Linear algebra

Jason Mak

May 13, 2025

Contents

1 Basics structures

1.1 Preliminary results in metric space

We skip the definition and basic properties of metric space and direct to some definition that will be used later.

Definition 1.1 (Bounded and totally bounded)**.** Given a metric space (M, d), a subset $S \subseteq M$ is bounded if

$$\exists d \in \mathbb{R}^+, x \in M,\ S \subseteq B(0, d)\ (B \text{ denote the open ball in } M)$$

S is totally bounded, with a stricter condition, if

$$\forall \epsilon \in \mathbb{R}^+,\ \exists \{B(x_i, \epsilon)\}_{i=1}^n \text{ such that } S \subseteq \bigcup_{i=1}^n B(x_i, \epsilon)$$

Definition 1.2 (Complete metric space)**.** A metric space (M, d) is (Cauchy) complete if for any Cauchy series $\{a_i \in M\}_{i=1}^n$, defined by

$$\forall \epsilon > 0,\ \exists i_0 \in \mathbb{N}^* \text{ such that } \forall i, j > i_0,\ d(a_i, a_j) < \epsilon$$

there exists $L \in M$ such that $\lim_{i\to\infty} a_i = L$. Such choice of L is unique if exists.

Definition 1.3 (Sequentially compactness)**.** Given a metric space (M, d), it is sequentially compact if any sequence $\{a_n \in M\}_{n=1}^\infty$ has a subsequence converging to a point in M.

The definition can be extended to general topology space, using the definition of convergence in topology.

Proposition 1.1 (Basic properties of sequentially compactness)**.** If (M, d) is a metric space and a subset $S \subseteq M$ (also a metric space with d) is sequentially compact, then

1. S is closed and bounded. The converse does't hold in general but in special cases (see example).
2. S is complete and totally bounded, and the converse holds.

Proof. First, compactness immediately implies closeness and completeness by the definition. Because a convergent (or Cauchy) sequence with a convergent subsequence converges to the same point.
Further, compactness implies totally boundedness (and thus boundedness). If not, pick the ϵ_0 such that the totally boundedness condition fails. Pick an arbitrary $a_1 \in S$. Then recursively pick $a_{n+1} \in S \setminus \bigcup_{i=1}^n B(a_i, \epsilon_0)$, given $\{a_i \in S\}_{i=1}^n$. The resulting sequence satisfies $d(a_i, a_j) \geq \epsilon_0\ \forall i, j \in \mathbb{N}^*$, it doesn't have a convergent subsequence in S.
Finally, we prove that completeness and total boundedness together imply compactness. Let $\{a_n \in S\}_{n=1}^\infty$ be a sequence, we construct its Cauchy subsequence, thus convergent in S. For $n \in \mathbb{N}^*$, let A_n be a finite subset of S such that $S \subseteq \bigcup_{x \in A_n} B(x, 2^{-n})$. Because of the finiteness of A_1, there exists $x_1 \in A_1$ such that the set

$$I_1 := \{i \in \mathbb{N}^* : a_i \in B(x_1, 2^{-1})\}$$

is infinite. Then recursively given a infinite set $I_n \subseteq \mathbb{N}^*$, pick $x_{n+1} \in A_{n+1}$ such that

$$I_{n+1} := \{i \in I_n : a_i \in B(x_{n+1}, 2^{-(n+1)})\}$$

is infinite. Let $\{a_{n_i}\}_{i=1}^\infty$ be the subsequence such that $n_i \in I_i\ \forall i \in \mathbb{N}^*$. Thus $\forall I \in \mathbb{N}^*$, $d(a_{n_i}, a_{n_j}) < 2^{1-I}\ \forall i, j \geq I$. The subsequence is therefore Cauchy and convergent in S. □

Example
When $M \subseteq \mathbb{R}^n$ for $n \in \mathbb{N}^*$ and d is the Euclidean distance, M is sequentially compact if and only if it is closed and bounded. It also applies to complex vector space with finite dimension.
It also holds for complex Euclidean space $\mathbb{C}^n$.

Proof. The "$\Leftarrow$" direction is stated by the Bolzano–Weierstrass theorem. Considering $n = 1$, we can always construct an either monotonic increasing or decreasing subsequence from $\{a_n \in M\}_{n=1}^{\infty}$. When $n \geq 0$, we can construct subsequence recursively to ensure monotonicity on every term. The final resulting subsequence converges and the limit is in M because M is closed and bounded. □

Theorem 1.1 (Equivalence of compact and sequentially compact in metric space)**.** Let S be a subset in a metric space (M, d). S is compact if and only if it is sequentially compact.
Recall the definition of compactness, i.e., for any open cover $\{U_i\}_{i\in I}$ of S, there exists a finite subset $J \subseteq I$ such that $S \subseteq \bigcup_{j\in J} U_j$.

Proof. First we prove the "$\Rightarrow$" direction by contradiction. If a sequence $\{a_n \in S\}_{n=1}^{\infty}$ satisfies $\mathrm{LIM}(a_n) \cap S = \emptyset$, then for any $x \in S$, there exists $\delta_x > 0$ such that $B(x, \delta_x)$ only contain finite terms of $\{a_n \in S\}_{n=1}^{\infty}$. Since $\{B(x, \delta_x)\}_{x\in S}$ is a open cover of S, take $\{B(x_j, \delta_{x_j})\}_{j=1}^{m}$ that covering S. The contradiction is that $\bigcup_{j=1}^{m} B(x_j, \delta_{x_j}) \supset S$ contains at most finite terms of the sequence.
Next, we prove the other direction. Let $\{U_i\}_{i\in I}$ be a open cover of S. We claim that there exists δ_0 such that $\forall x \in S, \exists i \in I$ such that $B(x, \delta_0) \subseteq U_i$.
Suppose that the claim does not hold. Recursively define the sequence $\{a_n \in S\}_{n=1}^{\infty}$ such that $B(a_n, \frac{1}{n})$ does not contain in any U_i. Let $\alpha \in S$ be the limit of one of its convergent subsequence. Suppose $p \in U_{i_0}$. We can show that there exists a large enough n such that $B(a_n, \frac{1}{n}) \subseteq U_{i_0}$, which gives the contradition.
Then note that A is totally bounded (Proposition 1.1). There exists a open ball cover $\{B(x_j, \delta_0)\}_{j=1}^{m}$. For each j, pick $i_j \in I$ such that $U_{i_j} \supset B(x_j, \delta_0)$. We have $\bigcup_{j=1}^{m} U_{i_j} \supset S$. □

With this theorem, we can use the term compact for sequentially compact.

Theorem 1.2 (Heine–Cantor theorem)**.** Let (M, d_M) and (N, d_N) be two metric spaces and $S \subseteq M$. If $f : S \mapsto N$ is continuous and S is compact, then f is uniformly continuous.

Proof. Fix $\epsilon > 0$. Since f is continuous, for any $x \in S$, there exists δ_x such that $d_M(x, y) < \delta_x \Rightarrow d_N(f(x), f(y)) < \frac{\epsilon}{2}$. Denote

$$B_x := B\left(x, \frac{\delta_x}{2}\right)$$

$\{B_x\}_{x\in S}$ is a open cover of M. We can take the finite sub-cover $\{B_j\}_{j=1}^{m} \subseteq \{B_x\}_{x\in S}$ such that $S \subseteq \bigcup_{j=1}^{m} B_j$. Set

$$\delta_0 := \min_{j=1}^{m} \frac{\delta_x}{2}$$

Then for any $x, y \in S$ such that $d(x, y) < \delta$. Suppose $x \in B_j$ for some j and x' is the centre of the ball B_j. We have

$$d_M(x', y) \leq d_M(x', x) + d_M(x, y) = \frac{\delta_{x'}}{2} + \delta \leq \delta_{x'}$$

The continuity and the choice of $\delta_{x'}$ implies $d_N(f(x'), f(y)) \leq \frac{\epsilon}{2}$ and $d_N(f(x'), f(x)) \leq \frac{\epsilon}{2}$. Therefore, $d_N(f(x), f(y)) \leq \epsilon$. The result follows because the above holds for any $\epsilon > 0$.

□

The following theorem will be useful in many proof.

Theorem 1.3 (Complete metric space is Baire space)**.** Recall some notation on topology space $(S, \mathcal{T})$. Let $A \subseteq S$ be a set

- $\mathrm{cl}(A)$ is the minimal closed set in S containing A. The minimal is given by the arbitrary intersection of all possible closed sets.
- $\mathrm{int}(A) = \{a \in S : \exists B \subseteq A \text{ such that } B \in \mathcal{T} \text{ and } a \in B\}$ is the interior of A. Apparently, the interior is a open set.

A is dense if any of the following equivalent conditions holds.

1. $\forall$non-empty $B \in \mathcal{T},\ B \cap A \neq \emptyset$
2. $\mathrm{cl}(A) = S$
3. $\mathrm{int}(A^c) = \emptyset$

Define the Baire property to be

$$\forall \{A_n \in \mathcal{T}\}_{n=1}^{\infty} \text{ such that } A_n \text{ are all dense, we have } \bigcap_{n=1}^{\infty} A_n \text{ is also dense}$$

i.e., any countable intersection of open dense sets is dense. Or equivalently,

$$\forall \{A_n \subseteq S\}_{n=1}^{\infty} \text{ such that } \mathrm{int}(A_n) = \emptyset \text{ and } A_n \text{ is closed for all, we have } \mathrm{int}\left(\bigcup_{n=1}^{\infty} A_n\right) = \emptyset$$

With such $(S, \mathcal{T})$ is a Baire space.
Our statement is that every complete metric space (S, d) is a Baire space, which is a special case of Baire category theorem.

Proof. Let $\{A_n \subseteq S\}_{n=1}^{\infty}$ be some open dense sets. To prove the intersection is dense, we show that for any nonempty open set $B \subseteq S$, $B \cap \bigcap_{n=1}^{\infty} A_n \neq \emptyset$.
Since $A_1 \cap B$ is open, there exists $x_0 \in A_1, r \in (0, 1)$ such that $\overline{B(x_1, 1)} \subseteq A_1 \cap B$. Use the axiom of choice to choose such (x_1, r_1), and recursively choose

$$(x_1, r_n) \text{ such that } r_n \in \left(0, \frac{1}{n}\right) \text{ and } \overline{B(x_1, r_n)} \subseteq B(x_{n-1}, r_{n-1}) \cap A_n$$

Then, $\{x_n\}_{n=1}^{\infty}$ is Cauchy sequence and thus converges to $x \in S$. We have

$$\forall n \in \mathbb{N}^*,\ x \in \overline{B(x_1, r_n)} \text{ thus } x \in \bigcap_{i=1}^{n} A_n \cap B$$

$\square$

1.2 Vector space

Definition 1.4 (Field)**.** A field is a set F equipped with two operators, addition $(+ : F \times F \mapsto F)$ and multiplication $(\times : F \times F \mapsto F$, notation may omitted) closed in F, on which

1. Both addition operator and multiplication operator support commutativity and associativity, i.e, $\forall a, b, c \in F$, $a + b = b + a$, $ab = ba$, $(a + b) + c = a + (b + c)$ and $(ab)c = a(bc)$
2. Distributivity of multiplication over addition, i.e., $\forall a, b, c \in F$, $a \times (b + c) = ab + ac$
3. A unique additive identity $0 \in F$ is defined, such that $\forall a \in F$, $a + 0 = a$ and there exists a unique additive inverse $-a \in F$ that $a + (-a) = 0$
4. A unique multiplicative identity $1 \in F$ is defined, such that $\forall a \in F$, $a \times 1 = a$ and if $a \neq 0$, there exists a unique multiplicative inverse $a^{-1} \in F$ that $a \times a^{-1} = 1$

The element in F is called scalar.

Any of $\mathbb{Q}$, $\mathbb{R}$ and $\mathcal{C}$ together with the arithmetic addition and multiplication forms a field.

Lemma 1.1. Let $\{\sum_{i=1}^{n} a_{j,i} \times x_i = 0\}_{j=1}^{m}$ be a system of equation with m equations and n variables, where both variables and coefficient are in a certain field. If $n > m$, it has non-trivial solution $(x_1, \cdots, x_n)$ that $x_i \neq 0$ for at least one i.

Sketch of proof: First show that the set of solutions does not change under the transformations of Gauss elimination. Then we can construct the required solution by standard equation solving procedures.

Definition 1.5 (Vector space)**.** A vector space is a set V over a field F and two operators, addition $(+ : V \times V \mapsto V)$ and scalar multiplication $(\times : F \times V \mapsto V$, where F is the field of scalar, notation may omitted) closed in V, on which

1. A unique $\mathbf{0} \in V$ is defined, such that $\forall \mathbf{x} \in V$, $\mathbf{x} + \mathbf{0} = \mathbf{x}$ and there exists a unique additive inverse $-\mathbf{x} \in V$ that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$
2. $1 \times \mathbf{x} = \mathbf{x}\ \forall \mathbf{x} \in V$, where 1 is the multiplicative identity in F
3. The addition operator supports commutativity and associativity
4. The multiplicative operator satisfies $a(b\mathbf{x}) = (ab)\mathbf{x}\ \forall a, b \in F, \mathbf{x} \in V$
5. Distributivity of scalar multiplication with respect to vector addition, and distributivity of scalar multiplication with respect to field addition

The element in V is called vector.

In most context, vector refers to the special case when $V = \mathbb{R}^n$ (or $\mathbb{C}^n$) $(n \in \mathbb{N}^*)$ and F is the corresponding $\mathbb{R}$ or $\mathbb{C}$.

Definition 1.6. Subspace, linear combination, span, linear dependency and basis

1. Given a vector space V over F, a subset $S \subseteq V$ is a subspace if it is a vector space with the same addition and multiplication operators. Observe that $\mathbf{0} \in S$ and the arbitrary intersection of subspaces is a subspace.

2. Given a vector space V over F and a finite collection of vectors $\{\mathbf{v}_i \in V\}_{i=1}^n$, a linear combination is $\sum_{i=1}^n c_i\mathbf{v}_i$, where $\{c_i \in F\}_{i=1}^n$ are arbitary scalars. The linear combination is in V.

3. Given a vector space V over F, the span of a subset $S \subseteq V$ is the minimal subspace that contains S, i.e.,

$$\text{span}[S] := \bigcap_{T\in\mathcal{C}} T, \text{ where } \mathcal{C} = \{T : S \subseteq T \subseteq V \text{ and } T \text{ is a subspace}\}$$

   Observe that $V \in \mathcal{C}$ thus $\mathcal{C}$ is not empty. Equivalently, we can define

$$\text{span}[S] := \{\mathbf{v} : \mathbf{v} \text{ is a linear combination of a finite collection of vectors in } S\}$$

   If $\text{span}[S] = V$, we say S generates V.

4. Given a vector space V over F, a subset $S \subseteq V$ is linear independent if for any finite subset $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\} \subseteq S$, the solution for $\sum_{i=1}^n c_i\mathbf{v}_i = \mathbf{0}$ is $c_i = 0\ \forall 1 \leq i \leq n$ only. Otherwise, S is linearly dependent.

5. Given a vector space V over F and a subset $S \subseteq V$, S is the (Hamel) basis of $\text{span}[S]$ if S is linearly independent. In particular, if $V = \mathbb{C}^n$ and $F = \mathbb{C}$, $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ denotes the standard basis.

Remark. A proper subspace $S \subset V$ is not an open set (just check the neighborhood of $\mathbf{0}$). In general, it is not closed either, unless S has finite dimension. Counter examples can be found when V is a Banach space with infinite dimension and S is the span of a countably infinite basis. To be mention later.

Proposition 1.2. Let V be a vector space over F. $S \subseteq V$ is a basis with finite cardinality that generates subspace T. If $S' \subseteq T$, any of the following two implies the remaining one

1. $\text{span}[S'] = T$
2. S' is linearly independent
3. $|S'| = |S|$

Sketch of proof: For 1&2 $\Rightarrow$ 3, we can prove that two bases with different finite cardinalities generate different subspaces. We may use Lemma 1.1. For 1&3 $\Rightarrow$ 2, prove it by contradiction that we can take the proper subset of S' that is a basis of S' as well as T. For 2&3 $\Rightarrow$ 1, first note that $\text{span}[S']$ is no larger than T since $S' \subseteq T$ and T is a subspace. Then, if $\text{span}[S'] \subsetneq T$, S' can become a basis of T by including sufficient vertices (at least one) in T but outside $\text{span}[S']$, then the contradiction arises.

Definition 1.7 (Dimension)**.** Let V be a vector space over F. If a subspace S has a finite basis, the dimension of S is the cardinality of its basis, written $\dim[S]$. Specifically, $\dim[\{\mathbf{0}\}] = 0$. If there not exists a finite space, S is of infinite dimension, i.e., $\dim[S] = \infty$.

Remark. Though not specifically written, dimension is dependent on the choice of the field F. One may write $\dim_F[S]$ In case of ambiguity.
Proposition 1.2 have shows that the dimension, if finite, is consistent regardless of the choice of basis.

The following proves that a linear independent set cannot be arbitrarily large, and the basis is always a maximal one, which always exists. We need the following result from set theory.

Theorem 1.4 (Hausdorff maximal principle)**.** Let $(S, \leq)$ be a partially order set. We call $C \subseteq S$ a chain if $(C, \leq)$ is a totally order set. There exists a maximal chain C_M such that every chain of S is contained in C.

The proof of Hausdorff maximal principle refers to *Real and Complex Analysis* (Rudin, Walter, 1921) page 395 (slide 410), using the axiom of choice.

Proposition 1.3 (Maximal basis)**.** Let V be a vector space over F and $S \subseteq V$ is a basis generating $T := \text{span}[S]$. Let $S' \subseteq T$ be an arbitrary linear independent subset. Then there exists a unique maximal set S^* such that $S^* \supseteq S'$, S^* is linear independent, and thus S^* is a basis for T. The maximality means that any subset of T which properly contains S^* is not a linear dependent set.

Proof. Let $\mathcal{S}$ be the collection of all linear independent subsets of T containing S'. Clearly, $S' \subseteq \mathcal{S}$ and $\mathcal{S}$ is a partially order space with subset $\subseteq$ relation. The Hausdorff maximal principle gives the existance of a maximal chain $\mathcal{C} \subseteq \mathcal{S}$ Let

$$S^* = \bigcup_{C \in \mathcal{C}} C$$

Clearly $S' \subseteq S^* \subseteq T$. S^* is linear independent because for any finite collection $\{\mathbf{u}_j\}_{j=1}^n \subseteq V$, there exist $C \in \mathcal{C}$ such that $\{\mathbf{u}_j\}_{j=1}^n \subseteq C$, and the linear independent condition is satisfied in C.
It remains to show that S^* is maximal and thus a basis of T. If otherwise $\text{span}[S^*] \neq T$, picking any $\mathbf{u} \in T \setminus \text{span}[S^*]$, $S^* \cup \{\mathbf{u}\} \supset S^*$ is a larger linear independent set. Then $\mathcal{C} \cup \{S^* \cup \{\mathbf{u}\}\}$ is a larger chain, contradicting to the maximality of $\mathcal{C}$. Therefor $\text{span}[S^*] = T$, and $S^* \cup \{\mathbf{u}\}$ is linear dependent for any $\mathbf{u} \in T$. □

Definition 1.8 (Linear operator)**.** Given two linear space V, W over the same field F, a function $f : V \mapsto W$ is a linear operator (aka. linear transformation) if

$$\forall u, v \in V, c \in F,\ f(u+v) = f(u) + f(v),\ f(cu) = cf(u)$$

- Define the preimage $f^{-1}(\{\mathbf{0}_W\})$ be the kernel of f, denoted $\ker(f)$, which is a subspace in W.
- Define $\text{rank}(f) := \dim_W(f(V))$, the dimension of the range, to be the rank of f.

Proposition 1.4. If T is a subspace in V with dimension $n \in \mathbb{N}^*$, and $S = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is its basis, then the relation from any $\mathbf{x} \in T$ to the choice of $\{c_i \in F\}_{i=1}^n$ such that $\sum_{i=1}^n c_i \mathbf{v}_i = \mathbf{x}$ is a bijective linear operator $T \mapsto F^n$.

Sketch of proof: We can first prove the uniqueness of the choice $\{c_i \in F\}_{i=1}^n$ and then prove the linearity conditions are satisfied. Note that T and F^n are defined over the same field F.

1.3 Norm

Definition 1.9 (Normed vector space)**.** A normed vector space is a vector space $V := (V, \|\cdot\|_V)$ over field $F \in \{\mathbb{R}, \mathbb{C}\})$ such that the norm operation is defined

$$\|\cdot\|_V : V \mapsto \mathbb{R}^+$$

with conditions

1. $\|\mathbf{x}\|_V \geq 0 \ \forall \mathbf{x} \in V$

2. $\|\mathbf{x}\|_V = 0 \iff \mathbf{x} = \mathbf{0}$

3. (Homogeneity) $\|\lambda\mathbf{x}\|_V = \|\lambda\|\|\mathbf{x}\|_V \ \forall \mathbf{x} \in V, \lambda \in F$. Note that $\|\lambda\|$ is the absolute value in either $\mathbb{R}$ or $\mathbb{C}$

4. (Triangle inequality) $\|\mathbf{x} + \mathbf{y}\|_V \leq \|\mathbf{x}\|_V + \|\mathbf{y}\|_V \ \forall \mathbf{x}, \mathbf{y} \in V$

A normed vector space naturally define a distance function $d(\mathbf{x}, \mathbf{y}) := \|x - y\|_V$, which makes it a metric space. Also note that the triangle inequality makes the norm function continuous.

Analysis of subspace and basis will be more feasible in the normed vector space.

Lemma 1.2 (Riesz's lemma)**.** Let $(V, \|\cdot\|_V)$ be a normed vector space of nonzero dimension, and $W \subset V$ be a closed and proper subspace. For any $\epsilon \in (0,1)$, there exists $\mathbf{x} \in V$ with $\|\mathbf{x}\|_V = 1$ such that

$$1 - \epsilon \leq \operatorname{dist}(\mathbf{x}, W) \leq 1 \quad \text{where } \operatorname{dist}(\mathbf{x}, W) = \inf\{d(\mathbf{x}, \mathbf{u}) : \mathbf{x} \in W\}$$

Proof. $\operatorname{dist}(\mathbf{x}, W) \leq 1$ is obvious because $\mathbf{0} \in W$ and $d(\mathbf{x}, \mathbf{0}) = 1$. Fix an arbitary $\epsilon \in (0,1)$ and $\mathbf{y} \in V \setminus W$. Since W is closed, $d := \operatorname{dist}(\mathbf{y}, W) > 0$. By the definition, there exists $\mathbf{u}_0 \in W$ such that

$$d \leq d(\mathbf{y}, \mathbf{u}_0) \leq \frac{d}{1-\epsilon}$$

Let

$$\mathbf{x} = \frac{\mathbf{y} - \mathbf{u}_0}{\|\mathbf{y} - \mathbf{u}_0\|_V} \text{ such that } \|\mathbf{x}\|_V = 1$$

We prove that $\mathbf{x}$ is what we need by $\|\mathbf{x} - \mathbf{u}\|_V \geq 1 - \epsilon \ \forall \mathbf{u} \in W$.

$$\begin{aligned}
\forall \mathbf{u} \in W, \ \|\mathbf{x} - \mathbf{u}\|_V &= \left\| \frac{\mathbf{y} - \mathbf{u}_0}{\|\mathbf{y} - \mathbf{u}_0\|_V} - \mathrm{u} \right\|_V \\
&= \frac{1}{\|\mathbf{y} - \mathbf{u}_0\|_V} \left\| \mathbf{y} - (\mathbf{u}_0 + \|\mathbf{y} - \mathbf{u}_0\|_V \mathbf{u}) \right\|_V \\
&\geq \frac{d}{\|\mathbf{y} - \mathbf{u}_0\|_V} \text{ (Since } (\mathbf{u}_0 + \|\mathbf{y} - \mathbf{u}_0\|_V \mathbf{u}) \in W) \\
&\geq 1 - \epsilon \text{ (by the choice of } \mathbf{u}_0)
\end{aligned}$$

□

Proposition 1.5. Let W be a subspace of normed vector space $(V, \|\cdot\|_V)$ over $F \in \{\mathbb{R}, \mathbb{C}\}$.

1. If $\dim[W] < \infty$, any closed and bounded subset is compact.

2. If $\dim[W] = \infty$,

$$W \cap \overline{B(\mathbf{0}, r)}$$

   is not compact for any $r \in (0, \infty)$.

Proof. If $k := \dim[W] < \infty$, let $B := \{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ be a basis of W and $f : W \mapsto F^k$ be the linear and bijective map of vector to the coordinate discribed in Proposition 1.4. Let $S \subseteq W$ be a closed and bounded subset. We can prove that $f(S)$ is closed and bounded, and $\{\mathbf{x}_n\}_{n=1}^{\infty}$ converges in W if and only if $\{f(\mathbf{x}_n)\}_{n=1}^{\infty}$ in F^n. Since the compactness in $F^k \in \{\mathbb{R}^k, \mathbb{C}^k\}$ is equivalent to closed

and boundedness, so as S.
For 2, it suffices to let $r = 1$. Since W is infinite dimensional, let $\{\mathbf{u}_i \in W : i \in \mathbb{N}^*\}$ be a countable linear independent set such that $\|\mathbf{x}_i\|_V = 1$. Let $W_0 = \{\mathbf{0}\}$ and $W_n = \text{span}[\{\mathbf{u}_i\}_{i=1}^n]$ for any $n \in \mathbb{N}^*$, such that W_n is a proper closed subspace of W_{n+1}. By Riesz's lemma, for any $n \in \mathbb{N}^*$, pick $\mathbf{x}_n \in W_n$ such that $\|\mathbf{x}_n\|_{W_n} = \|\mathbf{x}_n\|_V = 1$ and $\text{dist}(\mathbf{x}_n, W_{n-1}) \geq \frac{1}{2}$.
We have $\{\mathbf{x}_n\}_{n=1}^\infty \subset W \cap \overline{B(\mathbf{0}, 1)}$. If otherwise $W \cap \overline{B(\mathbf{0}, 1)}$ is compact, there exists a convergent subsequence $\{\mathbf{x}_{n_i}\}_{i=1}^\infty$. However, this is impossible because $\forall i, j \in \mathbb{N}^*$ such that $i < j$,

$$\mathbf{x}_{n_i} \in W_{n_j - 1}, \text{ thus } \|\mathbf{x}_{n_j} - \mathbf{x}_{n_i}\|_V \geq \frac{1}{2} \text{ by the choice of } \mathbf{x}_{n_j}$$

□

We continue to discuss the operators on normed vector space.

Definition 1.10 (Operation norm)**.** Let f be a function mapping two normed vector spaces $(V, \|\cdot\|_V) \mapsto (W, \|\cdot\|_W)$. The norm of f is defined to be

$$\text{Norm}(f) := \sup \left\{ \frac{\|f(\mathbf{x})\|_W}{\|\mathbf{x}\|_V} : \mathbf{x} \in V \right\}$$

We say f is bounded if $\text{Norm}(f) < \infty$.

Proposition 1.6. When the transformation f is linear, then f is bounded if and only if f is continuous.

Proof. Because of the linearity, f is continuous if and only if f is continuous at $\mathbf{x} = \mathbf{0}_V$. To prove the "$\Rightarrow$" direction, note that for any $\epsilon > 0$, taking $\delta := \frac{\epsilon}{\text{Norm}(f)}$ gives the continuity statement. For the other direction, the continuity of f gives that $\exists \delta_1 > 0,\ \|\mathbf{x}\|_V <\leq delta_1 \ \Rightarrow\ \|f(\mathbf{x})\|_W < 1$. Then,

$$\text{Norm}(f) = \sup \left\{ \frac{\|f(\mathbf{x})\|_W}{\delta_1} : \mathbf{x} \in V \text{ and } \|\mathbf{x}\|_V = \delta_1 \right\} < \frac{1}{\delta_1}$$

. □

Example of normed vector space
From the above proposition, the set of all linear and continuous $V \mapsto W$ functions, $\mathcal{F}(V \mapsto W)$, is a normed vector space with the norm defined to be

$$\|f\|_{V \to W} := \text{Norm}(f) \ \forall f \in \mathcal{F}(V \mapsto W)$$

We can check that $\mathcal{F}(V \mapsto W)$ is a vector space with $\mathbf{0}$ being the zero function, and that the norm $\|\cdot\|_{M \to N}$ satisfies the four conditions.

The following defines a special case, which will be use in later.

Definition 1.11 (Dual space)**.** Let V be a normed vector space over $F \in \{\mathbb{C}, \mathbb{R}\}$. The dual space of V, $V^* := \mathcal{F}(V, \mathbb{R})$, is defined to be the normed vector space of all linear and continuous (or bounded) $M \mapsto F$ functions.

The following type of operator defined on normed vector space is also important.

Definition 1.12 (Compact operator)**.** Let f be a map between two normed vector spaces $(V, \|\cdot\|_V) \mapsto (W, \|\cdot\|_W)$. Operator f is compact if there exists a neighborhood A of $\mathbf{0}_V$ (i.e., open set A containing $\mathbf{0}_V$) such that the image $f(A)$ is relatively compact (i.e. the closure $\overline{f(A)}$ is compact) in W.

Proposition 1.7. For linear map f, the followings are equivalent definitions for f being compact

1. $\overline{f[B(\mathbf{0}_V, 1)]} \subseteq Y$ is compact
2. $\forall$bounded $S \subset V$, $\overline{f(S)} \subset Y$ is compact
3. For any bounded sequence $\{\mathbf{x}_n \in V\}_{n=1}^{\infty}$, $\{f(\mathbf{x}_n) \in V\}_{n=1}^{\infty}$ has a convergent subsequence.

