Functional Analysis Notes
Metric Space

Zheng Xie

June 2, 2018

1 Contraction Mapping Principle

Definition 1.1 (metric space)
Let $\mathcal{X}$ be a non-empty set. We call $\mathcal{X}$ a **metric space** if there is a real-value function $\rho(x, y)$ defined on $\mathcal{X}$ such that
(1) $\rho(x, y) \geq 0$, and the equality holds if and only if $x = y$;
(2) $\rho(x, y) = \rho(y, x)$;
(3) $\rho(x, z) \leq \rho(x, y) + \rho(y, z) \quad , \forall x, y, z \in \mathcal{X}$.
where ρ is called a **metric** on $\mathcal{X}$. The metric space $\mathcal{X}$ with metric ρ is written as $(\mathcal{X}, \rho)$

Example 1.1 (Euclidean space)
Metric on Euclidean space is defined by

$$\rho(x, y) = \sqrt{< x - y, x - y >}$$

Example 1.2 (continuous functions on $[a, b]$)
We denote the set of all continuous functions on $[a, b]$ by $C[a, b]$.
It's a metric space with metric

$$\rho(x, y) = \max_{a \leq t \leq b} |x(t) - y(t)|$$

Definition 1.2 (convergence)
Sequence $\{x_n\}$ in $(\mathcal{X}, \rho)$ is **convergent** to x_0 if $\rho(x_n, x_0) \to 0$ as $n \to \infty$. We write $\lim\limits_{n \to \infty} x_n = x_0$ or $x_n \to x_0$.

Definition 1.3 (closed set)
A subset E of metric space $\mathcal{X}$ is **closed** if $\forall \{x_n\} \subset E$, if $x_n \to x_0$ then $x_0 \in E$.

Definition 1.4 (Cauchy sequence)
A sequence $\{x_n\}$ in metric space $\mathcal{X}$ is called a Cauchy sequence if $\rho(x_n, x_m) \to 0$ as $n, m \to \infty$. i.e. for any $\epsilon > 0$, there exists $N(\epsilon)$ such that for any $m, n \geq N(\epsilon)$, $\rho(x_n, x_m) < \epsilon$.

Definition 1.5 (completeness)
A metric space $\mathcal{X}$ is **complete** if and only if all the Cauchy sequence in $\mathcal{X}$ is convergent(to some point in $\mathcal{X}$).

Example 1.3
Euclidean space $\mathbb{R}^n$ is complete.
Proof.
Suppose $\{x_n\}$ is a Cauchy sequence in $\mathbb{R}^n$. It's easy to see that $\{x_n\}$ is bounded. By Bolzano-Weierstrass theorem, there is a convergent subsequence and since $\{x_n\}$ is Cauchy, the limit of the convergent subsequence is exactly the limit of $\{x_n\}$. □

Example 1.4
$(C[a,b],\rho)$ is complete.
Proof.
Suppose $\{x_n\}$ is a Cauchy sequence in $C[a,b]$.
By definition $\rho(x_n,x_m)\to 0$ as $n,m\to\infty$.
Hence given any $t\in[a,b]$, $|x_n(t)-x_m(t)|\to 0$ as $n,m\to\infty$. Since $\mathbb{R}$ is complete, $x_n(t)\to x_0(t)$ as $n\to\infty$.
For any $\epsilon>0$, there exists $N(\epsilon)$ such that $|x_n(t)-x_m(t)|<\epsilon$ for all $n,m>N(\epsilon)$. Fix t and let $m\to\infty$ we have $|x_n(t)-x_0(t)|<\epsilon$ when $n>N(\epsilon)$. Thus $x_n(t)$ converges to $x_0(t)$ uniformly and therefore $x_0(t)\in C[a,b]$.
So $\{x_n\}$ converges. □

Definition 1.6 (continuous mapping)
Let $T:(\mathcal{X},\rho)\to(\mathcal{Y},\gamma)$ be a map. T is **continuous** if for any $\{x_n\}\subset\mathcal{X}$, $x_0\in\mathcal{X}$,

$$\rho(x_n,x_0)\to 0\Rightarrow\gamma(Tx_n,Tx_0)\to 0$$

Proposition 1.1
Let $T:(\mathcal{X},\rho)\to(\mathcal{Y},\gamma)$ be a map.
Then T is continuous if and only if $\forall\epsilon>0$, $\forall x_0\in\mathcal{X}$, $\exists\delta=\delta(x_0,\epsilon)>0$, such that

$$\rho(x,x_0)<\delta\Rightarrow\gamma(Tx,Tx_0)<\epsilon,\forall x\in\mathcal{X}$$

