Functional Analysis
My study note

Mengqi Lin

May 2021

1 Some useful prerequisites

1.1 topological space

(Wiki) In mathematics, a topological space is, roughly speaking, a geometrical space in which closeness is defined but, generally, cannot be measured by a numeric distance. More specifically, a topological space is a set of points, along with a set of neighbourhoods for each point, satisfying a set of axioms relating points and neighbourhoods.

A topological space is the most general type of a mathematical space that allows for the definition of limits, continuity, and connectedness. Other space, such as Euclidean spaces, metric spaces, and manifolds are topological spaces with extra structures, properties or constraints.

Definition 1.1 (topological space defined via neighborhood). *A topological space $(X, \mathcal{N})$ is a set X with a function $\mathcal{N}$ that assigns each point $x \in X$ a non-empty collection $\mathcal{N}(x)$ of subsets of X. The elements of $\mathcal{N}(x)$ will be called the neighborhoods of x w.r.t $\mathcal{N}$. The function $\mathcal{N}$ is called a neighborhood topology, if the following axioms are satisfied, and then $(X, \mathcal{N})$ is called a topological space.*

1. *If N is a neighborhood of x, i.e $N \in \mathcal{N}(x)$, then $x \in N$. In other words, each point belongs to every one of its neighborhood.*
2. *If N is a subset of X and includes a neighborhood of x, then N is a neighborhood of x. i.e, every superset of a neighborhood of x is again a neighborhood of x.*
3. *The intersection of two neighborhoods of x is a neighborhood of x.*
4. *Any neighborhood N of x includes a neighborhood M of x s.t N is a neighborhood of each point of M.*

The fourth axiom has a very important use in the structure of the theory, that of linking together the neighborhood of different points of X.

Given such structure, a subset $U \subset X$ is open if U is a neighborhood of every point of U.

Definition 1.2 (topological space defined via open set). *A topological space is an ordered pair (X, τ), where X is a set and τ is a collection of subsets of X, satisfying the following axioms:*

1. *The empty set and X belongs to τ*
2. *Any union(finite of infinite) of the members of τ belongs to τ.*
3. *The intersection of any finite members of τ belongs to τ.*

The elements of τ are called the open sets, and the collection τ is called a topology on X. A subset $C \subset X$ is closed iff its complement $X\mathbb{C}$ is an element of τ.

Definition 1.3 (continuous function). *A function $f : X \to Y$ between two topological spaces is continuous if for every neighborhood N of $f(x)$, there is a neighborhood M of x s.t $f(M) \subset N$.*

1.2 measure theory

Definition 1.4 (Borel set, Borel algebra). *In a topological space X, a **Borel set** is any set that can be formed from open sets through the operations of countable union, countable intersection, and relative complement. The collection of all the Borel sets on X forms a σ-algebra, known as **Borel algebra** or **Borel σ-algebra**. The Borel algebra on X is the smallest σ-algebra containing all open sets.*

Definition 1.5 (measurable space). *A set X and a σ-algebra $\mathcal{A}$ on X $(X, \mathcal{A})$ is called a measurable space.*

Usually, if X is finite countably infinite, the σ-algebra is most often the power set on X, so $\mathcal{A} = \mathcal{P}(X)$. This leads to the measurable space $(X, \mathcal{P}(X))$. If X is a topological space, the σ-algebra is most commonly the Borel σ-algebra. This leads to the $(X, \mathcal{B}(X))$.
Another similar definition is measure space, where measure space requires additional definition of measure.

Definition 1.6 (measure). *A measure is a extended real valued function μ defined on a measurable space $(X, \mathcal{A})$, s.t*

1. ***Non-negative**: for $\forall E \in \mathcal{A},\ \mu(E) \geq 0$*
2. ***Null empty set** : $\mu(\emptyset) = 0$*
3. ***countable additivity**: For all pairwise disjoint sets*

$$\{E_k\}_{k=1}^{\infty} \in \mathcal{A} : \mu(\sum_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} \mu(E_k)$$

Definition 1.7 (measure space). *A measure space is a triple $(X, \mathcal{A}, \mu)$, where X is a set, $\mathcal{A}$ is the σ-algebra on X, μ is the measure on $(X, \mathcal{A})$.*

Definition 1.8 (σ finite measure). *Let $(X, \mathcal{A}, \mu)$ be a measure space, the measure μ is σ finite measure if the set X can be covered with at most countably many measurable sets with finite measure. i.e*

$$X \subset \bigcup_{k=1}^{\infty} E_k \quad s.t \;\; \mu(E_k) < \infty$$

Definition 1.9 (measurable function).

The next definition(essential infimum and essenial supremum) is realted to the definition of infimum and supremum, but adapted to measure theory and functional analysis, where one often deals with statement that are not valid for **all** elements in a set, but rather **almost everywhere**, i.e except on a set of measure zero. Intuitively, the essential supremum of a function is the smallest value that is larger or equal than the function values everywhere when allowing for ignoring what the function does at a set of points of measure zero.

Definition 1.10 (upper bound and supremum). *Let $f : X \to \mathbb{R}$ be a real valued function defined on a set X, a real number a is called an **upper bound** of f if $f(x) \leq a,\ \forall x \in X$. Or equivalently, the set*

$$f^{-1}(a, \infty) = \{x \in X : f(x) > a\} = \emptyset$$

Let

$$U_f = \{a \in \mathbb{R} : f^{-1}(a, \infty) = \emptyset\}$$

*be the set of upper bounds of f, then the **supremum** of f is defined by*

$$\sup f = \begin{cases} \inf U_f & U_f \neq \emptyset \\ \infty & U_f = \emptyset \end{cases}$$

Now we further assume that f is a measuable function in measure space $(X, \mathcal{A}, \mu)$

Definition 1.11 (essential upper(lower) bound and essential supremum(infimum)). *A* ***essential upper bound*** *of A is a real number a if*

$$\mu\left(f^{-1}(a, \infty)\right) = 0$$

And let

$$U_f^{\text{ess}} = \{a \in \mathbb{R} : \mu\left(f^{-1}(a, \infty)\right) = 0\}$$

be the set of essential upper bounds of f, the the essential supremum of f is defined by

$$\operatorname{ess\,sup} f = \begin{cases} \inf U_f^{\text{ess}} & U_f \neq \emptyset \\ \infty & U_f^{\text{ess}} = \emptyset \end{cases}$$

Similarly we can define **essential bounded** function that are bounded up to a set of measure zero.

1.3 Matrix Calculation

Definition 1.12 (Frobenius norm).

Let $\lambda_1, \lambda_2, \ldots \lambda_n$ be the eigenvalues of matrix $T^\mathsf{T} T$.

$$\|T\|_F = \sqrt{\langle T, T\rangle} = \sqrt{\operatorname{tr}(T^\mathsf{T} T)} = \sqrt{\sum_{i=1}^{n} \lambda_i}$$

So then

$$\|T^\mathsf{T} T\|_F = \sqrt{\sum_{i=1}^{n} \lambda_i^2}$$

hence

$$\|T^\mathsf{T} T\|_F \leq \|T\|_F^2$$

1.4 some facts

Lemma 1.1. *A Cauchy sequence is convergent iff it has a convergent subsequence.*

2 Metric Spaces

2.1 Contraction mapping theorem

Definition 2.1 (metric space). *A metric space (M, d) is a set M together with a metric d on the set, where the metric is a non-negative function $d : M \times M \to \mathbb{R}$. s.t for $x, y, z \in M$.*

1. $d(x, y) = 0 \equiv x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, z) \leq d(x, y) + d(y, z)$

Given the above three axioms, we also have $d(x,y) \geq 0$ for any $x, y \in M$, this is deduced as follows

$$\begin{aligned} d(x,y) + d(y,x) &\geq d(x,x) \\ d(x,y) + d(x,y) &\geq d(x,x) \\ d(x,y) + d(x,y) &\geq 0 \\ 2d(x,y) &\geq 0 \\ d(x,y) &\geq 0 \end{aligned}$$

Example 1 (metric space).

- Let $C[a,b]$ be the space of all continuous functions on interval $[a,b]$, and the metric is defined by

  $$d(x,y) = \max_{t \in [a,b]} |x(t) - y(t)|$$

  Or the metric defined by

  $$d_1(x,y) = \int_a^b |x(t) - y(t)| dt$$

  (but this one isn't complete)

- The discrete metric: $d(x,y) = 0$ if $x = y$, $d(x,y) = 1$ otherwise. This metric is simple but important, and can be defined on any set.

The reason to introduce metric is to characterize convergence.

Definition 2.2 (convergence). *A sequence $\{x_n\}$ in metric space (M,d) is said to converge to x_0 if $\lim_{n\to\infty} d(x_n, x_0) \to 0$, short for $x_n \to x_0$*

Every metric space is a topological space in a natural manner, and therefore all definitions and theorems in topo space all apply to metric space.

Definition 2.3 (open ball). *For any point $x \in M$, we define an open ball with radius $r > 0$ about x as the set*

$$B(x,r) = \{y \in M : d(x,y) < r\} \tag{1}$$

Definition 2.4 (open set). *A subset $U \subset M$ is open if for all $x \in U$, there exists an open ball $B(x, r_x)$, s.t $B(x, r_x) \subset U$.*

Definition 2.5 (closed set). *The complement of open set is closed set. A subset $U \subset M$ is said to be closed iff every convergent sequence $\{x_n\} \in U$ has its limit x_0 in U.*

Definition 2.6 (Cauchy sequence). *A sequence $\{x_n\} \in M$ is a Cauchy sequence if $d(x_n, x_m) \to 0$ as both n and m go to infinity independently. Namely, for any $\epsilon > 0, \exists N$, s.t for $\forall n, m > N$, $d(x_n, x_m) < \epsilon$.*

Definition 2.7 (complete space). *A metric space is said to be complete, if every Cauchy sequence in the space is convergent.*

Example 2. *The space $(C[a,b], d)$ in Example (1) is complete. But when the metric is given by*

$$d_1(x,y) = \int_a^b |x(t) - y(t)| dt$$

then the space isn't complete. The counterexample can be given by sequence $\{x_n(t)\}$:

$$x_n(t) = \begin{cases} 1 - nt & 0 \leq t < \frac{1}{n} \\ 0 & \frac{1}{n} \leq t \leq 1 \end{cases}$$

Definition 2.8 (continuous mapping). *We say that a mapping $T : (M_2, d_1) \to (M_2, d_2)$ is continuous if for any sequence $\{x_n\}$ and x_0 in M_1 , we have*

$$d_1(x_n, x_0) \to 0 \Rightarrow d_2(Tx_n, Tx_0) \to 0 \quad (n \to \infty)$$

Equivalent definition is the following proposition

Proposition 2.1. *T is continuous iff for $\forall \epsilon > 0$, there exists some $\delta > 0$ s.t*

$$d_1(x, x_0) < \delta \Rightarrow d_2(Tx, Tx_0) < \epsilon$$

**Fixed point problem

The fixed point problem arises in many areas and has a lot of application in mathematics and CS theory.

Example 3 (the root of an equation). *Let ϕ be a real valued function on $\mathbb{R}$, the root of the equation*

$$\phi(x) = 0$$

can be viewed as a fixed point problem:

Let $f(x) = x - \phi(x)$, then the root of the equation is equivalent to the fix point of $f(x)$: $f(x) = x$.

Example 4 (Initial value problem in ODE). $\begin{cases} \dfrac{dx}{dt} = F(t, x) \\ x(0) = \xi \end{cases}$

Or equivalently we are interested in finding the continuous function that satisfies

$$x(t) = \xi + \int_0^t F(\tau, x(\tau))d\tau \tag{2}$$

which can be also viewed as a fix point problem: We consider the metric space $C[-h, h]$ and the mapping

$$Tx(t) = \xi + \int_0^t F(\tau, x(\tau))d\tau$$

Then equation (2) is equivalent to search a fix point in $C[-h, h]$ s.t $x = Tx$.

Definition 2.9 (contraction mapping). *Mapping $T : (M, d) \to (M, d)$ is a contraction mapping if there exists $0 < \alpha < 1$ s.t. $d(Tx, Ty) \leq \alpha d(x, y)$*

It's easy to see that a contraction mapping is continuous.

Example 5. *Let $M = [0, 1]$ and $T(x)$ is a differentiable function on M s.t*

$$T(x) \in [0, 1]$$

$$|T'(x)| \leq \alpha < 1$$

Then $T(x)$ is a contraction mapping.

Question: For such contraction mapping, do we always have fix points? If so, how many?

