation $f(x) = \lambda \int_a^x K(x, y)\ f(y)\ dy + \phi(x)$ has a unique solution $f(x)$ for **any** *λ.*

Proof. For $f \in \mathcal{C}([a, b], \mathbb{R})$, we define
$(\Phi(f))(x) = \lambda \int_a^x K(x, y)\ f(y)\ dy + \phi(x)$

1. As in 5.7.4, Φ maps from $\mathcal{C}([a, b], \mathbb{R})$ to $\mathcal{C}([a, b], \mathbb{R})$.
2. Let $M = \sup_{x,y\in[a,b]} |K(x, y)|$. Then,

$$\begin{aligned} |\Phi(f)(x) - \Phi(g)(x)| &= |\lambda| \left| \int_a^x K(x, y)(f(y) - g(y))dy \right| \\ &\leq |\lambda||x - a|M\|f - g\| \end{aligned}$$

Proof of 5.7.4

3. From 2,

$$
\begin{aligned}
|\Phi^2(f)(x) - \Phi^2(g)(x)| &= |\lambda| \left| \int_a^x K(x,y)(\Phi(f)(y) - \Phi(g)(y))dy \right| \\
&\leq |\lambda| \left| \int_a^x M|y-a||\lambda|M\|f-g\|dy \right| \\
&\leq |\lambda|^2 M^2 \frac{|b-a|^2}{2!}\|f-g\|
\end{aligned}
$$

4. Inductively, we have

$$\|\Phi^n(f) - \Phi^n(g)| \leq \frac{|\lambda|^n M^n |b-a|^n}{n!}\|f-g\|$$

5. By the ratio test, $\sum \frac{|\lambda|^n M^n |b-a|^n}{n!}$ converges.
6. Hence, we can choose N so that $\frac{|\lambda|^N M^N |b-a|^N}{N!} < 1$. $\therefore$ Φ^N **is a contraction!**

Proof of 5.7.4

7. From 6, $\exists$ **unique** $f_* \in \mathcal{C}([a,b],\mathbb{R})$ **s.t.** $\Phi^N(f_*) = f_*$**.**
8. From 7, $\Phi^{N+1}(f_*) = \Phi(f_*)$.
9. From 8, $\Phi(f_*)$ is a fixed point of Φ^N.
10. From 7, 9, and **the uniqueness of the fixed point**, it must be $f_* = \Phi(f_*)$.

What a CUTE IDEA is!

Examples

- **Example 5.7.5.** Let $\Phi : \mathbb{R} \to \mathbb{R}$ be defined by $\Phi(x) = x + 1$. $|\Phi(x) - \Phi(y)| = |x - y| \not\leq k|x - y|$ for any $k \in [0, 1)$, and Φ does not have a unique fixed point.
- **Example 5.7.6.** Solve $x'(t) = x(t),\ x(0) = 1.$
  Solution. Let $\Phi(\phi)(t) = 1 + \int_0^t \phi(s)ds$. Let $\phi_0 = 1$ and $\phi_{n+1} = \Phi(\phi_n), n = 0, 1, \cdots$. Then $\phi_n(t) = \sum_{k=0}^{n} \frac{1}{k!} t^k$. Hence, $\phi_n(t) \to e^t$.
- **Example 5.7.7.** Solve $x'(t) = t\ x(t)$ for t near 0 and $x(0) = 3.$
  Solution. Let $\Phi(\phi)(t) = 3 + \int_0^t \phi(s)ds$. Let $\phi_0 = 3$ and $\phi_{n+1} = \Phi(\phi_n), n = 0, 1, \cdots$. Then $\phi_n(t) = 3\sum_{k=0}^{n} \frac{1}{k!} \left(\frac{t^2}{2}\right)^k$. Hence, $\phi_n(t) \to 3e^{t^2/2}$.

Examples

- **Example 5.7.5.** Consider the integral equation

$$f(x) = a + \int_0^x xe^{-xy} f(y)\, dy$$

Check directly on which intervals $[0, r]$ we get a contraction.
Solution. Let $K(x, y) = xe^{-xy}$ and let
$\Phi(f)(x) = a + \int_0^x xe^{-xy} f(y)\, dy$. Then

$$\begin{aligned} \|\Phi(f) - \Phi(g)\| &= \sup_{x\in[0,r]} \left| \int_0^x K(x, y)(f(y) - g(y)\, dy \right| \\ &\leq \sup_{x\in[0,r]} \left| \int_0^x K(x, y)\, dy \right| \|f - g\| \\ &= \sup_{x\in[0,r]} \left| 1 - e^{-x^2} \right| \|f - g\| \end{aligned}$$

Since $0 < 1 - e^{-r^2} < 1$ for any r, Φ is a contraction for any r.

5.8 The Stone-Weierstrass Theorem

Aim of Weierstrass Theorem is to show that any continuous function can be uniformly approximated by a function that has more easily managed properties, such as a polynomial.

Theorem. (5.8.1: Weierstrass-Bernstein)

Let $f \in \mathcal{C}([0,1], \mathbb{R})$. There exist a sequence of polynomial p_n such that $\lim_{n\to\infty} \|p_n - f\| = 0$. In fact,

$$p_n(x) = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} x^k (1-x)^{n-k} f(k/n) \quad \to \quad f \quad \textit{unformly}$$

- **Meaning of $r_k(x) := \frac{n!}{k!(n-k)!} x^k (1-x)^{n-k}$**: Imagine a coin with probability x of getting heads and, consequently, with probability $1-x$ of getting tails. In n tosses, the probability of getting exactly k heads is that quantity.

Rough proof: Weierstrass-Bernstein

- $\sum_{k=0}^{n} r_k(x) = 1$ and $\sum_{k=0}^{n} (k/n - x)^2 r_k(x) = x(1-x)$.

$$\lim_{n\to\infty} \sum_{|\frac{k}{n}-x|>\delta} r_k(x) = 0, \qquad \text{for any } \delta > 0$$

and

$$\lim_{n\to\infty} \sum_{|\frac{k}{n}-x|<\delta} r_k(x) = 1, \qquad \text{for any } \delta > 0$$

- Suppose that in gambling game called *n-tosses*, $f(k/n)$ dollars is paid out when exactly k heads turn up when n tosses are made.m The average amount (after a lo∼∼ong evening of playing *n-tosses*) paid out when n tosses are made is

$$p_n(x) = \sum_{k=0}^{n} r_k(x)\ f(k/n) \quad \approx \quad f(x)$$

The Weierstrass-Bernstein theorem can be applied to $\mathcal{C}([a,b],\mathbb{R})$ because

$$g \in \mathcal{C}([a,b],\mathbb{R}) \quad \Rightarrow \quad f(x) = g(x(b-a)+a) \in \mathcal{C}([a,b],\mathbb{R}).$$

Theorem. (5.8.2: Stone-Wierstrass)

Let M be a metric space, $A \subset M$ a compact set, and $\mathcal{B} \subset \mathcal{C}(A,\mathbb{R})$ satisfies the following:

1. *$\mathcal{B}$ is algebra: $f, g \in \mathcal{B}$ $\&\alpha \in \mathbb{R} \Rightarrow f+g, fg, \ \alpha g \in \mathcal{B}$*
2. $1 \in \mathcal{B}$
3. $\forall x, y \in A, \ x \neq y, \quad \exists\, f \in \mathcal{B} \ \ s.t. \ \ f(x) \neq f(y).$

Then $\mathcal{B}$ is dense in $\mathcal{C}(A,\mathbb{R})$, that is, $\overline{\mathcal{B}} = \mathcal{C}(A,\mathbb{R})$.