Proof.
$\Rightarrow$:
Suppose not, then $\exists\epsilon>0$, $x_0\in\mathcal{X}$ and $\forall\delta>0$, there exists $x\in\mathcal{X}$ such that

$$\rho(x,x_0)<\delta,\gamma(Tx,Tx_0)\geq\epsilon$$

Let $\delta_n=\frac{1}{n}$, then we get a sequence $\{x_n\}$ such that

$$\rho(x_n,x_0)<\frac{1}{n}\to 0,\gamma(Tx_n,Tx_0)\geq\epsilon$$

leading a contradiction since T is continuous.
$\Leftarrow$:
Pick $x_0\in\mathcal{X}$, and let $\{x_n\}\subset\mathcal{X}$ be any sequence converges to x_0.
Now we have $\rho(x_n,x_0)\to 0$, we show that $\gamma(Tx_n,Tx_0)\to 0$.
$\forall\epsilon>0,\exists\delta=\delta(x_0,\epsilon)>0$ such that

$$\rho(x,x_0)<\delta\Rightarrow\gamma(Tx,Tx_0)<\epsilon,\forall x\in\mathcal{X}$$

Fix δ above, there exists N such that if $n>N$, then $\rho(x_n,x_0)<\delta$ and therefore $\gamma(Tx_n,Tx_0)<\epsilon$, i.e. $\gamma(Tx_n,Tx_0)\to 0$ as $n\to\infty$. □

Let ϕ be a real-value function defined on $\mathbb{R}$. It's clear that the root of the equation $\phi(x)=0$ is also the fixed point of $f:\mathbb{R}\to\mathbb{R},f(x)=x-\phi(x)$.

Consider integral equation

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

and metric space $C[-h,h]$.
Let $T:C[-h,h]\to C[-h,h]$ be a map defined by

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

Then the integral equation is equivalent to $x=Tx$. (x is the fixed point of T)

Definition 1.7 (contraction mapping)
Let $T : (\mathcal{X}, \rho) \to (\mathcal{X}, \rho)$ be a map. T is called a contraction mapping if there exists $\alpha \in (0, 1)$ such that $\rho(Tx, Ty) \le \alpha\rho(x, y)$, $\forall x, y \in \mathcal{X}$.

Example 1.5
Let $\mathcal{X} = [0, 1]$, $T(x)$ defined on $[0, 1]$ is a differentiable function, satisfying

$$T(x) \in [0, 1], |T'(x)| \le \alpha < 1, \forall x \in [0, 1]$$

Then $T : \mathcal{X} \to \mathcal{X}$ is a contraction mapping.
Proof.

$$\begin{aligned}\rho(Tx, Ty) &= |T(x) - T(y)| \\ &= |T'(\theta x + (1 - \theta)y)(x - y)| \\ &\le \alpha|x - y| = \alpha\rho(x, y), \forall x, y \in \mathcal{X}, 0 < \theta < 1\end{aligned}$$

□

Theorem 1.1 (Banach fixed point theorem)
Let $(\mathcal{X}, \rho)$ be a complete metric space, $T : \mathcal{X} \to \mathcal{X}$ is a contraction mapping, then there exists unique fixed point of T in $\mathcal{X}$.
Proof.
Pick any point $x_0 \in \mathcal{X}$, we can obtain a sequnce $\{x_n\}$ by iteration $x_{n+1} = Tx_n$. Then we have

$$\begin{aligned}\rho(x_{n+1}, x_n) &= \rho(T^n x_1, T^n x_0) \\ &\le \alpha^n \rho(x_1, x_0), \forall n \in \mathbb{Z}_+\end{aligned}$$

And for any $m \in \mathbb{Z}_+$

$$\begin{aligned}\rho(x_{n+m}, x_n) &= \sum_{k=0}^{m-1} \rho(x_{n+k+1}, x_{n+k}) \\ &\le \sum_{k=0}^{m-1} \alpha^{n+k}\rho(x_1, x_0) \\ &= \alpha^n \frac{1 - \alpha^m}{1 - \alpha}\rho(x_1, x_0) \\ &< \frac{\alpha^n}{1 - \alpha}\rho(x_1, x_0), \forall n \in \mathbb{Z}_+\end{aligned}$$