Theorem 2.2 (Banach fix point theorem or contraction mapping theorem). *Let (M, d) be a non-empty complete metric space, and $T : M \to M$ a contraction mapping. Then T admits a unique fixed point x^* in M s.t $T(x^*) = x^*$. Furthermore, x^* can be found as follows: start with an arbitrary point $x_0 \in M$ and define a sequence $\{x_n\}$ by $x_n = T(x_{n-1})$ then $x_n \to x^*$*

Proof. It's easy to see that the above sequence satisfies

$$d(x_{n+1}, x_n) = d(Tx_n, Tx_{n-1}) \leq \alpha d(x_n, x_{n-1}) \ldots \leq \alpha^n d(x_1, x_0)$$

Therefore

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \ldots + d(x_{n+1}, x_n) \\ &\leq (\alpha^{m-1} + \alpha^{m-2} + \ldots + \alpha^n) \cdot d(x_1, x_0) \\ &\leq \frac{\alpha^n}{1-\alpha} \cdot d(x_1, x_0) \to 0 \end{aligned}$$

So x_n is a Cauchy sequence and since the space is complete, we know that the sequence converges. And suppose x^* is its limit.

$$x^* = \lim_{n\to\infty} x_n = \lim_{n\to\infty} T(x_{n-1}) = T(\lim_{n\to\infty} x_{n-1}) = T(x^*)$$

Therefore, x^* is the fix point. Now we shall proceed to prove that it's unique. Suppose x^{**} also satisfies that $T(x^{**}) = x^{**}$ then

$$d(x^{**}, x^*) = d(T(x^{**}), T(x^*)) \leq \alpha d(x^{**}, x^*)$$

whence $x^{**} = x^*$. □

2.2 Completeness

Suppose (M_1, d_1) and (M_2, d_2) are two metric spaces, if there exists a mapping $\phi : M_1 \to M_2$ s.t

1. ϕ is onto.
2. $d_1(x, y) = d_2(\phi x, \phi y)$

Then we say that (M_1, d_1) are $(M_2, d_@)$ are **isometry** and the mapping ϕ is called the isometry mapping.

From 2 we can also infer that ϕ is also one to one.

Note: We shall not distinguish any two metric spaces that are isometry.

2.3 Sequentially compact

Definition 2.10 (Bounded spaces). *A metric space M is bounded if there exists $r > 0$ s.t for every point $x, y \in M$, $d(x, y) \leq r$. The smallest possible such r is called the diameter of M.*

Definition 2.11 (precompact or totally bounded space). *A metric space M is precompact/totally bounded if for every $\epsilon > 0$ there exists finitely many open balls with radius ϵ whose unions cover M. i.e*

$$M \subset \bigcup_{x\in A_\epsilon} B(x, \epsilon)$$

where A_ϵ is a finite subset.

An equivalent definition of totally bounded space is through epsilon net:

Definition 2.12 ((finite) epsilon net). *Suppose $A \subset M$, given $\epsilon > 0$, The ϵ-net A_ϵ for set A is defined as follows : for $\forall x \in A$, there $\exists y \in A_\epsilon$, s.t $d(x, y) < \epsilon$. And if A_ϵ is finite, then the net is called finite ϵ-net.*

Definition 2.13 (totally bounded set). *A subset A of a metric space is totally bounded if for any $\epsilon > 0$, there exists a finite ϵ-net A_ϵ of A.*

It's easy to see that a totally bounded space is bounded. But the converse is not true, for example, consider the metric space $M = \mathbb{R}^\infty$ and the metric to be the l_2 distance, and consider the ball $B = B(0,1) = \{x \in M : d(x,0) = \|x\|_2 \le 1\}$ which is clearly bounded. But is not totally bounded, we proceed the proof by showing that we can't cover a subset of $B(0,1)$ with finitely many balls with radius $\epsilon < \frac{\sqrt{2}}{2}$. And therefore can't cover the whole $B(0,1)$. Take the subset of $B(0,1)$ to be the points e_j. (the standard vector basis) and we have that $d(e_i,e_j) = \sqrt{2}$ for $i \neq j$, and for sure that we can't find finitely many of the ϵ balls to cover it.

Theorem 2.3. *A subset $E \subset M$ is totally bounded iff every sequence in E has a Cauchy subsequence.*

Proof. $\Rightarrow$ Suppose $\{x_n\} \subset E$, for radius $= 1$, there is a finite set of points $y_{11}, y_{12}, \dots, y_{1k_1}$, s.t

$$E \subset \bigcup_{j=1}^{k_1} B(y_{1j}, 1)$$

And at least one ball $B(y_{1j}, 1)$ contains infinitely subsequence (x_{n1}) of (x_n), for simplicity, let's assume $B(y_{11}, 1)$ will do. And for radius $= 1/2$, there is a finite set of points $y_{21}, y_{22}, \dots, y_{2k_2}$ s.t

$$B(y_{11},1) \subset \bigcup_{j=1}^{k_2} B(y_{2j}, 1/2)$$

And at least one ball $B(y_{2j}, 1)$ contains infinitely subsequence (x_{n2}) of (x_{n1}), for simplicity, let's assume $B(y_{22}, 1/2)$ will do. Proceed the procedure we have subsequence (x_{nm}) of $(x_{n(m-1)})$ which is contained in a ball of radius $1/2^{(m-1)}$.
Claim: The sequence (x_{nn}) is a Cauchy sequence. Indeed, for $\forall p > 0$, both $x_{(n+p)(n+p)}$ and x_{nn} are contained in the ball of radius $1/2^{(n-1)}$, therefore

$$d(x_{(n+p)(n+p)}, x_{nn}) < \frac{1}{2^{(n-1)}} \to 0$$

$\Leftarrow$ Conversely, assume that every sequence has a Cauchy sequence and that E is not totally bounded, then there exists $\epsilon > 0$, s.t for a point $x_1 \in E$, there $\exists x_2 \in E$ s.t $d(x_1,x_2) \ge \epsilon$, otherwise we have that for all points $y \in \mathbb{E}$ we should have $d(x_1, y) < \epsilon$ and so x_1 will be the proper ϵ-net. Similarly, there also exists x_3 s.t

$$d(x_1,x_3) \ge \epsilon, \quad d(x_2,x_3) \ge \epsilon$$

Otherwise for all other points $y \in E$ we should have either $d(x_1,x_3) < \epsilon$ or $d(x_2,x_3) < \epsilon$ and then x_1, x_2 will be the proper ϵ-net. Proceed the procedure, we construct a sequence with

$$d(x_n, x_m) > \epsilon$$

which is in no way Cauchy. □

Definition 2.14 (separable space). *A metric space is separable if it has a countable dense subset. (where the dense subset is defined as follows)*

Definition 2.15 (dense subset). *A subset $E \subset M$ is dense if for any point $x \in M$ and any $\epsilon > 0$, there exists a point $y \in E$, s.t $d(x,y) < \epsilon$. Or equivalently, $\forall x \in M, \exists \{x_n\} \subset E$, s.t $x_n \to x$. Or equivalently, $\bar{E} = M$*

Theorem 2.4. *Totally bounded space is separable.*

Proof. Let N_n to be the $\frac{1}{n}$-net, then $\bigcup_{n=1}^{\infty} N_n$ is the countable dense subset. □

Definition 2.16 (sequentially compact space). *A metric space M is sequentially compact if every sequence in the space has a convergent subsequence converging to a point in M.*

Proposition 2.5. *Sequentially compact space is closed.*

Proposition 2.6. *In $\mathbb{R}^n$, any bounded sequence is sequentially compact.*

Proposition 2.7. *Sequentially compact space must be complete. [the proof uses lemma(1.1)].*

Proposition 2.8. *Sequentially compact space must be totally bounded. The proof follows from theorem(2.3).*

Proposition (2.7) and proposition (2.8) tell us two necessary conditions for a space to be sequentially compact, and actually they are also sufficient conditions. It can be proved by using the theorem(2.3).

Theorem 2.9. *A metric space is compact iff it's complete and totally bounded.*

In more general topological space, we have compact space, defined as:

Definition 2.17 (compact space). *A space M is compact if for any open cover of M, we can find finitely many of the open sets to cover it.*

Theorem 2.10. *For metric space, sequentially compact is equivalent to compact.*

Proof. $\Rightarrow$ Suppose metric space M is sequentially compact, and if that M isn't compact. That means, there exists an open cover $\{G_\alpha\}$ of M, which we can't find finite of them to cover M. Since M is sequentially compact, then for

$$\forall n \in \mathbb{N},\ M \subset \bigcup_{j=1}^{k_n} B(y_{nj}, 1/n)$$

And for each n, at least one ball $B(y_{nj}, 1/n)$ needs to be covered by infinitely many of the $\{G_\alpha\}$. WLOG let's assume that $B(y_{n1}, 1/n)$ will do. For sequence $\{y_{n1}\}$, since M is sequentially compact, there exists a convergent subsequence of $\{y_{n1}\}$. Let's suppose $y_{m1} \to y_0,\ m \in [n]$. And $y_0 \in G_{\alpha_0}$, and since G_{α_0} is open, for sufficiently large m, $B(y_{m1}, 1/m) \subset G_{\alpha_0}$ since $y_{m1} \to y_0$. And this suggests that for sufficiently large m, $B(y_{m1}, 1/m)$ can be covered by finite of $\{G_\alpha\}$. Hence, contradiction!

$\Leftarrow$ Suppose that M is compact, and suppose that it's not sequentially compact. Then there exists a sequence (x_n) that has no convergent subsequence. So there are infinitely many of (x_n) are distinct. WLOG assume (x_n) are all distinct. For $\forall x \in M$, exists $\epsilon_x > 0$ s.t $B(x, \epsilon_x)$ contains no points from (x_n), except possibly x itself. (Otherwise, for $\forall \epsilon > 0, B(x, \epsilon)$ contains point from (x_n) then (x_n) will have convergent subsequence converging to x.) And these balls $B(x, \epsilon_x)$ consists an open cover of M, but since there are infinitely many of distinct (x_n), therefore, it's impossible to select finite of $B(x, \epsilon_x)$ to cover M, this contradicts to the fact that M is compact. So we are done. □

Example 6 (compact space).

- Closed interval $[a, b]$.
- Cantor set.(by Heine–Borel theorem)
- Space with finite points.

Lots of results on bounded closed intervals $[a, b]$ can be generalized to results on compact space. For example, the extreme value theorem for closed interval can be generalized.

Proposition 2.11 (Continuous function preserves compactness). *Let M be a compact space and $u : M \to W$ be a continuous function on M, then the image $u(M)$ is compact.*

Proof. Take a sequence $\{y_n\} \subset u(M)$, then exists $\{x_n\} \subset M$ s.t $u(x_n) = y_n$, since M is compact, there exists a convergent subsequence $\{x_{n_k}\}$ of $\{x_n\} : x_{n_k} \to x_0$, by the continuity of u, we have $u(x_{n_k}) = y_{n_k} \to u(x_0) = y_0$ And so we've found a convergent subsequence $\{y_{n_k}\}$. Therefore $u(M)$ is compact. □

Corollary 2.11.1 (Extreme value theorem for continuous function on compact space). *As a special case of proposition (2.11), take $u : M \to \mathbb{R}$, then the compactness of $u(M)$ tells us that the maximum and minimum of $u(M)$ are attainable.*

Let $C(M)$ be the collection of all continuous functions defined on compact space M. Since the maximum and minimum are attainable, it's reasonable to define the following metric on $C(M)$

$$d(u, v) = \max_{x \in M} |u(x) - v(x)|$$

Proposition 2.12. *$(C(M), d)$ is a metric space. And also complete space.*

What is the condition for $C(M)$ to be compact?

Definition 2.18 (uniformly bounded). *A subset F of $C(M)$ is uniformly bounded if*

$$|u(x)| \leq M_1, \ \forall x \in M, \ \forall u \in F$$

Definition 2.19 (uniformly equicontinuous). *A subset F of $C(M)$ is uniformly equicontinuous if for $\forall \epsilon > 0$, $\exists \delta_\epsilon > 0$, s.t $\forall x, y \in M$ and $\forall u \in F$, as long as $|x - y| < \delta_\epsilon$*

$$|u(x) - u(y)| < \epsilon$$

Fact 2.13. *The continuous functions defined on compact set are uniformly continuous.*

Theorem 2.14 (Arzela-Ascoli: compact C(M)). *A subset $F \subset C(M)$ is compact iff F is uniformly bounded and uniformly equicontinous.*

Proof. Since $C(M)$ is complete, so by theorem (2.9) F is compact iff F is totally bounded.