Proof. For 1, clearly f is compact if such $B(\mathbf{0}_V, 1)$ exists. We prove the other direction that if the desired open set A exists such that $f(A)$ is relative compact, then so is $f[B(\mathbf{0}_V, 1)]$. Let $r > 0$ to be small such that $B(\mathbf{0}_V, r) \subseteq A$. Clearly

$$\overline{f[B(\mathbf{0}_V, r)]} \text{ is closed and } \overline{f[B(\mathbf{0}_V, 1)]} \subseteq \overline{f(A)} \subseteq W$$

Therefore $\overline{f[B(\mathbf{0}_V, r)]} \subseteq W$ and is compact. Since f is linear, we have $f[B(\mathbf{0}_V, r)] = \{r\mathbf{y} : \mathbf{y} \in f[B(\mathbf{0}_V, 1)]\}$ and then

$$\overline{f[B(\mathbf{0}_V, r)]} = \left\{r\mathbf{y} : \mathbf{y} \in \overline{f[B(\mathbf{0}_V, 1)]}\right\}$$

There exists continuous mapping between $\overline{f[B(\mathbf{0}_V, r)]}$ and $\overline{f[B(\mathbf{0}_V, 1)]}$. Thus the latter one is also compact.
1 implies 2, given that f map any open balls to a relatively compact set, and that there exists $r > 0$ such that $S \subseteq B(\mathbf{0}_V, r)$. 2 also implies 1 obviously.
2 clearly implies 3. For the other direction, fix a bounded subset $S \subset V$ and let $\{\mathbf{y}_n \in \overline{f(S)}\}_{n=1}^{\infty}$ be any sequence. For each $n \in \mathbb{N}^*$, there exists $\mathbf{x}_n \in V$ such that $\|\mathbf{y}_n - f(\mathbf{x}_n)\| < 2^{-n}$. If, by the premise, there exists a subsequence such that $\{f(\mathbf{x}_{n_i})\}_{i=1}^{\infty}$ converges in $\overline{f(S)}$, $\{\mathbf{y}_{n_i}\}_{i=1}^{\infty}$ also converges in $\overline{f(S)}$. □

Proposition 1.8. Let f, g be some linear and bounded maps between two normed vector spaces $(V, \|\cdot\|_V) \mapsto (W, \|\cdot\|_W)$ (both over $F \in \{\mathbb{R}, \mathbb{C}\}$).

1. If f, g are compact, $f + g$ is compact
2. If f is compact and $c \in F$, cf is compact
3. If f is bounded and $\text{rank}(f) < \infty$, f is compact.

Proof. 1 and 2 can be verified using the third equivalent definition (in Proposition 1.7) of compact operator. To prove 3, since $\dim[f(V)] < \infty$, $f(V) = \overline{f(V)}$ and $\overline{f(B(\mathbf{0}_V, 1))} \subseteq \overline{f(V)}$ is compact by Proposition 1.5. □

2 Banach space

2.1 A complete normed vector space

Proposition 2.1 (Completeness for normed vector space)**.** Given a normed vector space V over $F = \{\mathbb{R}, \mathbb{C}\}$, it is complete if and only if all the absolutely convergent series converge. Note that a series $\{\sum_{i=1}^{n} \mathbf{x}_i\}_{i=1}^{\infty}$, where $\mathbf{x}_i \in M\ \forall i \in \mathbb{N}^*$, is absolutely convergence if and only if

$$\sum_{n=1}^{\infty} \|\mathbf{x}_i\| < \infty$$

Proof. For the "$\Rightarrow$" direction, we can show that every absolutely convergent series is Cauchy, and thus converges by the assumption.
For the other direction, given a Cauchy sequence $\{\mathbf{x}_n\}_{n=1}^{\infty}$, we can construct a subsequence $\{\mathbf{x}_{n_i}\}_{i=1}^{\infty}$ such that

$$\mathbf{x}_{i_0} = \mathbf{0},\ \forall i \in \mathbb{N}^* \text{ and } i > 1,\ \|\mathbf{x}_{n_{i+1}} - \mathbf{x}_{n_i}\| \leq \frac{1}{2^i}$$

The sequence is meanwhile an absolutely convergent series of $\{\mathbf{x}_{n_i} - \mathbf{x}_{n_{i-1}}\}_{i=1}^{\infty}$, thus converges. The Cauchy sequence therefore also converges. □

Definition 2.1 (Banach space)**.** A Banach space is a normed vector space that is also a complete metric space. It is immediate that any closed subspace of a Banach space is a Banach subspace. The above proposition suggests another way to determine a Banach space.

Example of Banach space

1. the finite dimension Euclidean space $\mathbb{R}^n$ with Euclidean norm.
2. ($\mathcal{L}^p$ with finite dimension) more generally, given $n \in \mathbb{N}^*$ and $p \in [1, \infty)$, the vector space $\mathbb{R}^n$ over $\mathbb{R}$ with $\mathcal{L}^p$ norm, defined as

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \quad \forall \mathbf{x} = [x_1, \cdots, x_n]^T \in \mathbb{R}^n$$

3. similarly, $\mathbb{C}^n$ over $\mathbb{C}$ with $\mathcal{L}^p$ norm

$$\|\mathbf{z}\|_p = \left(\sum_{i=1}^{n} \|z_i\|^p\right)^{\frac{1}{p}} \quad \forall \mathbf{z} = [z_1, \cdots, z_n]^T \in \mathbb{C}^n$$

Proof. Note that under $\mathcal{L}^p$ norm, a sequence of finite dimension real or complex vector is Cauchy if and only if it is element-wise Cauchy. Then the Cauchy completeness of $\mathbb{R}$ or $\mathbb{C}$ gives the completeness of $\mathbb{R}^n$ and $\mathbb{C}^n$. The following show that $\mathcal{L}^p$ norm satisfies the triangle inequality, while other conditions trivially hold.
For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, to prove $\|\mathbf{x} + \mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p$ it suffices to restrict $RHS = 1$, because the case $RHS = 0$ is trivial, and the norm can be scaled by homogeneity. Let

$$\lambda := \|\mathbf{x}\|_p,\ 1 - \lambda = \|\mathbf{y}\|_p \text{ and } \mathbf{x}' := \frac{\mathbf{x}}{\lambda},\ \mathbf{y}' := \frac{\mathbf{y}}{1-\lambda},$$

when $\lambda \in (0, 1)$, otherwise the case $\lambda = 0$ or $\lambda = 1$ is trivial. The inequality becomes

$$\|\lambda\mathbf{x}' + (1-\lambda)\mathbf{y}'\|_p \leq 1 \text{ under } \|\mathbf{x}'\| = \|\mathbf{y}'\| = 1$$

Because when $p \in [1, \infty)$, the mapping $x \mapsto |x|^p$ (or $z \mapsto \|z\|^p$ for complex z) is convex, we have

$$\forall i \in \{1, \cdots, n\},\ \|\lambda x_i' + (1-\lambda)y_i'\|_p^p \leq \lambda\|x_i'\|_p^p + (1-\lambda)\|y_i'\|_p^p$$

where x_i' and y_i' are elements of $\mathbf{x}'$ and $\mathbf{y}'$. Summing up the above from $i = 1$ to $i = n$, the result follows.

□

The following generalise $\mathcal{L}^p$ space to the space of sequences. The full version of $\mathcal{L}^p$ space based on Lebesgue integral will be discussed in section 2.2.

Example: $\mathcal{L}^p$ space of sequence
Let $F \in \{\mathbb{R}, \mathbb{C}\}$ and $p \in [1, \infty)$. Define $\mathcal{L}^p(\mathbb{N}^*)$ to be the set of all sequence $\{a_n \in F\}_{n=1}^{\infty}$ such that

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} \|a_i\|^p \text{ converges}$$

$\mathcal{L}^p(\mathbb{N}^*)$ is a normed vector space over F with $\mathcal{L}^p$ norm

$$\|a\|_p = \left(\sum_{n=1}^{\infty} \|a_i\|^p \right)^{\frac{1}{p}} \quad \forall a := \{a_n\}_{n=1}^{\infty} \in \mathcal{L}^p(\mathbb{N}^*)$$

Here $\mathbb{N}^*$ specifies that the index of a is positive integers. We may write $\mathcal{L}^p(\mathbb{Z})$ for bi-directional sequence or simply $\mathcal{L}^p(n)$ for n dimensional $\mathcal{L}^p$ space.
Clearly, what we defined is a normed vector space, with the triangle inequality given by taking the limit of finite dimension cases. It remains to prove that it is a complete metric space.

Proof. Let $\{a^{(i)} \in \mathcal{L}^p(\mathbb{N}^*)\}_{i=1}^{\infty}$ be a Cauchy sequence such that each element $a^{(i)} = \{a_n^{(i)}\}_{n=1}^{\infty}$ is a sequence. First note that for any $n \in \mathbb{N}^*$, the scalar sequence $\{a_n^{(i)}\}_{i=1}^{\infty}$ is Cauchy and thus converge to, say, $a_n \in F$. We claim that $\{a^{(i)}\}_{i=1}^{\infty}$ converges to $a := \{a_n\}_{n=1}^{\infty}$ as $i \to \infty$ and $a \in \mathcal{L}^p(\mathbb{N}^*)$
Fix a arbitrary $\epsilon > 0$, set $I \in \mathbb{N}^*$ such that $\forall i, j \geq I,\ \|a^{(i)} - a^{(j)}\|_p < \frac{\epsilon}{2}$. Then

$$\begin{aligned}
\forall i \geq I,\ \|a^{(i)} - a\|_p^p = \lim_{N\to\infty} \sum_{n=1}^{N} \|a_n^{(i)} - a_n\|^p &= \lim_{N\to\infty} \lim_{j\to\infty} \sum_{n=1}^{N} \|a_n^{(i)} - a_n^{(j)}\|^p \\
&\leq \lim_{N\to\infty} \lim_{j\to\infty} \|a^{(i)} - a^{(j)}\|_p^p \\
&\leq \lim_{N\to\infty} \lim_{j\to\infty} \left(\frac{\epsilon}{2}\right)^p = \frac{\epsilon^p}{2^p} < \epsilon^p
\end{aligned}$$

gives the convergence. The above also shows $a^{(i)} - a \in \mathcal{L}^p(\mathbb{N}^*)$ because its norm is finite. By additivity, $a \in \mathcal{L}^p(\mathbb{N}^*)$. □

Proposition 2.2 (Dual space is Banach)**.** Let V, W be two normed vector spaces and $\mathcal{F}(V, W)$ is the normed vector space of bounded and linear $V \mapsto W$ functions. If W is Banach, then $\mathcal{F}(V, W)$ is Banach.
Specially, dual space is Banach space since $W = \mathbb{R}$ or $W = \mathbb{C}$.

Proof. Let $\{f_i \in \mathcal{F}(V, W)\}_{i=1}^{\infty}$ be Cauchy sequence. Since W is complete, for any $x \in V$, $\{f_i(x)\}_{i=1}^{\infty}$ converges to, say, $y(x) \in W$ since $\{f_i(x)\}_{i=1}^{\infty}$ is Cauchy. Define the function

$$f : V \to W \text{ by } f(x) := y(x) = \lim_{i\to\infty} f_i(x)$$

f is a linear function because each f_i is linear. f is also bounded because there exists $N \in \mathbb{N}^*$ such that

$$\forall i \geq N,\ \|f_i - f_N\|_{V \mapsto W} \leq 1 \ \Rightarrow\ \|f_i\|_{V \mapsto W} \leq \|f_N\|_{V \mapsto W} + 1$$

then

$$\sup\left\{\frac{\|f(\mathbf{x})\|_W}{\|\mathbf{x}\|_V} : \mathbf{x} \in V\right\} = \sup\left\{\lim_{i \to \infty} \frac{\|f_i(\mathbf{x})\|_W}{\|\mathbf{x}\|_V} : \mathbf{x} \in V\right\} \leq \max\{\|f_i\|_{V \mapsto W} : i \in \{1 \cdots, N\}\} + 1$$

Thus f is bounded, and $f \in \mathcal{F}(V, W)$. It remains to prove that $\{f_i\}_{i=1}^{\infty}$ converges to f. We use the argument similar to the previous example.

$$\|f_i - f\|_{V \mapsto W} = \sup\left\{\lim_{j \to \infty} \frac{\|(f_i - f_j)(\mathbf{x})\|_W}{\|\mathbf{x}\|_V} : \mathbf{x} \in V\right\} \leq \sup\left\{\lim_{j \to \infty} \frac{\epsilon\|\mathbf{x}\|_V}{\|\mathbf{x}\|_V} : \mathbf{x} \in V\right\} = \epsilon$$

where $\epsilon > 0$ is arbitrary and i is large enough such that $\|f_i - f_j\|_{V \mapsto W} \leq \epsilon$ for large j. □

The following proposition show that the set of compact operators mapping to Banach space is closed.

Proposition 2.3. Let $(V, \|\cdot\|_V)$ be a normed vector space and $(W, \|\cdot\|_W)$ be a Banach space. If $\{f_n : V \mapsto W\}_{n=1}^{\infty}$ is a sequence of compact operators converging to $f : V \mapsto W$ in operator norm, then f is also compact.

Proof. Let $B := B(\mathbf{0}_V, 1)$. Since $\overline{f(V)} \subseteq W$ is complete, by Proposition 1.1, it suffices to prove that $\overline{f(B)}$ is totally bounded. For any $\epsilon > 0$, pick $n \in \mathbb{N}^*$ such that $\|f_n - f\|_{V \mapsto W} < \frac{\epsilon}{2}$. Since $\overline{f_n(B)}$ is compact thus totally bounded, there exists

$$\left\{\mathbf{x}_i \in \overline{f_n(B)}\right\}_{i=1}^{m} \text{ such that } \overline{f_n(B)} \subseteq \bigcup_{i=1}^{m} B\left(\mathbf{x}_i, \frac{\epsilon}{2}\right)$$

We can prove that $\overline{f(B)} \subseteq \bigcup_{i=1}^{m} B(\mathbf{x}_i, \epsilon)$. □

2.2 $\mathcal{L}^p$ space of function

This section discuss the $\mathcal{L}^p$ space of measurable functions base on a measure space.

Definition 2.2 ($\mathcal{L}^p$ norm)**.** Let $(S, \mathcal{B}, \mu)$ be a measure space and $p \in [1, \infty)$. The $\mathcal{L}^p$ norm of a measurable function $f : S \mapsto F$, where $F \in \{\mathbb{R}, \mathbb{C}, \bar{\mathbb{R}}\}$ (not necessarily a field in the third case), is defined as

$$\|f\|_p = \left(\int_S \|f\|^p \mathrm{d}\mu\right)^{\frac{1}{p}}$$

Let $\mathcal{L}^p := \mathcal{L}^p(\mu) := \mathcal{L}^p(\mu \mapsto F)$ be the set of all the measurable functions such that the inner integral is finite. The set $\mathcal{L}^1$ match all the integrable functions.

Some lemmas are required to show that the set $\mathcal{L}^p$ is a Banach space with $\mathcal{L}^p$ norm.

Lemma 2.1 (Young's inequality)**.** Let $p, q \in (1, \infty)$ be arbitrary such that $\frac{1}{p} + \frac{1}{q} = 1$. We have

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}\ \forall a, b \in \mathbb{R}$$

The equality holds if and only if $a^p = b^q$.

Proof. Since the function exp is convex, we have

$$\forall \lambda \in [0,1] \text{and } x_1, x_2 \in \mathbb{R},\ \exp[\lambda x_1 + (1-\lambda)x_2] \leq \lambda \exp(x_1) + (1-\lambda)\exp(x_2)$$

Putting $\lambda = \frac{1}{p}$, $x_1 := p \ln a$ and $x_2 := q \ln b$, the result follows. □

Lemma 2.2 (Hölder's inequality)**.** Let $p, q \in (1, \infty)$ be arbitrary such that $\frac{1}{p} + \frac{1}{q} = 1$. For any two measurable function f, g on the measure space $(S, \mathcal{B}, \mu)$, we have

$$\|fg\|_1 \leq \|f\|_p \|g\|_q$$

If $RHS < \infty$, the equality holds if and only if $\exists c \in [0, \infty),\ \|f(\cdot)\|^p \overset{a.s.}{=} c\|g(\cdot)\|^q$.
The inequality shows that if $f \in \mathcal{L}^p(\mu)$ and $g \in \mathcal{L}^q$, then $fg \in \mathcal{L}^1$.

Proof. We only need to check check the case that $\|f\|_p < \infty$ and $\|g\|_q < \infty$, in which $\|f\| \overset{a.s.}{<} \infty$ and $\|g\| \overset{a.s.}{<} \infty$, hold. Let $S' \in \mathcal{B}$ be the set such that $\mu(S^c) = 0$ and $\|f(x)\|, \|g(x)\| < \infty\ \forall x \in S'$. For any $x \in S'$, putting $a := \|f(x)\|$ and $b := \|g(x)\|$ into Young's inequality, we have

$$\|f(x)g(x)\| = \|f(x)\|\|g(x)\| \leq \frac{\|f(x)\|^p}{p} + \frac{\|g(x)\|^q}{q}$$

Integrating over S (or S', equivalently)

$$\|fg\|_1 = \int_S \|f(x)g(x)\| \mathrm{d}\mu \leq \int_S \frac{\|f(x)\|^p}{p} \mathrm{d}\mu + \int_S \frac{\|g(x)\|^q}{q} \mathrm{d}\mu = \frac{\|f\|_p^p}{p} + \frac{\|g\|_q^q}{q}$$

If $\|f\|_p = 1$ and $\|g\|_q = 1$, the above gives $\|fg\|_1 \leq \frac{1}{p} + \frac{1}{q} = \|f\|_p \|g\|_q$. Also if either $\|f\|_p = 0$ or $\|g\|_q = 0$, one of them almost surely equals to 0 and the inequality holds. Otherwise, let $f' = \frac{f}{\|f\|_p}$ and $g' = \frac{g}{\|g\|_q}$, we have

$$\|fg\|_1 = \|f\|_p \|g\|_q \|f'g'\|_1 \leq 1 \text{ since } \|f'\|_p = \|g'\|_q = 1$$

The equality holds if and only if $f' \overset{a.s.}{=} g'$. □

Lemma 2.3 (Minkowski's Inequality)**.** Given measure space $(S, \mathcal{B}, \mu)$ $p \in [1, \infty)$ and $p \in [1, \infty)$, we have

$$\forall f, g \in \mathcal{L}^p(\mu),\ \|f + g\|_p \leq \|f\|_p + \|g\|_p$$

It states the triangle inequality of $\mathcal{L}^p$ norm.

Proof. First, $LHS < \infty$ because

$$\int_S \|f + g\|^p \mathrm{d}\mu \leq \int_S 2^p \max\{\|f\|, \|g\|\}^p \mathrm{d}\mu \leq \int_S 2^p \|f\|^p + 2^p \|g\|^p \mathrm{d}\mu = 2^p(\|f\|_p^p + \|g\|_p^p) < \infty$$

If $p = 1$, the results follow from the point-wise triangle inequality. Considering $p > 1$, we have

$$\begin{aligned}
\int_S \|f+g\|^p \mathrm{d}\mu &\leq \int_S (\|f\| + \|g\|) \cdot \|f+g\|^{p-1} \mathrm{d}\mu \text{ (point-wise triangle inequality)} \\
&= \int_S \left\|f \cdot \|f+g\|^{p-1}\right\| \mathrm{d}\mu + \int_S \left\|g \cdot \|f+g\|^{p-1}\right\| \mathrm{d}\mu \\
&= \left\|f \cdot \|f+g\|^{p-1}\right\|_1 + \left\|g \cdot \|f+g\|^{p-1}\right\|_1 \\
&\leq \|f\|_p \cdot \left\|\|f+g\|^{p-1}\right\|_q + \|g\|_p \cdot \left\|\|f+g\|^{p-1}\right\|_q \text{ (Hölder's, } q := \left(1 - p^{-1}\right)^{-1} \text{)} \\
&= (\|f\|_p + \|g\|_p|) \cdot \left(\int_S \|f+g\|^{q(1-p)} \mathrm{d}\mu\right)^{\frac{1}{q}} \\
&= (\|f\|_p + \|g\|_p|) \cdot \left(\int_S \|f+g\|^{p} \mathrm{d}\mu\right)^{\frac{1}{q}} \text{ (since } p = q(1-p)\text{)}
\end{aligned}$$

Note that $RHS < \infty$ because $\int_S \|f+g\|^p \mathrm{d}\mu < \infty$. If $\int_S \|f+g\|^p \mathrm{d}\mu = 0$, then $f \overset{a.s.}{=} 0$ and $g \overset{a.s.}{=} 0$ and the inequality holds. Otherwise, divide both side by $\left(\int_S \|f+g\|^p \mathrm{d}\mu\right)^{\frac{1}{q}}$ and the result follows. $\square$

We are ready to define the corresponding normed vector space.

Definition 2.3 (General $\mathcal{L}^p$ space)**.** Let $(S, \mathcal{B}, \mu)$ be a measure space and define the equivalent relation in $\mathcal{L}^p(\mu) := \mathcal{L}^p(\mu, F)$ $(F \in \{\mathbb{R}, \mathbb{C}\})$,

$$\forall f, g \in \mathcal{L}^p(\mu),\ f \sim g \text{ iff } f - g \overset{a.s.}{=} 0$$

Let $\mathcal{L}^p(\mu)'$ be the set of all distinct equivalent classes dividing $\mathcal{L}^p(\mu)$ under $\sim$, called quotient set. Define $\mathcal{L}^p(\mu)'$ to be $\mathcal{L}^p$ space over μ. It is a normed vector space over F with $\mathcal{L}^p$ norm.

Remark. Validity of the definition

1. From the lemmas, it is simple to check that $\mathcal{L}^p(\mu)'$ is a vector space, and $\mathcal{L}^p$ norm is a valid norm.

2. The use if $\mathcal{L}^p(\mu)'$ instead of $\mathcal{L}^p(\mu)$ is to ensure $f = 0$ being unique if $\|f\|_p = 0$. However, to simplify the notation, the following will still use $\mathcal{L}^p(\mu)$ instead, then the equality holds almost everywhere, and zeros are no longer unique and equal almost everywhere.

3. F can be $\bar{\mathbb{R}}$, in which case we ignore set of zero measure such that $\|f(\cdot)\| = \infty$. The vector space is still over $\mathbb{R}$.

4. The examples following 2.1 are special cases when S is discrete or countable and μ is the counting measure.

5. When (and only when) $p = 2$, a inner product can be equipped, defined as

$$\langle f, g \rangle = \int_S f\bar{g} \mathrm{d}\mu \ \forall f, g \in \mathcal{L}^2(\mu)$$

   which makes it a inner product space (and also a Hilber space)

6. $\mathcal{L}^p$ space is a Banach space. The following prove its completeness.

Proposition 2.4. $\mathcal{L}^p$ space is complete and is thus a Banach space.

Proof. Let $\{F_n := \sum_{i=1}^n f_i\}_{n=1}^\infty$ be a absolutely convergent series of $\{f_i \in \mathcal{L}^p(\mu)\}_{i=1}^\infty$, such that

$$\sum_{n=1}^{\infty} \|f_i\|_p \leq B < \infty$$

We prove that $\{F_n\}_{n=1}^\infty$ converges. First define

$$\forall n \in \mathbb{N}^* G_n : S \to \mathbb{R} \text{ by } G_n(x) = \sum_{i=1}^{n} \|f(x)\|$$

By Minkowski's inequality,

$$\|G_n\|_p \leq \sum_{i=1}^{n} \big\|\|f_i\|\big\|_p = \sum_{i=1}^{n} \|f_i\|_p \leq B$$

Let $G := \lim_{n=1}^\infty G_n$ be a measurable function. By monotone convergence theorem,

$$\|G\|_p^p = \int_S \left(\lim_{n\to\infty} G_n\right)^p \mathrm{d}\mu = \lim_{n\to\infty} \int_S {G_n}^p \mathrm{d}\mu = \lim_{n\to\infty} \|G_n\|_p^p \leq B^p < \infty$$

Therefore $G \in \mathcal{L}^p(\mu)$, and $G \overset{a.s.}{<} \infty$. Since $\forall x \in S$,

$$\|F_m(x) - F_n(x)\| \leq \sum_{i=n+1}^{m} \|f(x)\| = G_m(x) - G_n(x) \ \forall n, m \in \mathbb{N}^* \text{ and } m \geq n$$

$\{F_n(\cdot)\}_{n=1}^\infty$ is Cauchy almost everywhere. There exists a measurable function F on S such that $\lim_{n\to\infty} F_n \overset{a.s.}{=} F$. Since $\|F_n(x)\| \leq \|G_n(x)\| \ \forall n \in \mathbb{N}^*, x \in S$, we have $\|F(\cdot)\| \overset{a.s.}{\leq} \|G(\cdot)\|$ and then $\|F\|_p \leq \|G\|_p$, which gives $F \in \mathcal{L}^p(\mu)$.
It remains to prove the $\mathcal{L}^p$ convergence. Since

$$\forall n \in \mathbb{N}^* \text{ and } x \in S \text{ almost everywhere, } \|F(x) - F_n(x)\|^p = \left\|F(x) - \sum_{i=1}^{n} f_i(x)\right\|^p \leq \|2G(x)\|^p \in \mathcal{L}^1$$

the dominated convergence theorem gives the result

$$\lim_{n\to\infty} \|F_n - F\|_p^p = \lim_{n\to\infty} \int_S \|F_n - F\|^p \mathrm{d}\mu = \int_S \lim_{n\to\infty} \|F_n - F\|^p \mathrm{d}\mu = \int_S \left\|\lim_{n\to\infty} F_n - F\right\|^p \mathrm{d}\mu = 0$$

□

Next, we discuss the limit of $\mathcal{L}^p$ norm when $p \to \infty$.

Definition 2.4 ($\mathcal{L}^\infty$ norm)**.** Let $(S, \mathcal{B}, \mu)$ be a measure space. Define the $\mathcal{L}^\infty$ norm (aka. the essential supreme) of measurable function $f : S \mapsto F$, $F \in \{\mathbb{R}, \mathbb{C}, \bar{\mathbb{R}}\}$ to be

$$\|f\|_\infty = \inf\{a \in [0, \infty] : \|f\| \overset{a.s.}{\leq} a\}$$

and the set $\mathcal{L}^\infty := \mathcal{L}^\infty(\mu)$ to be such function with finite $\mathcal{L}^\infty$ norm.

Proposition 2.5. Properties of $\mathcal{L}^\infty$ norm

1. $\mathcal{L}^\infty(\mu)$ is a normed vector space with $\mathcal{L}^\infty$ norm (up to almost surely sense). Futher it is a Banach space.
2. For $\{f_n \in \mathcal{L}^\infty\}_{n=1}^\infty$ and $f \in \mathcal{L}^p$, $\lim_{n\to\infty} \|f_n - f\|_\infty = 0$ if and only if $\{f_n(\cdot)\}_{n=1}^\infty$ uniformly converges to $f(\cdot)$ almost surely.

The proofs are easy. The completeness of $\mathcal{L}^\infty$ space can be shown in the same way as the $\mathcal{L}^p$ one.

Proposition 2.6 (Relationship between $\mathcal{L}^p$ spaces)**.** Suppose $(S, \mathcal{B}, \mu)$ satisfies $\mu(S) < \infty$,

1. If $1 \le p < q \le \infty$, then $\mathcal{L}^p(\mu) \subseteq \mathcal{L}^q(\mu)$, and

$$\forall f \in \mathcal{L}^q(\mu),\ \frac{\|f\|_p}{\mu(A)^{\frac{1}{p}}} \le \frac{\|f\|_q}{\mu(A)^{\frac{1}{q}}}$$

2. If $f \in \mathcal{L}^\infty(\mu)$, then

$$\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$$

Proof. If $q = \infty$, 1 follows from

$$\|f\|_p^p = \int_S \|f\|^p \mathrm{d}\mu \le \int_S \|f\|_\infty{}^p \mathrm{d}\mu = \|f\|_\infty{}^p \mu(S)$$

Otherwise let $r = \frac{q}{p} > 1$ and $r^* = \left(1 - r^{-1}\right)^{-1}$. Hölder's inequality gives 1

$$\|f\|_p^p = \big\|\|f\|^p \cdot 1\big\|_1 \le \big\|\|f\|^p\big\|_r \big\|1\big\|_{r^*} = \left(\int_S \|f\|^{pr} \mathrm{d}\mu\right)^{\frac{1}{r}} \mu(A)^{\frac{1}{r^*}} = \mu(A)^{1-\frac{p}{q}} \|f\|_q^p$$

2 follows from the monotonic relation of 1 directly. □

Proposition 2.7 (Dense subset in $\mathcal{L}^p$ space)**.** Let $(S, \mathcal{B}, \mu)$ be a measure space and $p \in [1, \infty]$. Under $\mathcal{L}^p$ norm,

1. the set of all (complex) simple function $\mathcal{E}'$ such that

$$\mu(\{x \in S : \phi(x) \neq 0\}) < \infty$$

   is dense in $\mathcal{L}^p(\mu)$
2. if $(S, \mathcal{B}, \mu)$ is the Lebesgue measure space of $\mathbb{R}^n$ (in fact, it applies to any locally compact Hausdorff space) and $p < \infty$, the set of all continuous functions $\mathcal{C}(\mathbb{R}^k)$ with finite $\mathcal{L}^p$ norm is dense.

Proof. Using the construction of the measurability theorem (*Probability theory*, Section 4.1.1), for any almost surely bounded measurable function, there exists a sequence of simple functions uniformly converging to it almost surely.
For 2, *Probability theory*, Section 5.1.2 proves the special case that $p = 1$. The construction is valid for other p. Note that it does not hold for $p = \infty$. A simple step function $\mathbf{1}_{x>0}$ is a counter example. □

Remark. $(\mathcal{C}(\mathbb{R}^k), \|\cdot\|_p)$ is a rigorous normed vector space without the need to passing the functions to equivalent classes. Because two different continuous function must differs in a open set, which has a positive measure. Specially,

$$\mathcal{L}^\infty(f) = \sup\{f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^k\} \ \forall f \in \mathcal{C}(\mathbb{R}^k)$$

and the space $(\mathcal{C}(\mathbb{R}^k), \mathcal{L}^\infty)$ is complete, therefore a Banach space.

Proposition 2.8 (Weierstrass approximation theorem)**.** Let $I := [a, b]$ be a closed interval and $\mathcal{C}(I)$ be the set of continuous $I \mapsto \mathbb{R}$ function. The set of polynomial

$$\mathcal{P} := \left\{ \sum_{i=0}^{n} a_i x^i : n \in \mathbb{N}^*, \{a_i \in \mathbb{R}\}_{i=0}^n \right\}$$

is dense in the Banach space $(\mathcal{C}([a, b]), \mathcal{L}^\infty)$ over $\mathbb{R}$. This is a special case of Weierstrass approximation theorem for real function on a closed interval.

Proof. For simplicity, assume $I = [0, 1]$. Consider the class of Bernstein polynomials of $f \in \mathcal{C}(I)$,

$$\forall n \in \mathbb{N}^*,\ B_f^{(n)}(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$$

, we can prove that (refer the Wikipedia for details)

$$\lim_{n\to\infty} B_f^{(n)} = f \text{ uniformly}$$

□

2.3 Dual space of $\mathcal{L}^p$ space

This section shows that the dual space of $\mathcal{L}^p$ space is isometric to a $\mathcal{L}^q$ space for some $q \geq 1$.