The proof is easy (just technical). I just provide a rough insight.

1. Since $\mathcal{B}$ is algebra, $f \in \overline{\mathcal{B}} \Rightarrow p_n(f) \in \overline{\mathcal{B}}$.
2. Assume that A is a finite set. Then the proof is trivial.
3. Use the concept of finite $\delta-$net for the compact set A.

Differentiable Mappings

Definition: Let A be an open set in $\mathbb{R}^n$. A mapping $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ is said to be differentiable at $\mathbf{x}_0 \in A$ if $\exists$ a linear function ($m \times n$ matrix) $Df(\mathbf{x}_0) : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - Df(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)\|}{\|\mathbf{x}-\mathbf{x}_0\|} = 0$$

- **Theorem 6.2.2.** If $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable, then $\frac{\partial f_j}{\partial x_i}$ exist, and

$$Df(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \cdots & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \qquad \text{(called Jacobian matrix)}$$

- **1-Dimension.** If $f : (a,b) \to \mathbb{R}$ is differentiable at $\mathbf{x}_0$, then $\exists$ a number $m = f'(x_0)$ such that

$$\lim_{x\to x_0} \frac{\|f(x) - f(x_0) - m(x-x_0)\|}{\|x-x_0\|} = 0 \quad \text{or} \quad \lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0} = m$$

Thm 6.1.2. If $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{a}$, then f is continuous at $\mathbf{a}$ and $Df(\mathbf{a})$ is uniquely determined.

Proof of uniqueness. Let L_1 and L_2 be two $m \times n$ matrix (or linear mappings) satisfying

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\|f(\mathbf{x}) - f(\mathbf{a}) - L_1(\mathbf{x}-\mathbf{a})\|}{\|\mathbf{x}-\mathbf{a}\|} = 0 = \lim_{\mathbf{x}\to\mathbf{a}} \frac{\|f(\mathbf{x}) - f(\mathbf{a}) - L_2(\mathbf{x}-\mathbf{a})\|}{\|\mathbf{x}-\mathbf{a}\|}$$

It suffices to prove that $\|L_1\mathbf{e}_j - L_2\mathbf{e}_j\| = 0$ for $j = 1, \cdots, n$.

$$\begin{aligned}
\|L_1\mathbf{e}_j - L_2\mathbf{e}_j\| &= \tfrac{1}{|h|}\|L_1(h\mathbf{e}_j) - L_2(h\mathbf{e}_j)\| = \tfrac{\|L_1(h\mathbf{e}_j) - L_2(h\mathbf{e}_j)\|}{\|h\mathbf{e}_j\|} \\
&= \tfrac{\|f(\mathbf{a}+h\mathbf{e}_j) - f(\mathbf{a}) - L_1(h\mathbf{e}_j) - [f(\mathbf{a}+h\mathbf{e}_j) - f(\mathbf{a}) - L_2(h\mathbf{e}_j)]\|}{\|h\mathbf{e}_j\|} \\
&\le \tfrac{\|f(\mathbf{a}+h\mathbf{e}_j) - f(\mathbf{a}) - L_1(h\mathbf{e}_j)\|}{\|h\mathbf{e}_j\|} + \tfrac{\|f(\mathbf{a}+h\mathbf{e}_j) - f(\mathbf{a}) - L_2(h\mathbf{e}_j)\|}{\|h\mathbf{e}_j\|} \\
&\to 0 \text{ as } h \to 0
\end{aligned}$$

- Proof of continuity: Since $\lim_{\mathbf{y}\to\mathbf{a}} \|f(\mathbf{y}) - f(\mathbf{a}) - Df(\mathbf{a})(\mathbf{y}-\mathbf{a})\| = 0$, $\lim_{\mathbf{y}\to\mathbf{a}} \|f(\mathbf{y}) - f(\mathbf{a})\| = 0$.

Thm 6.2.2. Assume $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}$ and $Df(\mathbf{x}) = [a_{ij}]$. Then, $\frac{\partial f_j}{\partial x_i}$ exist and $a_{ij} = \frac{\partial f_j}{\partial x_i}$.

Proof. Denote $\mathbf{e}_1 = (1, 0, \cdots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \cdots, 0)$, $\mathbf{e}_n = (0, \cdots, 0, 1)$. We have

$$\lim_{\mathbf{y}\to\mathbf{x}} \frac{\|f(\mathbf{y}) - f(\mathbf{x}) - Df(\mathbf{x})(\mathbf{y}-\mathbf{x})\|}{\|\mathbf{y}-\mathbf{x}\|} = 0$$

$$\Rightarrow \quad \lim_{h\to 0} \frac{\|f(\mathbf{x}+h\mathbf{e}_i) - f(\mathbf{x}) - Df(\mathbf{x})(h\mathbf{e}_i)\|}{|h|} = 0, \qquad \imath = 1, 2, \cdots, n$$

$$\Rightarrow \quad \lim_{h\to 0} \frac{\sqrt{\sum_{j=1}^{m} |f_j(\mathbf{x}+h\mathbf{e}_i) - f_j(\mathbf{x}) - a_{ij}(h\mathbf{e}_i)|^2}}{|h|} = 0, \quad \jmath = 1, 2, \cdots, n$$

$$\Rightarrow \quad \frac{\partial f_j}{\partial x_i} \text{ exists and } a_{ij} = \frac{\partial f_j}{\partial x_i}$$

Thm 6.4.1. Let $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$. If each $\frac{\partial f_j}{\partial \mathbf{x}_i}$ exist and **continuous** on A, then f is differentiable on A.

[Proof for the case $n = 2, m = 1$.] Let $Df(\mathbf{x}) = \left[\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x})\right]$, $\mathbf{x} \in A$. From the mean value theorem,

$$\begin{aligned} f(\mathbf{y}) - f(\mathbf{x}) &= f(y_1, y_2) - f(x_1, y_2) + f(x_1, y_2) - f(x_1, x_2) \\ &= \frac{\partial f}{\partial x_1}(u_1, y_1)\,(y_1 - x_1) + \frac{\partial f}{\partial x_2}(x_1, u_2)\,(y_2 - x_2) \end{aligned}$$

for some u_i between x_i and y_i. Hence,

$$f(\mathbf{y}) - f(\mathbf{x}) - Df(\mathbf{x})(\mathbf{y} - \mathbf{x}) = \alpha\,(y_1 - x_1) + \beta\,(y_2 - x_2)$$

where $\alpha := \left[\frac{\partial f}{\partial x_1}(u_1, y_2) - \frac{\partial f}{\partial x_1}(\mathbf{x})\right]$ and $\beta := \left[\frac{\partial f}{\partial x_2}(x_1, u_2) - \frac{\partial f}{\partial x_2}(\mathbf{x})\right]$.