$\Rightarrow$ F is totally bounded therefore bounded, so F is uniformly bounded. Now prove that F is also uniformly equicontinuous. i.e For $\forall \epsilon > 0$, we wish to find $\delta = \delta(\epsilon) > 0$, s.t for $\forall x, \ y \in M$ and $d(x, y) < \delta$, we have $|\phi(x) - \phi(y)| < \epsilon$ for all $\phi \in C(M)$. Since F is totally bounded, there exists finite $\frac{\epsilon}{3}$-net of F:

$$N(\frac{\epsilon}{3}) = \{\phi_1, \phi_2, \ldots \phi_m\}$$

Since it's finite, there exists some common δ, s.t as long as $d(x, y) < \delta$,

$$|\phi_i(x) - \phi_i(y)| < \frac{\epsilon}{3}, \ \ i \in [m]$$

Now for any $\phi \in F$, $\exists \phi_i \in N(\frac{\epsilon}{3})$, s.t

$$|\phi(x) - \phi_i(x)| < \frac{\epsilon}{3}, \quad \forall x \in M$$

Therefore, for $\forall x, y \in M$, $s.t \ \ d(x, y) < \delta$,

$$\begin{aligned} |\phi(x) - \phi(y)| &< |\phi(x) - \phi_i(x)| + |\phi_i(x) - \phi_i(y)| + |\phi_i(y) - \phi(y)| \\ &< \epsilon \end{aligned}$$

$\Leftarrow$ Suppose F is uniformly bounded and uniformly equicontinuous. For $\forall \epsilon > 0$, we wish to find the ϵ-net. Since F is equicontinuous, $\exists \delta > 0$, s.t as long as $d(x, y) < \delta$,

$$|\phi(x) - \phi(y)| < \frac{\epsilon}{3}$$

And for such δ, let the finite δ-net on M be

$$N(\delta) = \{x_1, x_2, \dots, x_n\}$$

And we define the mapping $T : F \to \mathbb{R}^n$

$$T\phi \equiv (\phi(x_1), \phi(x_2), \dots, \phi(x_n))$$

Then we claim that $\tilde{F} \equiv T(F)$ is a bounded set in $\mathbb{R}^n$, this is because

$$\|T\phi\|_2 \le \sqrt{n} \max_{x \in M} |\phi(x)| \le \sqrt{n} M_1$$

Therefore $\tilde{F}$ is sequentially compact. And let the $\epsilon/3$-net of $\tilde{F}$ be

$$\tilde{N}(\frac{\epsilon}{3}) = \{T\phi_1, T\phi_2, \dots, T\phi_m\}$$

Now we claim that $\{\phi_1, \phi_2, \dots, \phi_m\}$ is the ϵ-net of F we desire to find. Since $\tilde{N}$ is the $\epsilon/3$-net of $\tilde{F}$, for $\forall \phi \in F$, $\exists i \in [m]$, s.t

$$d(T\phi_i, T\phi) < \frac{\epsilon}{3}$$

And $\forall x \in M,\ \exists x_j \in N(\delta),\ s.t\ \ d(x, x_j) < \delta$. Therefore

$$|\phi(x) - \phi_i(x)| \le |\phi(x) - \phi(x_j)| + |\phi(x_j) - \phi_i(x_j)| + |\phi_i(x_j) - \phi_i(x)| \le \epsilon$$

□

Example 7. *Let $\Omega \subset \mathbb{R}^n$ be a bounded open convex set and $\bar{\Omega}$ be its closure. And $M_1,\ M_2 > 0$. The set*

$$F \equiv \{\phi \in C^1(\bar{\Omega}) : |\phi| \le M_1,\ |\nabla\phi| \le M_2\}$$

Then F is sequentially compact.

Proof.

$$|\phi(x) - \phi(y)| = \nabla\phi(\theta x + (1-\theta)y) \cdot |x - y| \le M_2|x - y|$$

Therefore F is uniformly equicontinuous. And that F is uniformly bounded is obvious, so F is compact. □

2.4 Normed Linear Spaces

It's not enough for metric space to only have topological structure, since in the analysis of functional spaces, we are not only interested in the convergence but also the algebraic operations.

2.4.1 Linear space

Definition 2.20 (Linear Space). *A linear space is a quadruple, $(X, \mathbb{F}, +, \cdot)$ where X is a nonempty set over field $\mathbb{F}$, with two operations $(+, \cdot)$ and the two operations satisfy the 8 properties.*

The followings are the important basic concepts about linear spaces.

- **Linear isomorphism**: Suppose V and V_1 are two linear spaces, $T : V \to V_1$ is said to be a linear isomorphism if: 1.It's bijective. 2.It's linear. i.e $T(ax + by) = aT(x) + bT(y)$.
- **Linear subspace**: A subset $E \subset V$ is also a linear space itself, then E is a linear subspace of V.
- **Linear independent**
- **Linear basis**
- **Dimension**
- **Linear hull**
- **Sums and product sums**

2.4.2 Metric on linear space

We now combine the topological structure – metric d and the algebraic structure – linear space to a space. If the metric is translation-invariant (or equivalently continuous w.r.t operation " $+$ ") and continuous w.r.t operation " $\cdot$ ", then we may define a prenorm via $\|x\| = d(x, 0)$.

Definition 2.21 (prenorm). *A prenorm on a linear space V is a function $\| \bullet \|$ defined on V s.t*

1. $\|x\| \geq 0,\ \|x\| = 0$ *iff* $x = 0$
2. $\|x + y\| \leq \|x\| + \|y\|$
3. $\| - x\| = \|x\|$
4. $\lim_{a_n \to 0} \|a_n x\| = 0,\ \lim_{\|x_n\| \to 0} \|a x_n\| = 0$

Definition 2.22 (F^* space). *A prenormed linear space V defines convergence via*

$$\|x_n - x\| \to 0$$

Then we say that the space is F^* space.

Definition 2.23 (Frechet Space). *The complete F^* space is called Frechet space, shor for F space.*

Example 8. *The space $C(M)$ with prenorm defined by*

$$\|\phi\| = \max_{x \in M} \phi(x)$$

Example 9 (Sequence space S). *Let S be the space made up of all sequence $x = (x_1, x_2, \ldots x_n, \ldots)$ with the natural operation " $+$ ", " $\cdot$ ", and define*

$$\|x\| = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|x_n|}{1 + |x_n|}$$

Then this is a prenorm. And actually the space with this prenorm is also complete, therefore this is a F space.

Proof. It's easy to see that the (1), (3) condition for the prenorm are satisfied.
For condition (2): notice the inequality

$$\frac{\alpha + \beta}{1 + \alpha + \beta} \leq \frac{\alpha}{1 + \alpha} + \frac{\beta}{1 + \beta}, \quad (\alpha,\ \beta > 0)$$

Therefore

$$\|x + y\| = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|x_n + y_n|}{1 + |x_n + y_n|} \leq \sum_{n=1}^{\infty} \frac{1}{2^n} \left(\frac{|x_n|}{1 + |x_n|} + \frac{|y_n|}{1 + |y_n|}\right) = \|x\| + \|y\|$$

For condition (4): notice the inequality

$$\frac{\alpha\beta}{1 + \alpha\beta} \leq \begin{cases} \alpha \frac{\beta}{1+\beta}, & (\alpha \geq 1,\ \beta \geq 0) \\ \frac{\beta}{1+\beta}, & (0 < \alpha < 1, \beta \geq 0) \end{cases}$$

$\to$

$$\frac{\alpha\beta}{1 + \alpha\beta} \leq \max(\alpha, 1) \frac{\beta}{1 + \beta}$$

Therefore

$$\|a x_n\| \leq \max(|a|, 1) \|x_n\| \to 0 \ \ (\|x_n\| \to 0)$$

And when $|a_m| \to 0$, firstly there exists some n_0, s.t $\sum_{n=n_0}^{\infty} \frac{1}{2^n} < \frac{\epsilon}{2}$, and fix this n_0, there exists some N, s.t for $m > N$ and $n \in [n_0]$

$$\frac{|a_m x_n|}{1+|a_m x_n|} \leq |a_m x_n| \leq |a_m| \max_{1 \leq i \leq n_0} |x_i| \leq \frac{\epsilon}{2}$$

Therefore

$$\|a_m x\| = \sum_{n=1}^{n=n_0} \frac{1}{2^n} \frac{|a_m x_n|}{1+|a_m x_n|} + \sum_{n=n_0}^{\infty} \frac{1}{2^n} \frac{|a_m x_n|}{1+|a_m x_n|} \leq \frac{\epsilon}{2} \sum_{n=1}^{n=n_0} \frac{1}{2^n} + \sum_{n=n_0}^{\infty} \frac{1}{2^n} \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

For the completeness of the space: If

$$\|x^{(m+p)} - x^{(m)}\| = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|x_n^{(m+p)} - x_n^{(m)}|}{1+|x_n^{(m+p)} - x_n^{(m)}|} \to 0 \quad (m \to \infty, \forall p)$$

Then for $\forall n$ we must have

$$\lim_{m \to \infty} |x_n^{(m+p)} - x_n^{(m)}| \to 0$$

[Note: this is also saying that for a point in space S to converge to 0 iff every coordinate of the point converges to 0.]

So for every n there exists x_n^* s.t

$$x_n^{(m)} \to x_n^*, \quad m \to \infty$$

Let $x^* = (x_1^*, x_2^*, \dots, x_n^*, \dots)$ then

$$\|x^{(m)} - x^*\| = \sum_{n=1}^{n=n_0} \frac{1}{2^n} \frac{|x_n^{(m)} - x_n^*|}{1+|x_n^{(m)} - x_n^*|} + \sum_{n=n_0}^{\infty} \frac{1}{2^n} \frac{|x_n^{(m)} - x_n^*|}{1+|x_n^{(m)} - x_n^*|}$$

Similar to the previous proof, we first find n_0 s.t $\sum_{n=n_0}^{\infty} \frac{1}{2^n} < \frac{\epsilon}{2}$ and then find m large enough s.t for all $n \in [n_0]$

$$\frac{|x_n^{(m)} - x_n^*|}{1+|x_n^{(m)} - x_n^*|} \leq |x_n^{(m)} - x_n^*| \leq \frac{\epsilon}{2}$$

Therefore

$$\|x^{(m)} - x^*\| \leq \frac{\epsilon}{2} \sum_{n=1}^{n=n_0} \frac{1}{2^n} + \sum_{n=n_0}^{\infty} \frac{1}{2^n} \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Hence S is complete. □

Example 10. *Let $C(\mathbb{R}^n)$ be the space of continuous functions on $\mathbb{R}^n$ and let*

$$\|\phi\| = \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{\max_{|x| \leq k} |\phi(x)|}{1+\max_{|x| \leq k} |\phi(x)|}$$

Then $\| \bullet \|$ is a prenorm, together with $C(\mathbb{R}^n)$ is a F space.

2.4.3 Norm and Banach Space

For prenorm that also satisfies

$$\|ax\| = |a|\|x\|$$

We call it norm.

Definition 2.24 (norm). *A norm on a linear space V is a function $\| \bullet \|$ defined on V s.t*

1. $\|x\| \geq 0,\ \|x\| = 0$ *iff* $x = 0$
2. $\|x + y\| \leq \|x\| + \|y\|$
3. $\|ax\| = |a|\|x\|$

And spaces V with norm are called normed spaces, or B^* space. And if the space is also complete, we call it Banach space, short for B space.

Example 11 ($L^p(\Omega, \mu)(p \geq 1)$ space). *Suppose $(\Omega, \mathcal{B}, \mu)$ is a measure space, μ is the measure, and $L^p(\Omega, \mu)$ is the set of all measurable functions on Ω, s.t $|u(x)|^p$ is measurable. And define*

$$\|u\| = \left(\int_\Omega |u(x)|^p d\mu \right)^{\frac{1}{p}}$$

Then $\| \bullet \|$ is a norm this is because that condition (1) and (3) are obvious and that the (2) condition is exactly the famous Minkowski Inequality

$$\left(\int_\Omega |u(x) + v(x)|^p d\mu \right)^{\frac{1}{p}} \leq \left(\int_\Omega |u(x)|^p d\mu \right)^{\frac{1}{p}} + \left(\int_\Omega |v(x)|^p d\mu \right)^{\frac{1}{p}}$$

And by Riesz-Fisher theorem, $L^p(\Omega, \mu)$ is a B space.
Note that there are two special cases to this example.

1. When Ω is a measurable set in $\mathbb{R}^n$ and $d\mu$ is the lebesgue measure, then the space is denoted as $L^p(\Omega)$.
2. When $\Omega = \mathbb{N}$ and $\mu(\{n\}) = 1$, then $L^p(\Omega, d\mu)$ consists of sequences s.t $\sum_{n=1}^{\infty} |u_n|^p < \infty$. And denote such space by ℓ^p whose norm is given by

$$\|u\| = \left(\sum_{n=1}^{\infty} |u_n|^p \right)^{\frac{1}{p}}$$

If we further wish to generalize the above L^p space to L^∞ space, we will need to modify the definition a bit.

Example 12 ($L^\infty(\Omega, \mu)$ space). *Suppose $(\Omega, \mathcal{B}, \mu)$ is a measure space (1.7), and μ is σ-finite measure (1.8), the L^∞ space includes all essential bounded measurable functions (treat a.e equal functions as equal) for $f \in L^\infty$, we define the norm by its absolute essential supremum (1.11)*

$$\|f\| = \operatorname{ess\,sup} |f(x)|$$

Theorem 2.15 (The norm of L^∞). *is well defined.*

Proof. Check the 3 axioms. □

Special case: When $\Omega = \mathbb{N}$ the space corresponds to ℓ^∞, which consists of all bounded sequence with the norm defined by

$$\|u\| = \sup_{n \geq 1} |u_n|$$

2.4.4 norm equivalence on normed linear spaces

Definition 2.25 (norm equivalence). *Given two norms $\| \bullet \|_a$ and $\| \bullet \|_b$ we say that $\| \bullet \|_a$ is stronger than $\| \bullet \|_b$ if*

$$\| \bullet \|_a \to 0 \Rightarrow \| \bullet \|_b \to 0$$

And if $\| \bullet \|_a$ is stronger than $\| \bullet \|_b$ and $\| \bullet \|_b$ is stronger than $\| \bullet \|_a$, then we say that the two norms are equivalent.