Definition 2.5. (Homomorphism, Isometry and isomorphism)

- Homomorphism is a map between two spaces that preserves the structure. In the context of vector space, a linear map is a homomorphism. For normed vector space, a map is homomorphism if it is linear and preserves the norm.
- A map $f : (X, d_X) \mapsto (Y, d_Y)$ between two metric space is a (distance-preserving) isometry if

$$\forall x_1, x_2 \in X,\ d_Y(f(x_1), f(x_2)) = d_X(x_1, x_2)$$

- A map $f : (X, d_X) \mapsto (Y, d_Y)$ between two metric space is (isometric) isomorphism (aka. global isometry or congruence mapping) if it is an isometry, bijective, and the inverse is $Y \mapsto X$ isometric. We called X and Y are isometric if such map exists.

Remark. Some simplified conditions:

- To prove f is an isomorphism, it suffices to show that f is a surjective isometry, and the other conditions therefore hold.

- Due to the linearity, if X and Y are normed vector spaces, a homomorphism in between is an isometry. The homomorphism is continuous and bounded.
- If X and Y are normed vector spaces, a linear map $f : X \mapsto Y$ is an isomorphism if and only if it is a surjective map and a homomorphism, i.e,

$$\forall \mathbf{x} \in X, \ \|f(\mathbf{x})\|_Y = \|\mathbf{x}\|_X$$

Let $p, q \in [1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$, $(S, \mathcal{B}, \mu)$ be a measure space, $\mathcal{L}^p(\mu)$, $\mathcal{L}^q(\mu)$ be two $\mathcal{L}^p$ spaces over $\mathbb{C}$ and $\mathcal{L}^{p*}$ be the dual space of $\mathcal{L}^p(\mu)$ with operator norm. Given the Hölder's inequality, we have that $f \in \mathcal{L}^p$ and $g \in \mathcal{L}^q$ together implies $fg \in \mathcal{L}^1$. We therefore define the map $\Phi : \mathcal{L}^q(\mu) \mapsto \mathcal{L}^{p*}$,

$$\forall g \in \mathcal{L}^q(\mu), \ \Phi(g) : \mathcal{L}^p(\mu) \mapsto \mathbb{C} \text{ given by } \forall f \in \mathcal{L}^p, \ [\Phi(g)](f) = \int_S fg \mathrm{d}\mu \in \mathbb{C}$$

It is easy to show that Φ is well defined and bounded by the Hölder's inequality, and that both $\Phi : \mathcal{L}^q(\mu) \mapsto \mathcal{L}^{p*}$ and $\Phi(g) : \mathcal{L}^p(\mu) \mapsto \mathbb{C}$ are linear.

Lemma 2.4. (Φ is an isometry) The predefined map Φ satisfies

$$\forall g \in \mathcal{L}^q(\mu), \ \|\Phi(g)\|_{\mathcal{L}^p(\mu)\mapsto\mathbb{C}} = \|g\|_q$$

where $\|\cdot\|_{\mathcal{L}^p(\mu)\mapsto\mathbb{C}}$ is the operator norm. Together with the linearity, this shows that $\Phi(g) \in \mathcal{L}^{p*} \ \forall g \in \mathcal{L}^q$.

Proof. By the Hölder's inequality,

$$\begin{aligned}
\|\Phi(g)\|_{\mathcal{L}^p(\mu)\mapsto\mathbb{C}} &= \sup \{\| [\Phi(g)](f)\| : f \in \mathcal{L}^p(\mu) \text{ and } \|f\|_p = 1\} \\
&\le \sup \{\|fg\|_1 : f \in \mathcal{L}^p(\mu) \text{ and } \|f\|_p = 1\} \\
&\le \sup \{\|f\|_p\|g\|_q : f \in \mathcal{L}^p(\mu) \text{ and } \|f\|_p = 1\} = \|g\|_q
\end{aligned}$$

It suffices to find an $f \in \mathcal{L}^p(\mu)$ with $\|f\|_p = 1$ and $\| [\Phi(g)](f)\| = \|g\|_q$. Let $s_g : S \mapsto \mathbb{C}$,

$$\forall x \in S, \ s_g(x) = \mathbf{1}_{g(x)\neq 0} \times \frac{\|g(x)\|}{g(x)}$$

such that s_g is measurable, $\|g(\cdot)\| = g(\cdot)s_g(\cdot)$ and $\|s(\cdot)\| = \mathbf{1}_{g(\cdot)\neq 0}$. Pick

$$f(\cdot) = \frac{1}{\|g\|_q^{q-1}} s_g(\cdot)\|g(\cdot)\|^{q-1}$$

Given $p(q-1) = q$, we have

$$\|f\|_p^p = \frac{1}{\|g\|_q^{p(q-1)}} \int_S \left\|s_g(x)\|g(x)\|^{q-1}\right\|^p \mathrm{d}\mu(x) = \frac{1}{\|g\|_q^q} \int_S \|g(x)\|^{p(q-1)} \mathrm{d}\mu(x) = 1$$

and

$$\begin{aligned}
\| [\Phi(g)](f)\| = \left\| \int_S fg\mathrm{d}\mu \right\| &= \frac{1}{\|g\|_q^{q-1}} \left\| \int_S s_g(x)\|g(x)\|^{q-1} g(x)\mathrm{d}\mu(x) \right\| \\
&= \frac{1}{\|g\|_q^{q-1}} \left\| \int_S \|g(x)\|^q \mathrm{d}\mu(x) \right\| = \|g(x)\|_q
\end{aligned}$$

□

Lemma 2.5. If $(S, \mathcal{B}, \mu)$ is σ-finite, the predefined map Φ is surjective.

Proof. For any $\phi \in \mathcal{L}^{p*}$, i.e., $\phi : \mathcal{L}^p(\mu) \mapsto \mathbb{C}$ that is linear and bounded, we need to prove that there exists $g \in \mathcal{L}^q(\mu)$ such that $[\Phi(g)](f) = \phi(f)\ \forall f \in \mathcal{L}^p$. Write $\phi = \phi_1 + \mathrm{i}\phi_2$ where ϕ_1, ϕ_2 are real-valued functions, we have

$$\Phi(g) = \Phi(g_1) + \mathrm{i}\Phi(g_2) = \phi_1 + \mathrm{i}\phi_2 = \phi$$

whenever $\Phi(g_1) = \phi_1$ and $\Phi(g_2) = \phi_2$. It suffices to prove the case when ϕ is a linear and continuous (thus bounded) $\mathcal{L}^p(\mu) \mapsto \mathbb{R}$ functional.

Fix an arbitrary ϕ. We first look at the case $\mu(S) < \infty$. For any $A \in \mathcal{B}$, $\mathbf{1}_A \in \mathcal{L}^p$. Define $v : \mathcal{B} \mapsto \mathbb{R}$ by $v(A) = \phi(\mathbf{1}_A)$. v is a finite signed measure such that $v \ll \mu$, because

1. $v(\emptyset) = 0$ by linearity
2. For any countable collection $\{A_n \in \mathcal{B}\}_{n=1}^{\infty}$ of mutually disjoint subset,
$$v\left(\bigcup_{n=1}^{\infty} A_n\right) = \phi\left(\sum_{n=1}^{\infty} \mathbf{1}_{A_n}\right) = \sum_{n=1}^{\infty} \phi(\mathbf{1}_{A_n}) = \sum_{n=1}^{\infty} v(A_n)$$
We can put the limit of infinity summation out because ϕ is continuous
3. For any $A \in \mathcal{B}$ such that $\mu(A) = 0$, $\mathbf{1}_A$ almost surely equals to the zero function, thus $v(A) = 0$

By the Radon–Nikodym theorem (*Probability theory*, Section 5.2.1), there exists $g \in \mathcal{L}^1(\mu)$ such that

$$\forall A \in \mathcal{B},\ \phi(\mathbf{1}_A) = v(A) = \int_A g \mathrm{d}\mu = \int_S \mathbf{1}_A g \mathrm{d}\mu = [\Phi(g)](\mathbf{1}_A)$$

By proposition 2.6, $g \in \mathcal{L}^1(\mu)$ implies $g \in \mathcal{L}^q(\mu)$ given $\mu(S) < \infty$. By linearity, we can show that $[\Phi(g)](f) = \phi(f)$ for any simple function $f \in \mathcal{L}^p$. We can further use convergence theorems to verify the case for an arbitrary $f \in \mathcal{L}^p$.

When $\mu(S) = \infty$ and σ-finite, write $S = \bigcup_{j=1}^{\infty} S_j$, where $S_j \subseteq S_{j+1}$ and $\mu(S_j) < \infty$ for all $j \in \mathbb{N}^*$. For each j, we can find $g_j \in \mathcal{L}^q|_{S_j}$ such that for any $f \in \mathcal{L}^p$, $\phi(f\mathbf{1}_{S_j}) = [\Phi(g_j)](f\mathbf{1}_{S_j})$. Denote $\mathcal{L}^{p*}|_{S_j}$ to be the dual space of $\mathcal{L}^p(\mu)|_{S_j}$, and let

$$\phi|_{S_j} \in \mathcal{L}^{p*}|_{S_j} \text{ given by } \phi|_{S_j}(f) := \phi(f) = \phi(f\mathbf{1}_{S_j})\ \forall f \in \mathcal{L}^p(\mu)|_{S_j},\ \text{similar for } \Phi(g_j)|_{S_j}$$

For any $j' > j$, $A_j \subseteq A_{j'}$ and $\mathcal{L}^p|_{S_j} \subseteq \mathcal{L}^p|_{S_{j'}}$, thus

$$\forall f \in \mathcal{L}^p|_{S_j},\ \left[\Phi(g_j)|_{S_j}\right](f) = \phi|_{S_j}(f) = \phi|_{S_{j'}}(f) = \left[\Phi(g_{j'})|_{S_{j'}}\right](f) = \left[\Phi(g_{j'}\mathbf{1}_{S_j})|_{S_{j'}}\right](f)$$

Therefore $\Phi(g_j)|_{S_j} = \Phi(g_{j'}\mathbf{1}_{S_j})|_{S_{j'}}$ Since the map Φ is injective and so is restriction $\Phi(\cdot)|_{S_j}$, $g_j = g_{j'}\mathbf{1}_{S_j}$ (rigorously, $g_j \overset{a.s.}{=} g_{j'}\mathbf{1}_{S_j}$ if we do not consider $\mathcal{L}^q$ as a quotient set). Then, $g : S \to \mathbb{C}$ by

$$g(x) := g_{j(x)}(x) \quad \text{where } j(x) := \inf\{j \in \mathbb{N}^* : x \in S_j\}$$

is a pointwise limit of $\{g_j\}_{j=1}^{\infty}$, also a well defined $\mathcal{L}^q$ function such that $g|_{S_j} = g_j$ and $\Phi(g)|_{S_j} = \phi|_{S_j}$ for any $j \in \mathbb{N}^*$. For any $f \in \mathcal{L}^p$, $\|fg\| \in \mathcal{L}^1$ and $\|fg\mathbf{1}_{S_j}\| \leq \|fg\|\ \forall j \in \mathbb{N}^*$, we have that by the dominated convergence theorem,

$$\begin{aligned}[\Phi(g)](f) = \int_S \lim_{j\to\infty} (gf\mathbf{1}_{S_j})\mathrm{d}\mu = \lim_{j\to\infty} \int_S gf\mathbf{1}_{S_j}\mathrm{d}\mu &= \lim_{j\to\infty} \left[\Phi(g)|_{S_j}\right](f\mathbf{1}_{S_j}) \\ &= \lim_{j\to\infty} \phi|_{S_j}(f\mathbf{1}_{S_j}) = \phi(f)\end{aligned}$$

Therefore, for any $\phi \in \mathcal{L}^{p*}$, $\exists g \in \mathcal{L}^q$ such that $\Phi(g) = \phi$. □

The two lemmas conclude this theorem.

Theorem 2.1. Let $p, q \in [1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$, $(S, \mathcal{B}, \mu)$ be a σ-finite measure space, $\mathcal{L}^p(\mu)$, $\mathcal{L}^q(\mu)$ be two $\mathcal{L}^p$ spaces over $\mathbb{C}$ and $\mathcal{L}^{p*}$ be the dual space of $\mathcal{L}^p(\mu)$ with operator norm. We have that $\mathcal{L}^{p*}$ is isometric to $\mathcal{L}^q(\mu)$, given by the isomorphism Φ.
Note that strictly speaking, the bijectivity is based on $\mathcal{L}^q(\mu)$ being a quotient set.

3 Hilbert space

3.1 Inner product and Hilbert space

Definition 3.1 (Inner product space)**.** A vector space V over field $F \in \{\mathbb{R}, \mathbb{C}\}$ is a inner product space if the inner product $\langle \mathbf{u}, \mathbf{v} \rangle : V \times V \to \mathbb{C}$ is defined for all $\mathbf{u}, \mathbf{v} \in V$ with the following axioms

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$, where $\overline{z}$ denote the conjugation of $z \in F$
2. $\langle \mathbf{u} + \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle \ \forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V$
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle \ \forall c \in \ F$
4. $\langle \mathbf{u}, \mathbf{u} \rangle \in \mathbb{R}$ and $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$
5. $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ only if $\mathbf{u} = \mathbf{0}$

Especially, when field F is $\mathbb{R}$, then the conjugation is itself, and the inner product is real and all the scalar c mentioned above is restricted to be real, in which case we define a real inner product space.

Proposition 3.1. Basics result of inner product

1. $\langle \mathbf{u}, \mathbf{0} \rangle = \langle \mathbf{0}, \mathbf{u} \rangle = 0$
2. $\langle \mathbf{u}, c\mathbf{v} \rangle = \overline{c}\langle \mathbf{u}, \mathbf{v} \rangle \ \forall c \in F$
3. $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle \ \forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V$
4. Define the norm of $\mathbf{u} \in V$ to be $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle} \in [0, \infty)$, then $\|\mathbf{u}\|\|\mathbf{v}\| \geq \|\langle \mathbf{u}, \mathbf{v} \rangle\| \ \forall \mathbf{u}, \mathbf{v} \in V$, and the equality hold if and only if $\{\mathbf{u}, \mathbf{v}\}$ is linearly dependent (The Schwarz Inequality)
5. (The Triangle Inequality) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \ \forall \mathbf{u}, \mathbf{v} \in V$
6. (The parallelogram identity) $2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2 = \|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 \ \forall \mathbf{u}, \mathbf{v} \in V$

Proof. We prove 4 and 5.
The case $\|\mathbf{u}\| = 0$ or $\|\mathbf{v}\| = 0$ is trivial. Fix $\mathbf{u}, \mathbf{v} \in V \setminus \{\mathbf{0}\}$. By the property of conjugation, there exists $z \in \mathbb{C}$ such that $\|z\| = 1$ and $\|\langle \mathbf{u}, \mathbf{v} \rangle\| = z\langle \mathbf{v}, \mathbf{u} \rangle$. For any $a \in \mathbb{R}$, we have

$$\begin{aligned}
\langle \mathbf{u} - az\mathbf{v}, \mathbf{u} - az\mathbf{v} \rangle =& \langle \mathbf{u}, \mathbf{u} \rangle - az\langle \mathbf{v}, \mathbf{u} \rangle - a\overline{z}\langle \mathbf{u}, \mathbf{v} \rangle + a^2\langle \mathbf{v}, \mathbf{v} \rangle \\
=& \|\mathbf{u}\|^2 - az\langle \mathbf{v}, \mathbf{u} \rangle - \overline{az\langle \mathbf{v}, \mathbf{u} \rangle} + a^2\|\mathbf{v}\|^2 \\
=& \|\mathbf{u}\|^2 - a\|\langle \mathbf{u}, \mathbf{v} \rangle\| - \overline{a\|\langle \mathbf{u}, \mathbf{v} \rangle\|} + a^2\|\mathbf{v}\|^2 \\
=& \|\mathbf{u}\|^2 - 2a\|\langle \mathbf{u}, \mathbf{v} \rangle\| + a^2\|\mathbf{v}\|^2
\end{aligned}$$

LHS is real and non-negative, we have $\|\mathbf{u}\|^2 - 2a\|\langle \mathbf{u}, \mathbf{v} \rangle\| + a^2\|\mathbf{v}\|^2 \geq 0$ for any $a \in \mathbb{R}$. By the property of quadratic function, $\|\langle \mathbf{u}, \mathbf{v} \rangle\|^2 \leq \|\mathbf{u}\|^2\|\mathbf{v}\|^2$. The inequality follows.

We derive the condition for equality. Let $\mathbf{u}' = \frac{\mathbf{u}}{\|\mathbf{u}\|}$ and $\mathbf{v}' = \frac{\mathbf{v}}{\|\mathbf{v}\|}$ and take $z' \in \mathbb{N}^*$ such that $\|z'\| = 1$ and $\|\langle \mathbf{u}, \mathbf{v} \rangle\| = z\langle \mathbf{v}, \mathbf{u} \rangle$. We have

$$\begin{aligned} \|\langle \mathbf{u}, \mathbf{v} \rangle\|^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2 &\iff \|\langle \mathbf{u}', \mathbf{v}' \rangle\|^2 = 1 \\ &\iff \|\mathbf{u}' - z\mathbf{v}'\|^2 = \|\mathbf{u}'\|^2 + \|\mathbf{v}'\|^2 - 2\|\langle \mathbf{u}', \mathbf{v}' \rangle\| = 0 \\ &\iff \mathbf{u}' = z\mathbf{v}' \text{ i.e., } \mathbf{u} \text{ and } \mathbf{v} \text{ are linear dependent} \end{aligned}$$

Note that $\|\mathbf{u}' - z\mathbf{v}'\|^2 = \|\mathbf{u}'\|^2 + \|\mathbf{v}'\|^2 - 2\|\langle \mathbf{u}', \mathbf{v}' \rangle\|$ is the above when $a = 1$.
5 is given by

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle + \|\mathbf{u}\|^2 \leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 = (\|\mathbf{u}\| + \|\mathbf{v}\|)^2$$

□

Remark. It is immediate that an inner product space is also a normed vector space with the norm $\|x\| = \sqrt{\langle x, x \rangle}$.

Definition 3.2 (Hilbert space)**.** A Hilbert space $H := (H, \langle\cdot\rangle)$ is an inner vector space over field F and also a complete metric space with the above distance function. We also have complex Hilbert space ($F = \mathbb{C}$) and real Hilbert space ($F = \mathbb{R}$). Note that a Hilbert space is a special Banach space.

Example:

1. The set of n dimension real vector $\mathbb{R}^n$ ($n \in \mathbb{N}$) forms a real Hilbert space with the inner product being the dot product. It equips the real $\mathcal{L}^2$ space with an inner product.

2. The set of n dimension complex vector $\mathbb{C}^n$ ($n \in \mathbb{N}$) forms a complex Hilbert space with the inner product being the dot product defined as

$$\forall \mathbf{u} = [u_1, \cdots, u_n], \mathbf{v} = [v_1, \cdots, v_n] \in \mathbb{C}^n,\ \langle \mathbf{u}, \mathbf{v} \rangle := \mathbf{u} \cdot \mathbf{v} := \sum_{i=1}^{n} u_i \overline{v_i}$$

   It also equips the compact $\mathcal{L}^2$ space with an inner product.

3. In general, $\mathcal{L}^2(\mu, \{\mathbb{C}, \mathbb{R}\})$ space given the measure space $(S, \mathcal{B}, \mu)$, where the inner product is given by

$$\langle f, g \rangle := \int_S f\overline{g} \mathrm{d}\mu \ \forall f, g \in \mathcal{L}^2(\mu)$$

   We can check that it gives the same $\mathcal{L}^2$ norm.

Remark. Solving the equation of the parallelogram identity, we can find that a general $\mathcal{L}^p$ space can be extended into a Hilbert space only if $p = 2$.

Proposition 3.2. On a Hilbert space $(H, \langle\cdot\rangle)$ over $F \in \{\mathbb{R}, \mathbb{C}\}$, for any fixed $\mathbf{y} \in H$, the mappings

$$\mathbf{x} \to \langle \mathbf{x}, \mathbf{y} \rangle,\ \mathbf{x} \to \langle \mathbf{y}, \mathbf{x} \rangle,\ \mathbf{x} \to \|\mathbf{x}\|$$

are continuous.

3.2 Orthogonality

Definition 3.3 (Orthogonality)**.** Let V be an inner product space (either complex or real) and $\mathbf{u}, \mathbf{v} \in V$. We say $\mathbf{u}$ is orthogonal to $\mathbf{v}$ if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ (or equivalently $\langle \mathbf{v}, \mathbf{u} \rangle = 0$), write $\mathbf{v} \perp \mathbf{u}$
Fix $\mathbf{u} \in V$. Define $\mathbf{u}^\perp$ to be the orthogonal compliment, i.e. the set of all vector orthogonal to it.

$$\mathbf{u}^\perp := \{\mathbf{v} \in V : \mathbf{v} \perp \mathbf{u}\}$$

Observe that $\mathbf{u}^\perp$ is closed under vector addition and scalar multiplication, thus $\mathbf{u}^\perp$ is a subspace of V.
If $S \subseteq V$, define $S^\perp = \bigcap_{\mathbf{u} \in S} \mathbf{u}^\perp$ to be the orthogonal compliment. It is also a subspace.

Proposition 3.3. For any $\mathbf{u} \in V$ and $S \subseteq V$,

1. $\mathbf{0} \in \mathbf{u}^\perp$ and $\mathbf{0} \in S^\perp$, $\mathbf{0}^\perp = V$
2. $S^\perp$ is closed
3. $S^\perp = \operatorname{span}[S]^\perp = \overline{\operatorname{span}[S]}^\perp$ (The bar denote closure)
4. $\overline{\operatorname{span}[S]} \subseteq (S^\perp)^\perp$. If $\dim[V] < \infty$, $(S^\perp)^\perp = \operatorname{span}[S]$
5. $S^\perp \cap \overline{\operatorname{span}[S]} = \{\mathbf{0}\}$
6. if $\dim[S] = n \in \mathbb{N}$, $\dim[S^\perp] + \dim[S] = n$

Lemma 3.1. Let H be a Hilbert space. If $S \subseteq H$ is non-empty, closed and convex, there exists a unique vector in S with smallest norm. Note that by saying S is convex, it means that

$$\forall \mathbf{x}, \mathbf{y} \in S, \lambda \in (0,1),\ \lambda \mathbf{x} + (1-\lambda)\mathbf{y} \in S$$

Clearly any subspace is convex.

Proof. Let $d = \inf\{\mathbf{x} : \mathbf{x} \in S\}$. Apply the parallelogram identity to $\frac{\mathbf{x}}{2}$ and $\frac{\mathbf{y}}{2}$, where $\mathbf{x}, \mathbf{y} \in S$,

$$\frac{1}{4}\|\mathbf{x}-\mathbf{y}\|^2 = \frac{1}{2}\|\mathbf{x}\|^2 + \frac{1}{2}\|\mathbf{y}\|^2 - \left\|\frac{\mathbf{x}+\mathbf{y}}{2}\right\|^2 \le \frac{1}{2}\|\mathbf{x}\|^2 + \frac{1}{2}\|\mathbf{y}\|^2 - d \text{ (since } \frac{\mathbf{x}+\mathbf{y}}{2} \in S)$$

When $\|\mathbf{x}\| = \|\mathbf{y}\| = d$, we have $\|\mathbf{x}-\mathbf{y}\| = 0$, which gives the uniqueness. We continue to prove the existence.
By the definition of d, there exists a sequence $\{\mathbf{x}_n\}_{n=1}^\infty$ with $\{\|\mathbf{x}_n\|\}_{n=1}^\infty$ monotonic descreasing such that $\lim_{n\to\infty} \|\mathbf{x}_n\| = d$. For any $\epsilon > 0$, take $N \in \mathbb{N}^*$ such that $\|\mathbf{x}_n\|^2 < d + \frac{1}{4}\epsilon^2\ \forall n \ge N$. Then $\forall n, m \ge N$, $\|\mathbf{x}_n - \mathbf{x}_m\| \le \epsilon$ by the formula. Thus $\{\mathbf{x}_n\}_{n=1}^\infty$ is Cauchy. Given that S is closed, it converges to, say, $\mathbf{x}_0 \in S$. We have $\|\mathbf{x}_0\| = d$, which is what we need. □

Proposition 3.4 (Orthogonal projection)**.** Let W be a nonempty closed subspace of Hilbert space H. For any $\mathbf{x} \in H$, there exists a unique decomposition $\mathbf{x} = \mathbf{u} + \mathbf{v}$ such that $\mathbf{u} \in W$ and $\mathbf{v} \in W^\perp$, where $\mathbf{u}$ is the orthogonal projection of $\mathbf{x}$ onto W
Note that although a subspace in general is not closed, the closure of a subspace is still a subspace, where the statement applies.

Proof. Fix an arbitrary $\mathbf{x} \in H$. The set $W' := \{\mathbf{x} + \mathbf{y} : \mathbf{y} \in W\}$ is closed and convex. By the above lemma, let $\mathbf{v} \in W'$ be the one with the smallest norm, and $\mathbf{u} = \mathbf{x} - \mathbf{v}$. Clearly $\mathbf{u} \in W$. We show that $\mathbf{v} \in W^\perp$. For any $\mathbf{y} \in W$, take $\alpha := \frac{\langle \mathbf{v}, \mathbf{y} \rangle}{\|\mathbf{y}\|}$. We have

$$\begin{aligned} &\langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{v}\|^2 \leq \mathbf{v} - \alpha \|\mathbf{y}\|^2 = \langle \mathbf{v} - \alpha \mathbf{y}, \mathbf{v} - \alpha \mathbf{y} \rangle \text{ (since } \mathbf{v} \text{ has the smallest norm)} \\ \Rightarrow\ & 0 \leq \alpha \bar{\alpha} \|\mathbf{y}\|^2 - \alpha \langle \mathbf{y}, \mathbf{v} \rangle - \bar{\alpha} \langle \mathbf{v}, \mathbf{y} \rangle = \langle \mathbf{v}, \mathbf{y} \rangle \langle \mathbf{y}, \mathbf{v} \rangle - \langle \mathbf{v}, \mathbf{y} \rangle \langle \mathbf{y}, \mathbf{v} \rangle - \langle \mathbf{v}, \mathbf{y} \rangle \langle \mathbf{y}, \mathbf{v} \rangle = -\|\langle \mathbf{v}, \mathbf{y} \rangle\|^2 \end{aligned}$$

Thus, $\langle \mathbf{v}, \mathbf{y} \rangle = 0$. It remains to prove the uniqueness. Suppose $\exists \mathbf{u}, \mathbf{u}' \in S, \mathbf{v}, \mathbf{v}' \in S^\perp$ such that $\mathbf{x} = \mathbf{u} + \mathbf{v} = \mathbf{u}' + \mathbf{v}'$. Then $\mathbf{u} - \mathbf{u}' = \mathbf{v} - \mathbf{v}'$. Since $\mathbf{u} - \mathbf{u}' \in W$ and $\mathbf{v} - \mathbf{v}' \in W^\perp$. It is only possible that both are $\mathbf{0}$, which gives that $\mathbf{u} = \mathbf{u}'$ and $\mathbf{v} = \mathbf{v}'$. □

Corollary 3.1. Let W be a nonempty closed subspace of Hilbert space H. We have $W = (W^\perp)^\perp$. Further, when $S \subseteq W$ is nonempty, $(S^\perp)^\perp = \overline{\text{span}[S]}$.

Proof. It remains to show $(W^\perp)^\perp \subseteq W$. Let $\mathbf{x} \in (W^\perp)^\perp$ and $\mathbf{x} = \mathbf{u} + \mathbf{v}$ be the unique decomposition such that $\mathbf{u} \in W$ and $\mathbf{v} \in W^\perp$. Then

$$\|\mathbf{v}\|_H = \langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{x}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{v} \rangle = 0 - 0 = 0$$

Thus $\mathbf{x} = \mathbf{u} \in W$. The second result is because $\overline{\text{span}[S]}$ is closed and $S^\perp = \overline{\text{span}[S]}^\perp$. □

Theorem 3.1 (Riesz representation theorem)**.** Let $(H, \langle \cdot \rangle)$ be a Hilbert space over $F \in \{\mathbb{R}, \mathbb{C}\}$. For any continuous linear map (aka. continuous linear functional) $T : H \mapsto F$, there exists a unique $\mathbf{x}_T \in H$ such that

$$T(\mathbf{x}) = \langle \mathbf{x}, \mathbf{x}_T \rangle \ \forall \mathbf{x} \in H$$

Furthermore, the norm of the map equals to the Hilbert norm of $\mathbf{x}_T$, i.e., $\|T\|_{H \to F} = \|\mathbf{x}_T\|$.