Due to **continuity** of $\frac{\partial f}{\partial x_1}$ and $\frac{\partial f}{\partial x_2}$, $\alpha \to 0$ & $\beta \to 0$ as $\mathbf{y} \to \mathbf{x}$ and

$$\frac{|f(\mathbf{y}) - f(\mathbf{x}) - Df(\mathbf{x})(\mathbf{y} - \mathbf{x})|}{\|\mathbf{y} - \mathbf{x}\|} = \frac{\left|\alpha\,(y_1 - x_1) + \beta\,(y_2 - x_2)\right|}{\sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2}} \leq \sqrt{\alpha^2 + \beta^2} \to 0$$

as $\mathbf{y} \to \mathbf{x}$. This proves that $\lim_{\mathbf{y} \to \mathbf{x}} \frac{\|f(\mathbf{y}) - f(\mathbf{x}) - Df(\mathbf{x})(\mathbf{y} - \mathbf{x})\|}{\|\mathbf{y} - \mathbf{x}\|} = 0$.

Remark. About a differentiable map $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$.

- The proof of Thm 6.4.1 for the general case $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ is almost same as the special case $f : A \subset \mathbb{R}^2 \to \mathbb{R}$.

- Intuitively, $x \to f(\mathbf{x}_0) + Df(x_0)(x - x_0)$ is supposed to be the best affine approximation to f near x_0

- It should be noticed that the existence of $\frac{\partial f_j}{\partial x_i}$ does not imply that the derivative Df exist.

Directional Derivatives. Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ be real-valued function.

- Let $\mathbf{e} \in \mathbb{R}^n$ be a unit vector. $\frac{d}{dt} f(\mathbf{x} + t\mathbf{e})|_{t=0} = \lim_{t\to 0} \frac{f(\mathbf{x}+t\mathbf{e})-f(\mathbf{x})}{t}$ is called the directional derivative of f at $\mathbf{x}$ in the direction $\mathbf{e}$.

- If f is differentiable, then $\lim_{t\to 0} \frac{f(\mathbf{x}+te)-f(\mathbf{x})}{t} = Df(\mathbf{x}) \cdot e$.

- Note that the existence of all directional derivatives at a point need not imply differentiability.

  Example. Let $f(x, y) = \frac{xy}{x^2+y}$ for $x^2 \neq -y$ and $f(x, y) = 0$ if $x^2 = -y$. Note that f is not continuous at $(0, 0)$, since $\lim_{t\to 0} f(t, t^3 - t^2) = \lim_{t\to 0} \frac{t(t^3-t^2)}{t^2+t^3-t^2} = -1 \neq 0 = f(0, 0)$. But all directional derivative of f at $(0, 0)$ exist:

  $$\lim_{t\to 0} \frac{f(ta, tb)}{t} = \frac{1}{t} \frac{t^2 ab}{t^2 a^2 + tb} = a$$

  for any unit vector $\mathbf{e} = (a, b)$.

Chain Rule 6.5.1: Let $A \subset \mathbb{R}^n$ be open and let $f : A \to \mathbb{R}^m$ be differentiable. Let $B \subset \mathbb{R}^n$ be open, $f(A) \subset B$, and $g : B \to \mathbb{R}^p$ be differentiable. Then $h = g \circ f$ is differentiable on A and $Dh(\mathbf{x}) = Dg(f(\mathbf{x}))Df(\mathbf{x})$:

$$D(g \circ f)(\mathbf{x}) = \begin{pmatrix} \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_p}{\partial y_1} & \cdots & \frac{\partial g_p}{\partial y_m} \end{pmatrix} \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

Proof. From the assumption, it is easy to see that

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{\| \overbrace{(g \circ f)(\mathbf{x}) - (g \circ f)(\mathbf{x}_0) - Dg(f(\mathbf{x}_0))(f(\mathbf{x}) - f(\mathbf{x}_0))}^{:=\spadesuit} \|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0$$

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{\| \overbrace{f(\mathbf{x}) - f(\mathbf{x}_0) - Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)}^{:=\clubsuit} \|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0$$

Since $(g \circ f)(\mathbf{x}) - (g \circ f)(\mathbf{x}_0) - Dg(f(\mathbf{x}_0))Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = \spadesuit + Dg(f(\mathbf{x}_0))\clubsuit$, it follows from the above identities that

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{\| \overbrace{(g \circ f)(\mathbf{x}) - (g \circ f)(\mathbf{x}_0) - Dg(f(\mathbf{x}_0))Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)}^{= \spadesuit + Dg(f(\mathbf{x}_0))\clubsuit} \|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0.$$

Directional derivatives and examples

1. If $h(r,\theta) = f(r\cos\theta, r\sin\theta)$, then

$$\left(\frac{\partial h}{\partial r} \quad \frac{\partial h}{\partial \theta} \right) = \left(\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y} \right) \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

2. Consider a surface S defined by $f(\mathbf{x})$ =constant. Then $\nabla f(\mathbf{x})$ is orthogonal to this surface.

   Proof. Let $\mathbf{c} : [0,1] \to \mathbb{R}^n$ be a curve lying on S and $c(0) = \mathbf{x}_0$.

$$0 = \frac{d}{dt} f(\mathbf{c}(t)) = \nabla f(\mathbf{c}(t)) \cdot \mathbf{c}'(t).$$

   This means that $\nabla f(\mathbf{c}(t))$ is orthogonal to its tangent vector $\mathbf{c}'(t)$. Since this is true for arbitrary curve on S passing $\mathbf{x}_0$, $\nabla f(\mathbf{x}_0)$ is orthogonal to S at $\mathbf{x}_0$.

3. The direction of greatest rate of increase of $f(\mathbf{x})$ is $\nabla f(\mathbf{x})$.

6.7.1. Mean Value Theorem. Suppose $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable on an open set A. For any $\mathbf{x}, \mathbf{y} \in A$ such that the line segment joining $\mathbf{x}$ and $\mathbf{y}$ lies in A, $\exists c$ on that segment such that

$$f(\mathbf{y}) - f(\mathbf{x}) = Df(\mathbf{c}) \cdot (\mathbf{y} - \mathbf{x})$$

Proof. Define $h(t) = f((1-t)\mathbf{x} + t\mathbf{y})$. Then

$$\exists t_0 \in (0,1) \quad \text{such that} \quad h(1) - h(0) = h'(t_0)$$

and therefore

$$f(\mathbf{y}) - f(\mathbf{x}) = h(1) - h(0) = h'(t_0) = Df(\underbrace{(1-t_0)\mathbf{x} + t_0\mathbf{y}}_{=\mathbf{c}}) \cdot (\mathbf{y} - \mathbf{x})$$

- **Definition.** A bilinear map $B : R^m \times R^n \to R$ is $n \times m$ matrix such that

$$B(\mathbf{x}, \mathbf{y}) = \sum_{ij} a_{ij} x_i y_j = (x_1, \cdots, x_n) \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ v_m \end{pmatrix}$$

- **Definition 6.8.4.** For positive integer r, f is said to be of class C^r if all partial derivatives up to order r exist and continuous.

- Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is of class C^2. Then

$$D^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}$$

- If $D^2 f$ us continuous, $D^2 f$ is symmetric.

Taylor's Theorem 6.8.5.[Case:$f \in C^3$]. Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is of class C^3. Suppose $\mathbf{x} \in A$ and $\mathbf{x} + t\mathbf{h} \in A$ for $0 \le t \le 1$. Then $\exists \mathbf{c} = \mathbf{x} + t_0\mathbf{h}, 0 < t_0 < 1$, such that

$$
\begin{aligned}
f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}) h_i + \frac{1}{2!} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) h_i h_j \\
&\quad + \frac{1}{3!} \sum_{i,j,k=1}^{n} \left(\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(\mathbf{x} + t_0 \mathbf{h}) h_i h_j h_k \right)
\end{aligned}
$$

Proof.

$$
\begin{aligned}
&f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) = \int_0^1 \tfrac{d}{dt} f(\mathbf{x}+t\mathbf{h}) dt = \int_0^1 \sum_{i=1}^n \tfrac{\partial f}{\partial x_i}(\mathbf{x}+t\mathbf{h}) h_i dt \\
&\quad = \sum_{i=1}^n \int_0^1 \tfrac{\partial f}{\partial x_i}(\mathbf{x}+t\mathbf{h}) h_i \tfrac{d(t-1)}{dt} dt \qquad (\text{Why? } \tfrac{d(t-1)}{dt} = 1) \\
&\quad = \sum_{i=1}^n \left[\tfrac{\partial f}{\partial x_i}(\mathbf{x}) h_i - \int_0^1 \tfrac{d}{dt} \left(\tfrac{\partial f}{\partial x_i}(\mathbf{x}+t\mathbf{h}) h_i \right) (t-1)\, dt \right] \\
&\quad = \sum_{i=1}^n \tfrac{\partial f}{\partial x_i}(\mathbf{x}) h_i + R_1(\mathbf{h}, \mathbf{x})
\end{aligned}
$$

where

$$
R_1(\mathbf{h}, \mathbf{x}) = \sum_{i,j=1}^{n} \int_0^1 (1-t) \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}+t\mathbf{h}) h_i h_j \right) dt
$$

Using $\frac{d}{dt}\left(-\frac{(t-1)^2}{2!}\right) = (1-t)$ and integration by part,

$$\begin{aligned} R_1(\mathbf{h},\mathbf{x}) &= \textstyle\sum_{i,j=1}^{n} \int_0^1 \frac{d}{dt}\left(-\frac{(t-1)^2}{2!}\right)\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}+t\mathbf{h})h_i h_j\right)\, dt \\ &= \textstyle\frac{1}{2!}\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})h_i h_j + R_2(\mathbf{h},\mathbf{x}) \end{aligned}$$

where

$$R_2(\mathbf{h},\mathbf{x}) := \sum_{i,j,k=1}^{n} \int_0^1 \frac{(t-1)^2}{2!}\left(\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(\mathbf{x}+t\mathbf{h})h_i h_j h_k\right)\, dt$$

Recall the second mean value theorem for integral

$$\int_0^1 f(t)g(t)dt = g(t_0)\int_0^1 f(t)dt \qquad \text{for some } 0 < t_0 < 1.$$

Hence, $\exists t_0, 0 < t_0 < 1$ such that

$$R_2(\mathbf{h},\mathbf{x}) = \sum_{i,j,k=1}^{n}\left(\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(\mathbf{x}+t_0\mathbf{h})h_i h_j h_k\right)\underbrace{\int_0^1 \frac{(t-1)^2}{2!}dt}_{\frac{1}{3!}}.$$

One can proceed by using induction using the same method to get the general Taylor's theorem.

6.8.5. Taylor's Theorem [General Case:$f \in C^r$]. Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is of class C^r. Suppose $\mathbf{x} \in A$ and $\mathbf{x} + t\mathbf{h} \in A$ for $0 \le t \le 1$. Then

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + Df(\mathbf{x}) \cdot h + \cdots + \frac{1}{r!} D^{r-1} f(\mathbf{x}) \cdot (\mathbf{h}, \cdots, \mathbf{h}) + R_{r-1}(\mathbf{x}, \mathbf{h})$$

where $R_{r-1}(\mathbf{x}, \mathbf{h})$ is the remainder. Furthermore,

$$\frac{R_{r-1}(\mathbf{x}, \mathbf{h})}{\|\mathbf{h}\|^{r-1}} \to 0 \quad \text{as } \mathbf{h} \to 0$$

Another proof of Taylor's formula. Let $g(t) = f(\mathbf{x}+t\mathbf{h})$ for $t \in [0,1]$. Applying one-dimensional Taylor's formula, there exists $\tilde{t} \in (0,1)$ such that

$$g(1) = g(0) + \sum_{k=1}^{r-1} \frac{1}{k!} g^{(k)}(0) + \frac{1}{(r-1)!} g^{(k-1)}(\tilde{t})$$

Note that $R_{r-1}(\mathbf{x}, h) = +\frac{1}{r!} g^{(k-1)}(\tilde{t})$, $g(1) = f(\mathbf{x}+\mathbf{h})$, $g(0) = f(\mathbf{x})$,

$$g'(0) = Df(\mathbf{x}) \cdot \mathbf{h} = \textstyle\sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{x}) h_i$$
$$g''(0) = D^2 f(\mathbf{x}) \cdot (\mathbf{h}, \mathbf{h}) = \textstyle\sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) h_i h_j$$
$$g'''(0) = D^3 f(\mathbf{x})(\mathbf{h}, \mathbf{h}, \mathbf{h}) = \textstyle\sum_{i,j,k=1}^n \left(\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(\mathbf{x}) h_i h_j h_k \right)$$

Theorem 6.9.2. If $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable and $\mathbf{x}_0 \in A$ is an extreme point for f, then $Df(\mathbf{x}_0) = 0$.

Proof. Assume $Df(\mathbf{x}_0) \neq 0$. We try to prove that $f(\mathbf{x}_0)$ is not a local extreme value.

- Let $\mathbf{h} = \frac{Df(\mathbf{x}_0)}{\|Df(\mathbf{x}_0)\|}$. Since f is differentiable at $\mathbf{x}_0$,

$$\lim_{\lambda \to 0} \frac{1}{|\lambda|} \left| f(\mathbf{x}_0 + \lambda \mathbf{h}) - f(\mathbf{x}_0) - Df(\mathbf{x}_0) \cdot (\lambda \mathbf{h}) \right| = 0.$$

- Hence, (for given $\epsilon = \frac{\|Df(\mathbf{x}_0)\|}{2}$) there exist $\delta > 0$ such that

$$0 < |\lambda| < \delta \Rightarrow \left| f(\mathbf{x}_0 + \lambda \mathbf{h}) - f(\mathbf{x}_0) - Df(\mathbf{x}_0) \cdot (\lambda \mathbf{h}) \right| < \frac{\|Df(\mathbf{x}_0)\|}{2} |\lambda|$$

Since $Df(\mathbf{x}_0) \cdot \mathbf{h} = \|Df(\mathbf{x}_0)\|$, we have

$$-\frac{\|Df(\mathbf{x}_0)\|}{2} |\lambda| < f(\mathbf{x}_0 + \lambda \mathbf{h}) - f(\mathbf{x}_0) - \|Df(\mathbf{x}_0)\| \lambda < \frac{\|Df(\mathbf{x}_0)\|}{2} |\lambda|$$

This leads to the followings:

– for $0 < \lambda < \delta$, $\frac{\|Df(\mathbf{x}_0)\|}{2}\lambda < f(\mathbf{x}_0 + \lambda\mathbf{h}) - f(\mathbf{x}_0)$. Hence, $f(\mathbf{x}_0)$ is not local maximum.