Proposition 2.16. *$\| \bullet \|_a$ is stronger than $\| \bullet \|_b$ iff there exists some $C > 0$ s.t*

$$\| \bullet \|_b \leq C \| \bullet \|_a \tag{3}$$

Proof. $\Rightarrow$ Suppose (3) doesn't hold, i.e for $\forall n \in \mathbb{N}$, $\exists x_n$ s.t

$$\|x_n\|_b > n \|x_n\|_a$$

Then let $y_n = \dfrac{x_n}{\|x_n\|_b}$, then $\|y_n\|_b = 1$ and $\|y_n\|_a = \dfrac{\|x_n\|_a}{\|x_n\|_b} < \dfrac{1}{n} \to 0$. But $\|y_n\|_b = 1$ doesn't go to 0.

$\Leftarrow$ is obvious. ☐

Corollary 2.16.1. *For two norms $\| \bullet \|_a$ and $\| \bullet \|_b$ to be equivalent, iff there exists constants $C_1, C_2 > 0$ s.t*

$$C_1 \| \bullet \|_a \leq \| \bullet \|_b \leq C_2 \| \bullet \|_a$$

Theorem 2.17 (equivalent norms on finite dim linear spaces). *Suppose V is a n dimension linear space, then all norms on the space are equivalent.*

Proof. Since V is n dimension, for $\forall \boldsymbol{x} \in V$, $\boldsymbol{x}$ can be written as

$$\boldsymbol{x} = \sum_{i=1}^{n} a_i e_i$$

where $\{e_i\}$ is the basis for V. From Corollary(2.16.1), It suffice to show that for any given norm $\| \bullet \|_a$, there exists some $C_1, C_2 > 0$ s.t

$$C_1 \|\boldsymbol{x}\|_1 \leq \|\boldsymbol{x}\|_a \leq C_2 \|\boldsymbol{x}\|_1 \quad (\forall \boldsymbol{x} \in V)$$

Where $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |a_i|$, and since when $\boldsymbol{x} = 0$, then result is trivial, we may assume that $\|\boldsymbol{x}\|_1 > 0$, therefore equivalently, we need to show that

$$C_1 \leq \frac{\|\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_1} \leq C_2$$

$\Leftrightarrow$

$$C_1 \leq \|\boldsymbol{u}\|_a \leq C_2$$

where $\boldsymbol{u} = \dfrac{\boldsymbol{x}}{\|\boldsymbol{x}\|_1}$ and that $\|\boldsymbol{u}\|_1 = 1$. Therefore the vectors $\boldsymbol{u}$ consists of a compact set (a sphere w.r.t $\| \bullet \|_1$). And now we take the function $\| \bullet \|_a$ defined on this compact set. If we can prove that this function is continuous on this compact set. We can apply the extreme value theorem to obtain C_1 and C_2.

Now prove the continuity of $\| \bullet \|_a$:

Suppose $\boldsymbol{x} = \sum_{i=1}^{n} a_i e_i$, $\boldsymbol{x}' = \sum_{i=1}^{n} a_i' e_i$

$$|\|\boldsymbol{x}\|_a - \|\boldsymbol{x}'\|_a| \leq \|\boldsymbol{x} - \boldsymbol{x}'\|_a \leq \sum_{i=1}^{n} |a_i - a_i'| \|e_i\|_a \leq \max_{j \in [n]} \|e_j\|_a \sum_{i=1}^{n} |a_i - a_i'| = \max_{j \in [n]} \|e_j\|_a \|\boldsymbol{x} - \boldsymbol{x}'\|_1$$

Therefore, whenever $\|\boldsymbol{x}-\boldsymbol{x}'\|_1 \leq \dfrac{\epsilon}{\max_{j\in[n]}\|e_j\|_a}$, $|\|\boldsymbol{x}\|_a - \|\boldsymbol{x}'\|_a| \leq \epsilon$. So $\|\bullet\|_a$ is a continuous function on the compact set with norm $\|\bullet\|_1$. And since the set is compact, we can apply (2.11.1) to obtain C_1, C_2, i.e

$$C_1 \leq \|\boldsymbol{u}\|_a = \frac{\|\boldsymbol{x}\|_a}{\|\boldsymbol{x}\|_1} \leq C_2$$

And since $\|\boldsymbol{u}\|_1 = 1$, $\boldsymbol{u} \neq 0$, therefore $C_1 > 0$. So we are done!

Proof reference: https://math.mit.edu/ stevenj/18.335/norm-equivalence.pdf □

We've already known that for linear spaces with the same dimension are indistinguishable in algebraic structure. This theorem shows us that for linear spaces with the same dimension are also indistinguishable in topological structure.

Theorem 2.18. *Every finite dimensional normed linear space is complete.*

Proof. Suppose $(V, \|\bullet\|_a)$ is a normed linear space, and take a Cauchy sequence $\boldsymbol{x}_n$:

$$\|\boldsymbol{x}_n - \boldsymbol{x}_m\|_a \to 0, \quad (n, m \to \infty)$$

Then define $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |a_i|$ the sum of the absolute value of each coordinate, we have:

$$C_1\|\boldsymbol{x}_n - \boldsymbol{x}_m\|_1 \leq \|\boldsymbol{x}_n - \boldsymbol{x}_m\|_a \to 0, \quad (n, m \to \infty)$$

And since $(V, \|\bullet\|_1)$ is complete.

$$\|\boldsymbol{x}_n - \boldsymbol{x}_0\|_1 \to 0 \Leftrightarrow \boldsymbol{x}_n \to \boldsymbol{x}_0$$

Therefore

$$\|\boldsymbol{x}_n - \boldsymbol{x}_0\|_a \leq C_2\|\boldsymbol{x}_n - \boldsymbol{x}_0\|_1 \to 0$$

□

Corollary 2.18.1. *Every finite dimensional normed linear space is closed.*

Definition 2.26 (sublinear functional). *Let $P : V \to \mathbb{R}$ be a function defined on the linear space V s.t*

1. $P(\boldsymbol{x}+\boldsymbol{y}) \leq P(\boldsymbol{x}) + P(\boldsymbol{y})$
2. $P(\lambda\boldsymbol{x}) = \lambda P(\boldsymbol{x}), \quad (\forall \lambda > 0)$

Then P is called a sublinear functional on V.

Definition 2.27 (seminorm). *If further, P also satisfies that $P(\boldsymbol{x}) \geq 0$ and that $P(\lambda\boldsymbol{x}) = |\lambda|P(\boldsymbol{x})$ for all $\lambda \in \mathbb{R}$, P is called a seminorm on V*

Similar to the proof of theorem (2.17), we have

Theorem 2.19. *Let P be a sublinear functional on finite B^* space, $P(\boldsymbol{x}) \geq 0$, and $P(\boldsymbol{x}) = 0 \Leftrightarrow \boldsymbol{x} = 0$, then there exists $C_1, C_2 > 0$ s.t*

$$C_1\|\boldsymbol{x}\| \leq P(\boldsymbol{x}) \leq C_2\|\boldsymbol{x}\| \quad (\forall \boldsymbol{x} \in V)$$

2.4.5 Application: Best Approximation

Approximation theory is concerned with how function can best be approximated with simpler functions. In functional analysis language: Given a normed linear space V, and finite many vectors in $V : \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ and a given vector $\boldsymbol{x} \in V$, we are interested in

$$\underset{(\lambda_1,\lambda_2,\ldots,\lambda_n)}{\arg\min} f(\boldsymbol{\lambda}), \qquad \text{where } f(\boldsymbol{\lambda}) = \|\boldsymbol{x} - \sum_{i=1}^{n} \lambda_i \boldsymbol{e}_i\|$$

Our first question is, whether $\underset{(\lambda_1,\lambda_2,\ldots,\lambda_n)}{\arg\min} f(\boldsymbol{\lambda})$ exists or not?
WLOG, Let's suppose $\boldsymbol{e}_i$ are linearly independent. And notice that f is a continuous function on $\mathbb{F}^n$, and that

$$f(\boldsymbol{\lambda}) \geq \|\sum_{i=1}^{n} \lambda_i \boldsymbol{e}_i\| - \|\boldsymbol{x}\|$$

So as $|\boldsymbol{\lambda}| \to \infty,\ f(\boldsymbol{\lambda}) \to \infty$, hence the minimum of $f(\boldsymbol{\lambda})$ exists.

Theorem 2.20 (existence of best approximation). *Suppose V is a normed linear space, if $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ are given vectors in V, then for $\forall \boldsymbol{x} \in V$, there always exists a linear combination of $\boldsymbol{e}_i$ best approximated to $\boldsymbol{x}$. Equivalently, we can denote $M \triangleq \operatorname{span}\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$*

$$\rho(\boldsymbol{x}, M) \triangleq \inf_{\boldsymbol{y}\in M} \|\boldsymbol{x} - \boldsymbol{y}\| = \|\boldsymbol{x} - \boldsymbol{x}^*\|$$

Then ρ is well defined, and $\boldsymbol{x}^*$ is called the best approximation point.

The theorem tells us that in any normed linear space V, given any $\boldsymbol{x} \in V$ and any finite subspace M we can always find the best approximation of $\boldsymbol{x}$ to M

Our next question is: Is the best approximation point always unique? Apparently, to answer this question, we need to assume that $\{\boldsymbol{e}_i\}_{i=1}^n$ are linearly independent. Despite than this, the uniqueness of best approximation still relies on the property of the norm.

Definition 2.28 (strictly convex normed linear space). *Normed linear space $(V, \|\bullet\|)$ is strictly convex if $\forall \boldsymbol{x}, \boldsymbol{y} \in V,\ \boldsymbol{x} \neq \boldsymbol{y}$*

$$\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1 \Rightarrow\ \|\alpha\boldsymbol{x} + \beta\boldsymbol{y}\| < 1 \qquad (\forall \alpha, \beta > 0,\ \alpha + \beta = 1)$$

Theorem 2.21 (uniqueness of best approximation). *If the normed linear space $(V, \|\bullet\|)$ is strictly convex, and that the vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ are linearly independent, then the best approximation point in (2.20) is unique.*

Proof. Let $d = \rho(\boldsymbol{x}, M)$, where $M = \operatorname{span}\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$. Suppose

$$\|\boldsymbol{x} - \boldsymbol{y}\| = \|\boldsymbol{x} - \boldsymbol{z}\| = d$$

If $d = 0$, $\boldsymbol{x} = \boldsymbol{y} = \boldsymbol{z}$ then the best approximation point must be unique, so let's asssume $d > 0$, and let $\alpha + \beta = 1,\ \alpha > 0,\ \beta > 0$

$$\frac{1}{d}\|\boldsymbol{x} - (\alpha\boldsymbol{y} + \beta\boldsymbol{z})\| = \|\alpha\frac{(\boldsymbol{x} - \boldsymbol{y})}{d} + \beta\frac{(\boldsymbol{x} - \boldsymbol{z})}{d}\| < 1$$

This is saying $\|\boldsymbol{x} - (\alpha\boldsymbol{y} + \beta\boldsymbol{z})\| < d$, but d is the minimum, hence contradiction! □

This is why we are saying that the uniqueness of best approximation depends on the

Example 13 **($L^p(\Omega, \mu)$ $(1 < p < \infty)$ is strictly convex).** *Recall that the Minkowski inequality (p¿1)*

$$\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$$

is equal iff $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent. i.e $\boldsymbol{u} = k\boldsymbol{v},\ (k > 0)$. But when $\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1$ and $\boldsymbol{u} \neq \boldsymbol{v}$ this is impossible, so

$$\|\alpha\boldsymbol{u} + \beta\boldsymbol{v}\| < \alpha\|\boldsymbol{u}\| + \beta\|\boldsymbol{v}\| < 1$$

Hence (L^p) is strictly convex.

Example 14 **($C(M), L^1(\Omega, \mu)$ aren't strictly convex).**

For example $C[0,1]$ take $x(t) = 1,\ y(t) = t$. then $\|x\| = \|y\| = 1$ but $\|\dfrac{x+y}{2}\| = 1$.
For $L^1[0,1]$, take $x(t) = 1,\ y(t) = 2t,\ \ \|x\| = \|y\| = 1$, but $\|\dfrac{x+y}{2}\| = 1$.

2.4.6 charaterization of finite normed linear space

We've already known that in finite normed space, a unit sphere is compact. We'll show that conversely, if the unit sphere in a normed space is compact, this space must be finite.