Proof. If the kernel $\ker(f) = 0$, take $\mathbf{x}_T = \mathbf{0}$. Otherwise, we claim that

$$\mathbf{x}_T = \overline{T(\mathbf{x}_0)} \cdot \mathbf{x_0}, \text{ where } \mathbf{x}_0 \in \ker(f)^\perp \text{ is choosen arbitrary with } \|\mathbf{x}_0\| = 1$$

With such, $\mathbf{x}_T \in \ker(f)^\perp$, and since

$$T\left[\frac{T(\mathbf{x})}{T(\mathbf{x}_0)}\mathbf{x}_0 - \mathbf{x}\right] = 0, \text{i.e, } \frac{T(\mathbf{x})}{T(\mathbf{x}_0)}\mathbf{x}_0 - \mathbf{x} \in \ker(T)$$

We have
$$\left\langle \frac{T(\mathbf{x})}{T(\mathbf{x}_0)}\mathbf{x}_0 - \mathbf{x}, \mathbf{x}_T \right\rangle = 0, \text{ and thus } \left\langle \frac{T(\mathbf{x})}{T(\mathbf{x}_0)}\mathbf{x}_0, \mathbf{x}_T \right\rangle = \langle \mathbf{x}, \mathbf{x}_T \rangle$$

Then,

$$\forall \mathbf{x} \in H,\ T(\mathbf{x}) = \frac{T(\mathbf{x})}{T(\mathbf{x}_0)} T\left(\mathbf{x}_0\right) \langle \mathbf{x}_0, \mathbf{x}_0 \rangle = \frac{T(\mathbf{x})}{T(\mathbf{x}_0)} \langle \mathbf{x}_0, \overline{T(\mathbf{x}_0)}\mathbf{x}_0 \rangle = \left\langle \frac{T(\mathbf{x})}{T(\mathbf{x}_0)}\mathbf{x}_0, \mathbf{x}_T \right\rangle = \langle \mathbf{x}, \mathbf{x}_T \rangle$$

We then prove the uniqueness of $\mathbf{x}_T$. Suppose both $\mathbf{x}_T$ and $\mathbf{x}'_T$ fulfill the requirement, such that

$$\forall \mathbf{x} \in H,\ \langle \mathbf{x}, \mathbf{x}_T - \mathbf{x}'_T \rangle = 0$$

Taking $\mathbf{x} = \mathbf{x}_T - \mathbf{x}'_T$, it gives that $\|\mathbf{x}_T - \mathbf{x}'_T\| = 0$.
It remains to show the operator norm.

$$\|T\|_{H \mapsto F} = \sup \left\{ \|\langle \mathbf{x}, \mathbf{x}_T \rangle\| : \mathbf{x} \in H \text{ and } \|\mathbf{x}\| = 1 \right\} \geq \left\| \left\langle \frac{\mathbf{x}_T}{\|\mathbf{x}_T\|}, \mathbf{x}_T \right\rangle \right\| = \|\mathbf{x}_T\|$$

Also, $\|\langle \mathbf{x}, \mathbf{x}_T \rangle\| \leq \|\mathbf{x}\| \|\mathbf{x}_T\| = \|\mathbf{x}_T\|$ under $\|\mathbf{x}\| = 1$. The equation of norm follows. □

3.3 Orthonormal set and orthonormal basis

Definition 3.4 (Orthonormal set)**.** Let V be an inner product space (either real or complex). A non-empty subset $S \subseteq V$ is orthonormal if for any $\mathbf{u}, \mathbf{v} \in S$, $\|\mathbf{u}\| = 1$ and $\mathbf{u} \perp \mathbf{v}$ if $\mathbf{u} \neq \mathbf{v}$.
Observe that a orthonormal set is linearly independent. S is the orthonormal basis for span$[S]$.

Example: The Gram-Schmidt process to orthonormal basis in finite case
Let $S := \{\mathbf{u}_1, \cdots, \mathbf{u}_n\} \subseteq V$ be a basis for $T := \text{span}[S]$. Then $S' := \{\mathbf{u}'_1, \cdots, \mathbf{u}'_n\}$ defined by

$$\mathbf{u}'_1 := \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}, \text{ and for } 2 \leq i \leq n, \ \mathbf{u}'_i := \frac{1}{\left\|\mathbf{u}_i - \sum_{j=1}^{i-1}(\langle \mathbf{u}_i, \mathbf{u}'_j\rangle)\mathbf{u}'_j\right\|}\left(\mathbf{u}_i - \sum_{j=1}^{i-1}(\langle \mathbf{u}_i, \mathbf{u}'_j\rangle)\mathbf{u}'_j\right)$$

Definition 3.5 (Fourier coefficient)**.** Let $S := \{\mathbf{u}_i : i \in I\}$ be a orthonormal set in the Hlibert space H over $F \in \{\mathbb{R}, \mathbb{C}\}$. We associate each $\mathbf{x} \in H$ with some Fourier coefficients in F, defined as

$$\hat{\mathbf{x}}_i := \langle \mathbf{x}, \mathbf{u}_i \rangle \ \forall i \in I$$

Proposition 3.5 (Bessel inequality in the finite case)**.** Let $S := \{\mathbf{u}_i\}_{i=1}^n$ be a finite and nonempty orthonormal set in the Hlibert space H and $\mathbf{x} \in H$ is arbitrary.

1. The (unique) projection of $\mathbf{x}$ onto span$[S]$ is

$$\mathbf{u} := \sum_{i=1}^n \hat{\mathbf{x}}_i \mathbf{u}_i$$

2. $\sum_{i=1}^n \|\hat{\mathbf{x}}_i\|^2 \leq \|\mathbf{x}\|^2$. The equality holds if and only if $\mathbf{x} \in \text{span}[S]$

Proof. Clearly $\mathbf{u} \in \text{span}[S]$ and span$[S]$ is closed. Using proposition 3.4, it remains to check $\langle \mathbf{x} - \mathbf{u}, \mathbf{u}_i \rangle \ \forall i \in \{1, \cdots, n\}$, thus $\mathbf{x} - \mathbf{u} \in S^\perp = \text{span}[S]^\perp$. 2 follows from 1,

$$\|\mathbf{x}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{x} - \mathbf{u}\|^2 \geq \|\mathbf{u}\|^2 = \sum_{i=1}^n \|\hat{\mathbf{x}}_i\|^2 \|\mathbf{u}_i\|^2 = \sum_{i=1}^n \|\hat{\mathbf{x}}_i\|^2$$

The equality holds if and only if $\|\mathbf{x} - \mathbf{u}\|^2 = 0$ i.e, $\mathbf{x} = \mathbf{u} \in \text{span}[S]$. □

The following discuss extends the properties for orthogonal set of infinite cardinality, then gives the condition for the existence of countable orthonormal Hilbert basis. We will use this notation of infinite summation of the set $\{a_i \in [0, \infty) : i \in I\}$,

$$\sum_{i \in I} a_i := \sup \left\{ \sum_{j \in J} a_j : J \subset I \text{ is a finite subset} \right\} = \int_I a_i \mathrm{d}c(i)$$

where the integral is Lebesgue with the counting measure $c(\cdot)$ on the measurable space $(I, \mathcal{P}(I))$. Then the next theorem follows from its finite case.

Theorem 3.2 (Bessel inequality)**.** Let $S := \{\mathbf{u}_i : i \in I\}$ be a nonempty orthonormal set in the Hlibert space H over $F \in \{\mathbb{R}, \mathbb{C}\}$. For any $\mathbf{x} \in H$, we have

$$\sum_{i \in I} \|\langle \mathbf{x}, \mathbf{u}_i \rangle\|^2 = \sum_{i \in I} \|\hat{\mathbf{x}}_i\|^2 \leq \|\mathbf{x}\|^2$$

Let $W := \text{span}[S]$. If $\mathbf{x} \in \overline{W}$, the equality holds.

Proof. The inequality follows from its finite case and taking the supremum. If $\mathbf{x} \in W$, say, $\mathbf{x} = \sum_{j\in J} a_j \mathbf{u}_j$ for some finite $J \subset I$ and $\{a_j \in F : j \in J\}$, then

$$\sum_{i\in I} \|\langle \mathbf{x}, \mathbf{u}_i\rangle\|^2 = \sum_{i\in I} \left\| \left\langle \sum_{j\in J} a_j \mathbf{u}_j, \mathbf{u}_i \right\rangle \right\|^2 = \sum_{i\in I} \|a_i\|^2 \mathbf{1}_{i\in J} = \sum_{j\in J} \|a_j\|^2 = \|\mathbf{x}\|^2$$

If $\mathbf{x} \in \overline{W}$, suppose $\mathbf{x} = \lim_{n\to\infty} \mathbf{x}_n$ for some $\{\mathbf{x}_n \in W\}_{n=1}^{\infty}$. We have proved that

$$\|\mathbf{x}_n\|^2 = \sum_{i\in I} \|\hat{\mathbf{x}_{ni}}\|^2 = \int_I \|\hat{\mathbf{x}_{ni}}\|^2 \mathrm{d}c(I) < \infty \ \forall n \in \mathbb{N}^*$$

For any $n \in \mathbb{N}^*$, consider the function $f_n : I \mapsto [0, \infty)$ by

$$f_n(i) := \|\hat{\mathbf{x}_{ni}}\| \ \forall i \in I$$

We note that the integral form is exactly the squared $\mathcal{L}^2$ norm of f with counting measure space $(I, \mathcal{P}(I), c)$. We have $\|\mathbf{x}_n\|^2 = \|f_n\|_2^2$. By the continuity of norm and inner product operation,

$$\begin{aligned}\|\mathbf{x}\|^2 = \left\| \lim_{n\to\infty} \mathbf{x}_n \right\|^2 = \lim_{n\to\infty} \|\mathbf{x}_n\|^2 = \lim_{n\to\infty} \|f_n\|_2^2 = \left\| \lim_{n\to\infty} f_n \right\|_2^2 &= \left\| \lim_{n\to\infty} \|\langle \mathbf{x}_n, \mathbf{u}_i\rangle\| \right\|_2^2 \\ &= \int_I \|\langle \mathbf{x}, \mathbf{u}_i\rangle\|^2 \mathrm{d}c(i)\end{aligned}$$

The equality follows. $\square$

Similar to linearly independent set, the orthonormal space cannot be arbitrarily large by Hausdorff maximal principle (Theorem 1.4).

Proposition 3.6 (Maximal orthonormal set)**.** Let S be an orthonormal set of the inner product space V. There exists a maximal orthonormal set S' of V containing S, which means that it cannot include another vector to form a larger orthonormal set. Such S' is called a complete orthonormal set or an orthonormal basis.

Proof. Let $\mathcal{S}$ be the collection of all orthonormal set of V containing S. Clearly, $S \subseteq \mathcal{S}$ and $\mathcal{S}$ is a partially order space with subset $\subseteq$ relation. The Hausdorff maximal principle gives the existance of a maximal chain $\mathcal{C} \subseteq \mathcal{S}$ Let $S' = \bigcup_{C\in\mathcal{C}} C$. Clearly $S \subseteq S'$. We show that S' is a maximal orthonormal set.
For any $\mathbf{u}, \mathbf{v} \in S'$, there exists $S_1, S_2 \in \mathcal{C}$ such that $\mathbf{u} \in S_1$ and $\mathbf{v} \in S_2$. Suppose $S_1 \subseteq S_2$ (the other direction is similar). We have $\mathbf{u}, \mathbf{v} \in S_2$, then $\mathbf{u}, \mathbf{v}$ satisfy the orthonormal condition, i.e., $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$ and if $\mathbf{u} \neq \mathbf{v}$, $\mathbf{u} \perp \mathbf{v}$. Thus S' is an orthonormal set. By the maximality of $\mathcal{C}$, S' is also maximal in the sense that any set $S^* \subseteq V$ that properly contains S' cannot be an orthonormal set. $\square$

Remark. It makes sense to define orthonormal base for a subspace W of V by the maximality, i.e. $B \subset H$ is a orthonormal basis for W if it is an orthonormal set and

$$\forall \mathbf{x} \in V \text{ such that } \ \mathbf{x} \perp \mathbf{u} \ \forall \mathbf{u} \in B, \text{ then } \mathbf{x} \in W^{\perp}$$

is. Since $W^{\perp} = \overline{W}^{\perp}$, a orthonormal basis of W is also a orthonormal basis of $\overline{W}$. A orthonormal set $B \subset V$ is a orthonormal basis of span$[W]$ and $\overline{\text{span}[W]}$.

Corollary 3.2. Let W be a proper subspace of an inner product space V. If B_1 is an orthonormal basis for W and B_2 is an orthonormal basis for $W^\perp$, then $B_1 \cup B_2$ is an orthonormal basis for V.

Proof.

Clearly $B_1 \cup B_2$ is an orthonormal set, it is also maximal because if $\mathbf{x} \perp \mathbf{u}\ \forall \mathbf{u} \in B_1 \cup B_2$ for some $\mathbf{x} \in V$, we have $\mathbf{x} = \mathbf{0}$. □

Theorem 3.3 (Property of orthonormal basis)**.** Let $S := \{\mathbf{u}_i : i \in I\} \subseteq H$ be an orthonormal set of Hilbert space H. The followings are equivalent:

1. S is maximal, or it is an orthonormal basis
2. span$[S]$ is dense in H
3. (Bessel identity) For any $\mathbf{x} \in H$,
$$\sum_{i\in I} \|\hat{\mathbf{x}}_i\|^2 = \|\mathbf{x}\|^2$$
4. (Parseval's identity) For any $\mathbf{x}, \mathbf{y} \in H$,
$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i\in I} \hat{\mathbf{x}}_i \overline{\hat{\mathbf{y}}_j}$$

Proof. 1 implies 2, because if otherwise $H \setminus \overline{\text{span}[S]} \neq \emptyset$, proposition 3.4 gives that $\overline{\text{span}[S]}^\perp \neq \emptyset$, thus S is not maximal.
We have prove that 2 implies 3 in Theorem 3.2.
3 gives 4 by expressing the norm term of the following identity into the form of 3,

$$4\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2 + \mathrm{i}\|\mathbf{x} + \mathrm{i}\mathbf{y}\|^2 - \mathrm{i}\|\mathbf{x} - \mathrm{i}\mathbf{y}\|^2\ \forall \mathbf{x}, \mathbf{y} \in H \text{ (polarization identity)}$$

4 gives 1, because if otherwise $\exists \mathbf{u} \in S^\perp$, $\langle \mathbf{u}, \mathbf{u} \rangle$ do not satisfies the identity of 4. □

Note that orthonormal basis is not the (Hamel) basis that relies on finite linear combination. In fact, it can be proved that a Hamel basis of a Banach space has either finite or uncountably infinite cardinality. We may use another type of basis, which allows linear combination of infinite terms.

Definition 3.6 (Countable Schauder basis)**.** Let V be a vector space over F. A countable collection $\{\mathbf{e}_n\}_{n=1}^\infty$ is a Schauder basis of V if for any $\mathbf{x} \in V$, there exists a unique sequence $\{a_n \in F\}_{n=1}^\infty$ such that
$$\mathbf{x} = \sum_{n=1}^\infty a_n \mathbf{e}_n$$

Remark. The above defines for the countably infinite case. The finite case agrees on the Hamel basis, and extension to the uncountable case is also possible.
To satisfy convergence, the Schauder basis $\{\mathbf{e}_n\}_{n=1}^\infty$ should be ordered as a sequence.
The countable Schauder basis may not exist.

Proposition 3.7. Properties of countable Schauder basis

1. A Schauder basis $\{\mathbf{e}_n\}_{n=1}^\infty$ is necessarily linear independent

2. A subset $W \subseteq V$ is total if $\text{span}[W]$ is dense in V. When V is a normed vector space, $\{\mathbf{e}_n\}_{n=1}^{\infty}$ is necessarily total

3. If V is a normed vector space and has a countable Schauder basis, then V is separable, i.e., there exists a countable and dense subset of V.

Proof. 1 is obvious. Note that the span only count finite linear combination. For any $\mathbf{x} \in V$ and $\epsilon > 0$, we can truncate onto $\{1, \cdots, N\}$ to satisfy $\left\|\mathbf{x} - \sum_{n=1}^{N} \alpha_n \mathbf{e}_n\right\|_V < \epsilon$. 3 follows from 2, since the following set is countable and dense in real normed vector space,

$$\left\{\sum_{n=1}^{N} p_n \mathbf{e}_n : N \in \mathbb{N}^*, \{p_n \in \mathbb{Q}\}_{n=1}^{N}\right\}$$

For complex normed vector space, just replace p_n with $p_n + \mathrm{i}q_n$, where $p_n, q_n \in \mathbb{Q}$. □

Proposition 3.8. Let H be a Hilber space over $F \in \{\mathbb{R}, \mathbb{C}\}$. If H is separable, then

1. Any of its orthonormal set has at most countable cardinality

2. A orthonormal basis given by Hausdorff maximal principle is a Schauder basis

Proof. We first prove 1. First note that a separable space cannot contain an uncountable number of mutually disjoint open sets, since there exists a bijective map between any open set and one of its element in the dense set.
Suppose otherwise $B \subseteq H$ is an uncountable orthonormal set. The set

$$\left\{B\left(\mathbf{e}, \frac{1}{2}\right) : \mathbf{e} \in B\right\}$$

is a collection of uncountable open balls that are mutually disjoint, contradicting to the fact.
Therefore, when H is of infinite dimension, we can express its orthonormal basis as $\{\mathbf{u}_i\}_{i=1}^{\infty}$. For any $n \in \mathbb{N}^*$,

$$\mathbf{u}^{(n)} := \sum_{i=1}^{n} \hat{\mathbf{x}}_i \mathbf{u}_i$$

is the projection onto $\text{span}[\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}]$. We have $\lim_{n\to\infty} \mathbf{u}^{(n)}$ by the Bessel identity, then

$$\lim_{n\to\infty} \left\|\mathbf{x} - \mathbf{u}^{(n)}\right\| = \lim_{n\to\infty} \left(\|\mathbf{x}\| - \left\|\mathbf{u}^{(n)}\right\|\right) = 0$$

Thus,

$$\mathbf{x} = \lim_{n\to\infty} \mathbf{u}^{(n)} = \sum_{i=1}^{\infty} \langle \mathbf{x}, \mathbf{u}_i \rangle \mathbf{u}_i \text{ is the representation}$$

To prove the uniqueness, suppose $\mathbf{x} = \sum_{i=1}^{\infty} a_i \mathbf{u}_i = \sum_{i=1}^{\infty} b_i \mathbf{u}_i$ for some $\{a_i \in F\}_{i=1}^{\infty}$ and $\{b_i \in F\}_{i=1}^{\infty}$, then

$$0 = \left\|\sum_{i=1}^{\infty} (a_i - b_i) \mathbf{u}_i\right\|^2 = \sum_{i=1}^{\infty} \|a_i - b_i\|^2 \;\Rightarrow\; \forall i \in \mathbb{N}^*,\; a_i = b_i$$

□

4 Matrix

4.1 Matrix and determinant

Definition 4.1 (Matrix)**.** A matrix A is n rows and m columns of values in $\mathbb{C}$ (or $\mathbb{R}$, same for below, though we focus on the general case mostly), where $n, m \in \mathbb{N}^*$. Use $A_{i,j}$ to denote its (i,j)-th entry. It can be regarded as a collection of m column vectors of $\mathbb{C}^n$ or n row vectors of $\mathbb{C}^m$. A $\mathbb{C}^n$ vector is a $n \times 1$ matrix. Also,

- Let $\mathbb{C}^{n\times m}$ be the set of n by m matrices with complex entries, and $\mathbb{R}^{n\times m}$ be those with real entries ($n, m \in \mathbb{N}^*$)
- Define addition, multiplication between matrix and matrix, and between matrix and scalars. Define the transpose, diagonal and trace of a matrix.
- Use I_n, O_n (or simply I, O) to denote the n by n identity matrix and null matrix respectively
- A matrix $A \in \mathbb{C}^{n,m}$ can be regarded as a linear transformation $A : \mathbb{C}^m \mapsto \mathbb{C}^n$, defined by $f(\mathbf{x}) = A\mathbf{x}\ \forall \mathbf{x} \in \mathbb{C}^m$.
- Define the rank of a matrix $\text{rank}(A)$ to be the dimension of the subspace in $\mathbb{C}^n$ generated by the column vectors of A
- Define the nullity of a matrix $\text{null}(A)$ to be the dimension of the kernel of the linear transformation A

Proposition 4.1. about matrix rank and nullity

1. Column rank (the one we have defined) is equal to row rank (the dimension of the subspace in $\mathbb{R}^m$ (or $\mathbb{C}^m$) generated by the row vectors)
2. $\text{rank}(A) + \text{null}(A) = m$

Sketch of proof: we can prove that the step of Gauss elimination does not change the subspace generated by the row vectors, and preserves both the column rank and nullity. The statements are clear when the matrix is in its Reduced Row Echelon Form (*RREF*).

Definition 4.2 (Elementary matrices)**.** We define the following three types of elementary square matrices in $\mathbb{C}^{n\times n}$

- Type I: those corresponding to the transformation of interchanging row 1 and j ($j \neq 1$), i.e., $A_{i,i} = 1\ \forall i \in \{2, \cdots, n\} \setminus \{j\}$, $A_{1,1} = A_{j,j} = 0$, and $A_{j,1} = A_{1,j} = 1$.
- Type II: those corresponding to the transformation of scalar multipling row 1, i.e., $A_{i,i} = 1\ \forall i \in \{2, \cdots, n\}$ and $A_{1,1} = c$.
- Type III: those corresponding to the transformation of adding $c \in \mathbb{C} \setminus \{0\}$ times row j to row 1 ($j \neq 1$), $A_{i,i} = 1\ \forall i \in \{1, \cdots, n\}$ and $A_{1,j} = c$.

Definition 4.3 (Matrix inverse)**.** Given $A \in \mathbb{C}^{n\times n}$ if there exists $B \in \mathbb{C}^{n\times n}$ such that $AB = I = BA$, where I denotes the $n \times n$ identity matrix, then A is invertible, and B is the inverse of A, denoted as

$$B = A^{-1}, \text{ and conversely, } B = A^{-1}$$

If such B does not exists, A is singular.

Lemma 4.1. If $A, B \in \mathbb{C}^{n\times n}$, $BA = I$, and $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ is a basis for $\mathbb{C}^n$, then $\{A\mathbf{x}_1, \cdots, A\mathbf{x}_n\}$ is also its basis.

Proof. It suffices to prove the new set is linearly independent. Otherwise if $\sum_{i=1}^n c_i A\mathbf{x}_i = \mathbf{0}$ for some $\{c_i \in \mathbb{C}\}_{i=1}^n$, we have $\sum_{i=1}^n c_i BA\mathbf{x}_i = \mathbf{0}$ and thus $\sum_{i=1}^n c_i\mathbf{x}_i = \mathbf{0}$, contradicting to our assumption. $\square$

Corollary 4.1. Further from Lemma 4.1,

1. For any $\mathbf{y} \in \mathbb{C}^n$, there exists $\mathbf{x} \in \mathbb{C}^n$ such that $A\mathbf{x} = \mathbf{y}$. If $\mathbf{y} = \sum_{i=1}^n c_i A\mathbf{x}_i$, then take $\mathbf{x} = \sum_{i=1}^n c_i\mathbf{x}_i$, which is the unique choice given the basis $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$
2. $AB = I$. Because $\forall \mathbf{y} \in \mathbb{C}^n$, $\exists \mathbf{x} \in \mathbb{C}^n$ such that $AB\mathbf{y} = ABA\mathbf{x} = A\mathbf{x} = I\mathbf{y}$. Conversely, $AB = I$ also implies $BA = I$.
3. The inverse is unique, since $AB = BA = AC = CA = I \;\Rightarrow\; B = BAC = C$. Then the set of all invertible matrices of $\mathbb{C}^{n,n}$ forms a group with matrix multiplication.
4. if A is invertible, $\text{null}(A) = 0$. It follows from 1.

Proposition 4.2. about matrix inverse

1. Elementary matrices are all invertible. The inverse is also a elementary matrix of the same type.
2. If $A, B \in \mathbb{C}^{n,n}$ are invertible, $(cA)^{-1} = c^{-1}A^{-1}$ ($c \in \mathbb{C}$ and $c \neq 0$), $(A^T)^{-1} = (A^{-1})^T$ $(AB)^{-1} = B^{-1}A^{-1}$.
3. A matrix $A \in \mathbb{C}^{n,n}$ is invertible if and only if it is a product of a finite number of elementary matrices, say, $A = E_1E_2\cdots E_m$. And
$$A^{-1} = E_m^{-1}\cdots E_1^{-1}$$

1 and 2 gives $\Leftarrow$ of 3. And we have A is invertible $\Rightarrow\;$ null$[A] = 0 \;\;\Rightarrow \{E_j\}_{j=1}^m$ exists by Gauss elimination into *RREF*.

Definition 4.4 (Determinant)**.** Let $A \in \mathbb{C}^{n\times n}$. If $n = 1$, define the determinant of A to be $\det A := |A| := A_{1,1}$. If $n \geq 2$, suppose we have defined the determinant of any $(n-1)\times(n-1)$ matrix to be a value in $\mathbb{C}$. Define the (i, j)-cofactor of A to be $(-1)^{i+j}|A_{(i,j)}|$, where $A_{(i,j)} \in \mathbb{C}^{(n-1)\times(n-1)}$ is obtained by deleting the i-th row and j-th column of A. Then the collection of cofactors forms a cofactor matrix $C \in \mathbb{C}^{n\times n}$ with entries $C_{i,j} := (-1)^{i+j}|A_{(i,j)}|$. The determinant of A is defined by

$$\det A := |A| := \sum_{j=1}^n A_{1,j} \times C_{1,j}$$

Proposition 4.3. Determinant maps n row vectors $\{\mathbf{v}_i \in \mathbb{C}^n\}_{i=1}^n$of matrix A into $\mathbb{C}$.

1. If $\mathbf{v}_i = a\mathbf{u} + b\mathbf{w}$ for one $i \in \{1, \cdots, n\}$, where $a, b \in \mathbb{C}, \mathbf{u}, \mathbf{w} \in \mathbb{C}^n$

$$\det(A) = \det\begin{pmatrix}\mathbf{v}_1\\ \vdots\\ \mathbf{v}_i\\ \vdots\\ \mathbf{v}_n\end{pmatrix} = a\det\begin{pmatrix}\mathbf{v}_1\\ \vdots\\ \mathbf{u}\\ \vdots\\ \mathbf{v}_n\end{pmatrix} + b\det\begin{pmatrix}\mathbf{v}_1\\ \vdots\\ \mathbf{w}\\ \vdots\\ \mathbf{v}_n\end{pmatrix}$$

2. If $\mathbf{v}_i = \mathbf{v}_j$ for some $i \neq j$, $\det(A) = 0$

3. The determinant multiplies by -1 after swapping row i with row j $(i \neq j)$

4. The determinant does not change after adding a scalar multiple of row j into row i

5. If $\mathbf{v}_i = \mathbf{0}$ for some i, $\det(A) = 0$

6. For any $i \in \{1, \cdots, n\}$, $\det(A) = \sum_{j=1}^{n} A_{i,j} C_{i,j}$, where C is the cofactor matrix of A

Proof. 1 follows from the definition using induction. Note that 1&2 $\Rightarrow$ 3, since

$$0 = \det \begin{pmatrix} \vdots \\ \mathbf{v}_i + \mathbf{v}_j \\ \vdots \\ \mathbf{v}_i + \mathbf{v}_j \\ \vdots \end{pmatrix} = \det \begin{pmatrix} \vdots \\ \mathbf{v}_i \\ \vdots \\ \mathbf{v}_j \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ \mathbf{v}_j \\ \vdots \\ \mathbf{v}_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ \mathbf{v}_i \\ \vdots \\ \mathbf{v}_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ \mathbf{v}_j \\ \vdots \\ \mathbf{v}_j \\ \vdots \end{pmatrix}$$

The last two terms are 0, which implies the first two terms are additive inverse of each other.
We do induction on 2 and 3 together. Obviously, they hold on $n = 2$. Suppose they hold on $n = k$ $(k \geq 2)$, now we prove that they hold on $n = k + 1$. If the identical rows are below row 1, the determinant is 0 by our assumption and the definition. If row 1 is identical to row $i > 1$, swap row i with row 2 to obtain A', we have

$$\det(A') = \sum_{i=1}^{n} A'_{1,i} C'_{1,i} = \sum_{i=1}^{n} -A_{1,i} C_{1,i} = -\det(A)$$

by our assumption that 3 holds on $n = k$. It suffices to prove that the matrix A' with identical row 1 and row 2 has determinant 0. Observe

$$\det(A') = \sum_{i=1}^{n} A'_{1,i} \times (-1)^{1+i} \sum_{j \neq i} A'_{1,j} \times (-1)^{1+j-\mathbf{1}_{\{j>i\}}} \det\left(A'_{(1,i),(2,j)}\right)$$

where $A'_{(1,i),(2,j)}$ is the matrix obtain by deleting row $1, 2$ and column i, j of A'. We can show that the coefficient of $A'_{1,i} A'_{1,j}$ are all 0. The result follows.
4 follows from 1 and 2. 5 follows from 2 and 4. Permutate row index $\{1, \cdots, i\}$ into $\{i, 1, 2 \cdots, i-1\}$, which involves $i - 1$ swapping, 6 follows. $\square$

Theorem 4.1. Let $A, B \in \mathbb{C}^{n \times n}$ be two matrix. $\det(AB) = \det(A)\det(B)$. Also, $\det(A^T) = \det(A)$.

Proof. 1,3,4 of proposition 4.3 states how 3 types of transformation change the determinant, and we can compute the determinant of elementary matrices using the definition, which are 1 (type I), c (type II) and 1 (type III) We can observe that $\det(EA) = \det(E)\det(A)$ when E is a elementary matrix of any type. Transform both A and B into their *RREF* A' and B', we have

$$A = \prod_{i=1}^{m_A} E_i^A A' \text{ and } B = \prod_{j=1}^{m_B} E_j^B B'$$

where $\{E_i^A\}_{i=1}^{m_A}$, $\{E_j^A\}_{j=1}^{m_B}$ are elementary matrices. To cases happens (same for B), either $\mathbf{0}$ is a row vector of A', in which $\det(A) = \det(A') = 0$, or $A' = I$, in which $\det(A) = \prod_{i=1}^{m_i} E_i^A \neq 0$.

In the first case, $\mathbf{0}$ is also a row vector of $A'B$, thus $\det(AB) = 0$. In the second case, $AB = \prod_{i=1}^{m_A} E_i^A \prod_{j=1}^{m_B} E_j^B B'$. The result follows.
Observe that $\det(E^T) = \det(E)$ when E is an elementary matrix of any type, the second statement holds. □

The following is immediate from our above results.

Proposition 4.4. The following statements on a matrix $A \in \mathbb{C}^{n\times n}$ are equivalent.

1. $\operatorname{rank}(A) = n$
2. $\operatorname{null}(A) = 0$
3. A is invertible
4. $\det(A) \neq 0$

Theorem 4.2 (Obtaining matrix inverse)**.** Let $A \in \mathbb{C}^{n\times n}$ and C be its cofactor matrix. we have $AC^T = \det(A)I$. In particular, if A is invertible, $A^{-1} = \frac{1}{\det(A)} C^T$.
C^T is called the adjoint matrix of A.

Proof. The (i, j)-entry of AC^T is $\sum_{k=1}^{n} A_{i,k}C_{j,k}$. When $i = j$, it equals to $\det(A)$ by the statement 6 of proposition 4.3. When $i \neq j$, it equals to $\det(A')$, where A' is obtained by replacing row j with row i, thus having determinant 0. □

Theorem 4.3 (Effect of matrix transformation on Lebesgue measure)**.** Let $T \in \mathbb{R}^{n\times n}$ be a invertible matrix as well as a transformation $\mathbb{R}^n \mapsto \mathbb{R}^n$ and λ is the Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Then for any $A \in \mathcal{B}(\mathbb{R}^n)$, the image $T(A) := \{T(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\}$ is in $\mathcal{B}(\mathbb{R}^n)$, and

$$\lambda[T(A)] = |\det(A)|\lambda(A)$$

As a corollary, $\lambda[B(\mathbf{0}, r)] = r^n \lambda[B(\mathbf{0}, 1)]$.