– for $-\delta < \lambda < 0$, $f(\mathbf{x}_0 + \lambda\mathbf{h}) - f(\mathbf{x}_0) < \frac{\|Df(\mathbf{x}_0)\|}{2}\lambda$. Hence, $f(\mathbf{x}_0)$ is not local minimum.

Theorem 6.9.4. Suppose $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is a C^3-function and $\mathbf{x}_0$ is a critical point.

- If f has a local maximum at $\mathbf{x}_0$, then $H_{\mathbf{x}_0}(f)$ is negative semi-definite.
- If $H_{\mathbf{x}_0}(f)$ is negative (positive) definite, then f has a local maximum (minimum) at x_0

Indeed, this theorem holds true for $f \in C^2$.

Proof. Since $Df(\mathbf{x}_0) = 0$, Taylor's theorem gives

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}) = \frac{1}{2} D^2 f(\mathbf{x}_0)(\mathbf{h}, \mathbf{h}) + R_2(\mathbf{x}_0, \mathbf{h})$$

where $\lim_{\mathbf{h}\to 0} \frac{R_2(\mathbf{x}_0,\mathbf{h})}{\|\mathbf{h}\|} = 0$.

If $D^2 f(\mathbf{x}_0)$ is negative definite, then

$$\frac{1}{2} D^2 f(\mathbf{x}_0)(\mathbf{h}, \mathbf{h}) + R_2(\mathbf{x}_0, \mathbf{h}) < 0 \qquad \text{for sufficiently small } \mathbf{h}$$

and therefore $f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) < 0$ for sufficiently small $\mathbf{h}$. Hence, $f(\mathbf{x}_0)$ has a local maximum at $\mathbf{x}_0$.

- **Example 6.9.5.** The matrix $\mathbb{A} = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ is positive definite if

$$(x, y)\begin{pmatrix} a & b \\ b & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} > 0 \quad \text{if } (x, y) \neq (0, 0).$$

Hence, $\mathbb{A}$ is positive definite iff $ax^2 + 2bxy + dy^2 > 0$ for all all x, y. Therefore, $\mathbb{A}$ is positive definite iff $a > 0$ and $ad - b^2 > 0$.

- **Example 6.9.6.** Let $f(x, y) = x^2 - xy + y^2$. Then $Df(0, 0) = (0, 0)$ and $D^2f(0, 0) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$. Hence, the Hessian is positive definite. Thus f has a local minimum at $(0, 0)$.

Chapter 8. Integration

Definition. Let $A \subset \mathbb{R}^2$ be a bounded set and let $f : A \to \mathbb{R}$ be a bounded function.

- We enclose A in some rectangle $B = [a_1, b_1] \times [a_2, b_2]$ and extend f to the whole rectangle by defining it to be zero outside of A.

- Let $\mathcal{P}$ be a partition of B obtained by dividing $a_1 = x_0 < x_1 < \cdots < x_n = b_1$ and $a_2 = y_0 < y_1 < \cdots < y_m = b_2$:

$$\mathcal{P} = \{\underbrace{[x_i, x_{i+1}] \times [y_j, y_{j+1}]}_{=\text{subrectangle } R} : i = 0, 1, \cdots, n-1, \ j = 0, 1, \cdots, m-1\}.$$

- Define the upper sum of f:

$$U(f, \mathcal{P}) := \sum_{R \in \mathcal{P}} \sup\{f(x, y) \mid (x, y) \in R\} \times (\text{volume of } R)$$

- Define the lower sum of f:

$$L(f, \mathcal{P}) := \sum_{R \in \mathcal{P}} \inf\{f(x, y) \mid (x, y) \in R\} \times (\text{volume of } R)$$

- Define the upper integral of f on A by

$$\overline{\int_A} f = \inf \{ L(f, \mathcal{P}) \ : \ \mathcal{P} \text{ is a partition of } B \}$$

and the lower integral of f on A by

$$\underline{\int_A} f = \sup \{ L(f, \mathcal{P}) \ : \ \mathcal{P} \text{ is a partition of } B \}$$

- We say that f is **Riemann integrable or integrable** if

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

- If f is **integrable** on A, we denote

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

Volume and sets of measure zero.

Definition. Let A be a bounded set of $\mathbb{R}^n$.

- The characteristic function 1_A of A is the map defined by $1_A(\mathbf{x}) = 1$ if $\mathbf{x} \in A$ and $1_A(\mathbf{x}) = 1$ if $\mathbf{x} \notin A$.

- We say that A has volume if 1_A is Riemann integrable and the volume is the number

$$vol(A) = \int_A 1_A(\mathbf{x}) d\mathbf{x}.$$

- The set A is said to have measure zero if for every $\epsilon > 0$ there is a countable number of rectangles $R_1, R_2, \cdots$ such that

$$A \subset \cup_{n=1}^{\infty} R_n \quad \& \quad \sum_{n=1}^{\infty} vol(R_n) < \epsilon.$$

- **Examples:** The set of rational number has measured zero in $\mathbb{R}$. As a subset of $\mathbb{R}^2$, the real line has measure zero.

- **Lebesgue's monotone convergence theorem.** Let $g_n : [0,1] \to \mathbb{R}$ be integrable functions and $\left| \int_0^1 g_n(x) dx \right| < \infty$. Suppose that $0 \le g_{n+1} \le g_n$ and $g_n(x) \to 0$ for all $x \in [0,1]$. Then

$$\lim_{n\to\infty} \int_0^1 g_n(x) dx = 0.$$

- **Example:** $\lim_{n\to\infty} \int_0^1 e^{-nx^2} x^p dx = 0$ if $p > -1$.

- **Fubini's Theorem.** Let $A = [a,b] \times [c,d] \subset \mathbb{R}^2$, and let $f : A \to \mathbb{R}$ be continuous. Then

$$\int_A f = \int_a^b \left(\int_c^d f(x,y) dy \right) dx = \int_c^d \left(\int_a^b f(x,y) dx \right) dy$$

Chapter 10 Fourier Series. Fourier analysis arouse historically in connection with problems in mechanic such as heat conduction and wave motion.

- **Vibrating string.** Consider a string of length l with clamped ends that is free to vibrate when plunked. Let $y(t,x)$ is the displacement of the string at time t and $x \in [0,l]$.
  - y obeys the wave equation

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2}$$

    (Force=mass$\times$ acceleration = tension)
  - That the string has clamped ends entails that $y(t,0) = y(t,l) = 0$.