Theorem 2.22 **(normed space is finite iff its unit sphere is compact).**

Proof. $\Rightarrow$ Let the unit sphere be

$$S_1 \triangleq \{\boldsymbol{x} \in V \mid \|\boldsymbol{x}\| = 1\}$$

If the normed linear space V is infinite, take n linearly independent vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$. Since V is infinite, $M_n \triangleq \text{span}\{e_i\} \neq V$, pick a vector $\boldsymbol{y} \notin M_n$, then

$$d = \rho(\boldsymbol{y}, M_n) = \rho(\boldsymbol{y}, \boldsymbol{x}) > 0$$

for some $\boldsymbol{x} \in M_n$. Let $\boldsymbol{e}_{n+1} = (\boldsymbol{y} - \boldsymbol{x})/d$. Then $\|\boldsymbol{e}_{n+1}\| = 1$ so $\boldsymbol{e}_{n+1} \in S_1$ and

$$\|\boldsymbol{e}_{n+1} - \boldsymbol{e}_i\| \geq 1 \quad i \in [n]$$

Since V is inifinite, we can proceed the procedure to construct sequence $\{\boldsymbol{e}_m\}$ s.t

$$\|\boldsymbol{e}_n - \boldsymbol{e}_m\| \geq 1 \quad (\forall n, m \in \mathbb{N},\ n \neq m)$$

Such sequence is not compact. □

Definition 2.29 **(bounded set in normed linear space).** *A subset E in normed linear space V is bounded if there exists some constant $C > 0$ s.t*

$$\|\boldsymbol{x}\| \leq C \quad (\forall \boldsymbol{x} \in E)$$

Corollary 2.22.1. *A normed linear space is finite iff all its bounded sets are compact.*

The following lemma is very useful.

Theorem 2.23 **(Fisher Riesz lemma).** *If E is a* ***closed proper*** *subset of the normed linear space V, then for $\forall\ 0 < \epsilon < 1$, $\exists\ \boldsymbol{y} \in V$, s.t $\|\boldsymbol{y}\| = 1$ and*

$$\|\boldsymbol{y} - \boldsymbol{x}\| \geq 1 - \epsilon \quad (\forall\ \boldsymbol{x} \in E)$$

Proof. Take some $\boldsymbol{y}_0 \in V/E$ and let

$$d = \rho(\boldsymbol{y}_0, E) = \inf_{\boldsymbol{x}\in E} \|\boldsymbol{y}_0, \boldsymbol{x}\| > 0$$

Since E is closed, for $\forall\, \eta > 0$, $\exists \boldsymbol{x}_0 \in E$ s.t

$$d \leq \|\boldsymbol{y}_0 - \boldsymbol{x}_0\| < d + \eta$$

Let $\boldsymbol{y} = \dfrac{\boldsymbol{y}_0 - \boldsymbol{x}_0}{\|\boldsymbol{y}_0 - \boldsymbol{x}_0\|}$, $\|\boldsymbol{y}\| = 1$ and for $\forall \boldsymbol{x} \in E$

$$\|\boldsymbol{y} - \boldsymbol{x}\| > \frac{d}{d+\eta} = 1 - \frac{\eta}{d+\eta}$$

Therefore for $\forall\, \epsilon \in (0, 1)$, we can take $\eta = \frac{d\epsilon}{1-\epsilon}$ then we'll have $\|\boldsymbol{y} - \boldsymbol{x}\| \geq 1 - \epsilon$ □

2.5 convex set and fix point

Definition 2.30 (convex set). *In linear space V, a set $E \subset V$ is said to be convex if*

$$\lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in E \quad (\forall\, \boldsymbol{x}, \boldsymbol{y} \in E,\ \forall 0 \leq \lambda \leq 1)$$

Proposition 2.24. *If $\{E_\alpha \mid \alpha \in \Gamma\}$ is a family of convex sets, then $\bigcap E_\alpha$ is also a convex set.*

2.6 Inner product spaces

De. spite that B^* spaces allow concepts of length, convergence, it seems to be imperfect without the notion of "angle". But this notion is very natural in Euclidean spaces, Therefore, we hope to generalize such ideas to infinite dimension spaces.

Definition 2.31 (inner product, inner product space). *Let V be a linear space over field $\mathbb{F}$, an inner product $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{F}$ is a scalar-valued function defined on V s.t for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$ and $\alpha, \beta \in \mathbb{F}$*

1. $\langle \boldsymbol{x}, \boldsymbol{x}\rangle \geq 0$; $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = 0$ *iff* $\boldsymbol{x} = 0$
2. $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x}\rangle}$
3. $\langle \alpha\boldsymbol{x} + \beta\boldsymbol{y}, \boldsymbol{z}\rangle = \alpha\langle \boldsymbol{x}, \boldsymbol{z}\rangle + \beta\langle \boldsymbol{y}, \boldsymbol{z}\rangle$

An inner product space is a linear space together with an inner product $(V, \langle\cdot,\cdot\rangle)$

With the above three axioms, we also have

$$\langle \boldsymbol{x}, \boldsymbol{y}+\boldsymbol{z}\rangle = \overline{\langle \boldsymbol{y}+\boldsymbol{z}, \boldsymbol{x}\rangle} = \overline{\langle \boldsymbol{y}, \boldsymbol{x}\rangle + \langle \boldsymbol{z}, \boldsymbol{x}\rangle} = \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{x}, \boldsymbol{z}\rangle$$

Example 15 ($(\ell_2, \langle\cdot,\cdot\rangle)$). *The ℓ_2 space: all sequences $\boldsymbol{x} = (x_1, x_2, \ldots, x_n, \ldots)$ of real or complex numbers with $\sum_{i=1}^{\infty} |x_i|^2 < \infty$. For $\boldsymbol{x}, \boldsymbol{y} \in \ell_2$, define $\langle\cdot,\cdot\rangle$:*

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \sum_{i=1}^{\infty} x_i \bar{y}_i$$

We must show that this inner product is well-defined. i.e the series converges; Since

$$\sum_{i=1}^{\infty} |x_i \bar{y}_i| \leq \sum_{i=1}^{\infty} \frac{1}{2}(|x_i^2| + |y_i^2|) < \infty$$

Example 16 **($L^2(\Omega,\mu), \langle\cdot,\cdot\rangle$).** *The $L^2(\Omega,\mu)$ space (11): all measurable functions on Ω, s.t $|u(x)|^2$ is measurable. For $u(x), v(x) \in L^2(\Omega,\mu)$, define $\langle\cdot,\cdot\rangle$:*

$$\langle u(x), v(x)\rangle = \int_\Omega u(x)\overline{v(x)}dx$$

Example 17 **($C^k(\overline{\Omega})$).**

Inner products can induce norm, with Cauchy-Schwarz inequality

Proposition 2.25 (Cauchy-Schwarz inequality). *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space and let*

$$\|\boldsymbol{x}\| = \langle \boldsymbol{x}, \boldsymbol{x}\rangle^{\frac{1}{2}}$$

Then we have

$$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \le \|\boldsymbol{x}\|\|\boldsymbol{y}\|$$

and the equality holds iff $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly dependent.

Proof. The result is trivial for $\boldsymbol{x} = 0$ or $\boldsymbol{y} = 0$, so we may assume that $\boldsymbol{x} \neq 0,\ \boldsymbol{y} \neq 0$.

$$0 \le \langle \boldsymbol{x} - \alpha\boldsymbol{y}, \boldsymbol{x} - \alpha\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{x}\rangle - \alpha\langle \boldsymbol{y}, \boldsymbol{x}\rangle - \bar{\alpha}\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \alpha\bar{\alpha}\langle \boldsymbol{y}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{x}\rangle - \bar{\alpha}\langle \boldsymbol{x}, \boldsymbol{y}\rangle - \alpha[\langle \boldsymbol{y}, \boldsymbol{x}\rangle - \bar{\alpha}\langle \boldsymbol{y}, \boldsymbol{y}\rangle] \tag{4}$$

Now choosing $\alpha = \dfrac{\langle \boldsymbol{x}, \boldsymbol{y}\rangle}{\langle \boldsymbol{y}, \boldsymbol{y}\rangle}$, we have

$$0 \le \langle \boldsymbol{x}, \boldsymbol{x}\rangle - \frac{\langle \boldsymbol{y}, \boldsymbol{x}\rangle}{\langle \boldsymbol{y}, \boldsymbol{y}\rangle}\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{x}\rangle - \frac{|\langle \boldsymbol{y}, \boldsymbol{x}\rangle|^2}{\langle \boldsymbol{y}, \boldsymbol{y}\rangle} \tag{5}$$

whence

$$|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \le \|\boldsymbol{x}\|\|\boldsymbol{y}\|$$

Now suppose the equality holds, then by (4) and (5) we have $\langle \boldsymbol{x} - \alpha\boldsymbol{y}, \boldsymbol{x} - \alpha\boldsymbol{y}\rangle = 0$ i.e $\boldsymbol{x} = \alpha\boldsymbol{y}$. ☐

Theorem 2.26. *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space, define*

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

Then $\|\bullet\|$ defines a norm on the space V, and $(V, \|\bullet\|)$ is a B^* space.

Proof. Now check the three axioms of norm Definition(2.24).

1. $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle} \ge 0$ and $\|\boldsymbol{x}\| = 0$ iff $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = 0$ i.e $\boldsymbol{x} = 0$
2. $\|a\boldsymbol{x}\| = \sqrt{\langle a\boldsymbol{x}, a\boldsymbol{x}\rangle} = \sqrt{a\bar{a}\langle \boldsymbol{x}, \boldsymbol{x}\rangle} = |a|\sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle} = |a|\|\boldsymbol{x}\|$
3.

$$\begin{aligned}
\|\boldsymbol{x} + \boldsymbol{y}\| = \sqrt{\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y}\rangle} &= \sqrt{\|\boldsymbol{x}\|^2 + \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{y}, \boldsymbol{x}\rangle + \|\boldsymbol{y}\|^2} \\
&= \sqrt{\|\boldsymbol{x}\|^2 + 2Re\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2} \\
&\le \sqrt{\|\boldsymbol{x}\|^2 + 2|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| + \|\boldsymbol{y}\|^2} \\
&\le \sqrt{\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{x}\|\|\boldsymbol{y}\| + \|\boldsymbol{y}\|^2} \\
&= \sqrt{(\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2} \\
&= \|\boldsymbol{x}\| + \|\boldsymbol{y}\|
\end{aligned}$$

□

Proposition 2.27. *In inner product spaces $(V, \langle\cdot,\cdot\rangle)$ the inner product $\langle\cdot,\cdot\rangle$ is a continuous function w.r.t the norm $\|\bullet\|$*

Proof. Suppose $\|\boldsymbol{x}-\boldsymbol{x}_0\|<\epsilon$, $\|\boldsymbol{y}-\boldsymbol{y}_0\|<\epsilon$, therefore $\boldsymbol{x}, \boldsymbol{y}$ are bounded. And suppose they are all bounded by M

$$\begin{aligned}
|\langle\boldsymbol{x},\boldsymbol{y}\rangle-\langle\boldsymbol{x}_0,\boldsymbol{y}_0\rangle| &= |\langle\boldsymbol{x},\boldsymbol{y}\rangle-\langle\boldsymbol{x},\boldsymbol{y}_0\rangle+\langle\boldsymbol{x},\boldsymbol{y}_0\rangle-\langle\boldsymbol{x}_0,\boldsymbol{y}_0\rangle| \\
&\leq |\langle\boldsymbol{x},\boldsymbol{y}\rangle-\langle\boldsymbol{x},\boldsymbol{y}_0\rangle|+|\langle\boldsymbol{x},\boldsymbol{y}_0\rangle-\langle\boldsymbol{x}_0,\boldsymbol{y}_0\rangle| \\
&= |\langle\boldsymbol{x},\boldsymbol{y}-\boldsymbol{y}_0\rangle|+|\langle\boldsymbol{x}-\boldsymbol{x}_0,\boldsymbol{y}_0\rangle| \\
&\leq M\|\boldsymbol{y}-\boldsymbol{y}_0\|+M\|\boldsymbol{x}-\boldsymbol{x}_0\| \\
&\leq 2M\epsilon
\end{aligned}$$

□

Proposition 2.28. *Inner product space is strictly convex.*

Proof. $\forall 0<\lambda<1$ when $\|\boldsymbol{x}\|=\|\boldsymbol{y}\|=1$ and $\boldsymbol{x}\neq\boldsymbol{y}$

$$\|\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y}\|^2=\lambda^2\|\boldsymbol{x}\|^2+2\lambda(1-\lambda)Re\langle\boldsymbol{x},\boldsymbol{y}\rangle+(1-\lambda)^2\|\boldsymbol{y}\|^2<(\lambda+(1-\lambda))^2=1$$

The inequality in the last step is because that when $\|\boldsymbol{x}\|=\|\boldsymbol{y}\|=1$ and $\boldsymbol{x}\neq\boldsymbol{y}$ the equality is impossible. □

A natural question arises: When is it proper for a normed linear space to induce an inner product?