Remark. If A is not invertible, the image $T(\mathbb{R}^n)$ is a subspace with dimension $k < n$. We quote the fact that such subspace has Lebesgue measure 0. Thus any $T(A) \subseteq T(\mathbb{R}^n)$ is measurable and with measure 0. The above holds with $\det(A) = 0$

The proof need the following lemma.

Lemma 4.2. If a measure μ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is translation-invariant, which means that

$$\forall A \in \mathcal{B}(\mathbb{R}^n), \mathbf{t} \in \mathbb{R},\ \mu(A) = \mu(A + \mathbf{t}) := \mu(\{\mathbf{x} + \mathbf{t} : \mathbf{x} \in A\})$$

then there exists $c \in \mathbb{R}$ such that $\mu(A) = c\lambda(A)\ \forall A \in \mathcal{B}(\mathbb{R})^n$

Proof. For any $\delta > 0$, there is a countable collection of boxes with corners $\{\mathbf{x}_i \in \mathbb{R}^n\}_{i=1}^{\infty}$ such that

$$\bigcup_{i=1}^{\infty} [\mathbf{x}_{i1}, \mathbf{x}_{i1} + \delta) \times \cdots \times [\mathbf{x}_{in}, \mathbf{x}_{in} + \delta) = \mathbb{R}^n$$

and the boxes are mutually disjoint. Let $\mathcal{C}_\delta$ be the set of these boxes. Note that $\forall B_1, B_2 \in \mathcal{C}_\delta$, $B_1, B_2 \in \mathcal{B}(\mathbb{R}^k)$ and they are a translation of each other. Thus $\mu(B_1) = \mu(B_2)$. Also, $\lambda(B_1) = \lambda(B_2) = \delta^n$.
We can construct $\{\mathcal{C}_m := \mathcal{C}_{2^{-m}}\}_{m=1}^{\infty}$ such that $\mathcal{C}_{m+1}$ is a further decomposition of $\mathcal{C}_m$. Each $B \in \mathcal{C}_m$

is further divided into 2^n mutually translatable subsets to form $\mathcal{C}_{m+1}$. Therefore, $\forall B_1 \in \mathcal{C}_m, B_2 \in \mathcal{C}_{m+1}$, $\mu(B_1) = 2^n \mu(B_2)$. The set $\mathcal{C} = \bigcup_{m=1}^{\infty} \mathcal{C}_m$ is countable. And $\forall B \in \mathcal{C}$, $\frac{\mu(A)}{\lambda(A)}$ is fixed, say to be c.
We claim that every open set $A \subseteq \mathbb{R}^n$ is a union of sets in $\mathcal{C}$. For any $\mathbf{x} \in A$, choose the open ball B contained in A and including $\mathbf{x}$, there is a sufficiently large m such that $\mathbf{x} \in C_{\mathbf{x}}$ and $C_{\mathbf{x}} \subseteq B$ for a unique choice of $C_{\mathbf{x}} \in \mathcal{C}_m$. Thus $\bigcup_{\mathbf{x}\in A} C_{\mathbf{x}} = A$ gives our claim. Note that this is at most a countable union because $\mathcal{C}$ is countable.
Suppose A is open and $\bigcup_{j=1}^{\infty} C_j = A$. We have $\mu(A) = \sum_{j=1}^{\infty} \mu(C_j) = \sum_{j=1}^{\infty} c\lambda(C_j) = c\lambda(A)$. Then it can be generalize into for any $A \in \mathcal{B}(\mathbb{R}^n)$.

□

Proof of Theorem 4.3. Because T is bijective, $T(A)$ is a Borel set, and $\mu = \lambda \circ T$ is a measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. The measure is translate-invariant because

$$\forall A \in \mathcal{B}(\mathbb{R}^n), \mathbf{t} \in \mathbb{R},\ \mu(A + \mathbf{t}) = \lambda\left[T(A) + T(\mathbf{t})\right] = \lambda\left[T(A)\right] = \mu(A)$$

Using the previous lemma, $\frac{\mu(A)}{\lambda(A)}$ is fixed for any $A \in \mathcal{B}(\mathbb{R}^n)$. Take $A := [0,1)^n$ to compute the ratio. Because T is a product of elementary matrices, $\det(T_1T_2) = \det(T_1)\det(T_1)$, and $T_1T_2 = T_1 \circ T_2$ as transformation, it suffices to prove that $\lambda(E(A)) = |\det(E)|\lambda(A) = |\det(E)|$ for any elementary matrix E.
When E is of type I, $E(A) = A$ and $|\det(E)| = 1$. When E is of type II, $E(A) = [0,c) \times [0,1)^{n-1}$ and $|\det(E)| = c$ for some scalar $c \in \mathbb{R}^*$. Their results follows. When E is of type III, $\det(E) = 1$. We can write $E = (E_1^{-1}E_2^{-1}E_1)E_3(E_1^{-1}E_2E_1)$ where E_1 is of type I, E_2 is of type II, $(E_1^{-1}E_2E_1)$ multipy row j by $c \in \mathbb{R}^*$, and E_2 correspond to the transformation of adding row j to row 1 ($2 \leq j \leq n$). It suffices to prove that $\lambda[E_3(A)] = 1$. Note that

$$E_3(A) = \{[x_1, \cdots, x_n]^T : x_j \leq x_1 < 1 + x_j,\ 0 \leq x_i \leq 1\ \forall 2 \leq i \leq n\}$$

Take $B := E_3(A)\backslash A$ and $B' = B - [1, 0, \cdots, 0]^T$ be its translation. Observe that $A = B' \cup (E_3(A) \cap A)$ and $B' \cap (E_3(A) \cap A) = \emptyset$. We have

$$\lambda[E_3(A)] = \lambda(B) + \lambda(A \cap E_3(A)) = \lambda(B') + \lambda(A \cap E_3(A)) = \lambda(A) = 1$$

We have prove the theorem. □

4.2 Eigenvalue and eigenvector

Definition 4.5. Eigenvalue, eigenvector, characteristic polynomial and eigenspace

- Let $A \in \mathbb{C}^{n\times n}$. If $A\mathbf{x} = \lambda\mathbf{x}$ for some $\lambda \in \mathbb{C}$ and $\mathbf{x} \in \mathbb{C} \setminus \{\mathbf{0}\}$, λ is a eigenvalue of A, and $\mathbf{x}$ is a eigenvector of A corresponding to λ.
- The equation is equal to $(A - \lambda I)\mathbf{x} = \mathbf{0}$. $\mathbf{x}$ exists if and only if $\det(A - \lambda I) = 0$. Define $f(\lambda) = \det(A - \lambda I)$ to be the characteristic polynomial of A. One can show by induction that the polynomial is of degree n and the coeffision for λ^n is $(-1)^n$. The eigenvalues are the zeros of this polynomial.
- The fundamental theorem of algebra gives the factoriazation

  $$\det(A - \lambda I) = f(\lambda) = (-1)^n \prod_{i=1}^{m} (\lambda - \lambda_i)^{k_i},\ \text{where } \{\lambda_i \in \mathbb{C}\}_{i=1}^{m} \text{ are distinct eigenvalues}$$

  The power k_i is called the algebraic multiplicity of eigenvalue λ_i. One has $n = \sum_{i=1}^{m} k_i$.

- Given a eigenvalue λ, the kernel (excluding $\mathbf{0}$) of the linear transformation $(A-\lambda I)$ is the set of all eigenvectors corresponding to λ, which is a subspace, called the eigenspace E_λ corresponding to λ. ($\mathbf{0} \in E_\lambda$ to make it a subspace, though it is not a eigenvector.) Define $\dim(E_\lambda)$ to be the geometric multiplicity of eigenvalue λ.

Remark. The existence of at least one eigenvalue is given by the fundamental theorem of algebra. The factorization also gives

$$\det(A) = f(0) = \prod_{i=1}^{m} \lambda_i^{k_i}$$

Proposition 4.5. If $(\lambda_1, \mathbf{x}_1), \cdots, (\lambda_k, \mathbf{x}_k)$ are k pairs of eigenvalues and corresponding eigenvectors. If λ_i are all distinct $\forall 1 \leq i \leq k$, then $\{\mathbf{x}_1, \cdots, \mathbf{x}_k\}$ is linearly independent. Since $\mathbb{C}^n$ is of dimension n, A have at most distinct eigenvalues.

Proof. Suppose otherwise $\mathbf{x}_i$ is a linear combination of $\{\mathbf{x}_j : j \in J\}$, where $J \subseteq \{1, \cdots, k\} \setminus \{i\}$, i.e.,

$$\mathbf{x}_i = \sum_{j \in J} a_j \mathbf{x}_j, \text{ where } \{a_j \in \mathbb{C}\}_{j \in J}$$

Without loss of generality, further assume $\{\mathbf{x}_j : j \in J\}$ is linear independent and $a_j \neq 0 \ \forall j \in J$. Since

$$A\mathbf{x}_i = \lambda \mathbf{x}_i = \sum_{j \in J} \lambda a_j \mathbf{x}_j \text{ and } A\mathbf{x}_i = \sum_{j \in J} a_j A\mathbf{x}_j = \sum_{j \in J} a_j \lambda_j \mathbf{x}_j$$

By the uniqueness of expressing $\lambda \mathbf{x}_i$ with its basis $\{\mathbf{x}_j : j \in J\}$ (Proposition 1.4), $a_j \lambda = a_j \lambda_j \ \forall j \in J$, contradicting to λ_k being distinct. $\square$

Theorem 4.4 (Matrix norm)**.** Regard $A \in \mathbb{C}^{n \times n}$ as a linear $(\mathbb{C}^n, \|\cdot\|_2) \to (\mathbb{C}^n, \|\cdot\|_2)$ operator. The operator norm is the largest norm of its eigenvectors.
We can therefore define the norm for any $\mathbb{C}^{n \times n}$ matrix to be its operator norm.

Sketch of proof: Use Lagrange multiplier.

Definition 4.6. A matrix $A \in \mathbb{C}^{n \times n}$ is diagonalizable if there exists a invertible matrix $P \in \mathbb{C}^{n \times n}$ and a diagonal matrix $D \in \mathbb{C}^{n \times n}$ such that $A = PDP^{-1}$.

Theorem 4.5. A is diagonalizable if and only if it has n linearly independent eigenvector. Then the diagonal of D are the eigenvalues $\lambda_1, \cdots, \lambda_n$ (not necessarily distinct), and the column vector of P, $\mathbf{x}_1, \cdots, \mathbf{x}_n$, are the eigenvectors corresponding to their eigenvalues by indices.

Proof by definition. Note that A and D have the same characteristic polynomial if $A = PDP^{-1}$.

We now study the diagonalization of a certain group of matrices.

Definition 4.7 (Conjugate transpose and Hermitian matrix)**.** Given $A \in \mathbb{C}^{n \times m}$, the Conjugate transpose (aka. Hermitian transpose) of A is the matrix A^* defined by $A^*_{i,j} = \overline{A_{j,i}}$. A is a Hermitian matrix if $A = A^*$ (one should first have $n = m$).
Conjugate transpose applies to vertex, since it is a special matrix.

Note that when $A \in \mathbb{R}^{n \times n}$, the above concept is equivalent to matrix transpose and symmetric matrix. The following properties are also similar to the transpose version. They follows from the definition immediately.

Proposition 4.6. For any $A, B \in \mathbb{C}^n$ and $c \in \mathbb{C}$

1. $(A+B)^* = A^* + B^*$
2. $(cA)^* = \bar{c}A^*$
3. $(A^*)^* = A$
4. $(AB)^* = B^*A^*$
5. A^* is invertible if and only if A^* is. And $(A^*)^{-1} = (A^{-1})^*$
6. $\det(A^*) = \overline{\det(A)}$
7. For $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x}^*\mathbf{x} = \|x\|$

If A is a Hermitian matrix,

1. The diagonal of A is real.
2. $\det(A) \in \mathbb{R}$, since $\det(A^*) = \overline{\det(A)} = \det(A)$
3. For any $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x}^*A\mathbf{x} \in \mathbb{R}$, since $(\mathbf{x}^*A\mathbf{x})^* = \mathbf{x}^*A\mathbf{x}$.

Definition 4.8 (Unitary matrix)**.** A matrix $A \in \mathbb{C}^{n\times n}$ is unitary if $A^*A = I$, or equivalently, the column vectors of A form a orthogonal basis of $\mathbb{C}^n$.
The analogy for real matrix $A \in \mathbb{R}^{n\times n}$ is orthogonal matrix, in which case $A^TA = I$.

Lemma 4.3. For any matrix, $A \in \mathbb{C}^{n\times n}$, there exists a unitary matrix (which must be invertible) C such that $C^{-1}AC$ is an upper triangle matrix (i.e., the entries under the diagonal are 0).

Proof. We use induction. When $n = 1$, take $C = 1$. Suppose the statement holds for $n = k$. We show that it holds for $n = k + 1$.
Let $\mathbf{x}$ be a eigenvector of A corresponding to eigenvalue λ. Normalize $\mathbf{x}$ such that $\|\mathbf{x}\|$. Let $\{\mathbf{x}, \mathbf{v}_1, \cdots, \mathbf{v}_k\}$ be a orthogonal basis for $\mathbb{C}^{k+1}$, which is possible due to the Gram-Schmidt process. Let $U_1 = [\mathbf{x}, \mathbf{v}_1, \cdots, \mathbf{v}_k]$ be a unitary matrix. Examine the first column of $U_1^{-1}AU_1$,

$$U_1^{-1}AU_1\mathbf{e}_1 = U_1^{-1}A\mathbf{x} = \lambda U_1^{-1}\mathbf{x} = \lambda U_1^{-1}U_1\mathbf{e}_1 = \lambda\mathbf{e}_1$$

Thus the entries from row 2 to row $k + 1$ of column 1 zeros 0. Let B be the matrix formed by deleting row 1 and column 1 of $U_1^{-1}AU_1$. By our hypothesis, there exits a unitary matrix $U_2' \in \mathbb{C}^{k\times k}$ such that $U_2'^{-1}BU_2'$ is upper triangle. Let

$$U_2 = \begin{pmatrix} 1, \mathbf{0}^T \\ \mathbf{0}, U_2' \end{pmatrix}$$

be a $\mathbb{C}^{(k+1)\times(k+1)}$ matrix. Observe that U_2 is unitary, then so as U_1U_2
We claim that $C = U_1U_2$. Let

$$U_1^{-1}AU_1 = \begin{pmatrix} \lambda, A' \\ \mathbf{0}, B \end{pmatrix}$$

for some $A' \in \mathbb{C}^{1\times k}$, then

$$C^{-1}AC = \begin{pmatrix} 1, \mathbf{0}^T \\ \mathbf{0}, U_2' \end{pmatrix}^{-1} \begin{pmatrix} \lambda, A' \\ \mathbf{0}, B \end{pmatrix} \begin{pmatrix} 1, \mathbf{0}^T \\ \mathbf{0}, U_2' \end{pmatrix} = \begin{pmatrix} 1, \mathbf{0}^T \\ \mathbf{0}, U_2'^{-1} \end{pmatrix} \begin{pmatrix} \lambda, A' \\ \mathbf{0}, B \end{pmatrix} \begin{pmatrix} 1, \mathbf{0}^T \\ \mathbf{0}, U_2' \end{pmatrix} = \begin{pmatrix} \lambda, U_2'^{-1}A'U_2' \\ \mathbf{0}, U_2'^{-1}BU_2 \end{pmatrix}$$

is upper triangle. We have proved the result. □

Theorem 4.6 (Spectral theorem in finite dimension case)**.** If $A \in \mathbb{C}^{n\times n}$ is Hermitian, then the eigenvalues of A is real and A is diagonalizable.

Proof. By Lemma 4.3, choose P to be the unitary matrix that $P^{-1}AP$ is upper triangle. Observe that

$$(P^{-1}AP)^* = P^*A^*(P^{-1})^* = P^{-1}AP$$

Therefore, $P^{-1}AP$ is Hermitian. It is only possible that $D := P^{-1}AP$ is a diagonal matrix and the diagonal entries are real. We may write

$$A = PDP^{-1} = PDP^*$$

Thus A is diagonalizable with real eigenvalues. And PDP^* is called the eigendecomposition of A. □

4.3 Order of matrix

Definition 4.9 (Definite matrix)**.** A Hermitian matrix $A \in \mathbb{C}^{n\times n}$ is categorized as a

- Positive definite matrix, if $\mathbf{x}^*A\mathbf{x} > 0\ \forall \mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$
- Positive semi-definite matrix (aka. non-negative definite), if $\mathbf{x}^*A\mathbf{x} \geq 0\ \forall \mathbf{x} \in \mathbb{C}^n$
- Negative definite matrix, if $\mathbf{x}^*A\mathbf{x} < 0\ \forall \mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$
- Negative semi-definite matrix (aka. non-positive definite), if $\mathbf{x}^*A\mathbf{x} \leq 0\ \forall \mathbf{x} \in \mathbb{C}^n$
- Indefinite, if it does not falls on the above cases.

We have a version of the definition of definite real symmetric matrix $A \in \mathbb{R}^{n\times n}$, in which $\mathbf{x}$ is in $\mathbb{R}^n$ (in some context, symmetric is not even required). But we mostly focus on complex matrix.

Proposition 4.7. The following properties about positive definite and positive semi-definite may also have a version about negative (semi)-definite. Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian.

1. If A is positive definite, then it is positive semi-definite
2. If A is positive definite and B is positive semi-definite, $A + B$ is positive definite. If A and B are positive semi-definite, $A + B$ is positive semi-definite
3. If A is positive definite, the entries in the diagonal (which is real) are positive. If A is positive semi-definite, the entries in the diagonal (which is real) are non-negative.

Proof. 1 and 2 is easy. To prove 3, note that if the i-th diagonal entry is non-positive, $\mathbf{e}_i^*A\mathbf{e}_i < 0$. To show 4. □

Theorem 4.7 (Definite matrix and eigenvalue)**.** A Hermitian matrix $A \in \mathbb{C}^{n\times n}$ is

1. positive definite if all its eigenvalues are positive
2. positive semi-definite if all its eigenvalues are non-negative
3. negative definite if all its eigenvalues are negative
4. negative semi-definite if all its eigenvalues are non-positive

5. Indefinite otherwise

Proof. We prove that A is positive definite if and only if all its eigenvalues are positive. Others are similar. Suppose $A\mathbf{x} = \lambda\mathbf{x}$ for some $\lambda < 0$ and $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x}^*A\mathbf{x} = \lambda\|\mathbf{x}\| < 0$. Thus if A has negative eigenvalue, A is not positive definite.
Then suppose the eigenvalues are positive. Diagonalize $A = PDP^{-1}$, where D is the diagonal matrix with positive entries, thus D is positive definite. Since P is invertible, ranging $\mathbf{x}$ in $\mathbb{C}^n$ is equivalent to ranging $\mathbf{y} := P^{-1}\mathbf{x}$ in $\in \mathbb{C}^n$. Thus A is also positive definite. □

Corollary 4.2. A matrix $A \in \mathbb{C}^{n\times n}$ is Hermitian and positive semi-definite if and only if $\exists B \in \mathbb{C}^{n\times n}$, $A = BB^*$

Proof. If A is Hermitian and positive semi-definite. Let $A = UDU^*$ be the eigendecomposition where D is the diagonal matrix with non-negative value, and U is unitary. Let $D^{\frac{1}{2}}$ be the matrix formed by taking the non-negative square root of the diagonal entries of D. Then $D = D^{\frac{1}{2}}D^{\frac{1}{2}}$ and $A = UD^{\frac{1}{2}}D^{\frac{1}{2}}U^* = (UD^{\frac{1}{2}})(UD^{\frac{1}{2}})^*$.
If $A = BB^*$, we have $A^* = A$, and for any $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x}^*A\mathbf{x} = (B\mathbf{x})^*B\mathbf{x} = \|B\mathbf{x}\| \geq 0$. □

Definition 4.10 (Loewner order)**.** Given $n \in \mathbb{N}$, let $\mathcal{C}$ be the set of Hermitian matrix in $\mathbb{C}^{n\times n}$ (similar for $\mathbb{R}^{n\times n}$). For A, write $A \succeq 0$ if A is positive semi-definite. Similarly, using $\preceq$ to denote negative semi-definite. Observe that $A = 0$ if and only if $A \succeq 0$ and $A \preceq 0$. We can define a partial order between some $A, B \in \mathcal{C}$, i.e., $A \preceq B$ if $(B - A)$ is positive semi-definite, or equivalently, $(A - B)$ is negative semi-definite. $A \succeq B$ and $A \preceq B$ together implies $A = B$. Note that not every (A, B) pair have defined order.

5 Application: Expectation of random vector

Definition 5.1 (Expectation of random vector)**.** Let $\mathbf{X} = [X_1, \cdots, X_n]$ be a real or complex random vector on the probability space $(\Omega, \mathcal{B}, P)$. If $\mathbb{E}(X_i)$ exists for each $1 \leq i \leq n$, we write

$$\mathbb{E}(\mathbf{X}) := [\mathbb{E}(X_1), \cdots, \mathbb{E}(X_n)]^T \in \bar{\mathbb{R}}^n \text{ or } \mathbb{C}^n$$

Definition 5.2 (Covariance matrix)**.** Let $\mathbf{X} = [X_1, \cdots, X_n]$ and $\mathbf{Y} = [Y_1, \cdots, Y_n]$ be two complex random vectors on the probability space $(\Omega, \mathbf{B}, P)$. If for each $1 \leq i \leq n$ and $1 \leq j \leq m$,

$$\operatorname{cov}(X_i, Y_j) = \mathbb{E}\left[(X_i - \mathbb{E}(X_i))\overline{(Y_i - \mathbb{E}(Y_i))}\right]$$

exists, we define the covariance matrix $\operatorname{cov}(\mathbf{X}, \mathbf{Y}) \in \mathbb{C}^{n\times m}$ such that $\operatorname{cov}(\mathbf{X}, \mathbf{Y})_{i,j} = \operatorname{cov}(X_i, Y_j)$.
Define $\operatorname{var}(\mathbf{X}) = \operatorname{cov}(\mathbf{X}, \mathbf{X})$ to be the variance of vector $\mathbf{X}$.
The above applies for real random vectors.

Proposition 5.1. The following applies for compact random vectors in general. A similar or simplified version holds for real random vectors

1. Vector expectation is also linear, the same as normal expectation

2. If $\operatorname{cov}(\mathbf{X}, \mathbf{Y})$ exists,

$$\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = \mathbb{E}\left[(\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{Y} - \mathbb{E}(\mathbf{Y}))^*\right] = \mathbb{E}(\mathbf{X}\mathbf{Y}^*) - \mathbb{E}(\mathbf{X})\mathbb{E}(\mathbf{Y}^*)$$

3. $\operatorname{cov}(\mathbf{X}, \mathbf{Y}) = \overline{\operatorname{cov}(\mathbf{Y}, \mathbf{X})}$ (i.e., point-wise conjugate)

4. Let $A \in \mathbb{C}^{k\times n}$, $B \in \mathbb{C}^{k\times p}$, $\mathbf{a} = \mathbb{C}^k$, $\mathbf{b} = \mathbb{C}^p$.

$$\mathrm{cov}(A\mathbf{X}+\mathbf{a}, B\mathbf{Y}+\mathbf{b}) = \mathrm{cov}(A\mathbf{X}, B\mathbf{Y}) = A\mathrm{cov}(\mathbf{X},\mathbf{Y})B^*$$

   The existence of one gives the existence of another.

5. $\mathrm{var}(\mathbf{X})$ is Hermitian and non-negative definite, since $\forall \mathbf{v} \in \mathbb{C}^n$, $\mathbf{v}^T\mathrm{var}(\mathbf{X})\mathbf{v} = \mathrm{var}(\mathbf{v}^T\mathbf{X}) \in [0,\infty)$

6. The entries of $\mathbf{X}$ are independent if $\mathrm{var}(\mathbf{X})$ is diagnal matrix

7. $\mathrm{var}(\mathbf{X}+\mathbf{Y}) = \mathrm{var}(\mathbf{X}) + \mathrm{cov}(\mathbf{X},\mathbf{Y}) + \mathrm{cov}(\mathbf{Y},\mathbf{X}) + \mathrm{var}(\mathbf{Y})$

Lemma 5.1 (Trace-variance formula)**.** Let $A \in \mathbb{R}^{n\times n}$ be a fixed matrix and $\mathbf{X}$ is a n dimension real random vector on $(\Omega, \mathcal{B}, P)$. Then $\mathbf{X}^T A\mathbf{X}$ is a random variable. If $\mathbf{X}^T A\mathbf{X} \in \mathcal{L}^1$,

$$\begin{aligned}
\mathbb{E}(\mathbf{X}^T A\mathbf{X}) = \mathbb{E}[\mathrm{tr}(\mathbf{X}^T A\mathbf{X})] = \mathbb{E}[\mathrm{tr}(A\mathbf{X}\mathbf{X}^T)] &= \mathrm{tr}[A\mathbb{E}(\mathbf{X}\mathbf{X}^T)] \\
&= \mathrm{tr}[A\mathrm{var}(\mathbf{X}) + A\mathbb{E}(\mathbf{X})\mathbb{E}(\mathbf{X})^T] \\
&= \mathrm{tr}[A\mathrm{var}(\mathbf{X})] + \mathrm{tr}[\mathbb{E}(\mathbf{X})A\mathbb{E}(\mathbf{X})^T]
\end{aligned}$$

where $\mathrm{tr}(*)$ denote the trace, i.e., sum of diagonal entries. We used the fact that $\mathrm{tr}(PQ) = \mathrm{tr}(QP)$ for any $P \in \mathbb{C}^{n\times m}$ and $Q \in \mathbb{C}^{m\times k}$ (can be proved by arithmetic).

6 Spectral theory

6.1 Fundamentals for functional analysis

Theorem 6.1 (Uniform boundedness principle, aka. Banach–Steinhaus theorem)**.** Let $(X, \|\cdot\|_X)$ be a Banach space and $(Y, \|\cdot\|_Y)$ be a normed vector space. $\mathcal{B}(X \mapsto Y)$ is the set of bounded and linear $X \mapsto Y$ functions. If $\mathcal{F} \subseteq \mathcal{B}(X \mapsto Y)$ satisfies

$$\forall \mathbf{x} \in X,\ \sup\{\|f(\mathbf{x})\|_Y : f \in \mathcal{F}\} < \infty$$

then

$$\sup\{\|f\|_{X\mapsto Y} : f \in \mathcal{F}\} := \sup\{\|f(\mathbf{x})\|_Y : f \in \mathcal{F} \text{ and } \|\mathbf{x}\|_X = 1\} < \infty$$

Proof. From the premise, for any $n \in \mathbb{N}^*$, define

$$X_n = \{\mathbf{x} \in X : \sup\{f(\mathbf{x}) : f \in \mathcal{F}\} \le n\}$$

to be some closed set. We have $\bigcup_{n=1}^\infty X_n = X$. By Baire category theorem (Theorem 1.3), X, as a complete metric space, is a Baire space. It cannot be a countable union of closed subsets with empty interior, i.e., $\exists n_0 \in \mathbb{N}^*, \mathbf{x}_0 \in X_n, r_0 \in \mathbb{R}^+$ such that

$$\overline{B(\mathbf{x}_0, r_0)} \subseteq X_{n_0}$$

Then, for any $f \in \mathcal{F}$ and $\mathbf{x} \in X$ such that $\|\mathbf{x}\|_X = 1$. We have the desired result

$$\begin{aligned}
\|f(\mathbf{x})\|_Y = r_0^{-1}\|f(\mathbf{x}_0 + r_0\mathbf{x}) - f(\mathbf{x}_0)\| &\le \frac{1}{r_0}(\|f(\mathbf{x}_0 + r_0\mathbf{x})\|_Y + \|f(\mathbf{x}_0)\|_Y) \\
&\le \frac{1}{r_0}(n_0 + n_0) < \infty \text{ (by the definition of } X_{n_0})
\end{aligned}$$

□

Corollary 6.1. If the sequence $\{f_n \in \mathcal{B}(X \mapsto Y)\}_{n=1}^{\infty}$ converges pointwise in Y, then the limit function $f(\cdot) = \lim_{n\to\infty} f_n(\cdot)$ is a bounded and linear operator, i.e., $f \in \mathcal{B}(X \mapsto Y)$.

Theorem 6.2 (Hahn–Banach Theorem)**.** Let $(V, \|\cdot\|_V)$ be a nonempty normed vector space over $F \in \{\mathbb{R}, \mathbb{C}\}$ with its dual space V^*. For any subspace $W \subseteq V$ and linear and continuous functional $f : W \mapsto F$, there exists an extension $\tilde{f} \in V^*$ such that

$$\tilde{f}(\mathbf{x}) = f(\mathbf{x})\ \forall \mathbf{x} \in V \text{ and } \|\tilde{f}\|_{V\mapsto F} = \|f\|_{W\mapsto F}$$

The proof need the following lemma, then use the Hausdorff maximal principle (Theorem1.4).

Lemma 6.1. Suppose $F = \mathbb{R}$, W is a proper subspace of V and $\mathbf{x}_0 \in V \setminus W$. If $f : W \mapsto \mathbb{R}$ is linear and bounded, then there exists an extension $\bar{f} : U \mapsto \mathbb{R}$, where $U := \text{span}[W \cup \{\mathbf{x}_0\}]$, satisfying the requirements (i.e., $\bar{f}$ is linear and bounded, agrees with f on W and has the same operator norm as f).