- It is both important and remarkable that any solution $y(x,t)$ can be decomposed into harmonics:

$$y(x,t) = \sum_{n=1}^{\infty} c_n \underbrace{y_n(x,t)}_{\text{standing wave}} = \sum_{n=1}^{\infty} c_n \sin\left(\frac{n\pi}{l}x\right)\cos(\omega_n t), \quad \underbrace{\omega_n = \frac{n\pi c}{l}}_{\text{frequency}}$$

- Physically, a standing wave is a synchronous up-and-down motion that repeats its shape periodically after time $\frac{2\pi}{\omega}$, such as occurs when a string produces a pure note.

- Specific standing waves called fundamental solutions (a kind of basis) are given by

$$y_n(x,t) = \sin\left(\frac{n\pi}{l}x\right)\cos(\omega_n t), \quad n = 0, 1, 2, \cdots$$

- Thus a complicated-looking vibration is in reality **an infinite linear combination of harmonics**.

- The purpose of Fourier analysis is to carry out this procedure of decomposition using general method.

Exercise: Using separable variable, prove that any solution $y(x,t)$ can be decomposed into harmonics

$$y(x,t) = \sum_{n=1}^{\infty} c_n y_n(x,t) = \sum_{n=1}^{\infty} c_n \sin\left(\frac{n\pi}{l}x\right)\cos(\omega_n t), \quad \omega_n = \frac{n\pi c}{l}$$

10.1 Review: Inner Product in $\mathbb{R}^n$.

- For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, define inner product and norm:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{j=1}^{n} x(j)y(j), \quad \|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}.$$

- The distance (or metric) between $\mathbf{x}$ and $\mathbf{y}$ is defined by $\|\mathbf{x} - \mathbf{y}\|$, and hence $\|\mathbf{x} - \mathbf{y}\| = 0$ implies $\mathbf{x} = \mathbf{y}$.

- If $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, $\mathbf{x}$ and $\mathbf{y}$ are said to be orthogonal.

- $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ is said to be an orthonormal basis of $\mathbb{R}^n$ if

  1. $\mathbb{R}^n = \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$
  2. $\|\mathbf{e}_j\| = 1,\ j = 1, \cdots, n$
  3. $\langle \mathbf{e}_j, \mathbf{e}_i \rangle = 0$ if $i \neq j$.

- For example, $\mathbf{e}_1 = (1, 0, \cdots, 0), \mathbf{e}_2 = (0, 1, 0, \cdots, 0), \ldots.$

- If $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ is an orthonormal basis, then every $x \in \mathbb{R}^n$ can be represented uniquely by

$$\mathbf{x} = \sum_{j=1}^{n} \langle \mathbf{x}, \mathbf{e}_j \rangle \mathbf{e}_j$$

- If $V_m = span\{\mathbf{e}_1, \cdots, \mathbf{e}_m\}$, the element in V_m closest to $\mathbf{x}$ is

$$\mathbf{x}_m = \sum_{j=1}^{m} \langle \mathbf{x}, \mathbf{e}_j \rangle \mathbf{e}_j$$

  with the distance $\|\mathbf{x} - \mathbf{x}_m\| = \sqrt{\sum_{j=m}^{n} \langle \mathbf{x}, \mathbf{e}_j \rangle^2}$.

This useful dot product properties in Euclidean space can be generalized to infinite dimensional spaces by introducing Hilbert space.

10.1 Inner Product space $C[0, 2\pi]$

- Let A be the interval $(0, 2\pi)$.
- Let V be the space of all continuous functions $f : [0, 2\pi] \to \mathbb{C}$.
- For $f, g \in V$, we define **the inner product**

$$\langle f, g \rangle = \int_0^{2\pi} f(x)\overline{g(x)}\, dx$$

where $\overline{g(x)}$ denotes the complex conjugate of $g(x)$. The above inner product can be approximated by

$$\langle f, g \rangle \approx \sum_{j=1}^{n} f(x_j)\bar{g}(x_j)\, \Delta x.$$

where we divide the interval $[0, 2\pi]$ into n subintervals with endpoints $x_0 = 0 < x_1 < \cdots < x_n = 2\pi$ and equal width $\Delta x = \frac{2\pi}{n}$.

- Two functions f and g are said to be **orthogonal** if

$$\langle f, g \rangle = \int_0^{2\pi} f(x)\overline{g(x)} dx = 0.$$

- **The norm of f** is defined as

$$\|f\| = \sqrt{\langle f, f \rangle} = \sqrt{\int_0^{2\pi} |f(x)|^2 dx}.$$

- **The distance between f and g** is defined by

$$d(f, g) = \|f - g\|.$$

- If $\{\phi_n\}$ is an orthogonal set of functions on the interval A with the property that $\|\phi_n\| = 1$, then we call $\{\phi_n\}$ as **an orthonormal set**.

- **Example.**

$$\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}} \cos x, \frac{1}{\sqrt{\pi}} \sin x, \frac{1}{\sqrt{\pi}} \cos 2x, \frac{1}{\sqrt{\pi}} \sin 2x, \cdots\}$$

is an orthonormal set in V.

10.1 Inner Product space

Definition. Let V be a complex vector space V. An **inner product** on V is a mapping $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ with the following properties :

1. $\langle \alpha f + \beta g\, h\rangle = \alpha\langle f, h\rangle + \beta\langle g, h\rangle$ for all $f, g, h \in V$ and $\alpha, \beta \in \mathbb{C}$.

2. $\langle f, g\rangle = \overline{\langle g, f\rangle}$

3. $\langle f, f\rangle \geq 0$, and $\langle f, f\rangle = 0 \;\Rightarrow\; f = 0$

Theorem 10.1.2. The space V of the continuous functions $f : [a, b] \to \mathbb{C}$ forms an inner product space if we define

$$\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}dx.$$

10.1 Inner Product space $V = C[a,b]$ Consider the space V of the continuous functions $f : [a,b] \to \mathbb{C}$ with the inner product $\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}dx$.

- Define the norm of f by $\|f\| = \sqrt{\langle f, f \rangle}$.
- Define the distance between f and g by $d(f,g) = \|f - g\|$.

For $f, g, h \in V$, we have

- **Cauchy-Schwarz inequality.** $|\langle f, g \rangle| \leq \|f\|\|g\|$
- **Minkowski inequality.** $\|f + g\| \leq \|f\| + \|g\|$
- **Parallelogram law.** $\|f + g\|^2 + \|f - g\|^2 = 2\|f\|^2 + 2\|g\|^2$
- **Pythagorean Theorem.**

$$\text{If } \langle f, g \rangle = 0, \quad \text{then} \quad \|f + g\|^2 = \|f\|^2 + \|g\|^2$$

Cauchy-Schwarz inequality. $|\langle\, f, g\,\rangle| \leq \|f\|\|g\|$

Proof:

- Suppose $g \neq 0$. Let $h = \frac{g}{\|g\|}$. Then

$$|\langle\, f, g\,\rangle| \leq \|f\|\|g\| \quad \Leftrightarrow \quad |\langle\, f, h\,\rangle| \leq \|f\|$$

- Denote $\alpha = \langle\, f, h\,\rangle$. Then

$$\begin{aligned} 0 &\leq \|f - \alpha h\|^2 = \langle f - \alpha h, f - \alpha h \rangle \\ &= \|f\|^2 - \alpha \langle h, f \rangle - \bar{\alpha} \langle f, h \rangle + |\alpha|^2 \\ &= \|f\|^2 - |\alpha|^2 \end{aligned}$$

  Hence, $|\alpha| = |\langle\, f, h\,\rangle| \leq \|f\|$. This completes the proof.