Theorem 2.29 (polarization identity). *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space on the field $\mathbb{F}$, then*

$$\langle\boldsymbol{x},\boldsymbol{y}\rangle=\frac{\|\boldsymbol{x}+\boldsymbol{y}\|^2}{4}-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{4} \qquad (\mathbb{F}=\mathbb{R}) \tag{6}$$

$$\langle\boldsymbol{x},\boldsymbol{y}\rangle=\frac{\|\boldsymbol{x}+\boldsymbol{y}\|^2}{4}-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{4}+i\left(\frac{\|\boldsymbol{x}+i\boldsymbol{y}\|^2}{4}-\frac{\|\boldsymbol{x}-i\boldsymbol{y}\|^2}{4}\right) \qquad (\mathbb{F}=\mathbb{C}) \tag{7}$$

Theorem 2.30 (parallelogram identity). *In an inner product space, we have*

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2+\|\boldsymbol{x}-\boldsymbol{y}\|^2=2\|\boldsymbol{x}\|^2+2\|\boldsymbol{y}\|^2$$

The parallelogram identity turns out to be the sufficient condition for a normed linear space to be able to define an inner product. And the way to define such an inner product is via the polarization identity (2.29)

Theorem 2.31. *A normed linear space V over a field $\mathbb{F}$ is an inner product space iff the parallelogram identity holds.*

Proof. It suffice to prove for $\Leftarrow$. We define the inner product via (2.29). Now check the three axioms of inner product:

1. $\langle\boldsymbol{x},\boldsymbol{x}\rangle=\|\boldsymbol{x}\|\geq 0$ and $\langle\boldsymbol{x},\boldsymbol{x}\rangle=0$ iff $\boldsymbol{x}=0$
2. $\langle\boldsymbol{x},\boldsymbol{y}\rangle=\langle\boldsymbol{y},\boldsymbol{x}\rangle$
3. The hard part is to show the linearity of the function $\langle\cdot,\cdot\rangle$, here I present the proof for $\mathbb{F}=\mathbb{R}$:

Step1: prove that $\langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{z}\rangle = \langle \boldsymbol{x}, \boldsymbol{z}\rangle + \langle \boldsymbol{y}, \boldsymbol{z}\rangle$:
By the parallelogram identity:

$$\|\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}\|^2 = 2\|\boldsymbol{x}+\boldsymbol{z}\|^2 + 2\|\boldsymbol{y}\|^2 - \|\boldsymbol{x}+\boldsymbol{z}-\boldsymbol{y}\|^2$$
$$\|\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}\|^2 = 2\|\boldsymbol{y}+\boldsymbol{z}\|^2 + 2\|\boldsymbol{x}\|^2 - \|\boldsymbol{y}+\boldsymbol{z}-\boldsymbol{x}\|^2$$

Since $A = B = C$ imply that $A = \frac{1}{2}(B+C)$:

$$\|\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}\|^2 = \|\boldsymbol{y}+\boldsymbol{z}\|^2 + \|\boldsymbol{x}\|^2 + \|\boldsymbol{x}+\boldsymbol{z}\|^2 + \|\boldsymbol{y}\|^2 - \frac{1}{2}\|\boldsymbol{x}+\boldsymbol{z}-\boldsymbol{y}\|^2 - \frac{1}{2}\|\boldsymbol{y}+\boldsymbol{z}-\boldsymbol{x}\|^2$$

Replacing $\boldsymbol{z} = -\boldsymbol{z}$ in the above equation gives:

$$\|\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{z}\|^2 = \|\boldsymbol{y}-\boldsymbol{z}\|^2 + \|\boldsymbol{x}\|^2 + \|\boldsymbol{x}-\boldsymbol{z}\|^2 + \|\boldsymbol{y}\|^2 - \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{z}-\boldsymbol{y}\|^2 - \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{z}-\boldsymbol{x}\|^2$$

Applying $\|\boldsymbol{w}\| = \|-\boldsymbol{w}\|$ to the two negative terms in the last equation we have

$$\langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{z}\rangle = \frac{\|\boldsymbol{x}+\boldsymbol{y}+\boldsymbol{z}\|^2}{4} - \frac{\|\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{z}\|^2}{4} = \frac{\|\boldsymbol{y}+\boldsymbol{z}\|^2 - \|\boldsymbol{y}-\boldsymbol{z}\|^2}{4} + \frac{\|\boldsymbol{x}+\boldsymbol{z}\|^2 - \|\boldsymbol{x}-\boldsymbol{z}\|^2}{4} = \langle \boldsymbol{y}, \boldsymbol{z}\rangle + \langle \boldsymbol{x}, \boldsymbol{z}\rangle$$

Step2: prove that $\langle \lambda\boldsymbol{x}, \boldsymbol{y}\rangle = \lambda\langle \boldsymbol{x}, \boldsymbol{y}\rangle$:
By the last step we have for $\lambda = 2$: $\langle 2\boldsymbol{x}, \boldsymbol{y}\rangle = 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ and by induction we have that $\lambda \in \mathbb{N}$. And since for $\lambda = -1$ this clearly holds, so the result holds for $\lambda \in \mathbb{Z}$. If $\lambda = \frac{p}{q}$, $p, q \in \mathbb{Z}$, $q \neq 0$, let $\boldsymbol{x}' = \frac{\boldsymbol{x}}{q}$, then

$$\langle \lambda\boldsymbol{x}, \boldsymbol{y}\rangle = \langle p\boldsymbol{x}', \boldsymbol{y}\rangle = p\langle \boldsymbol{x}', \boldsymbol{y}\rangle$$

multiply q in both sides we have

$$q\langle \lambda\boldsymbol{x}, \boldsymbol{y}\rangle = p\langle \boldsymbol{x}, \boldsymbol{y}\rangle$$

divide both sides with q we have

$$\langle \lambda\boldsymbol{x}, \boldsymbol{y}\rangle = \lambda\langle \boldsymbol{x}, \boldsymbol{y}\rangle$$

Then this gives

$$\langle \lambda\boldsymbol{x}, \boldsymbol{y}\rangle = \lambda\langle \boldsymbol{x}, \boldsymbol{y}\rangle \quad (\lambda \in)$$

And since from (2.27) we know that the function $t \to \frac{1}{t}\langle t\boldsymbol{x}, \boldsymbol{y}\rangle$ is continuous, and is equal to $\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ for $t \in$ and therefore it holds for $t \in \mathbb{R}$

Proof reference: https://math.stackexchange.com/questions/21792/norms-induced-by-inner-products-and-the-parallelogram-law □

Example 18 ($\ell_p (p \neq 2)$ space is not an inner product space). *We show that the norm doesn't satisfy the parallelogram identity, just take:*

$$\boldsymbol{x} = (1, 0, 0 \ldots), \quad \boldsymbol{y} = (0, 1, 0, \ldots)$$

Example 19. *$C[a, b]$ with sup norm isn't an inner product space.*

Definition 2.32 (Hilert space). *A complete inner product space is called Hilbert space.*

Definition 2.33 (orthogonality). *The elements $\boldsymbol{x}, \boldsymbol{y}$ in an inner product space $(V, \langle\cdot,\cdot\rangle)$ are said to be orthogonal if*

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$$

denoted by $\boldsymbol{x} \perp \boldsymbol{y}$
If M is a subset of V s.t $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$ for all $\boldsymbol{y} \in M$, we say that $\boldsymbol{x}$ is orthogonal to M, denoted by $\boldsymbol{x} \perp M$. Further, we shall denote the set of all elements in V that are orthogonal to M by

$$M^{\perp} = \{\boldsymbol{x} \in V : \boldsymbol{x} \perp M\}$$

And the set $M^{\perp}$ is said called the orthogonal complement.

Proposition 2.32.

1. $\{0\}^{\perp} = V$, $V^{\perp} = \{0\}$
2. $M^{\perp}$ is a closed linear subspace of V

   Proof. The fact that $M^{\perp}$ is linear is obvious, now we show that it's closed:
   Suppose $\boldsymbol{x}_n \in M^{\perp}$ and $\lim\limits_{n\to\infty} \boldsymbol{x}_n \to \boldsymbol{x}$ then for $\forall \boldsymbol{y} \in M$: $\langle \boldsymbol{y}, \boldsymbol{x}_n \rangle = 0$ then by the continuity of the inner product we have

   $$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \lim_{n\to\infty} \langle \boldsymbol{x}_n, \boldsymbol{y} \rangle = 0$$

   □

Theorem 2.33 (pythagoras). *Let $(V, \langle , \rangle)$ be an inner product space on $\mathbb{F} = \mathbb{R}$, then $\boldsymbol{x} \perp \boldsymbol{y}$ iff*

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$$

Definition 2.34 (orthogonal set, standard orthogonal set, complete orthogonal set). *$(V, \langle , \rangle)$ an inner product space, A set $S = \{\boldsymbol{e}_\alpha, \alpha \in \Lambda\}$ of elements of V is called an **orthogonal set** if*

$$\boldsymbol{e}_\alpha \perp \boldsymbol{e}_\beta \quad (\alpha \neq \beta, \ \ \forall \alpha, \beta \in \Lambda)$$

And if further

$$\|\boldsymbol{e}_\alpha\| = 1 \quad (\forall \alpha \in \Lambda)$$

*We say the set S is a **standard orthogonal set**.*
*And if we have that: if $S \subset T$ where T is an orthogonal set, then $S = T$. i.e S is the maximal orthogonal set in V. Then we say that S is **complete**. Or equivalently, there exists no non-zero element in V orthogonal to S.*

Proposition 2.34 (existence of complete orthogonal set). *For proper($\neq \{0\}$) inner product space V, there always exists a complete orthogonal set.*

The proof is based on Zorn's lemma.

Definition 2.35 (orthogonal baisis, Fourier coefficients and Fourier series). *An orthogonal set in $(V, \langle , \rangle)$ is said to be an orthogonal basis if $\forall \boldsymbol{x} \in V$, $\boldsymbol{x}$ can be written as*

$$\boldsymbol{x} = \sum_{\alpha \in \Lambda} \langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle \boldsymbol{e}_\alpha$$

where the numbers $\langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle$ are called the Fourier coefficients of $\boldsymbol{x}$ w.r.t S, and the formal series $\sum_{\alpha \in \Lambda} \langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle \boldsymbol{e}_\alpha$ is called the Fourier series of $\boldsymbol{x}$.

Theorem 2.35 (Best Fit). *Inner product space $(V, \langle , \rangle)$ If $S = \{\boldsymbol{e}_1, \boldsymbol{e}_2 \ldots, \boldsymbol{e}_n\}$ is a finite standard orthogonal set, and let $M = Span\langle S \rangle$, then for $\forall \boldsymbol{x} \in V$, there exists $\boldsymbol{y} \in M$ s.t*

$$\|\boldsymbol{x} - \boldsymbol{y}\| = d(\boldsymbol{x}, M)$$

In fact,

$$\boldsymbol{y} = \sum_{i=1}^{n} \langle \boldsymbol{x}, \boldsymbol{e}_i \rangle \boldsymbol{e}_i$$

Proof. For any choice of the scalars $\lambda_1, \lambda_2, \dots, \lambda_n$,

$$\begin{aligned}\|\boldsymbol{x} - \sum_{i=1}^{n} \lambda_i \boldsymbol{e}_i\|^2 &= \|\boldsymbol{x}\|^2 - 2\sum_{i=1}^{n} \lambda_i \langle \boldsymbol{e}_i, \boldsymbol{x}\rangle + \|\sum_{i=1}^{n} \lambda_i \boldsymbol{e}_i\|^2 + \sum_{i=1}^{n} |\langle \boldsymbol{e}_i, \boldsymbol{x}\rangle|^2 - \sum_{i=1}^{n} |\langle \boldsymbol{e}_i, \boldsymbol{x}\rangle|^2 \\ &= \|\boldsymbol{x}\|^2 - \sum_{i=1}^{n} |\langle \boldsymbol{e}_i, \boldsymbol{x}\rangle|^2 + \sum_{i=1}^{n} (\lambda_i - \langle \boldsymbol{e}_i, \boldsymbol{x}\rangle)^2\end{aligned}$$

has minimal $\|\boldsymbol{x}\|^2 - \sum_{i=1}^{n} |\langle \boldsymbol{e}_i, \boldsymbol{x}\rangle|^2$ when $\lambda_i = \langle \boldsymbol{e}_i, \boldsymbol{x}\rangle$ □

Theorem 2.36 (Bessel Inequality). *Inner product space $(V, \langle, \rangle)$, and $S = \{\boldsymbol{e}_\alpha | \ \alpha \in \Lambda\}$ is a standard orthogonal set, then for $\forall \boldsymbol{x} \in V$, we have*

$$\sum_{\alpha \in \Lambda} |\langle \boldsymbol{x}, \boldsymbol{e}_i\rangle|^2 \leq \|\boldsymbol{x}\|^2 \tag{8}$$

Proof. By the proof of the last theorem (2.35), we know that for finite S, $(\dim(S) = n, \forall n \in \mathbb{N})$ we have

$$\|\boldsymbol{x}\|^2 \geq \sum_{i=1}^{n} |\langle \boldsymbol{x}, \boldsymbol{e}_i\rangle|^2$$

Let

$$S_n = \left\{ \boldsymbol{e}_i \in S | \ |\langle \boldsymbol{x}, \boldsymbol{e}_i\rangle| > \frac{1}{n} \right\}$$

Then we must have $|S_n| < \infty$ for all n. Therefore Let

$$S_0 = \{\boldsymbol{e}_i \in S | \ \langle \boldsymbol{x}, \boldsymbol{e}_i\rangle \neq 0\}$$

then $|S_0|$ must be countable, i.e there are at most countable many of the elements in Λ that satisfy $\langle \boldsymbol{x}, \boldsymbol{e}_i\rangle \neq 0$. Therefore the equation (8) turns out to be summing over an countable set. And we've already proved the equation holds for any integer, by letting $n \to \infty$, we have that the equation holds when Λ is a countable set. □

We know from (2.35) that for finite S, for any vector $\boldsymbol{x} \in V$, we can always find the best approximation to $\boldsymbol{x}$ in Span(S), which is $\sum_{i=1}^{n} \langle \boldsymbol{x}, \boldsymbol{e}_i\rangle \boldsymbol{e}_i$, but when S isn't finite, we can't always guarantee that such approximation is still in the space V. But luckily, for Hilbert space, this is true. And we will prove the result with Bessel Inequality(2.36).