Proof. If $\|f\|_{W\mapsto\mathbb{R}}$, $\bar{f} = 0$ is as required. Otherwise without loss of generality, assume $\|f\|_{W\mapsto\mathbb{R}} = 1$. By the definition, we can write

$$U = \{\mathbf{u} + \lambda\mathbf{x}_0 : \mathbf{u} \in W,\ \lambda \in \mathbb{R}\}$$

For some $\alpha \in \mathbb{R}$, define

$$\bar{f}(\mathbf{u} + \lambda\mathbf{x}_0) = f(\mathbf{u}) + \lambda\alpha\ \forall \mathbf{u} \in W, \lambda \in \mathbb{R}$$

Clearly $\bar{f} : U \mapsto \mathbb{R}$ is linear and agrees with f on W. It suffices to restrict α so that $\|\bar{f}\|_{U\mapsto\mathbb{R}} = 1$, or equivalently, given $\|f\|_{W\mapsto\mathbb{R}} = 1$,

$$|f(\mathbf{u}) + \lambda\alpha| \le \|\mathbf{u} + \lambda\mathbf{x}_0\|_V\ \forall \mathbf{u} \in W, \lambda \in \mathbb{R} \iff |\lambda||-f(\mathbf{u}') + \alpha| \le |\lambda|\|-\mathbf{u}' + \mathbf{x}_0\|_V\ \forall \mathbf{u}' \in W, \lambda \in \mathbb{R}$$

The second condition is then

$$f(\mathbf{u}') - \|\mathbf{u}' - \mathbf{x}_0\|_V \le \alpha \le f(\mathbf{u}') + \|\mathbf{u}' - \mathbf{x}_0\|_V\ \forall \mathbf{u}' \in W, \lambda \in \mathbb{R}$$

The following shows that such $\alpha \in \mathbb{R}$ exists,

$$\begin{aligned} \forall \mathbf{v}, \mathbf{w} \in W,\ f(\mathbf{v}) - f(\mathbf{w}) &\le |f(\mathbf{v} - \mathbf{w})| \le \|\mathbf{v} - \mathbf{w}\|_V \le \|\mathbf{v} - \mathbf{x}_0\|_V + \|\mathbf{w} - \mathbf{x}_0\|_V \\ \Rightarrow\ & f(\mathbf{v}) - \|\mathbf{v} - \mathbf{x}_0\|_V \le f(\mathbf{w}) - \|\mathbf{w} - \mathbf{x}_0\|_V \\ \text{and } & f(\mathbf{w}) - \|\mathbf{w} - \mathbf{x}_0\|_V \le f(\mathbf{v}) - \|\mathbf{v} - \mathbf{x}_0\|_V \end{aligned}$$

□

Proof. of Theorem 6.2
Still the case $\|f\|_{W\mapsto F} = 0$ is trivial, and we first consider $F = \mathbb{R}$. Let

$$\mathcal{S} = \left\{(U, \bar{f}_U) : U \text{ is a subspace of } V,\ U \supseteq W,\ \bar{f}_A \text{ is an extension of } f \text{ to } U \text{ with the requirements}\right\}$$

be a set with the partial order

$$(A, \bar{f}_A) \preceq (B, \bar{f}_A) \text{ if and only if } A \subseteq B \text{ and } \bar{f}_B\big|_A = \bar{f}_A$$

for any $(A, \bar{f}_A), (B, \bar{f}_A) \in \mathcal{S}$ whenever comparable. Note that $(W, f) \in \mathcal{S}$ thus it is not empty. Apply the Hausdorff maximal principle, say, $\mathcal{C} \subseteq \mathcal{S}$ is the maximal chain. Our definition gives that the union of the elements in $\mathcal{C}$, V', is a subspace of V. The extension $\tilde{f} : V' \mapsto \mathbb{R}$ of f is also well defined such that

$$\forall U \in \mathcal{C}, \mathbf{x} \in U, \;\; \tilde{f}(\mathbf{x}) := \bar{f}_U(\mathbf{x})$$

$\tilde{f}$ is therefore linear and has norm 1. If $V' \subset V$ is proper, our above lemma gives an extension further from V', which contradicts to the maximality of $\mathcal{C}$. Therefore, $V' = V$.
It remains to show the case when $F = \mathbb{C}$. Consider this property of complex number,

$$\forall z \in \mathbb{C}, \;\; z = \mathrm{Re}(z) - \mathrm{i}\mathrm{Re}(\mathrm{i}z)$$

Thus,

$$\forall \mathbf{x} \in W, \;\; f(\mathbf{x}) = \mathrm{Re}(f(\mathbf{x})) - \mathrm{i}\mathrm{Re}(\mathrm{i}f(\mathbf{x})) = [\mathrm{Re}(f)]\,(\mathbf{x})) - \mathrm{i}\,[\mathrm{Re}(f)]\,(\mathrm{i}\mathbf{x})$$

where $\mathrm{Re}(f) : W \mapsto \mathbb{R}$ is the real part. The complex vector space V is also a real vector space with the same norm, using $\mathbb{R}$ as scalar set, and $\mathrm{Re}(f)$ is linear in the real vector space V. Applying the above, we obtain the linear and bounded functional $f' : V \mapsto \mathbb{R}$, the extension of $\mathrm{Re}(f)$ such that $\|\mathrm{Re}(f)\|_{W \mapsto \mathbb{R}} = \|f'\|_{V \mapsto \mathbb{R}}$. Let

$$\tilde{f}(\mathbf{x}) := f'(\mathbf{x}) - \mathrm{i}f'(\mathrm{i}\mathbf{x}), \;\; \forall \mathbf{x} \in V$$

Note that $\tilde{f}(\mathrm{i}\mathbf{x}) = \mathrm{i}\tilde{f}(\mathbf{x}) \; \forall \mathbf{x} \in V$, $\tilde{f}$ is a linear function in the complex vector space V. The below lemma gives the desire equation in norm.

$$\|f\|_{W \mapsto \mathbb{C}} = \|\mathrm{Re}(f)\|_{W \mapsto \mathbb{R}} = \|f'\|_{V \mapsto \mathbb{R}} = \|\tilde{f}\|_{V \mapsto \mathbb{C}}$$

□

Lemma 6.2. Let $(V, \|\cdot\|_V)$ be a normed vector space over $\mathbb{C}$. If $f : V \mapsto \mathbb{C}$ is linear and bounded, $g : V \mapsto \mathbb{R}$ is bounded and they satisfies

$$f(\mathbf{x}) = g(\mathbf{x}) - \mathrm{i}g(\mathrm{i}\mathbf{x}), \;\; \forall \mathbf{x} \in V$$

Then $\|f\|_{V \mapsto \mathbb{C}} = \|g\|_{V \mapsto \mathbb{R}}$

Proof. Since $\forall \mathbf{x} \in V$, $\|f(\mathbf{x})\| = \sqrt{g(\mathbf{x})^2 + g(\mathrm{i}\mathbf{x})^2} \geq |g(\mathbf{x})|$, $\|f\|_{V \mapsto \mathbb{C}} \geq \|g\|_{V \mapsto \mathbb{R}}$. Also, for any $\mathbf{x} \in V$ such that $f(\mathbf{x}) \neq 0$ (it is trivial if such $\mathbf{x}$ not exist), take

$$c := \frac{\|f(\mathbf{x})\|}{f(\mathbf{x})} \text{ such that } \|c\| = 1$$

Then,

$$\|f(\mathbf{x})\| = cf(\mathbf{x}) = f(c\mathbf{x}) = g(c\mathbf{x}) \leq \|g\|_{V \mapsto \mathbb{R}}\|c\|\|\mathbf{x}\|_V = \|g\|_{V \mapsto \mathbb{R}}\|\mathbf{x}\|_V$$

where $f(c\mathbf{x}) = g(c\mathbf{x})$ is given by $f(c\mathbf{x}) \in \mathbb{R}$. Thus $\|f\|_{V \mapsto \mathbb{C}} \leq \|g\|_{V \mapsto \mathbb{R}}$ and the result follow. □

Theorem 6.3 (Open mapping theorem)**.** Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two Banach spaces and $\mathcal{B}(X \mapsto Y)$ be the collection of all bounded and linear $X \mapsto Y$ functions. Then a function $f \in B(X \mapsto Y)$ is surjective if and only if it is a open mapping, i.e.,

$$\forall \text{open set } A \subseteq X, \;\; f(X) \text{ is a open set in } Y$$

Proof. We fist prove the "$\Rightarrow$" direction. Since f is surjective,

$$Y = f(X) = f\left(\bigcup_{n=1}^{\infty} B(\mathbf{0}_X, n)\right) = \bigcup_{n=1}^{\infty} f(B(\mathbf{0}_X, n)) = \bigcup_{n=1}^{\infty} \overline{f(B(\mathbf{0}_X, n))}$$

Since Y is a Baire space, we have shown in the proof of Theorem 6.1 that there exists $n_0 \in \mathbb{N}^*, \mathbf{y}_0 \in Y, r \in \mathbb{R}$ such that

$$B(\mathbf{y}_0, r_0) \subseteq \overline{f(B(\mathbf{0}_X, n_0))}$$

Let $\mathbf{y} \in B(\mathbf{0}_Y, r_0)$. We have $\mathbf{y}_0, \mathbf{y}_0 - \mathbf{y} \in B(\mathbf{y}_0, r)$. There exists two sequence $\{\mathbf{x}_{1i} \in B(\mathbf{0}_X, n_0)\}_{i=1}^{\infty}$ and $\{\mathbf{x}_{2i} \in B(\mathbf{0}_X, n_0)\}_{i=1}^{\infty}$ such that

$$\mathbf{y}_0 = \lim_{i\to\infty} f(\mathbf{x}_{1i}) \text{ and } \mathbf{y}_0 - \mathbf{y} = \lim_{i\to\infty} f(\mathbf{x}_{2i}), \text{ thus } \mathbf{y} = \lim_{i\to\infty} f(\mathbf{x}_{1i} - \mathbf{x}_{2i})$$

where $\mathbf{x}_{1i} - \mathbf{x}_{2i} \in B(\mathbf{0}_X, 2n_0)\ \forall i \in \mathbb{N}^*$. We have prove that for any $\mathbf{y} \in Y$ with $\|\mathbf{y}\|_Y < r_0$ and $\epsilon > 0$, there exists $\mathbf{x} \in B(\mathbf{0}_X, 2n_0)$ such that $\|\mathbf{y} - f(\mathbf{x})\| < \epsilon$, that is,

$$B(\mathbf{0}_Y, r_0) \subseteq \overline{f(B(\mathbf{0}_X, 2n_0))}, \text{ which gives } \overline{B(\mathbf{0}_Y, r_0)} \subseteq \overline{f(B(\mathbf{0}_X, 2n_0))}$$

Next we prove that

$$B(\mathbf{0}_Y, r_0) \subseteq f(B(\mathbf{0}_X, 2n_0))$$

Fix an arbitrary $\mathbf{y} \in B(\mathbf{0}_Y, r_0)$. Let $\alpha := \frac{r_0}{2n_0} \in \mathbb{R}^+$. We have $\mathbf{y}' := \frac{r_0}{\|\mathbf{y}\|_Y}\mathbf{y} \in \overline{B(\mathbf{0}_Y, r_0)}$. Set $\epsilon_1 := \frac{r_0 - \|\mathbf{y}\|}{2^2}$. Take $\mathbf{x}_1' \in B(\mathbf{0}_X, 2n_0)$ and $\mathbf{x}_1 = \frac{\|\mathbf{y}\|_Y}{r_0}\mathbf{x}_1'$ such that

$$\|\mathbf{y}' - f(\mathbf{x}_1')\|_Y < \epsilon_1' := \frac{r_0\epsilon_1}{\|\mathbf{y}\|_Y}, \text{ or } \frac{r_0}{\|\mathbf{y}\|_Y}\|\mathbf{y} - f(\mathbf{x}_1)\|_Y < \epsilon_1', \text{ or } \|\mathbf{y} - f(\mathbf{x}_1)\|_Y < \epsilon_1$$

We also have $\|\mathbf{x}_1\|_X < 2n_0\frac{\|\mathbf{y}\|_Y}{r_0} = \alpha^{-1}\|\mathbf{y}\|_Y$. Recursively, given $\{\mathbf{x}_j\}_{j=1}^{k}$, we do the same for $\mathbf{y} - \sum_{j=1}^{k} f(\mathbf{x_j})$ in the place of $\mathbf{y}$, i.e, to pick $\mathbf{x}_{k+1} \in X$ such that

$$\left\|\mathbf{y} - \sum_{j=1}^{k} f(\mathbf{x_j}) - f(\mathbf{x}_{k+1})\right\|_Y < \epsilon_{k+1} := \frac{r_0 - \|\mathbf{y}\|_Y}{2^{k+2}} \text{ and } \|\mathbf{x}_{k+1}\|_X < \alpha^{-1}\left\|\mathbf{y} - \sum_{j=1}^{k} f(\mathbf{x_j})\right\|_Y$$

The sequence $\{\mathbf{x}_k\}_{k=1}^{\infty}$ satisfies

$$\forall K \in \mathbb{N}^*,\ \sum_{k=1}^{K}\|\mathbf{x}_k\|_X < \alpha^{-1}\sum_{k=1}^{K}\left\|\mathbf{y} - \sum_{j=1}^{k-1} f(\mathbf{x_j})\right\|_Y < \alpha^{-1}\left(\|\mathbf{y}\|_Y + \sum_{k=2}^{K}\epsilon_{k-1}\right)$$

Since

$$\sum_{k=1}^{\infty}\epsilon_k = \sum_{k=1}^{\infty}\frac{r_0 - \|\mathbf{y}\|_Y}{2^{k+1}} = \frac{r_0 - \|\mathbf{y}\|_Y}{2} < \infty$$

, those series absolutely converges. Therefore, $\mathbf{y} = \sum_{k=1}^{\infty} f(\mathbf{x}_k)$ and, say, $\mathbf{x} := \sum_{k=1}^{\infty}\mathbf{x}_k$. We have

$$f(\mathbf{x}) = \mathbf{y} \text{ and } \|\mathbf{x}\|_X \le \sum_{k=1}^{\infty}\|\mathbf{x}_k\|_X \le \alpha^{-1}\left(\|\mathbf{y}\|_Y + \sum_{k=1}^{\infty}\epsilon_k\right) = \frac{2n_0}{r_0}\left(\|\mathbf{y}\|_Y + \frac{r_0 - \|\mathbf{y}\|_Y}{2}\right) < 2n_0$$

Since $\mathbf{y} \in B(\mathbf{0}_Y, r_0)$ is arbitrary, we have prove that $B(\mathbf{0}_Y, r_0) \subseteq f(B(\mathbf{0}_X, 2n_0))$. Since f is linear, the image of any open ball (in X) is an open ball (in Y). For a general open set $A \subseteq X$ and any point $\mathbf{y} \in f(A)$, there exists $\mathbf{x} \in A, r \in \mathbb{R}^+$ such that $f(\mathbf{x}) = \mathbf{y}$ and $B(\mathbf{x}, r) \subseteq A$. Then $f(B(\mathbf{x}, r))$ is an neighborhood of $\mathbf{y}$ contained in $f(A)$. Thus $f(A)$ is open.
The proof of the "$\Leftarrow$" direction is simpler. Since f is linear and open, $f(X) \subseteq Y$ is a open subspace. There exists $B(\mathbf{y}_0, r) \subseteq f(Y)$. By linearity, $B(\mathbf{0}_Y, r) \subseteq f(Y)$. Then, for any $\mathbf{y} \in Y$,

$$\frac{r}{2\|\mathbf{y}\|_Y}\mathbf{y} \in B(\mathbf{0}_Y, r), \text{ thus } \frac{r}{2\|\mathbf{y}\|_Y}\mathbf{y} \in f(X)$$

By linearity, $\mathbf{y} \in f(X)$. Thus f is surjective. □

Corollary 6.2 (Bounded inverse theorem)**.** If $f \in B(X \mapsto Y)$ is bijective, then its inverse is continuous.

Proof. Since f is surjective, for any open set $A \subseteq X$,

$$\left(f^{-1}\right)^{-1}(A) = f(A) \subseteq Y \text{ is open}$$

which meets the definition of continuity for f^{-1}. Clearly $f^{-1} : Y \mapsto X$ is linear, then it is also bounded. □

6.2 Spectrum of bounded linear operator

Definition 6.1 (Eigenvalue and eigenvector of linear operator)**.** Let V be a vector space over $\mathbb{C}$ and $f : V \mapsto V$ is a linear function. If $f(\mathbf{x}) = \lambda f(\mathbf{x})$ for some $\mathbf{x} \in V \setminus \{\mathbf{0}\}$ and $\lambda \in \mathbb{C}$, then λ is an eigenvalue of f and $\mathbf{x}$ is a corresponding eigenvector.
Given a eigenvector λ of f, the subspace $\ker(f - \lambda I)$ is the eigenspace (aka. characteristic space) of λ that contains all the corresponding eigenvectors, where $I : V \mapsto V$ is the identity transformation. The dimension of the eigenspace is the geometric multiplicity.

Proposition 6.1. Two properties of eigenvector

1. If $\lambda \neq 0$ is a eigenvector, the eigenspace $\ker(f - \lambda I)$ is of finite dimension.
2. Let $S := \{\lambda_i : i \in I\}$ be the set of all distinct eigenvalues of f and $\{\mathbf{x}_i : i \in I\}$ is the corresponding eigenvectors (associated by index), then $\{\mathbf{x}_i : i \in I\}$ is linear independent.

Proof. Let $A := \ker(f - \lambda I) \cap \overline{B(\mathbf{0}, 1)}$ be a bounded set. Since $f(A)$ is compact and $f(A) = \{\lambda\mathbf{u} : \mathbf{u} \in A\} = \overline{B(\mathbf{0}, \|\lambda\|)} \cap \ker(f - \lambda I)$. This is possible only if $\ker(f - \lambda I)$ is a finite dimensional subspace (by proposition 1.5).
The proof is the same as the matrix case Proposition 4.5, since linear dependency only consider finite subset. □

Note that λ is the eigenvalue of f if and only if the linear map $(f - \lambda I)$ is not injective. The following extend this concept for linear and bounded operator on a complex Banach space.

Definition 6.2 (Spectrum of linear and bounded operator)**.** Let $(V, \|\cdot\|_V)$ be a Banach space over $\mathbb{C}$ and $f : V \mapsto V$ is a bounded and linear function. The spectrum of f is defined as

$$\sigma(f) = \{\lambda \in \mathbb{C} : (f - \lambda I) \text{ is not bijective}\}$$

and the resolvent set of f is

$$\rho(f) := \{\lambda \in \mathbb{C} : \text{the inverse } (f - \lambda I)^{-1} \text{ exists and is bounded}\}$$

Remark. By the bounded inverse theorem (Corollary 6.2), the inverse of a linear and bounded operator between two Banach spaces is linear and bounded if exists, thus the second condition is redundant, i.e.,

$$\rho(f) = \mathbb{C} \setminus \sigma(f)$$

The spectrum contains all the eigenvalues of f.

If $f : V \mapsto V$ is an operator, we will use the notation f^n for that operator that applies f for n times. Formally, $f^0 = I$ and $\forall n \in \mathbb{N}^*$, $f^n = f \circ f^{n-1}$, not to confuse $f^n(\cdot)$ with $f(\cdot)^n$.

Lemma 6.3 (Neumann series)**.** Let $g : V \mapsto V$ be a linear operator such that $\|g\|_{V \mapsto V} < 1$. Then $h := I - g$ is bijective with inerse

$$h^{-1}(\cdot) = \sum_{n=0}^{\infty} g^n(\cdot)$$

Proof. Observe that

$$\forall N \in \mathbb{N}^*, \mathbf{x} \in V,\ h \circ \left[\sum_{n=0}^{N} g^n\right](\mathbf{x}) = \left[\sum_{n=0}^{N} g^n\right](\mathbf{x}) - \left[\sum_{n=1}^{N+1} g^n\right](\mathbf{x}) = I(\mathbf{x}) - g^{N+1}(\mathbf{x})$$

Since for any $\mathbf{x} \in V$, $\|g(\mathbf{x})\|_V < \|\mathbf{x}\|_V$, thus $\|g^N(\mathbf{x})\|_V < {\|\mathbf{x}\|_V}^N$ $\forall N \in \mathbb{N}^*$ by induction, which gives $\lim_{N\to\infty} \|g^N(\mathbf{x})\|_V = 0$. Further by the continuity of the $\|\cdot\|_V$ operation,

$$\left\|\lim_{N\to\infty} g^N(\mathbf{x})\right\|_V = \lim_{N\to\infty} \|g^N(\mathbf{x})\|_V = 0 \ \Rightarrow \ \lim_{N\to\infty} g^N(\mathbf{x}) = \mathbf{0}\ \forall \mathbf{x} \in V$$

Therefore,

$$h(\cdot) \circ \left[\sum_{n=0}^{\infty} g^n(\cdot)\right] = \lim_{N\to\infty} h \circ \left[\sum_{n=0}^{N} g^n\right] = I(\cdot) - \lim_{N\to\infty} g^{N+1}(\cdot) = I(\cdot)$$

□

Proposition 6.2. Properties of spectrum and resolvent set

1. $\rho(f)$ is an open set, thus $\sigma(f)$ is closed.
2. For any $\lambda \in \rho(f)$,

$$\left\|(f - \lambda I)^{-1}\right\|_{V \to V} \geq \frac{1}{\operatorname{dist}(\lambda, \sigma(f))},\ \text{where } \operatorname{dist}(\lambda, \sigma(f)) = \inf\{d(\lambda, \lambda') : \lambda' \in \sigma(f)\}$$

3. Fix an arbitrary $\mathbf{x} \in V$, the function $h_{\mathbf{x}} : \rho(f) \mapsto \mathbb{C}$ by $h_{\mathbf{x}}(\lambda) = (f - \lambda I)^{-1}(\mathbf{x})$ is analytic, i.e., for any $\lambda_0 \in \rho(f)$, there exists a neighborhood B of λ_0 and $\{a_n \in \mathbb{C}\}_{n=0}^{\infty}$ such that

$$h_{\mathbf{x}}(\lambda) = \sum_{n=0}^{\infty} a_n(\lambda - \lambda_0)^n\ \forall \lambda \in B, \text{ which is the form of the Taylor Expansion}$$

4. Let $R(f) = \sup\{\|\lambda\| : \lambda \in \sigma(f)\}$ be the radius of spectrum. We have $R(f) \leq \|f\|_{V \mapsto V}$, thus $\sigma(\lambda)$ is bounded.
5. Since $\sigma(f) \subset \mathbb{C}$, 1 and 4 together implies $\sigma(f)$ is compact.

6. If $V \neq \{\mathbf{0}\}$, then $\sigma(f) \neq \emptyset$

Proof. We first prove 1-3 together. Fix a $\lambda_0 \in \rho(f)$ and let $\epsilon := \frac{1}{\|(f-\lambda_0 I)^{-1}\|_{V \mapsto V}}$. For any $\lambda \in B(\lambda_0, \epsilon)$, we have

$$[f - \lambda I](\cdot) = [f - \lambda_0 I](\cdot) + (\lambda_0 - \lambda)I(\cdot) = [f - \lambda_0 I] \circ \left[I - (\lambda_0 - \lambda)(f - \lambda_0 I)^{-1}\right](\cdot)$$

Let $g := (\lambda_0 - \lambda)(f - \lambda_0 I)^{-1}$, we have $\|g\|_{V \mapsto V} \leq \epsilon \|(f - \lambda_0 I)^{-1}\|_{V \mapsto V} < 1$. By the above lemma, g is inversible, and so is $(f - \lambda I) = (f - \lambda_0 I) \circ (I - g)$. Thus $\lambda \in \rho(f)$. We have proved that $\rho(f)$ is open, and

$$\operatorname{dist}(\lambda_0, \sigma(f)) \geq \epsilon = \frac{1}{\|(f - \lambda_0 I)^{-1}\|_{V \mapsto V}}$$

is also immediate. Further,

$$(f - \lambda I)^{-1} = (I - g)^{-1} \circ (f - \lambda_0 I)^{-1} = \sum_{n=0}^{\infty} g^n \circ (f - \lambda_0 I)^{-1} = \sum_{n=0}^{\infty} (\lambda_0 - \lambda)^n I \circ \left[(f - \lambda_0 I)^{-1}\right]^{n+1}$$

It gives that for any $\mathbf{x} \in V$, the function $h_{\mathbf{x}}$ is analytic.
To prove 4, let $\lambda \in \mathbb{C}$ such that $\|\lambda\| > \|f\|_{V \mapsto V}$. The operator $g' := \frac{f}{\lambda}$ thus have norm smaller than 1. Applying the above lemma,

$$(I - g') \text{ is invertible, then } (f - \lambda I) = (-\lambda I) \circ (I - g') \text{ is invertible}$$

Thus $\|\lambda\| > \|f\|_{V \mapsto V}$ implies $\lambda \in \rho(f)$.
The proof of 6 requires Liouville's theorem. ◻

Lemma 6.4. Let $p : \mathbb{C} \mapsto \mathbb{C}$ be a polynomial of order $n \in \mathbb{N}^*$ with coefficient $\{a_k\}_{k=0}^n$. Denote $p(f)$ as the linear transformation

$$[p(f)](\cdot) = \left[\sum_{k=0}^{n} a_i f^n\right](\cdot) = \sum_{k=0}^{n} a_i f^n(\cdot) : V \mapsto V$$

We have $p(\sigma(f)) := \{p(\lambda) : \lambda \in \sigma(f)\} \subseteq \sigma(p(f))$.
A simple corollary gives that for any $\lambda \in \sigma(f)$, $\|\lambda\|^n \leq \|f^n\|_{V \mapsto V}$ $\forall n \in \mathbb{N}^*$.

Proof. Fix an arbitrary $\forall \lambda \in \mathbb{C}$. $p(\cdot) - p(\lambda) : \mathbb{C} \mapsto \mathbb{C}$ is a polynomial with a zero at λ. Thus there exists a polynomial $q : \mathbb{C} \mapsto \mathbb{C}$ such that $p(x) - p(\lambda) = (x - \lambda)q(x)$ $\forall x \in \mathbb{C}$. Then we can write (it requires the linearity of f, such that operator product is distributive among addition)

$$[p(f) - p(\lambda)I](\cdot) = [p(f)](\cdot) - [p(\lambda)I](\cdot) = [(f - \lambda I) \circ q(f)](\cdot) = [q(f) \circ (f - \lambda I)](\cdot)$$

Using this simple fact,

$$\forall t, s : V \mapsto V, \text{ if } t \circ s = s \circ t \text{ is bijective, then } s \text{ and } t \text{ are bijective}$$

we have that if $[p(f) - p(\lambda)I]$ is bijective, so is $(f - \lambda I)$. Thus, if $\lambda \in \sigma(f)$, $p(\lambda) \in \sigma(p(f))$.
We therefore have that for any $n \in \mathbb{N}^*$, $\lambda \in \sigma(f)$ implies $\lambda^n \in \sigma(f^n)$. By property 4 of spectrum,

$$\|\lambda\|^n = \|\lambda^n\| \leq \|f^n\|_{V \mapsto V}$$

◻

Theorem 6.4 (Gelfand formula for spectral radius)**.** Let $f : V \mapsto V$ be a linear and bounded map on a complex Banach space $(V, \|\cdot\|_V)$. We have

$$R(f) = \lim_{n\to\infty} (\|f^n\|_{V\mapsto V})^{\frac{1}{n}} = \inf\left\{(\|f^n\|_{V\mapsto V})^{\frac{1}{n}} : n \in \mathbb{N}^*\right\}$$

where $R(f)$ is the spectrum radius.

Proof. Note that the sequence is monotonic decreasing, thus the limit and the infimum are equal. The above lemma gives that $R(f) \leq (\|f^n\|_{V\mapsto V})^{\frac{1}{n}}\ \forall n \in \mathbb{N}^*$. □

We will explore more desirable properties of spectrum of compact operator. The Riesz's lemma (Lemma 1.2) and the following lemma are stated first.

Lemma 6.5. Let $(V, \|\cdot\|_V)$ be a complex Banach space and $f : V \mapsto V$ is a compact linear operator. For any $\gamma \in \mathbb{C} \setminus \{0\}$, the image of $(\gamma I - f)$, i.e., $[\gamma I - f](V)$, is closed.

Proof. Denote $g := \gamma I - f$ and $W := \ker(g)$. For any $\{\mathbf{x}_n \in V\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} g(\mathbf{x}_n) = \mathbf{y} \in V$, we need to prove that $\mathbf{y} \in g(V)$. We first prove that the result holds if $\{\text{dist}(\mathbf{x}_n, W) : n \in \mathbb{N}^*\}$ is bounded in $\mathbb{R}$, in which case there exists $B \in [0, \infty)$ such that

$$\forall n \in \mathbb{N}^*,\ \exists \mathbf{u}_n \in W \text{ such that } d(\mathbf{x}_n, \mathbf{u}_n) < B \text{ i.e. } \|\mathbf{x}_n - \mathbf{u}_n\| < B$$

Thus, $\{\mathbf{x}_n - \mathbf{u}_n : n \in \mathbb{N}^*\}$ is bounded. Since f is a compact operator, $\overline{f(\{\mathbf{x}_n - \mathbf{u}_n : n \in \mathbb{N}^*\})}$ is compact, and there exists a subsequence $\{f(\mathbf{x}_{n_i} - \mathbf{u}_{n_i})\}_{i=1}^{\infty}$ convergent to, say, $\mathbf{y}' \in V$, i.e.,

$$\mathbf{y}' = \lim_{i\to\infty} f(\mathbf{x}_{n_i} - \mathbf{u}_{n_i}) = \lim_{i\to\infty} f(\mathbf{x}_{n_i}) - f(\mathbf{u}_{n_i}) = \lim_{i\to\infty} f(\mathbf{x}_{n_i}) - \gamma\mathbf{u}_{n_i} \text{ (since } \mathbf{u}_{n_i} \in W\ \forall n \in \mathbb{N}^*)$$

Given $\lim_{i\to\infty} g(\mathbf{x}_{n_i}) = \lim_{i\to\infty} \gamma\mathbf{x}_{n_i} - f(\mathbf{x}_{n_i}) = \mathbf{y}$, we have

$$\mathbf{y} + \mathbf{y}' = \lim_{i\to\infty} \gamma\mathbf{x}_{n_i} - \gamma\mathbf{u}_{n_i}, \text{ then } g\left(\frac{\mathbf{y} + \mathbf{y}'}{\gamma}\right) = \lim_{i\to\infty} g(\mathbf{x}_{n_i}) - g(\mathbf{u}_{n_i}) = \lim_{i\to\infty} g(\mathbf{x}_{n_i}) = \mathbf{y}$$

Since $\frac{\mathbf{y}+\mathbf{y}'}{\gamma} \in V$, $\mathbf{y} \in g(V)$ as desired.
Next we prove that the condition $\{\text{dist}(\mathbf{x}_n, W) : n \in \mathbb{N}^*\}$ being bounded must hold. If not, let $\{\mathbf{x}_{n_j}\}_{j=1}^{\infty}$ be a subsequence such that $\text{dist}(\mathbf{x}_{n_j}, W) > j\ \forall j \in \mathbb{N}^*$. Let

$$\mathbf{z}_{n_j} = \frac{\mathbf{x}_{n_j}}{\text{dist}(\mathbf{x}_{n_j}, W)}\ \forall j \in \mathbb{N}^*$$

be another sequence such that $\forall j \in \mathbb{N}^*$,

$$\text{dist}(\mathbf{z}_{n_j}, W) = \inf\left\{d\left(\frac{\mathbf{x}_{n_j}}{\text{dist}(\mathbf{x}_{n_j}, W)}, \mathbf{u}\right) : \mathbf{u} \in W\right\} = \frac{1}{\text{dist}(\mathbf{x}_{n_j}, W)} \inf\left\{d\left(\mathbf{x}_{n_j}, \mathbf{u}'\right) : \mathbf{u}' \in W\right\} = 1$$

Then $\{\text{dist}(\mathbf{z}_{n_j}, W) : j \in \mathbb{N}^*\}$ is bounded, and

$$\left\|\lim_{j\to\infty} g(\mathbf{z}_{n_j})\right\|_V = \lim_{j\to\infty} \|g(\mathbf{z}_{n_j})\|_V = \lim_{j\to\infty} \frac{\|g(\mathbf{x}_{n_j})\|_V}{\text{dist}(\mathbf{x}_{n_j}, W)} = \lim_{j\to\infty} \frac{\|\mathbf{y}\|_V}{\text{dist}(\mathbf{x}_{n_j}, W)} = 0$$

which gives $\lim_{j\to\infty} g(\mathbf{z}_{n_j}) = \mathbf{0}$. We have proved that there exists a further subsequence $\{\mathbf{z}_{n_{j_k}}\}_{k=1}^{\infty}$ and $\{\mathbf{v}_k \in W\}_{k=1}^{\infty}$ such that

$$\mathbf{y}_{\mathbf{z}} := \lim_{k\to\infty} g(\mathbf{z}_{n_{j_k}}) = \mathbf{0},\ \mathbf{y}'_{\mathbf{z}} := \lim_{k\to\infty} f(\mathbf{z}_{n_{j_k}} - \mathbf{v}_k) \in V,\ \mathbf{y}_{\mathbf{z}} + \mathbf{y}'_{\mathbf{z}} = \lim_{k\to\infty} \gamma(\mathbf{z}_{n_{j_k}} - \mathbf{v}_k)$$

$$\text{and } \frac{1}{\gamma} g(\mathbf{y}'_{\mathbf{z}}) = g\left(\frac{\mathbf{y}_{\mathbf{z}} + \mathbf{y}'_{\mathbf{z}}}{\gamma}\right) = \mathbf{y}_{\mathbf{z}} = \mathbf{0}, \text{ thus } \lim_{k\to\infty} \gamma(\mathbf{z}_{n_{j_k}} - \mathbf{v}_k) = \mathbf{y}'_{\mathbf{z}} \in \ker(W)$$

which contradicts to our construction that $\forall j \in \mathbb{N}^*$ and $\mathbf{v} \in \ker(g)$, $d(\mathbf{z}_{n_j}, \mathbf{v}) \geq 1$. □

Theorem 6.5 (Properties of spectrum of compact operator)**.** Let $(V, \|\cdot\|_V)$ be a complex Banach space and $f : V \mapsto V$ is a compact and linear function (thus it is bounded) with spectrum $\sigma(f)$.