Minkowski inequality. $\|f+g\| \leq \|f\| + \|g\|$

Proof:

$$\begin{aligned}
\|f+g\|^2 &= \langle f+g, f+g\rangle = \|f\|^2 + \langle f,g\rangle + \langle g,f\rangle + \|g\|^2 \\
&\leq \|f\|^2 + 2\|f\|\|g\| + \|g\|^2 \\
&= (\|f\| + \|g\|)^2
\end{aligned}$$

Definition of convergence in an inner product space V. Let V be an inner product space and let f_n be a sequence in V. We say that f_n converges to f (in mean) and write $f_n \to f$ if $\|f_n - f\| = 0$, that is,

$$\forall \epsilon > 0,\ \exists N \quad s.t. \quad n \geq N \Rightarrow \|f_n - f\| < \epsilon.$$

Similarly, a series $\sum_{k=1}^{n} g_k$ converges to f if

$$\lim_{n\to\infty} \|\sum_{k=1}^{n} g_k - f\| = 0.$$

Examples: Let $V = C([0,1])$, the space of continuous functions $f : [0,1] \to \mathbb{C}$.

- Let $f_n = nx\chi_{[0,\frac{1}{n}]} + (2 - nx)\chi_{(\frac{1}{n},\frac{2}{n}]}$. Then $f_n \to 0$ in mean, that is, $\int_0^1 |f_n(x) - 0|^2 dx \to 0$.
- Let $f_n = n^2x\chi_{[0,\frac{1}{n}]} + (2n - n^2x)\chi_{(\frac{1}{n},\frac{2}{n}]}$. Then

$$\lim_{n\to\infty} f_n(x) = 0 \ (\forall x \in \mathbb{R}) \quad \& \quad \lim_{n\to\infty} \int_0^1 |f_n(x) - 0|^2 dx = \infty.$$

Definition of Cauchy sequence. A sequence f_n in an inner product space is said to be a **Cauchy sequence** when

$$\forall \epsilon > 0, \exists N \quad s.t. \quad n, m \geq N \Rightarrow \|f_n - f_m\| < \epsilon.$$

An inner product space is called **complete** if every Cauchy sequence in V converges. A complete inner product space is called a Hilbert space.

Remark: The inner product space $V = C([0,2])$ is not complete.

- Let $f_n(x) = x^n$ for $0 \leq x \leq 1$ and $f_n(x) = 1$ for $1 \leq x \leq 2$.
- Then f_n is Cauchy sequence since $\|f_n - f_m\|^2 = \int_0^1 |x^n - x^m|^2 dx \to 0$ as $n, m \to \infty$.
- However, $f_n \to f$ where $f(x) = 0$ for $0 \leq x \leq 1$ and $f(x) = 1$ for $1 \leq x \leq 2$. $f \notin V$.

A complete inner product space. To make the inner product space $V = C([a, b])$ complete, we need the following theorem and measure theory:

Theorem 8.3.4 If $g(x)$ is integrable, $g \geq 0$, and $\int_a^b g(x)dx = 0$, then the set $\{x \in [a, b] \ : \ g(x) \neq 0\}$ has measure zero.

Proof. TA

♣ For any integrable function f, theorem 8.3 leads to

$$\int_a^b |f(x)|^2 dx = 0 \Rightarrow f = 0 \text{ except for those } x \text{ in a set of measure zero.}$$

Regarding such a f as equivalent to zero, we have the following theorem:

Theorem 10.1.6 Let $V = L^2([a, b])$ be the space of functions $f : [a, b] \to \mathbb{C}$ that $|f|^2$ is integrable. Then V is an inner product space with inner product $\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}dx$ and norm $\|f\| = \sqrt{\langle f, f \rangle}$.

Proof of Theorem 8.3.4: *If $g(x)$ is integrable, $g \geq 0$, and $\int_a^b g(x)dx = 0$, then the set $\{x \in [a,b] \ : \ g(x) \neq 0\}$ has measure zero.*

- We first show that a set $A_m = \{x \in A \ : \ g(x) > 1/m\}$ has measure zero.

- Recall $\int_a^b g(x) = \inf\{U(f,P) \ : \ P \text{ is any partition}\}$.

- Let $\epsilon > 0$ be given. There exist a partition P such that $U(g,P) < \frac{\epsilon}{m}$.

- Let $I_1, \cdots, I_k$ be the subintervals of the partition P such that $I_i \cap A_m \neq \emptyset$. Then

$$\sum_{i=1}^{k} |I_i| \leq \sum_{i=1}^{k} \left(m \sup_{I_i} g(x) |I_i| \right) \leq m\, U(g,P) < \epsilon.$$

  where $|I_i|$ is the length of the interval I_i.

- Since $A_m \subset \cup_{i=1}^k I_i$ and $\sum_{i=1}^k |I_i| < \epsilon$, A_m has measure zero.

- Since $\{x \in [a,b] \ : \ g(x) \neq 0\} \subset \cup_{m=1}^{\infty} A_m$, the set has measure zero.

Proof of Theorem 10.1.6: Prove that $V = L^2([a,b])$ is an inner product space.

- If $\|f\| = 0$, $\int_a^b |f(x)|^2 dx = 0$. From Theorem 8.3.4, $f = 0$ since we are identifying functions that agree except on a set of measure zero.

- It is easy to see that $\langle f, g \rangle$ satisfies all the other rules of inner product space. We only need to prove that $|\langle f, g \rangle| < \infty$ for all $f, g \in V$.

- If we split f and g into real and imaginary part, and into positive and negative part, we are reduce to the case in which f and g are real and positive.

- From Lebesgue monotone convergence theorem (page 467), it suffices to show that

$$\lim_{M\to\infty} \int_a^b (fg)_M < \infty \qquad \text{(see page 462)}$$

- Note that $0 \leq (fg)_M \leq f_{\sqrt{M}} g_{\sqrt{M}} + f^2_{\sqrt{M}} + g^2_{\sqrt{M}}$.

- $\int_a^b (fg)_M \leq \|f_{\sqrt{M}}\| \|g_{\sqrt{M}}\| + \|f_{\sqrt{M}}\|^2 + \|g_{\sqrt{M}}\|^2$.

- Hence, $\int_a^b (fg)_M \leq \|f\| \|g\| + \|f\|^2 + \|g\|^2 < \infty$.

Example 10.1.8. If $f_1, \cdots, f_n$ are orthonormal in an inner product space V, prove that $f_1, \cdots, f_n$ are linearly independent.