Corollary 2.36.1. *Suppose $(V, \langle, \rangle)$ is a **Hilbert** space, and $S = \{\boldsymbol{e}_\alpha | \ \alpha \in \Lambda\}$ is a standard orthogonal set, then for $\forall \boldsymbol{x} \in V$,*

$$\sum_{\alpha \in \Lambda} \langle \boldsymbol{x}, \boldsymbol{e}_\alpha\rangle \boldsymbol{e}_\alpha \in V$$

Further,

$$\|\boldsymbol{x} - \sum_{\alpha \in \Lambda} \langle \boldsymbol{x}, \boldsymbol{e}_\alpha\rangle \boldsymbol{e}_\alpha\|^2 = \|\boldsymbol{x}\|^2 - \sum_{\alpha \in \Lambda} |\langle \boldsymbol{x}, \boldsymbol{e}_\alpha\rangle|^2$$

Proof. From the proof of (2.36), we know that we can actually substitute Λ to be $\mathbb{N}$, and by Bessel Inequality, we know that the series converges :

$$\sum_{n=1}^{\infty} |\langle \boldsymbol{x}, \boldsymbol{e}_n\rangle|^2 \leq \|\boldsymbol{x}\|^2 < \infty$$

whence

$$\sum_{n=m+1}^{m+p} |\langle \boldsymbol{x}, \boldsymbol{e}_n \rangle|^2 \to 0; \quad (m \to \infty)$$

Therefore define $\boldsymbol{x}_m = \sum_{n=1}^{m} \langle \boldsymbol{x}, \boldsymbol{e}_n \rangle \boldsymbol{e}_n$, then

$$\|\boldsymbol{x}_{m+p} - \boldsymbol{x}_m\|^2 = \| \sum_{n=m+1}^{m+p} \langle \boldsymbol{x}, \boldsymbol{e}_n \rangle \boldsymbol{e}_n \|^2 = \sum_{n=m+1}^{m+p} |\langle \boldsymbol{x}, \boldsymbol{e}_n \rangle|^2 \to 0$$

Therefore $\boldsymbol{x}_m$ is a Cauchy sequence, and

$$\sum_{n=1}^{\infty} \langle \boldsymbol{x}, \boldsymbol{e}_n \rangle \boldsymbol{e}_n = \lim_{m \to \infty} \boldsymbol{x}_m \in V$$

And since $\boldsymbol{x} - \sum_{n=1}^{\infty} \langle \boldsymbol{x}, \boldsymbol{e}_n \rangle \boldsymbol{e}_n \perp \sum_{n=1}^{\infty} \langle \boldsymbol{x}, \boldsymbol{e}_n \rangle \boldsymbol{e}_n$

$$\|\boldsymbol{x} - \sum_{\alpha \in \Lambda} \langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle \boldsymbol{e}_\alpha\|^2 = \|\boldsymbol{x}\|^2 - \sum_{\alpha \in \Lambda} |\langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle|^2$$

□

When does $\boldsymbol{x} = \sum_{\alpha \in \Lambda} \langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle \boldsymbol{e}_\alpha$?

Theorem 2.37. *$(V, \langle , \rangle)$ the Hilbert space, if $S = \{\boldsymbol{e}_\alpha \mid \alpha \in \Lambda\}$ is a standard orthogonal set. Then the following statements are equivalent:*

1. *S is complete. i.e $S^\perp = 0$*
2. *S is an orthogonal **basis***
3. *Parseval identity holds. i.e*
$$\|\boldsymbol{x}\|^2 = \sum_{\alpha \in \Lambda} |\langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle|^2$$

Proof. (1) ⇒ (2): By the last corollary (2.36.1), we know that $\sum_{\alpha \in \Lambda} \langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle \boldsymbol{e}_\alpha \in V$ and now we wish to show that $\boldsymbol{x} = \sum_{\alpha \in \Lambda} \langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle \boldsymbol{e}_\alpha$:

$$\langle \boldsymbol{x} - \sum_{\alpha \in \Lambda} \langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle \boldsymbol{e}_\alpha, \boldsymbol{e}_\alpha \rangle = 0, \quad \forall \boldsymbol{e}_\alpha \in S$$

And since S is complete, therefore $\boldsymbol{x} = \sum_{\alpha \in \Lambda} \langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle \boldsymbol{e}_\alpha, \boldsymbol{e}_\alpha$

(2) ⇒ (3): By the last corollary (2.36.1),

$$0 = \|\boldsymbol{x} - \sum_{\alpha \in \Lambda} \langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle \boldsymbol{e}_\alpha\|^2 = \|\boldsymbol{x}\|^2 - \sum_{\alpha \in \Lambda} |\langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle|^2$$

(3) ⇒ (1): Suppose that S is not complete, $\exists\ \boldsymbol{x} \neq 0 \in V$ s.t

$$\langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle = 0, \quad (\forall \boldsymbol{e}_\alpha \in S)$$

But by the last corollary we have

$$\boldsymbol{x} = \sum_{\alpha \in \Lambda} \langle \boldsymbol{x}, \boldsymbol{e}_\alpha \rangle \boldsymbol{e}_\alpha = 0$$

Therefore, contradiction! □

Now we give some examples of standard orthogonal **basis**:

Example 20. *On the space $L^2[0,\ 2\pi]$,*

$$\boldsymbol{e}_n(t) = \frac{1}{\sqrt{2\pi}} e^{int} \quad (n = 0, \pm 1, \pm 2, \ldots)$$

is a standard orthogonal basis. and for $\forall \boldsymbol{u} \in L^2[0, 2\pi]$, its Fourier coefficients are

$$\langle \boldsymbol{u}, \boldsymbol{e}_n \rangle = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} \boldsymbol{u}(t) e^{-\boldsymbol{int}} dt$$

(be careful! The field is $\mathbb{C}$)

Theorem 2.38 (Gram-Schmidt Orthonormalisation Procedure). *If $\{\boldsymbol{x}_k\}_1^\infty$ is a linearly independent set in an inner product space, then there exists an orthogonal set $\{\boldsymbol{e}_k\}_1^\infty$ s.t*

$$Span\langle \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \rangle = Span\langle \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n \rangle \quad \forall n \in \mathbb{N}$$

Proof. Set $\boldsymbol{y}_1 = \boldsymbol{x}_1,\ \boldsymbol{e}_1 = \dfrac{\boldsymbol{y}_1}{\|\boldsymbol{y}_1\|}$. And $\boldsymbol{y}_2 = \boldsymbol{x}_2 - \langle \boldsymbol{x}_2, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1,\ \ \boldsymbol{e}_2 = \dfrac{\boldsymbol{y}_2}{\|\boldsymbol{y}_2\|}$ ....

$$\boldsymbol{y}_n = \boldsymbol{x}_n - \sum_{i=1}^{n-1} \langle \boldsymbol{x}_n, \boldsymbol{e}_i \rangle \boldsymbol{e}_i,\ \ \boldsymbol{e}_n = \frac{\boldsymbol{y}_n}{\|\boldsymbol{y}_n\|}$$

□

Definition 2.36 (isomorphic, isometry). *Two linear spaces V_1, V_2 on field $\mathbb{F}$ are said to be* ***isomorphic*** *if there is a one-to-one linear map T from V_1 to V_2 s.t*

$$T(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha T(\boldsymbol{x}) + \beta T(\boldsymbol{y})$$

If the two spaces are also inner product and the T further satisfies

$$\langle T(\boldsymbol{x}), T(\boldsymbol{y}) \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle \quad (\forall \boldsymbol{x}, \boldsymbol{y} \in V_1)$$

Then we say V_1 and V_2 are ***isometry****.*

Theorem 2.39. *A standard orthogonal set S in a separable inner product space $(V, \langle, \rangle)$ is at most countable.*

Proof. If S is finite, then there's no need to prove, now suppose that S is infinite. Observe that for $\boldsymbol{x}, \boldsymbol{y} \in S$, $\|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{2}$ since $\boldsymbol{x} \perp \boldsymbol{y}$. Let $D = \{\boldsymbol{\xi}_n |\ n \in \mathbb{N}\}$ be a countable dense subset of V. Then for each $\boldsymbol{x} \in S$, there exists $\boldsymbol{\xi}_n \in D$ s.t $\|\boldsymbol{x} - \boldsymbol{\xi}_n\| < \dfrac{\sqrt{2}}{4}$. And we claim that for distinct $\boldsymbol{x} \in S$, the corresponding $\boldsymbol{\xi}_n$ are distinct too. Otherwise suppose there exists some $\boldsymbol{\xi}_n$ s.t $\|\boldsymbol{x} - \boldsymbol{\xi}_n\| < \dfrac{\sqrt{2}}{4}$ and $\|\boldsymbol{y} - \boldsymbol{\xi}_n\| < \dfrac{\sqrt{2}}{4}$ for some $\boldsymbol{x} \neq \boldsymbol{y},\ \boldsymbol{x}, \boldsymbol{y} \in S$. then we have $\|\boldsymbol{x} - \boldsymbol{y}\| < \frac{\sqrt{2}}{2}$ which is impossible. Therefore, we must have a one-to-one correspondence between the elements of S and a subset of $\mathbb{N}$. □

Theorem 2.40. *Every separable Hilbert space has a countable orthogonal basis.*

Proof. □

2.6.1 Best approximation in Hilbert Space

Recall in section (2.4.5) where we concluded that we can always find any vector $\boldsymbol{x}$'s best approximation to a finite subspace. But for infinite closed subspace it's unclear whether the best approximation point does exist. [See exercise 1.4.14]
This section we will show that this not only happens for Hilbert Space, but also the closed subspace can be generalized to closed convex set. [Note that linear spaces are all convex set.]

Let's start with $\boldsymbol{x} = 0$. And we wish to seek the best approximation of $\boldsymbol{x}$ to any closed convex set C.

Theorem 2.41. *If C is a closed convex set in Hilbert Space V, then there exists a unique vector $\boldsymbol{x}_0 \in C$ s.t $\|\boldsymbol{x}_0\|$ is the minimum.*

Proof. **existence** □

3 Linear operator and Linear functional

3.1 Linear operator

3.1.1 Concepts of linear operator and linear functional

Definition 3.1 (linear operator). *Let $\mathcal{X}, \mathcal{Y}$ be two linear spaces. A linear operator from $\mathcal{X}$ to $\mathcal{Y}$ is a mapping $T : \mathcal{X} \to \mathcal{Y}$ s.t*

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y) \quad (\forall x, y \in \mathcal{X})$$

Simply put, a linear operator is a linear mapping that preserves the structure of the underlying linear spaces.

More generally, we can restrict T defined on a subset $D \subset \mathcal{X}$ and we say that D is the domain of T denote as $D(T)$, similarly we denote $R(T)$ to be the range of the linear operator

$$R(T) = \{Tx | \forall x \in D\}$$

We shall denote by $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ the set of all linear operators from $\mathcal{X}$ to $\mathcal{Y}$, and $\mathcal{L}(\mathcal{X})$ for $\mathcal{L}(\mathcal{X}, \mathcal{X})$.