1. $\forall \lambda \in \sigma(f) \setminus \{0\}$, λ is an eigenvalue of f, i.e., $(f - \lambda I)$ is not injective. We have proved that the eigenspace $\ker(f - \lambda I)$ is of finite dimension
2. $\forall \lambda \in \sigma(f) \setminus \{0\}$, there exists $m \in \mathbb{N}^*$ such that $\ker[(\lambda I - f)^m] = \ker[(\lambda I - f)^{m+1}]$, which means the sequence $\{\ker[(\lambda I - f)^n]\}_{n=1}^{\infty}$ is eventually constant
3. If $\sigma(f) \setminus \{0\}$ is a infinite set, for any complex sequence $\{\lambda_n \in \sigma(f) \setminus \{0\}\}_{n=1}^{\infty}$, we have $\lim_{n\to\infty} \lambda_n = 0$ whenever it converges
4. $\sigma(f) \setminus \{0\}$ can be empty or finite, but is at most countable
5. If $\dim[V] = \infty$, $0 \in \sigma(f)$. Combining with 1, $\sigma(f) = \{0\} \cup \{\text{eigenvalues of } f\}$ is at most countable

Proof. We first prove 1 by contradiction. Suppose otherwise $\lambda \in \sigma(f) \setminus \{0\}$ while λ is not an eigenvalue. It must be that $(I - \lambda f)$ is injective but not surjective. Letting $\gamma := \frac{1}{\lambda}$, $(\gamma I - f) = (\gamma I) \circ (I - \lambda f)$ is also injective but not surjective. By lemma 6.5, $[\gamma I - f](V)$ is closed. Further, $[\gamma I - f](W_1) \subset W_1$ is proper because $(\gamma I - f)$ is injective. Let $W_n := [(\gamma I - f)^n](V)$ $\forall n \in \mathbb{N}$. We can prove that

$$\forall n \in \mathbb{N}^*,\ W_n \text{ is a proper closed subspace of } W_{n-1}$$

Applying Lemma 1.2, we can take $\mathbf{x}_n \in W_n$ for $n \in \mathbb{N}^*$ such that $\|\mathbf{x}_n\|_{W_n} = \|\mathbf{x}_n\|_V = 1$ and $\text{dist}(\mathbf{x}_n, W_{n+1}) > \frac{1}{2}$. Since f is a compact operator and $\{\mathbf{x}_n\}_{n=1}^{\infty}$ is bounded, there exists a convergent subsequence $\{f(\mathbf{x}_{n_i})\}_{i=1}^{\infty}$ of $\{f(\mathbf{x}_n)\}_{n=1}^{\infty}$. However, $\forall i, j \in \mathbb{N}^*$ such that $j > i$,

$$\begin{aligned}\left\|f(\mathbf{x}_{n_i}) - f(\mathbf{x}_{n_j})\right\|_V =& \left\|\gamma\mathbf{x}_{n_i} - [\gamma I - f](\mathbf{x}_{n_i}) + [\gamma I - f](\mathbf{x}_{n_j}) - \gamma\mathbf{x}_{n_j}\right\|_V \\ =& \|\gamma\| \left\|\mathbf{x}_{n_i} - \frac{1}{\gamma}\left\{[\gamma I - f](\mathbf{x}_{n_i}) - [\gamma I - f](\mathbf{x}_{n_j}) + \gamma\mathbf{x}_{n_j}\right\}\right\|_V \geq \frac{\|\gamma\|}{2}\end{aligned}$$

because $\left\{[\gamma I - f](\mathbf{x}_{n_i}) - [\gamma I - f](\mathbf{x}_{n_j}) + \gamma\mathbf{x}_{n_j}\right\} \in W_{n_i+1}$. The sequence is not Cauchy, contradicting to its being convergent.

Next we prove 2. We note that for two linear and bounded $V \mapsto V$ operator f, g, $\ker(f)$ is closed and $\ker(g) \subseteq \ker(f \circ g)$. Fix $\lambda \in \sigma(f) \setminus \{0\}$. $\{\ker[(\lambda I - f)^n]\}_{n=0}^{\infty}$ is a monotonic increasing sequence of closed subspaces. We claim that

$$\exists m \in \mathbb{N} \text{ such that } \ker[(\lambda I - f)^m] = \ker[(\lambda I - f)^{m+1}]$$

If otherwise for all $n \in \mathbb{N}$, $\ker[(\lambda I - f)^n] \subset \ker[(\lambda I - f)^{n+1}]$, we can take $\{\mathbf{y}_n \in \ker[(\lambda I - f)^n]\}_{n=1}^{\infty}$ such that $\|\mathbf{y}_n\|_V = 1$ and $\text{dist}(\mathbf{y}_n, \ker[(\lambda I - f)^{n-1}]) > \frac{1}{2}$ by Lemma 1.2. As before, take the subsequence such that $\{f(\mathbf{y}_{n_j})\}_{j=1}^{\infty}$ converges. A similar contradiction can be derived.

To prove 3, suppose otherwise there exists a sequence $\{\lambda_i \in \sigma(f) \setminus \{0\}\}_{i=1}^{\infty}$ with corresponding eigenvector $\{\mathbf{u}_i\}_{i=1}^{\infty}$ such that $\|\lambda_i\| > \epsilon_0$ for some $\epsilon_0 > 0$. Let $Z_0 = \{\mathbf{0}\}$ and $Z_n = \text{span}[\{\mathbf{u}_i\}_{i=1}^{n}]$. Since $\{\mathbf{u}_i\}_{i=1}^{\infty}$ is linear independent, Z_n is a proper subspace of Z_{n+1} for $n \in \mathbb{N}$. The same as the proof of 1 and 2, We use Lemma 1.2 and construct sequence to arrive at contradiction.

4 follows from 3, because $\sigma(f) \setminus B\left(0, \frac{1}{n}\right)$ must be a finite set for any $n \in \mathbb{N}^*$, thus

$$\sigma(f) \setminus \{0\} = \bigcup_{n=1}^{\infty} \left[\sigma(f) \setminus B\left(0, \frac{1}{n}\right)\right]$$

is at most countable. There are cases in finite matrix that the eigenvalues are finite or 0 only.
Finally we prove 5. With infinite dimension, there exists a countable linear independent set $\{\mathbf{v}_i\}_{i=1}^{\infty}$ and $\{V_n := \text{span}[\{\mathbf{v}_i\}_{i=1}^{n}]\}_{n=0}^{\infty}$ be a strictly monotonic increasing sequence of closed subspaces. We still construct a sequence $\{\mathbf{z}_n \in V_n\}_{n=1}^{\infty}$ in the same way that all the entries have norm 1 and non of its subsequence converges. We still have a subsequence such that $\{f(\mathbf{z}_{n_k})\}_{k=1}^{\infty}$ converges to, say $\mathbf{y} \in V$.
If $0 \neq \sigma(f)$, f is bijective. Then there exists $\mathbf{z} \in V$ such that $f(\mathbf{z}) = \mathbf{y}$ and $\lim_{k\to\infty} \mathbf{z}_{n_k} = \mathbf{z}$ (by taking f^{-1} on each term), which is impossible. □

6.3 The spectrum theorem

We have known the spectrum theorem for the finite dimension case (Theorem 4.6). To explore the general case, we first define the analogy of Hermitian for linear operator.

Definition 6.3 (Hermitian adjoint)**.** Let $(H, \langle\cdot\rangle)$ be a complex Hilbert space and $f : H \mapsto H$ is a linear and bounded operator. The (Hermitian) adjoint (aka. Hermitian conjugate) of f is a linear operator $f^* : H \mapsto H$ such that

$$\langle f(\mathbf{x}), \mathbf{y}\rangle = \langle \mathbf{x}, f^*(\mathbf{y})\rangle \ \forall \mathbf{x}, \mathbf{y} \in H$$

The existence and uniqueness of f^* is given by the the Riesz representation theorem, because fixing a $\mathbf{y} \in H$, $T(\cdot) := \langle f(\cdot), \mathbf{y}\rangle$ is linear and bounded, then there exists a unique $\mathbf{y}_T$ such that $T(\cdot) = \langle \cdot, \mathbf{y}_T\rangle$. f^* is chosen to map $\mathbf{y}$ to $\mathbf{y}_T$, which is also linear and bounded.

Proposition 6.3 (Properties of operator adjoint)**.** Let f and g be bounded linear $H \mapsto H$ operator.

1. $(f^*)^* = f$
2. $(f + g)^* = f^* + g^*$
3. $\forall a \in \mathbb{C}$, $(af)^* = \bar{a}f^*$
4. $(f \circ g)^* = g^* \circ f^*$
5. $f(H)^\perp = \ker(f^*)$ and $\ker(f)^\perp = \overline{f^*(H)}$

They follows immediately form the definition by the properties of inner product.

Definition 6.4 (Adjoint operator and normal operator)**.** The linear and bounded operator $f : H \mapsto H$ is

- self-adjoint if $f^* = f$.
- skew-adjoint if $f^* = -f$
- normal if $f^* \circ f = f \circ f^*$
- unital if $f^* \circ f = I$. Then one has $f \circ f^* = I = f^* \circ f$, i.e., it is normal

It is clear that self-adjoint operators and skew-adjoint operators are normal.

Remark. In finite dimension Euclidean space, the adjoint of a matrix operator is taking its conjugate transpose. Hermitian matrix is a self-adjoint operator.

Proposition 6.4. If $f : H \mapsto H$ is a normal transformation on Hilbert space $(H, \langle\cdot\rangle)$, then the spectrum radius is $R(f) = \|f\|_{H\mapsto H}$

Proof. For any $\mathbf{x} \in H$ such that $\|\mathbf{x}\|_H = 1$,

$$\|f(\mathbf{x})\|_H^2 = \langle f(\mathbf{x}), f(\mathbf{x})\rangle = \langle f^* \circ f(\mathbf{x}), \mathbf{x}\rangle \le \|f^* \circ f(\mathbf{x})\|_H \|\mathbf{x}\|_H, \text{ then } \|f\|_{H\mapsto H}^2 \le \|f^* \circ f\|_{H\mapsto H}$$

Together with $\|f^* \circ f\|_{H\mapsto H} \le \|f\|_{H\mapsto H}^2$, we have $\|f^* \circ f\|_{H\mapsto H} = \|f\|_{H\mapsto H}^2$.

If f is self-adjoint, $\|f^2\|_{H\mapsto H} = \|f\|_{H\mapsto H}^2$. By induction, $\left\|f^{2^k}\right\|_{H\mapsto H} = \left\|f\right\|_{H\mapsto H}^{2^k}$ $\forall k \in \mathbb{N}^*$. Using the Gelfand formula,

$$R(f) = \lim_{n\to\infty} (\|f^n\|_{H\mapsto H})^{\frac{1}{n}} = \lim_{k\to\infty} \left(\|f^{2^k}\|_{H\mapsto H}\right)^{\frac{1}{2^k}} = \|f\|_{H\mapsto H}$$

When f is normal, we can prove that $f^* \circ f$ is self-adjoint. Since $\|f^* \circ f\|_{H\mapsto H} = \|f\|_{H\mapsto H}^2$, The result still follows from the Gelfand formula. □

Proposition 6.5. Let $f : H \mapsto H$ be a self-adjoint operator.

1. If λ is one of its eigenvalues, then $\lambda \in \mathbb{R}$
2. If $\lambda_1, \lambda_2 \in \mathbb{R}$ are two distinct eigenvalues of f corresponding to eigenvectors $\mathbf{u}_1, \mathbf{u}_2 \in H \setminus \{\mathbf{0}\}$, then $\mathbf{u}_1 \perp \mathbf{u}_2$. In other words, their eigenspaces are orthogonal, $\ker(f - \lambda_1 I) \perp \ker(f - \lambda_2 I)$.
3. $R(f) = \|f\|_{H\mapsto H} = \sup\{\|\langle f(\mathbf{u}), \mathbf{u}\rangle\| : \mathbf{u} \in H,\ \|\mathbf{u}\|_H = 1\}$

Proof. Let $\mathbf{u} \in H \setminus \{\mathbf{0}\}$ be an eigenvector corresponding to λ. Then,

$$\lambda\langle\mathbf{u}, \mathbf{u}\rangle = \langle\lambda\mathbf{u}, \mathbf{u}\rangle = \langle f(\mathbf{u}), \mathbf{u}\rangle = \langle\mathbf{u}, f(\mathbf{u})\rangle = \langle\mathbf{u}, \lambda\mathbf{u}\rangle = \bar{\lambda}\langle\mathbf{u}, \mathbf{u}\rangle$$

Since $\langle\mathbf{u}, \mathbf{u}\rangle \in \mathbb{R}^*$, it is only possible that $\lambda \in \mathbb{R}$. 2 is given in a similar way.

$$\lambda_1\langle\mathbf{u}_1, \mathbf{u}_2\rangle = \langle f(\mathbf{u})_1, \mathbf{u}_2\rangle = \langle\mathbf{u}_1, f(\mathbf{u}_2)\rangle = \lambda_2\langle\mathbf{u}_1, \mathbf{u}_2\rangle, \text{ thus } \langle\mathbf{u}_1, \mathbf{u}_2\rangle = 0 \text{ since } \lambda_1 \ne \lambda_2$$

For 3, the first half is Proposition 6.4. It remains to prove that $\|f\|_{H\mapsto H} = \sup\{\langle f(\mathbf{u}), \mathbf{u}\rangle : \|\mathbf{u}\|_H = 1\}$. For one direction, we have

$$\sup\{\|\langle f(\mathbf{u}), \mathbf{u}\rangle\| : \|\mathbf{u}\|_H = 1\} \le \sup\{\|f(\mathbf{u})\|_{H\mapsto H}\|\mathbf{u}\|_H : \|\mathbf{u}\|_H = 1\} = \|f(\mathbf{u})\|_{H\mapsto H}$$

The case $\|f\|_{H\mapsto H}$ is trivial. Otherwise let $\mathbf{x} \in H$ such that $\|\mathbf{x}\|_H = 1$ and $\mathbf{y} := f(\mathbf{x}) \ne \mathbf{0}$. Define $\mathbf{y}' = \frac{\mathbf{y}}{\|\mathbf{y}'\|_H}$. We have

$$\|\mathbf{y}\|_H = \frac{\langle\mathbf{y}, \mathbf{y}\rangle}{\|\mathbf{y}\|_H} = \langle\mathbf{y}', f(\mathbf{x})\rangle = \langle f(\mathbf{x}), \mathbf{y}'\rangle = \langle f(\mathbf{y}'), \mathbf{x}\rangle = \langle\mathbf{x}, f(\mathbf{y}')\rangle \in \mathbb{R}$$

Observe that

$$\begin{aligned}
\|\mathbf{y}\|_H =& \frac{1}{4}\left(\langle f(\mathbf{x}+\mathbf{y}'), \mathbf{x}+\mathbf{y}'\rangle - \langle f(\mathbf{x}-\mathbf{y}'), \mathbf{x}-\mathbf{y}'\rangle\right) \\
\le& \frac{1}{4}\left(\|\langle f(\mathbf{x}+\mathbf{y}'), \mathbf{x}+\mathbf{y}'\rangle\| + \|\langle f(\mathbf{x}-\mathbf{y}'), \mathbf{x}-\mathbf{y}'\rangle\|\right) \\
\le& \frac{1}{4} \times \sup\{\|\langle f(\mathbf{u}), \mathbf{u}\rangle\| : \|\mathbf{u}\|_H = 1\} \times (\|\mathbf{x}+\mathbf{y}'\|_H + \|\mathbf{x}-\mathbf{y}'\|_H) \\
=& \frac{1}{4} \times \sup\{\|\langle f(\mathbf{u}), \mathbf{u}\rangle\| : \|\mathbf{u}\|_H = 1\} \times 2(\|\mathbf{x}\|_H + \|\mathbf{y}'\|_H) \\
=& \sup\{\|\langle f(\mathbf{u}), \mathbf{u}\rangle\| : \|\mathbf{u}\|_H = 1\}
\end{aligned}$$

Taking the supremum, we have the other side, $\|f\|_{H\mapsto H} = \sup\{\|\langle f(\mathbf{u}), \mathbf{u}\rangle\| : \|\mathbf{u}\|_H = 1\}$ □

Theorem 6.6 (Spectrum Theorem)**.** Let $(H, \langle\cdot\rangle)$ be a Hilbert space and $f : H \mapsto H$ be a compact and self-adjoint bounded linear operator. Then $\sigma(f) \subset \mathbb{R}$, and there exists orthonormal basis of H whose elements are all eigenvetors of f.

Proof. If $\|f\|_{H\mapsto H} = 0$, any nonzero vector is an eigenvector corresponding to eigenvalue 0. Otherwise, Since $R(f) = \|f\|_{H\mapsto H}$ (Proposition 6.5), $\sigma(f) \setminus \{0\} \neq \emptyset$.
Proposition 6.5 gives that $\sigma(f) \setminus \{0\}$ are at most countable number of eigenvalues, and that if there are infinite number of eigenvalues, the sequence of eigenvalues converges to 0. Proposition 6.5 states that the eigenvalues are real numbers and the eigenspaces are mutually orthogonal. We can order the nonzero eigenvalues

$$\{\lambda_n\}_{n=1}^N \text{ such that } |\lambda_n| > |\lambda_{n+1}|$$

It is possible that $N \in \mathbb{N}^*$ or $N = \infty$. Define the corresponding eigenspaces are

$$E_n = \ker(f - \lambda_n I)\ \forall n \in [0, N] \cap \mathbb{N}$$

We know in Proposition 6.5 that each E_n is a closed subspace with finite dimension, at which an finite orthogonal basis B_n can be chosen. If 0 is an eigenvalue, $E_0 := \ker(f)$ is a closed subspace with an orthogonal basis, say, B_0. We claim that such collection of eigenvalue,

$$\bigcup_{n=0}^{N} B_n \text{ (let } B_0 = \emptyset \text{ if 0 is not an eigenvalue)}$$

is a basis for H. Because $E_0^\perp = \ker(f)^\perp = \overline{f(H)}$ and $\ker(f)$ is closed, it suffices (by Corollary 3.2) to prove that $B := \bigcup_{n=1}^N B_n$ is an orthogonal basis for $\overline{f(H)}$. We know B is an orthogonal set in Proposition 6.5. Also $B \subseteq f(H)$ because for any nonzero eigenvector $\mathbf{u}$ corresponding to nonzero eigenvalue λ, $\mathbf{u} = f\left(\frac{\mathbf{u}}{\lambda}\right)$. It remains to show that B is maximal, such that $\mathbf{y} \perp \mathbf{u}\ \forall \mathbf{u} \in B$ for some $\mathbf{y} \in \overline{f(H)}$ implies $\mathbf{y} = 0$.
Let $W_0 = \mathbf{0}$ and $W_n := \operatorname{span}\left[\bigcup_{i=1}^n B_i\right]$ for each n. We first prove this lemma

$$\forall n \in [0, N] \cap \mathbb{N},\ \sup\left\{\langle f(\mathbf{x}), \mathbf{x}\rangle : \mathbf{x} \in W_n^\perp,\ \|\mathbf{x}\|_H = 1\right\} = \mathbf{1}_{n<N} \|\lambda_{n+1}\|$$

For each $n \in [1, N] \cap \mathbb{N}$, define

$$f_n(\cdot) := f(\cdot) - \sum_{i=1}^{n} \lambda_i \sum_{\mathbf{u}\in B_i} \langle \cdot, \mathbf{u}\rangle \mathbf{u} : H \mapsto H$$

Since λ_i are real, we can show that f_n is self-adjoint. f_n is also compact (by Proposition 1.8) because it is the sum of a compact operator f and some finite-rank operators $\lambda_i\langle\cdot, \mathbf{u}\rangle\mathbf{u}$. Observe that $\forall \mathbf{x} \in W_n$, $f_n(\mathbf{x}_n) = 0$, and that $\forall \mathbf{x} \in W_n^\perp$, $f_n(\mathbf{x}) = f(\mathbf{x})$. We discuss two cases.

1. When $\|f_n(\cdot)\|_{H\mapsto H} = 0$, we have that

$$\forall \mathbf{x} \in H,\ f(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i \sum_{\mathbf{u}\in B_i} \langle \mathbf{x}, \mathbf{u}\rangle \mathbf{u}$$

   Therefore, $W_n = f(H) = \overline{f(H)}$ is of finite dimension with orthogonal basis $\bigcup_{i=1}^n B_i$. Indeed, $N = n$ and the chain of eigenvalues terminate.

2. When $\|f_n(\cdot)\|_{H \mapsto H} > 0$, f_n has at least one nonzero eigenvalue λ' with eigenvector $\mathbf{u}' \in W_n^{\perp}$. Since $f_n(\mathbf{x}) = f(\mathbf{x}) \ \forall \mathbf{x} \in W_n^{\perp}$, λ' is also an eigenvector of f and $\lambda' \notin \{\lambda_i\}_{i=1}^n$ because $\mathbf{u}' \in W_n^{\perp}$. Thus we have $N > n$. Also, by Proposition 6.5,

$$
\begin{aligned}
0 < R(f_n) &= \sup\{\|\langle f_n(\mathbf{u}), \mathbf{u}\rangle\| : \mathbf{u} \in H, \ \|\mathbf{u}\|_H = 1\} \\
&= \sup\{\|\langle f_n(\mathbf{u}), \mathbf{u}\rangle\| : \mathbf{u} \in W_n^{\perp}, \ \|\mathbf{u}\|_H = 1\} \\
&= \sup\{\|\langle f(\mathbf{u}), \mathbf{u}\rangle\| : \mathbf{u} \in W_n^{\perp}, \ \|\mathbf{u}\|_H = 1\} = \|\lambda_{n+1}\|
\end{aligned}
$$

We have proved the lemma. If $N < \infty$, case 1 shows that $\overline{f(H)} = \mathrm{span}[\bigcup_{n=1}^N B_n]$ and $B = \bigcup_{n=1}^N B_n$ is a finite orthogonal basis. If $N = \infty$, we have

$$
\lim_{n\to\infty} \|f_n\|_{H\mapsto H} = \lim_{n\to\infty} R(f_n) = \lim_{n\to\infty} \|\lambda_{n+1}\| = 0
$$

If there exists $\mathbf{x}_0 \in \overline{H}$ such that $\mathbf{y}_0 := f(\mathbf{x}_0) \perp \mathbf{u} \ \forall \mathbf{u} \in B$, we have $f_n(\mathbf{x}_0) = f(\mathbf{x}_0) \ \forall n \in \mathbb{N}^*$. Thus

$$
0 \leq \|f(\mathbf{x}_0)\|_H = \lim_{n\to\infty} \|f_n(\mathbf{x}_0)\|_H \leq \lim_{n\to\infty} \|f_n\|_{H\mapsto H}\|\mathbf{x}_0\|_H = 0, \text{ i.e., } \mathbf{y}_0 = 0
$$

Therefore, B is an orthogonal basis for $f(H)$, so as for $\overline{f(H)}$. □

Corollary 6.3. If there exists a compact and self-adjoint bounded linear operator $f : H \mapsto H$ on Hilbert space H such that $\ker(f)$ has a countable orthogonal basis B_0, then $B_0 \cup B$ defined above is a countable orthogonal basis, and thus H is separable.

Proposition 6.6 (Definite operator)**.** Let $(H, \langle\cdot\rangle)$ be a separable Hilbert space and $f : H \mapsto H$ be a compact and self-adjoint bounded linear operator.

1. $\langle f(\mathbf{x}), \mathbf{x}\rangle \geq 0 \ \forall \mathbf{x} \in H$ if and only if all the eigenvalues of f are non-negative, in which case f is a positive semi-definite operator.
2. $\langle f(\mathbf{x}), \mathbf{x}\rangle > 0 \ \forall \mathbf{x} \in H \setminus \{\mathbf{0}\}$ if and only if all the eigenvalues of f are positive, in which case f is a positive definite operator.

Similar properties and definition apply to negative definite operator and negative semi-definite operator.

Note that if $\mathbf{u}$ is a non-zero eigenvector corresponding to eigenvalue λ,

$$
\langle f(\mathbf{u}), \mathbf{u}\rangle = \lambda\|\mathbf{u}\|_H
$$

Thus $\langle f(\mathbf{x}), \mathbf{x}\rangle \geq 0 \ \forall \mathbf{x} \in H$ gives that $\lambda \geq 0$, and $\langle f(\mathbf{x}), \mathbf{x}\rangle > 0 \ \forall \mathbf{x} \in H$ gives that $\lambda > 0$.
We then prove the other direction. Suppose $\dim[H] = \infty$ (the finite dimension case is similar). Let $\{\mathbf{u}_n\}_n^{\infty}$ be an orthonormal basis for H with each element being a eigenvector corresponding to

eigenvalue λ_n. For any $\mathbf{x} \in H$, we have (by the continuity of f and $\langle \cdot, \cdot \rangle$)

$$\begin{aligned}
\langle f(\mathbf{x}), \mathbf{x} \rangle &= \left\langle f\left(\sum_{n=1}^{\infty} \langle \mathbf{x}, \mathbf{u}_n \rangle \mathbf{u}_n\right), \sum_{n=1}^{\infty} \langle \mathbf{x}, \mathbf{u}_n \rangle \mathbf{u}_n \right\rangle \\
&= \left\langle \sum_{n=1}^{\infty} \lambda_n \langle \mathbf{x}, \mathbf{u}_n \rangle \mathbf{u}_n, \sum_{n=1}^{\infty} \langle \mathbf{x}, \mathbf{u}_n \rangle \mathbf{u}_n \right\rangle \\
&= \sum_{n=1}^{\infty} \lambda_n \langle \mathbf{x}, \mathbf{u}_n \rangle \left\langle \mathbf{u}_n, \sum_{n=1}^{\infty} \langle \mathbf{x}, \mathbf{u}_n \rangle \mathbf{u}_n \right\rangle \\
&= \sum_{n=1}^{\infty} \lambda_n \| \langle \mathbf{x}, \mathbf{u}_n \rangle \|^2
\end{aligned}$$

Therefore, if $\lambda_n \geq 0 \ \forall n \in \mathbb{N}^*$, $\langle f(\mathbf{x}), \mathbf{x} \rangle \geq 0$, and if $\lambda_n > 0 \ \forall n \in \mathbb{N}^*$ and $\mathbf{x} \neq \mathbf{0}$, $\langle f(\mathbf{x}), \mathbf{x} \rangle > 0$.

7 Application: Low rank approximation

7.1 Arzelà–Ascoli theorem

Definition 7.1 (Uniformly equicontinuous and uniformly bounded)**.** Let $\{f_i : i \in I\}$ be a collection of continuous mappings between two metric spaces $(X, d_X) \mapsto (Y, d_Y)$.

1. The set is uniformly equicontinuous if

   $\forall \epsilon > 0, \ \exists \delta > 0$ such that $\forall i \in I, x_1, x_2 \in X,$ we have $d(x_1 - x_2) < \delta \ \Rightarrow \ d(f(x_1) - f(x_2)) < \epsilon$

   That is, the choice of δ is independent of i and x, y.

2. The set is uniformly bounded if the function are bounded by the same constant, i.e.,

$$\sup\{|f_i(x)| : i \in I, x \in X\} < \infty$$

Theorem 7.1 (Arzelà–Ascoli theorem)**.** Let I be a compact rectangle in $\mathbb{R}^d$ with positive volume and $\{f_n : I \mapsto \mathbb{C}\}_{n=1}^{\infty}$ be a sequence of continuous real functions. If the sequence is uniformly equicontinuous and uniformly bounded, then there exists a subsequence that converges uniformly. And conversely, if every subsequence of $\{f_n\}_{n=1}^{\infty}$ has a further subsequence that converges uniformly, then $\{f_n\}_{n=1}^{\infty}$ is uniformly equicontinuous and uniformly bounded.
The statement implies the following. Let $(\mathcal{C}(I \to \mathbb{C}), \| \cdot \|_{\infty})$ be a $\mathcal{L}^{\infty}$ complex normed vector space for continuous $I \to \mathbb{C}$ function. For any $A \subseteq \mathcal{C}(I \to \mathbb{C})$, A is compact if and only if A is closed, bounded and uniformly equicontinuous.
The above holds if we replace $\mathbb{C}$ with $\mathbb{R}$.