- **Definition.** $f_1, \cdots, f_n$ are said to be linearly independent if

$$\sum_{i=1}^{n} c_i f_i = 0 \quad \Rightarrow \quad c_1 = \cdots = c_n = 0.$$

- Assume that $\sum_{i=1}^{n} c_i f_i = 0$. We want to prove $c_1 = \cdots = c_n = 0$.

- Due to orthogonality, we have

$$c_k = c_k \|f_k\|^2 = \left\langle \sum_{i=1}^{n} c_i f_i, f_k \right\rangle = \langle 0, f_k \rangle = 0.$$

Example 10.1.8. Let V be an inner product space. Define the **project of f on g** to be the vector

$$h = \frac{\langle f, g \rangle}{\|g\|^2} g$$

Show that h and $f - h$ are orthogonal, and interpret this result geometrically.

Proof: First, let us prove it when $\|g\| = 1$:

$$\langle h, f - h \rangle = \langle h, f \rangle - \|h\|^2 = \langle \langle f, g \rangle g, f \rangle - |\langle f, g \rangle|^2 = 0.$$

For the general case, repeat the above procedure.

10.2 Orthogonal family of functions

- Throughout this section, we assume that V is an inner product space with an inner product $\langle \cdot, \cdot \rangle$.

- A vector $\phi \in V$ is called normalized if $\|\phi\| = \sqrt{\langle \phi, \phi \rangle} = 1$.

- f and g are called orthogonal if $\langle f, g \rangle = 0$.

- **Definition.** An orthonormal family $\phi_0, \phi_1, \cdots$ in V is called **complete** if every $f \in V$ can be written

$$f = \sum_{k=0}^{\infty} c_k \phi_k \qquad (c_k = \langle f, \phi_k \rangle)$$

  We call $f = \sum_{k=0}^{\infty} c_k \phi_k$ the **Fourier series of f** with respect to $\phi_0, \phi_1, \cdots$ and $c_k = \langle f, \phi_k \rangle$ the **Fourier coefficients.**

- An orthonormal family $\{\phi_k\}$ in V is **complete** iff for every $f \in V$,

$$\lim_{n \to \infty} \|f - \sum_{k=0}^{n} \langle f, \phi_k \rangle \phi_k\| = 0.$$

Theorem 10.2.1: Suppose $f = \sum_{k=0}^{\infty} c_k \phi_k$ for an orthonormal family $\phi_0, \phi_1, \cdots$ in V (convergence in mean). Then $c_k = \langle f, \phi_k \rangle = \overline{\langle f, \phi_k \rangle}$.

Proof.

- Set $s_n = \sum_{k=0}^{n} c_k \phi_k$, so that $\|s_n - f\| \to 0$.
- Hence, $| \langle f - s_n, \phi_i \rangle | \leq \|f - s_n\| \to 0$ as $n \to \infty$.
- If $n \geq i$, then $\langle s_n, \phi \rangle = \sum_{k=0}^{n} \langle c_k \phi_k, \phi_i \rangle = c_i$.
- If $n \geq i$, $|\langle f - s_n, \phi_i \rangle| = |\langle f, \phi_i \rangle - c_i| \leq \|f - s_n\| \to 0$ as $n \to \infty$.
- Hence, $\langle f, \phi_i \rangle = c_i$.

Examples of complete orthonormal families :

- Let $V = L^2([0, 2\pi])$ be the inner product space in Theorem 10.1.6.

- The exponential system $\{\phi_n(x) = \frac{e^{inx}}{\sqrt{2\pi}} : \ n = 0, \pm 1, \pm 2\}$ is a complete orthonormal system in the space V, that is, Fourier series for $f \in V$ for this family is given by

$$f = \sum_{k=-\infty}^{\infty} \frac{c_k e^{ikx}}{\sqrt{2\pi}}, \qquad c_k = \langle f, \phi_k \rangle = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x) e^{-ikx} dx.$$

- The trigonometric system $\frac{1}{\sqrt{2\pi}}, \frac{\cos mx}{\sqrt{2\pi}}, \frac{\sin nx}{\sqrt{2\pi}}, m, n = 1, 2, \cdots$ is complete orthonormal system in V.

Proof. See Mean completeness theorem 10.3.1. (optional)

Gram-Schmidt process :

- Let $g_0, g_1, g_2, \cdots$ be an linearly independent functions in an inner product space V.

- We can form a corresponding orthonormal system $\phi_0, \phi_1, \cdots$ as follows

$$\phi_0 = \frac{g_0}{\|g_0\|}$$

$$\phi_1 = \frac{\tilde{\phi}_1}{\|\tilde{\phi}_1\|} \qquad \tilde{\phi}_1 = g_1 - \langle g_1, \phi_0 \rangle \phi_0$$

$$\phi_{k+1} = \frac{\tilde{\phi}_k}{\|\tilde{\phi}_k\|} \qquad \tilde{\phi}_k = g_k - \sum_{i=0}^{k} \langle g_1, \phi_i \rangle \phi_i$$

Theorem: Bessel inequality: *Let $\phi_0, \phi_1, \cdots$ be an orthonomal system in an inner product space V. For each $f \in V$, the real series $\sum_{i=0}^{\infty} |\langle f, \phi_i \rangle|^2$ converges and*

$$\sum_{i=0}^{\infty} |\langle f, \phi_i \rangle|^2 \leq \|f\|^2.$$

Proof.

- Set $s_n = \sum_{k=0}^{n} c_k \phi_k$ where $c_k = \langle f, \phi_k \rangle$.
- **Key idea 1:** $f - s_n$ and s_n are orthogonal.
- **Key idea 2:** Apply Pythagoras' theorem: $\|f\|^2 = \|f - s_n\|^2 + \|s_n\|^2$.
- Hence, $\|s_n\|^2 \leq \|f\|^2$.
- Since ϕ_i are orthogonal, $\|s_n\|^2 = \sum_{i=0}^{n} |\langle f, \phi_i \rangle|^2$.

Parseval's Theorem : Let $\phi_0, \phi_1, \cdots$ be an orthonomal system in an inner product space V. Then $\phi_0, \phi_1, \cdots$ is complete iff for every $f \in V$, we have

$$\sum_{i=0}^{\infty} |\langle f, \phi_i \rangle|^2 = \|f\|^2.$$

Proof.

- Set $s_n = \sum_{k=0}^{n} c_k \phi_k$ where $c_k = \langle f, \phi_k \rangle$.

- Then $\|f\|^2 = \|f - s_n\|^2 + \|s_n\|^2$.

- If $\phi_0, \phi_1, \cdots$ is complete, $\|f - s_n\|^2 \to 0$. Therefore, letting $n \to \infty$,

$$\|f\|^2 = \lim_{n \to \infty} \left\{ \|f - s_n\|^2 + \|s_n\|^2 \right\} = 0 + \sum_{i=0}^{\infty} |\langle f, \phi_i \rangle|^2$$

- Conversely, if $\sum_{i=0}^{\infty} |\langle f, \phi_i \rangle|^2 = \|f\|^2$, then $\|f\|^2 - \|s_n\|^2 \to 0$, and so $\|f - s_n\|^2 \to 0$.