Fact 3.1 (Some facts you should know about linear operator). *Letthe null space or the kernel of T is the set*

$$ker(T) = \{x \in \mathcal{X} : Tx = 0\} = T^{-1}(0)$$

1. *$T(0) = 0$*
2. *T is one-to-one if $ker(T) = \{0\}$ and onto if $ran(T) = \mathcal{Y}$*
3. *If T is one-to-one then there exists a map $T^{-1} : ran(T) \to D(T)$ and the map is called the inverse of T. A map T is called invertible if $T : \mathcal{X} \to \mathcal{Y}$ has an inverse.*
4. *If T is invertible, then T^{-1} is also invertible, and $(T^{-1})^{-1} = T$.*
5. *$TT^{-1} = I_{\mathcal{Y}}, \; T^{-1}T = I_{\mathcal{X}}$*
6. *T^{-1} is a linear operator*

Example 21. *Let $\mathcal{X} = \mathbb{R}^n, \; \mathcal{Y} = \mathbb{R}^m, \; T \in \mathbb{R}^{m \times n}$ is a matrix.*

$$\boldsymbol{x} \to T\boldsymbol{x} \in \mathbb{R}^m \;\; (\forall \boldsymbol{x} \in \mathbb{R}^n)$$

is a linear operator

Example 22. *Let $\mathcal{X} = \mathcal{Y} = C^\infty(\bar{\Omega})$, and*

$$P(\partial_x) = \sum_{k=0}^{m} a_k(x) \partial_x^k \quad (a_k(x) \in C^\infty(\bar{\Omega}))$$

then $T : u(x) \to P(\partial_x) u(x) \;\; (\forall u \in \mathcal{X})$ Then T is a linear operator.

Example 23. *Let $\mathcal{X} = L^1(\mathbb{R})$, and*

$$T : u(x) \to \int_{-\infty}^{\infty} e^{i\xi x} u(\xi) d\xi$$

then $T \in \mathcal{L}(\mathcal{X})$

Definition 3.2 (linear functional). *If $\mathcal{Y} = \mathbb{R}/\mathbb{C}$ then we name the linear operator as real/complex linear functional*

Example 24 (integration). *Let $\mathcal{X} = C(\bar{\Omega})$ and*

$$f(u) = \int_{\Omega} u(\xi)d\xi \ \ (\ \forall u \in C(\bar{\Omega}) \ )$$

then f is a linear functional

Example 25 (derivative at some point). *Let $\mathcal{X} = C^{\infty}(\Omega)$, for some $k \in \mathbb{N}$ and $\xi_0 \in \Omega$*

$$f(u) = \partial_x^k u(\xi_0) \quad (\ \forall u \in C^{\infty}(\Omega) \ )$$

then f is a linear functional

3.1.2 continuity and boundedness of linear operator

Definition 3.3 (continuous linear operator). *If $\mathcal{X}$ and $\mathcal{Y}$ are two F^* (B^*), A linear operator $T : \mathcal{X} \to \mathcal{Y}$ is continuous at x_0 if*

$$\|x_n - x_0\| \to 0 \quad \Rightarrow \quad \|Tx_n - Tx_0\| \to 0$$

Proposition 3.2. *For linear operator T to be continuous in $D(T)$, it suffice to have T continuous at 0*

Proof. Let $x_n, \ x_0 \in D(T), \ x_{n0}$ Then

$$x_n - x_0 \to 0 \Rightarrow T(x_n - x_0) = T(x_n) - T(x_0) \to T0 = 0$$

□

Definition 3.4 (bounded linear operator). *If $\mathcal{X}$ and $\mathcal{Y}$ are two B^* spaces, linear operator T is said to be bounded if there exists some constant $M > 0$ s.t*

$$\|Tx\|_{\mathcal{Y}} \leq M\|x\|_{\mathcal{X}} \ \ (\forall x \in \mathcal{X})$$

Proposition 3.3 (linear operator continuous iff bounded on B^*). *If $\mathcal{X}$ and $\mathcal{Y}$ are two B^* spaces, linear operator T is bounded iff T is continuous.*

Proof. $\Leftarrow$ Otherwise suppose T is not bounded then for $\forall n \in \mathbb{N}$ there exists $x_n \in \mathcal{X}$ s.t

$$\|Tx_n\| > n\|x_n\|$$

Then let $y_n = \dfrac{x_n}{n\|x_n\|}$ then $y_n \to 0$, but $\|Ty_n\| > 1$ therefore T is impossible to be continuous, whence T must be bounded.
$\Rightarrow$ From Proposition (3.2), it suffice to show that T is continuous at 0. Let $x_n \to 0$ Since

$$\|Tx_n\| \leq M\|x_n\| \to 0$$

Hence T is continuous at 0. □

Definition 3.5 (operator norm). *Let $\mathcal{B}(\mathcal{X}, \mathcal{Y})$ be the set of all bounded linear operator from $\mathcal{X}$ to $\mathcal{Y}$, define*

$$\|T\| = \inf\{M : \|Tx\| \leq M\|x\|\}$$

to be the operator norm of T, denoted by $\|T\|$

Theorem 3.4 (equivalent definition of operator norm). *The following definitions of the operator norm are equivalent:*

$$\|T\| = \sup_{x \in \mathcal{X} - \{0\}} \frac{\|Tx\|}{\|x\|} = \sup_{\|x\|=1} \|Tx\| = \sup_{\|x\| \leq 1} \|Tx\|$$

Proof. Let $\alpha = \sup_{x \in \mathcal{X}-\{0\}} \frac{\|Tx\|}{\|x\|}$, $\beta = \sup_{\|x\|=1} \|Tx\|$, $\gamma = \sup_{\|x\|\leq 1} \|Tx\|$, we proceed the prove by showing that $\|T\| = \alpha \leq \beta \leq \gamma \leq \|T\|$:
By definition, $\frac{\|Tx\|}{\|x\|} \leq \alpha \Rightarrow \|Tx\| \leq \alpha\|x\|$, hence $\|T\| \leq \alpha$, and since $\|Tx\| \leq \|T\|\|x\|$ we have $\|$ and in particular for $\forall x \neq 0$, $\frac{\|Tx\|}{\|x\|} \leq T$ and $\alpha = \sup_{x \in \mathcal{X}-\{0\}} \frac{\|Tx\|}{\|x\|} \leq \|T\|$.
Hence, $\alpha = \|T\|$
And since

$$\left\{\frac{\|Tx\|}{\|x\|} \ : \ x \neq 0\right\} = \left\{\|T\left(\frac{x}{\|x\|}\right)\| \ : x \ \neq 0\right\} \subset \{\|Tx\| \ : \|x\| = 1\} \subset \{\|Tx\| \ : \|x\| \leq 1\}$$

Therefore $\alpha \leq \beta \leq \gamma$, and since

$$\frac{\|Tx\|}{\|x\|} \leq \alpha \Rightarrow \|Tx\| \leq \|x\|\alpha \leq \alpha \ \ (\forall \ \|x\| \leq 1 \ )$$

So $\gamma \leq \alpha$. Then we are done. □

Theorem 3.5 ($\mathcal{B}(\mathcal{X}, \mathcal{Y})$ is B^* space). *Suppose $\mathcal{X}$ and $\mathcal{Y}$ are two B^* spaces, then $\mathcal{B}(\mathcal{X}, \mathcal{Y})$ is obviously a linear space. And the operator norm given above is a norm on the space $\mathcal{B}(\mathcal{X}, \mathcal{Y})$.*

Proof. Check the 3 axioms of norm:

1. $\|T\| \geq 0$ is obvious, and if $\|T\| = 0$ then $Tx = 0$ $(\forall x \in \mathcal{X}) \Leftrightarrow T = 0$
2. $\|T_1 + T_2\| = \sup_{\|x\|=1} \|(T_1 + T_2)x\| \leq \sup_{\|x\|=1} \|T_1 x\| + \sup_{\|x\|=1} \|T_2 x\| = \|T_1\| + \|T_2\|$
3. $\|aT\| = \sup_{\|x\|=1} \|aTx\| = |a| \sup_{\|x\|=1} \|Tx\| = |a|\|T\|$

□

In fact, further if $\mathcal{Y}$ is B space (complete) then $\mathcal{B}(\mathcal{X}, \mathcal{Y})$ with the norm is also complete. Before giving this theorem, we shall give two definitions.

Definition 3.6 (uniformly operator convergent). *A sequence $\{T_n\}$ in $\mathcal{B}(\mathcal{X}, \mathcal{Y})$ is said to be uniformly operator convergent if*

$$\lim_{n\to\infty} \|T_n - T\| = 0$$

*This is also referred to as **convergence in the uniform topology** or **convergence in the operator norm topology** of $\mathcal{B}(\mathcal{X}, \mathcal{Y})$. In this case, T is called the **uniform operator limit** of the sequence. And $T \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$*

Definition 3.7 (strongly operator convergent). *If*

$$\lim_{n\to\infty} \|T_n x - Tx\| = 0 \quad (\forall x \in \mathcal{X} \ )$$

*And T is called the **strong operator limit**, but T isn't necessarily bounded.*

It's easy to see that a uniformly operator convergent sequence is also a strong operator convergent sequence.

Theorem 3.6. *Suppose $\mathcal{X}, \mathcal{Y}$ are two B^* spaces, then $\mathcal{B}(\mathcal{X}, \mathcal{Y})$ is a Banach space if $\mathcal{Y}$ is Banach space.*

Proof. We've shown that Let's firstly be clear about what we need to show: Suppose $\{T_n\}$ is a Cauchy sequence in $\mathcal{B}(\mathcal{X},\mathcal{Y})$, then there exists some T s.t for $\forall\epsilon > 0$, there exists some N s.t for $\forall n > N$

$$\|T_n - T\| \leq \epsilon$$

whence

$$\|T_n x - Tx\| \leq \epsilon\|x\| \quad (\forall x \in \mathcal{X}) \tag{9}$$

Now since $\{T_n\}$ is a Cauchy sequence, we have for $\forall\epsilon > 0$, there exists some N s.t for $\forall n > N$, $\forall p > 0$ we have

$$\|T_{n+p} - T_n\| \leq \epsilon$$

and

$$\|T_{n+p}x - T_n x\| \leq \epsilon\|x\| \quad (\forall x \in \mathcal{X}) \tag{10}$$

So for each $x \in \mathcal{X}$, sequence $T_n x$ is a Cauchy sequence in $\mathcal{Y}$, therefore there exists some $y \in \mathcal{Y}$ s.t

$$\lim_{n\to\infty} T_n x \to y$$

We shall define $y = Tx$ for each of x. Now we proceed to show that $T \in \mathcal{B}(\mathcal{X},\mathcal{Y})$ and that it is indeed the limit of T_n.

1. T is linear operator:

$$T(\alpha x_1 + \beta x_2) = \lim_{n\to\infty} T_n(\alpha x_1 + \beta x_2) = \lim_{n\to\infty} \alpha T_n(x_1) + \beta T_n(x_2) = \alpha T(x_1) + \beta T(x_2)$$

2. T is the limit of T_n:
   Take $p \to \infty$ in (10)

$$\|Tx - T_n x\| \leq \epsilon\|x\| \quad (\forall x \in \mathcal{X})$$

   which is exactly (9), hence T is indeed the limit.

3. T is bounded. Take $n = N + 1$

$$\|T\| \leq \|T - T_n\| + \|T_n\| \leq \epsilon + \|T_n\|$$

   whence T is bounded.

□

Theorem 3.7. *Suppose T is a linear operator from B^* space $\mathcal{X}$ to B^* space $\mathcal{Y}$. And both $\mathcal{X},\mathcal{Y}$ are finite. Then T must be continuous.*

Proof. By (2.4.4) we know that all norms on finite dimension linear space are equivalent, so it suffice to take $\mathcal{X} = \mathbb{F}^n$, $\mathcal{Y} = \mathbb{F}^m$. Furthermore, due to the linearity of T, T can be expressed by matrix. With a bit abused notation, we denote $T \in \mathbb{R}^{m\times n}$ to be the mapping matrix.

$$\|Tx\|_2^2 = \langle Tx, Tx\rangle = x^\top T^\top Tx = \mathrm{tr}(T^\top Txx^\top) = \langle T^\top T, xx^\top\rangle \leq \|T^\top T\|_F\|xx^\top\|_F = \|T^\top T\|_F\|x\|_2^2$$

By the submultiplity of Frobebinus norm, we have

$$\|T^\top T\|_F\|x\|_2^2 \leq \|T\|_F^2\|x\|_2^2$$

Therefore, T is continuous. □

Example 26.

3.2 Riesz Theorem and its Application

Suppose $\mathcal{X}$ is a Hilbert space, and for $\forall \boldsymbol{y} \in \mathcal{X}$, define

$$f_{\boldsymbol{y}}(x) : \boldsymbol{x} \to \langle \boldsymbol{x}, \boldsymbol{y} \rangle$$

Then $f_{\boldsymbol{y}} \in \mathcal{X}^*$, where $\mathcal{X}^*$ is the space of all bounded linear functional of $\mathcal{X}$.
In fact,

$$|f_{\boldsymbol{y}}(x)| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$$

So $\|f_{\boldsymbol{y}}\| \leq \|\boldsymbol{y}\|$. And actually since

$$|f_{\boldsymbol{y}}(\boldsymbol{y})| = \|\boldsymbol{y}\|^2$$

So $\|f_{\boldsymbol{y}}\| = \|\boldsymbol{y}\|$.
The following theorem tells us that the converse is also true.

Theorem 3.8 (F.Riesz theorem). *Suppose f is a bounded linear functional for the Hilbert space $\mathcal{X}$. i.e $f \in \mathcal{X}^*$, then there must exist some $\boldsymbol{y}_f \in \mathcal{X}$ s.t*

$$f(x) = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$$

Proof. □

3.3 category and open mapping theorem