Proof. Suppose $\{f_n\}_{n=1}^{\infty}$ is uniformly equicontinuous and uniformly bounded, we show that it has a required subsequence. Let $\{q_k\}_{k=1}^{\infty}$ be a enumeration of $\mathbb{Q}^k \cap I$. Let $B := \sup\{\|f_n\|_{\infty} : n \in \mathbb{N}^*\} < \infty$ be the bound. Then $\sup\{f_n(x_1) : n \in \mathbb{N}^*\} < B$. Let $I_1 \subseteq \mathbb{N}^*$ be the index set of a subsequence of $\{f_n\}_{n=1}^{\infty}$ such that

$$\{f_n(x_1)\}_{n \in I_1} \text{ converges to, say, } L_1 \in \mathbb{C}$$

Recursively, given the subsequence indexed by I_k, we can construct a further subsequence

$$\{f_n\}_{n \in I_{k+1}}, \ I_{k+1} \subseteq I_k \text{ such that } \{f_n(x_{k+1})\}_{n \in I_{k+1}} \text{ converges to, say, } L_{k+1} \in \mathbb{C}$$

Then, for every $k \in \mathbb{N}^*$, $\{f_n(\cdot)\}_{n \in I_k}$ converges at points $\{q_1, \cdots, q_k\}$. Now define the diagonal sequence $\{f_{kk}\}_{k=1}^{\infty}$, such that f_{kk} is the k-th terms of the sequence $\{f_n\}_{n \in I_k}$. By our construction, $\{f_{kk}(\cdot)\}_{k=1}^{\infty}$ converges at every point in $\mathbb{Q} \cap I$.
Choose any $\epsilon > 0$. For any $q \in \mathbb{Q} \cap I$, there exists $N_q \in \mathbb{N}^*$ such that

$$\|f_{j_1 j_1}(q) - f_{j_2 j_2}(q)\| < \frac{\epsilon}{3} \; \forall j_1, j_2 \geq N_q$$

Since $\{f\}_{n=1}^{\infty}$ is uniformly equalcontinuous, for every $x \in I$, there exists a open neighborhood $B_x \ni x$ such that

$$\|f_n(x_1) - f_n(x_2)\| \leq \frac{\epsilon}{3} \; \forall n \in \mathbb{N}^*, x_1, x_2 \in B_x$$

$\{B_x : x \in I\}$ is a open cover of the compact set I. There exists a finite sub-cover, say $\{B_{x_i}\}_{i=1}^{m}$. Then there exists $K \in \mathbb{N}^*$ such that for any B_{x_i}, $\exists k \in \{1, \cdots, K\}$, $q_k \in B_{x_i}$. Therefore, for any $t \in I$, there exist $i \in \{1, \cdots, m\}$ such that $t \in B_{x_i}$ and there exists $k \in \{1, \cdots, K\}$ such that $q_k \in B_{x_i}$. Further,

$$\begin{aligned}
\|f_{j_1 j_1}(t) - f_{j_2 j_2}(t)\| &\leq \|f_{j_1 j_1}(t) - f_{j_1 j_1}(q_k)\| + \|f_{j_1 j_1}(q_k) - f_{j_2 j_2}(q_k)\| + \|f_{j_2 j_2}(t) - f_{j_2 j_2}(q_k)\| \\
&\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \\
&\forall j_1, j_2 \geq N := \max\{N_{q_1}, \cdots, N_{q_K}\}
\end{aligned}$$

This shows that $\{f_{kk}\}_{k=1}^{\infty}$, a subsequence of $\{f_n\}_{n=1}^{\infty}$ is uniformly Cauchy. The pointwise limit $f = \lim_{k \to \infty} f_{kk}$ exists, and

$$\forall t \in I, \; |f_{kk}(t) - f(t)| = \lim_{j \to \infty} |f_{kk}(t) - f_{jj}(t)| \leq \epsilon \text{ whenever } k \geq N$$

Therefore, $\{f_{kk}\}_{k=1}^{\infty}$ converges uniformly to f.
We can prove by taking the limit that f is continuous and bounded, thus $f \in \mathcal{C}(I \to \mathbb{C})$. With this, if a set $A \subseteq \mathcal{C}(I \to \mathbb{C})$ is closed, bounded and uniformly equicontinuous, any its sequence have a subseqence convergent in A, thus A is compact.
We prove the other direction. Suppose $\{f_n\}_{n=1}^{\infty}$ is not uniformly bounded. We can easily construct a subsequence such that any further subsequence of it do not converges uniformly. It remains to show that the existence of convergent further subsequence implies that $\{f_n\}_{n=1}^{\infty}$ is uniformly equicontinuous. First, $\{f_n : n \in \mathbb{N}^*\}$ is totally bounded, that is, for any $\epsilon > 0$, there exists a finite open cover $\{B(f_{n_j}, \frac{\epsilon}{3})\}_{j=1}^{\infty}$ such that

$$\{f_n : n \in \mathbb{N}^*\} \subseteq \bigcup_{j=1}^{m} B\left(f_{n_j}, \frac{\epsilon}{3}\right) \text{ and } B\left(f_{n_j}, \frac{\epsilon}{3}\right) \text{ is open ball in } (\mathcal{C}(I \mapsto \mathbb{C}), \mathcal{L}^{\infty})$$

which can be proved by contradiction (similar to compactness implying totally boundedness). We also have that $\forall n \in \mathbb{N}^*$, f_n is uniformly bounded because $[a, b]$ is compact. Then, for any $j \in \{1, \cdots, m\}$ there exits $\delta_k > 0$ such that

$$\begin{aligned}
&\forall f \in B\left(f_{n_j}, \frac{\epsilon}{3}\right) \text{ and } x_1, x_2 \in I, \\
&\quad |x_1 - x_2| < \delta_j \;\Rightarrow\; \left|f_{n_j}(x_1) - f_{n_j}(x_2)\right| < \frac{\epsilon}{3} \\
&\quad \phantom{|x_1 - x_2| < \delta_j} \;\Rightarrow\; |f(x) - f(y)| < |f(x) - f_{n_j}(x)| + |f_{n_j}(x) - f_{n_j}(y)| + |f(y) - f_{n_j}(y)| < \epsilon
\end{aligned}$$

Letting $\delta = \min\{\delta_1, \cdots, \delta_m\}$, we can show that $\{f_n\}_{n=1}^\infty$ is uniformly equicontinuous.
For compact $A \subseteq \mathcal{C}(I \to \mathbb{C})$, we have show that it is closed and totally bounded. We can show that A is uniformly equicontinuous by applying the above reasoning, i.e., substituting $\{f_n : n \in \mathbb{N}^*\}$ with A. □

Example:
As a subset of $\mathcal{C}([a,b] \mapsto \mathbb{R})$, the set

$$\mathcal{F} := \{f : [a,b] \mapsto \mathbb{R} \text{ is continuous differentiable} : |f(x)| < B_1 \text{ and } |f(x)'| < B_2 \ \forall x \in [a,b]\}$$

where $B_1, B_2 \in \mathbb{R}^+$ are bounds, is compact.

Proof. Clearly the set $\mathcal{F}$ is bounded. For any uniformly convergent sequence $\{f_n \in \mathcal{F}\}_{n=1}^\infty$, f_n' are Lebesgue integrable because they are continuous and bounded. Thus by dominated convergence theorem, $\lim_{n\to\infty} f_n'$ exists and is continuous, which gives that the limit of $\{f_n \in \mathcal{F}\}_{n=1}^\infty$ is in $\mathcal{F}$. Therefore, $\mathcal{F}$ is closed. It is further uniformly equicontinuous since the rate of change is bounded by B_2 by the mean value theorem. □

7.2 Hilbert–Schmidt integral operator

Definition 7.2 (Hilbert–Schmidt integral operator)**.** Let $A_1 \subseteq \mathbb{R}^{k_1}$, $A_2 \subseteq \mathbb{R}^{k_2}$ be two compact rectangles with positive volume and μ_1, μ_2 are Lebesgue measures on $\mathbb{R}^{k_1}$, $\mathbb{R}^{k_2}$ respectively. Let $\mathcal{F}_1 := (\mathcal{C}(A_1 \mapsto \mathbb{C}), \|\cdot\|_\infty)$ and $\mathcal{F}_2 := (\mathcal{C}(A_2 \mapsto \mathbb{C}), \|\cdot\|_\infty)$ be two complex Banach spaces. Given a bounded and continuous function $K \in \mathcal{C}(A_1 \times A_2 \mapsto \mathbb{C})$ (called kernel), define

$$T_K : \mathcal{F}_1 \mapsto \mathcal{F}_2 \text{ by } \forall f \in \mathcal{F}_1,\ T_K(f) \in \mathcal{F}_2 \text{ such that } \forall x \in A_2,\ [T_K(f)](x) = \int_{A_1} K(x,t) f(t) \mathrm{d}\mu_1(t)$$

to be a Hilbert–Schmidt integral operator. It is easy to show that the definition is valid by proving that $T_K(f)$ is continuous and bounded. We also see that T_K is determined by the uniquely K.
We can define Hilbert–Schmidt integral operator in the same way when $\mathcal{F}_1, \mathcal{F}_2$ are real Banach space of real functions.

Proposition 7.1. $T_K : \mathcal{F}_1 \mapsto \mathcal{F}_2$ is a compact linear operator, even if we change the norm of $\mathcal{F}_1, \mathcal{F}_2$ into $\mathcal{L}^2$ norm.
The follows discusses the case when $H := \mathcal{F}_1 = \mathcal{F}_2$ (i.e., $A := A_1 = A_2$ and $\mu := \mu_1 = \mu_2$) is a complex $\mathcal{L}^2$ Hilbert space.

1. If $K(s,t) = \overline{K(t,s)}\ \forall s,t \in A$, then T_K is a self-adjoint operator
2. Further from 1, if

$$\forall \{\alpha_i \in \mathbb{C}\}_{i=1}^n, \{s_i \in A\}_{i=1}^n, \{t_i \in A\}_{i=1}^n, \text{ we have} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\alpha_j} K(s_j, t_i) > 0$$

   ($LHS \in \mathbb{R}$ because T_K is self-adjoint), then T_k is a positive semi-definite operator, i.e., $\langle T_k(f), f\rangle > 0\ \forall f \in H$
3. Further from 1, $\sup\{|K(t,t)| : t \in A\} \le \sup\{|\langle T_K(f), f\rangle| : f \in H, \|f\|_H = 1\}$ (Note that $K(t,t) \in \mathbb{R}\ \forall t \in A$ and $\langle T_K(f), f\rangle \in \mathbb{R}\ \forall f \in H$)

We refer to T_K whose Kernel K satisfies the premise of 1 and 2 as a non-negative self-adjoint Hilbert–Schmidt integral operator.

Proof. Clearly T_K is linear and bounded. It remains to show that

$$S := \overline{T_K\left[B(\mathbf{0}_{\mathcal{F}_1}, 1)\right]} \subseteq \mathcal{F}_2 \text{ is compact}$$

by Arzelà–Ascoli theorem. Obviously, S is closed and bounded. It is further uniformly equicontinuous, because

$$\forall f \in B(\mathbf{0}_{\mathcal{F}_1}, 1), \forall s, t \in A_2, \ \left\| \left[T_K(f)\right](s) - \left[T_K(f)\right](t) \right\| \leq \int_{A_1} \left\| K(s,v) - K(t,v) \right\| \|f\|_\infty \mathrm{d}\mu_1(v)$$

For any $\epsilon > 0$, the RHS can be bounded by ϵ if we choose $\delta > 0$ such that $\|K(s,v) - K(t,v)\|$ is small when $\|s - t\| < \delta$. Such choice also apply to the boundary of S. Thus S is uniformly equicontinuous. The theorem gives that S is compact with respect to $\mathcal{L}^\infty$ norm. Since $\mu_2(A_2) < \infty$, convergence in $\mathcal{L}^\infty$ norm implies convergence in $\mathcal{L}^2$ norm. Thus S is still compact with respect to $\mathcal{L}^2$ norm.
When $K(s,t) = \overline{K(t,s)}$ $\forall s, t \in A$, we have $\forall f \in H$,

$$\begin{aligned} \langle T_K(f), f\rangle = \int_A \int_A K(s,t) f(t) \overline{f(s)} \mathrm{d}\mu(s) \mathrm{d}\mu(t) &= \int_A \int_A f(t) \overline{K(t,s) f(s)} \mathrm{d}\mu(s) \mathrm{d}\mu(t) \\ &= \int_A f(t) \overline{[T_K(f)](t)} \mathrm{d}\mu(t) = \langle f, T_K(f)\rangle \end{aligned}$$

1 is proven. Note that if the premise of 2 hold, any Riemann sum approaching

$$\int_A \int_A K(s,t) f(t) \overline{f(s)} \mathrm{d}\mu(s) \mathrm{d}\mu(t)$$

is non-negative, and that the Lebesgue integral agrees on Riemann integral because $A \times A$ is compact and $K(s,t) f(t) \overline{f(s)}$ is continuous and bounded, thus $\langle T_K(f), f\rangle \geq 0$.
To prove 3, since A is compact, we can take $t_0 \in A$ such that $|K(t_0, t_0)| = \sup\{|K(t,t)| : t \in A\}$. Suppose t_0 is not at the boundary of K for simplicity (otherwise take the union with A if necessary for the followings).
Fix an arbitrary $\epsilon > 0$. Since K is continuous, there exists $\delta > 0$ such that $|\mathrm{Re}[K(s,t)]| \geq |K(t_0,t_0)| - \epsilon$ $\forall s, t \in B(t_0, \delta)$. There also exists $f \in H$ such that $f(t) = 0$ $\forall t \notin B(t_0, \delta)$ and $f(t) \in [0, \infty)$ $\forall t \in B(t_0, \delta)$ with $\|f\|_H = 1$. Then,

$$\begin{aligned} |\langle T_K(f), f\rangle| &= \left| \int_S \int_S K(s,t) f(t) \overline{f(s)} \mathrm{d}\mu(t) \mathrm{d}\mu(s) \right| \\ &= \left| \mathrm{Re} \left[\int_S \int_S K(s,t) f(t) \overline{f(s)} \mathrm{d}\mu(t) \mathrm{d}\mu(s) \right] \right| \ (\text{since } \langle T_K(f), f\rangle \in \mathbb{R}) \\ &= \int_{B(t_0,\delta)} \int_{B(t_0,\delta)} \mathrm{Re}\left[|K(s,t)|\right] f(t) f(s) \mathrm{d}\mu(t) \mathrm{d}\mu(s) \\ &\geq \int_{B(t_0,\delta)} \int_{B(t_0,\delta)} \left[|K(t_0,t_0)| - \epsilon\right] f(t) f(s) \mathrm{d}\mu(t) \mathrm{d}\mu(s) \\ &= \left[|K(t_0,t_0)| - \epsilon\right] \|f\|_H = |K(t_0,t_0)| - \epsilon \end{aligned}$$

Therefore, $\sup\{|\langle T_K(f), f\rangle| : f \in H, \|f\|_H = 1\} \geq |K(t_0,t_0)| - \epsilon$. Taking $\epsilon \to 0$, we have $\sup\{|K(t,t)| : t \in A\} \leq \sup\{|\langle T_K(f), f\rangle| : f \in H, \|f\|_H = 1\}$ ∎

Theorem 7.2 (Mercer's theorem). Let $A \subset \mathbb{R}^d$ be a compact rectangle with positive volumn and $H := (\mathcal{C}(A \mapsto \mathbb{C}), \langle \cdot, \cdot \rangle)$ be the complex $\mathcal{L}^2$ Hilbert space. Let $T_K : H \mapsto H$ be a non-negative self-adjoint Hilbert–Schmidt integral operator with kernel $K : A \times A \mapsto \mathbb{C}$. There exists a countable number of eigenvectors $\{u_i \in H\}_{i=1}^{\infty}$ (i.e., eigenfunctions since $u_i : A \mapsto \mathbb{C}$) of T_K corresponding to eigenvectors $\{\lambda_i\}_{i=1}^{\infty}$ respectively, which constitute an orthogonal basis of H. Further,

$$\forall s, t \in A,\ K(s,t) = \lim_{n \to \infty} \sum_{i=1}^{n} \lambda_i u_i(s) \overline{u_i(t)}$$

The convergence is absolute and uniform among $(s,t) \in A \times A$.

Proof. Note that $\dim[H] = \infty$, and H is separable because the polynomials with rational coefficients comprise a countable dense set. Therefore, the orthogonal basis is countable. Since we have proved that T_K is compact, self-adjoint and positive semi-definite, the orthogonal basis using eigenvectors $\{u_i \in H\}_{i=1}^{\infty}$ exists by spectrum theorem, and $\lambda_i \in [0, \infty)\ \forall i \in \mathbb{N}^*$. For any $f \in H$, we have

$$f = \sum_{i=1}^{\infty} \langle f, u_i \rangle u_i, \text{ thus } T_K(f) = \sum_{i=1}^{\infty} \lambda_i \langle f, u_i \rangle u_i, \text{ and } \langle T_K(f), f \rangle = \sum_{i=1}^{\infty} \lambda_i \| \langle f, u_i \rangle \|^2$$

For $n \in \mathbb{N}^*$, define Hilbert–Schmidt integral operators T_{K_n}, T_{R_n} generated from

$$K_n, R_n : A \times A \mapsto \mathbb{C} \text{ by } K_n(s,t) = \sum_{i=1}^{n} \lambda_i u_i(s) \overline{u_i(t)} \text{ and } R_n(s,t) = K(s,t) - K_n(s,t)\ \forall s, t \in A$$

which is continuous and bounded. T_{K_n}, T_{R_n} are also self-disjoint. Then,

$$[T_{K_n}(f)](\cdot) = \sum_{i=1}^{n} \lambda_i u_i(\cdot) \int_A f(t) \overline{u_i(t)} \mathrm{d}\mu(t) = \sum_{i=1}^{n} \lambda_i \langle f, u_i \rangle u_i(\cdot), \text{ and } \langle T_{K_n}(f), f \rangle = \sum_{i=1}^{n} \lambda_i \| \langle f, u_i \rangle \|^2$$

and thus,

$$\langle T_{R_n}(f), f \rangle = \langle T_K(f), f \rangle - \langle T_{K_n}(f), f \rangle = \sum_{i=n+1}^{\infty} \lambda_i \| \langle f, u_i \rangle \|^2 \geq 0$$

Thus T_{R_n} is positive semi-definite. $R_n(t,t) \in \mathbb{R}\ \forall t \in A$ by the definition. We further prove that $R_n(t,t) \geq 0\ \forall t \in A$. Otherwise if $\exists t_0 \in A$ (suppose not at the boundary, otherwise similar) with $R_n(t_0, t_0) < 0$, by the continuity, there exists δ such that $\forall s, t \in B(t_0, \delta)$, $\mathrm{Re}[R_n(s,t)] < 0$. Then there exists $f \in H$ such that $f(t) \in [0, \infty)\ \forall t \in B(t_0, \delta)$ and $f(t) = 0\ \forall t \notin B(t_0, \delta)$ with

$$\begin{aligned} \langle T_{R_n}(f), f \rangle &= \int_A \int_A R_n(s,t) f(t) \overline{f(s)} \mathrm{d}\mu(t) \mathrm{d}\mu(s) \\ &= \int_{B(t_0,\delta)} \int_{B(t_0,\delta)} \mathrm{Re}[R_n(s,t)] f(t) f(s) \mathrm{d}\mu(t) \mathrm{d}\mu(s) < 0 \end{aligned}$$

Therefore, we have proved that

$$\begin{aligned} &\forall t \in A, \forall n \in \mathbb{N}^*,\ K(t,t) \geq K_n(t,t) = \sum_{i=1}^{n} \lambda_i u_i(t) \overline{u_i(t)} = \sum_{i=1}^{n} \lambda_i \| u_i(t) \|^2 \\ &\Rightarrow \sum_{i=1}^{\infty} \lambda_i \| u_i(t) \|^2 \leq M := \sup\{K(t,t) : t \in A\} \end{aligned}$$

By Cauchy-Schwartz inequality,

$$\forall s,t \in A,\ \left\|\sum_{i=1}^{\infty}\left[\sqrt{\lambda_i}u_i(s)\right]\left[\overline{u_i(t)}\sqrt{\lambda_i}\right]\right\|^2 \leq \left[\sum_{i=1}^{\infty}\lambda_i\|u_i(s)\|^2\right]\left[\sum_{i=1}^{\infty}\lambda_i\|u_i(t)\|^2\right] \leq M^2$$

Thus, we have prove that the series is uniformly bounded, and further point-wise absolute convergence because it is point-wise monotonically increasing

$$\forall s,t \in A,\ \lim_{n\to\infty}\|K_n(s,t)\| = \sum_{i=1}^{\infty}\lambda_i\left\|u_i(s)\overline{u_i(t)}\right\| = \left\|\sum_{i=1}^{\infty}\left[\sqrt{\lambda_i}u_i(s)\right]\left[\overline{u_i(t)}\sqrt{\lambda_i}\right]\right\| \leq M$$

It remains to show that the convergence of $\{K_n\}_{n=1}^{\infty}$ to K, or equivalently, $\{R_n\}_{n=1}^{\infty}$ to zero function, is uniform on $A \times A$. Note that for any $n \in \mathbb{N}^*$, by the above proposition,

$$\begin{aligned}
\sup\{|R_n(t,t)| : t \in A\} \leq & \sup\{\langle T_{R_n}(f), f\rangle : f \in H, \|f\|_H = 1\} \\
= & \sup\left\{\sum_{i=n+1}^{\infty}\lambda_i\|\langle f, u_i\rangle\|^2 : f \in H, \|f\|_H = 1\right\} \\
= & \sup\left\{\sum_{i=n+1}^{\infty}\lambda_i\|\langle f, u_i\rangle\|^2 : f \in \overline{\operatorname{span}[\{u_{n+1}, u_{n+2}\cdots\}]}, \|f\|_H = 1\right\} \\
= & \sup\{\lambda_{n+1}, \lambda_{n+2}, \cdots\} \text{ (since } \sum_{i=n+1}^{\infty}\|\langle f, u_i\rangle\|^2 = 1)
\end{aligned}$$

Using Cauchy-Schwartz inequality again,

$$\begin{aligned}
\forall s,t \in A,\ \|R_n(s,t)\| = \sum_{i=n+1}^{\infty}\lambda_i\left\|u_i(s)\overline{u_i(t)}\right\| & \leq \left[\sum_{i=n+1}^{\infty}\lambda_i\left\|u_i(s)\overline{u_i(s)}\right\|\right]^{\frac{1}{2}}\left[\sum_{i=n+1}^{\infty}\lambda_i\left\|u_i(t)\overline{u_i(t)}\right\|\right]^{\frac{1}{2}} \\
& = |R_n(t,t)|^{\frac{1}{2}}|R_n(s,s)|^{\frac{1}{2}} \\
& \leq \sup\{\lambda_{n+1}, \lambda_{n+2}, \cdots\} < \infty
\end{aligned}$$

The RHS goes to 0 as $n \to \infty$ by the property of eigenvalue and it is independent to s, t. Thus $\lim_{n\to\infty} R(s,t) = 0$ uniformly among $(s,t) \in A \times A$. □

Corollary 7.1. The Mercer's theorem gives

$$K(t,t) = \sum_{i=1}^{\infty}\lambda_i u_i(t)\overline{u_i(t)}\ \forall t \in A$$

taking the integral, we have

$$\int_A K(t,t)\mathrm{d}\mu(t) = \int_A\sum_{i=1}^{\infty}\lambda_i u_i(t)^2\mathrm{d}\mu(t) = \sum_{i=1}^{\infty}\lambda_i\int_A u_i(t)\overline{u_i(t)}\mathrm{d}\mu(t) = \sum_{i=1}^{\infty}\lambda_i\|u_i\|_H = \sum_{i=1}^{\infty}\lambda_i$$

We call $\int_A K(t,t)\mathrm{d}\mu(t)$ (or $\sum_{i=1}^{\infty}\lambda_i$) the trace of T_K.

7.3 Accelerating Gaussian Process prediction

Definition 7.3 (Gaussian Process)**.** Given a probability space $(\Omega, \mathcal{B}, P)$, suppose there exists a collection of real random variables $\{X_t\}_{t\in A}$, where the index set A is a complex rectangle in $\mathbb{R}^d$ (usually $d = 1$ for time series, and $d = 2$ for spatial data). $\{X_t\}_{t\in A}$ is a Gaussian Process with mean function $\mu : A \to \mathbb{R}$ and covariance function $\gamma : A \times A \to \mathbb{R}$ if for any finite collection $\{t_i \in A\}_{i=1}^n$,

$$[X_{t_1}, \cdots, X_{t_n}]^T \sim N_n\left([\mu(t_1), \cdots, \mu(t_n)]^T, \Sigma\right), \text{ where } \Sigma \in \mathbb{R}^{n\times n}, \ \Sigma_{i,j} = \gamma(t_i, t_j)$$

Some requirements are imposed on μ and γ,

- μ and γ are bounded and continuous on A
- $\gamma(s,t) = \gamma(t,s) \ \forall s,t \in A$
- For any $\{\alpha_i \in \mathbb{R}\}_{i=1}^m$, $\{s_i \in \mathbb{C}\}_{i=1}^m$, $\{t_i\}_{i=1}^m$,

$$\sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j K(s_i, t_j) \geq 0$$

The requirements guarantee the covariance matrix to be symmetric and positive semi-definite. Note that γ is then a kernel for real non-negative self-adjoint Hilbert–Schmidt integral operator. It is possible to extend it to complex Gaussian Process, where any finite subset of $\{X_t\}$ admits a complex Normal distribution, and γ is a kernel for complex non-negative self-adjoint Hilbert–Schmidt integral operator.

Theorem 7.3 (Karhunen–Loève theorem)**.** Let A be a complex rectangle in $\mathbb{R}^k$ with Lebesgue measure μ and $\{X_t\}_{t\in A}$ be a Gaussian Process with zero mean function and covariance function γ (it also holds for any continuous stochastic process as long as μ and γ satisfies the above). Then there exists a basis $\{e_n\}_{n=1}^\infty$ for $\mathcal{C}(A \mapsto \mathbb{R})$ and a collection of mutually uncorrelated random variable $\{Z_n\}_{n=1}^\infty$ such that

$$\forall t \in A, \ X_t = \sum_{n=1}^\infty Z_n e_n(t)$$

Proof. Given γ, we can define a real non-negative self-adjoint Hilbert–Schmidt integral operator mapping $\mathcal{C}(A \mapsto \mathbb{R}) \mapsto \mathcal{C}(A \mapsto \mathbb{R})$. Let $\{e_n\}_{n=1}^\infty$ be its eigenfunctions corresponding to eigenvalues $\{\lambda_n\}_{n=1}^\infty$, as well as a orthogonal basis for $\mathcal{C}(A \mapsto \mathbb{R})$. Then,

$$\forall \omega \in \Omega, \ \forall s \in A, \ X_s(\omega) = \sum_{n=1}^\infty e_n(s) \int_{t\in A} X_t(\omega) e_n(t) \mathrm{d}\mu(t)$$

For any $n \in \mathbb{N}^*$, define

$$Z_n = \int_{t\in A} X_t e_n(t) \mathrm{d}\mu(t)$$

which is a random variable because all the Riemann sums are random variables. We have $\forall t \in A, \ X_t(\omega) = \sum_{n=1}^\infty e_n(t) Z_n$ as desired and

$$\mathbb{E}(Z_n) = \int_{t\in A} \mathbb{E}(X_t) e_n(t) \mathrm{d}\mu(t) = 0$$

By Mercer's theorem,

$$\forall s,t \in A,\ \gamma(s,t) = \sum_{i=1}^{\infty} \lambda_i e_i(s) e_i(t) =$$

Therefore, $\forall n,m \in \mathbb{N}^*$,

$$\begin{aligned}
\mathbb{E}(Z_n Z_m) =& \mathbb{E}\left[\left(\int_{t\in A} X_t e_n(t) \mathrm{d}\mu(t)\right)\left(\int_{s\in A} X_t e_m(s) \mathrm{d}\mu(s)\right)\right] \\
=& \int_{t\in A}\int_{s\in A} \mathbb{E}(X_t X_s) e_n(t) e_m(s) \mathrm{d}\mu(s) \mathrm{d}\mu(t) \\
=& \int_{t\in A}\int_{s\in A} \gamma(t,s) e_n(t) e_m(s) \mathrm{d}\mu(s) \mathrm{d}\mu(t) \\
=& \int_{t\in A}\int_{s\in A} \sum_{i=1}^{\infty} \lambda_i e_i(s) e_i(t) e_n(t) e_m(s) \mathrm{d}\mu(s) \mathrm{d}\mu(t) \\
=& \int_{t\in A} \sum_{i=1}^{\infty} \lambda_i e_i(t) e_n(t) \langle e_i, e_m \rangle \mathrm{d}\mu(t) \\
=& \int_{t\in A} \lambda_m e_m(t) e_n(t) \mathrm{d}\mu(t) \\
=& \lambda_n \times \mathbf{1}_{n=m}
\end{aligned}$$

Note that $\gamma(t,s) = \mathrm{var}(X_t, X_s) = \mathbb{E}(X_t X_s)$ is by the definition. □

Corollary 7.2. If $\{X_t\}_{t\in A}$ is a Gaussian Process with zero mean function and covariance function γ, we can represent

$$X_t = \sum_{n=1}^{\infty} \sqrt{\lambda_n} e_n(t) Z_i$$

where $\{\lambda_n, e_n\}_{n=1}^{\infty}$ are eigenvalues and eigenfunctions stated above and $\{Z_i\}_{i=1}^{\infty}$ are i.i.d Standard Normal variables.

Remark. The idea of Low-Rank Approximation is to use $\sum_{n=1}^{M} \sqrt{\lambda_n} e_n(t) Z_i$ to represent X_t, where M is large enough. The eigenvalues and eigenfunctions are estimated throught eigendecomposition of an $M \times M$ matrix after discretizing the integral. Such approximation enables faster computation of the inverse of covariance matrix.