Hence $\{x_n\}$ is a Cauchy sequence. Since $(\mathcal{X}, \rho)$ is complete, there exists $x \in \mathcal{X}$ such that $x_n \to x$.
Then take the limit of both sides of the iteration $x_{n+1} = Tx_n$ we get $x = Tx$, i.e. x is a fixed point of T.
Suppose there is another fixed point x', then

$$\rho(x, x') = \rho(Tx, Tx') \le \alpha\rho(x, x')$$

This implies that $x' = x$. Therefore the fixed point of T is unique. □

2 Completion

Definition 2.1 (isometry)
Let $(\mathcal{X}, \rho)$, $(\mathcal{Y}, \gamma)$ be two metric spaces.
If there is a map $\phi : \mathcal{X} \to \mathcal{Y}$ satisfying
(1) ϕ is onto.
(2) $\rho(x, y) = \gamma(\phi x, \phi y)$, $\forall x, y \in \mathcal{X}$.
Then we call $(\mathcal{X}, \rho)$ and $(\mathcal{Y}, \gamma)$ are **isometric**. And ϕ is called an **isometry**.

Note.
(2) implies that ϕ is injective.

If metric space $(\mathcal{X}, \rho)$ is isometric to a subspace of another metric space $(\mathcal{Y}, \gamma)$. Then we say that $(\mathcal{X}, \rho)$ can be embedded in $(\mathcal{Y}, \gamma)$. Usually written as $(\mathcal{X}, \rho) \subset (\mathcal{Y}, \gamma)$

Definition 2.2 (dense)
Let $(\mathcal{X}, \rho)$ be a metric space. $E \subset \mathcal{X}$ is called a dense subset of $\mathcal{X}$ if $\forall x \in \mathcal{X}$, $\forall \epsilon > 0$, $\exists z \in E$ such that $\rho(x, z) < \epsilon$. In other words, $\forall x \in \mathcal{X}$, $\exists \{x_n\} \subset E$ such that $x_n \to x$.

Example 2.1
Denote the set of all polynomials on $[a, b]$ by $P[a, b]$. By Weierstrass theorem, $P[a, b]$ is dense in $C[a, b]$.

Definition 2.3 (completion)
The smallest complete metric space of given metric space $(\mathcal{X}, \rho)$ is called the completion of $\mathcal{X}$.

Proposition 2.1
If $(\mathcal{X}, \rho)$, $(\mathcal{X}_1, \rho_1)$ are metric spaces. $(\mathcal{X}, \rho) \subset (\mathcal{X}_1, \rho_1)$, $(\mathcal{X}_1, \rho_1)$ is complete. $\rho_1|_{\mathcal{X}\times\mathcal{X}} = \rho$ and $\mathcal{X}$ is dense in $\mathcal{X}_1$, then $\mathcal{X}_1$ is the completion of $\mathcal{X}$.
Proof.
$\forall \xi \in \mathcal{X}_1$, $\exists \{x_n\} \subset \mathcal{X}$ such that $\rho_1(x_n, \xi) \to 0$.
If there is another complete metric space, say $(\mathcal{X}_2, \rho_2)$, of which $(\mathcal{X}, \rho)$ is the subspace. (Naturally we have $\rho_2|_{\mathcal{X}\times\mathcal{X}} = \rho$).
Note that

$$\rho_2(x_n, x_m) = \rho_1(x_n, x_m) \to 0 \text{ as } n, m \to \infty$$

Thus $\exists \hat{\xi} \in \mathcal{X}_2$ such that $\rho_2(x_n, \hat{\xi}) \to 0$.
Then define a map $T : \mathcal{X}_1 \to \mathcal{X}_2$, $T\xi = \hat{\xi}$, it suffices to show that T is an isometry.
Since $\forall \eta \in \mathcal{X}_1$, $\exists y_n \in \mathcal{X}$ such that $\rho_1(y_n, \eta) \to 0$,

$$\rho_1(\xi, \eta) = \lim_{n\to\infty} \rho_1(x_n, y_n) = \lim_{n\to\infty} \rho_2(x_n, y_n) = \rho_2(\hat{\xi}, \hat{\eta})$$

Hence T is an isometric embedding, i.e. $(\mathcal{X}_1, \rho_1) \subset (\mathcal{X}_2, \rho_2)$ $\square$

Theorem 2.1
Every metric space has a completion.
Proof.
Let $(\mathcal{X}, \rho)$ be a metric space.
Step 1. Construct a metric space $\mathcal{X}_1$ containing $\mathcal{X}$
First we define a relation $\sim$ on all the Cauchy sequence in $\mathcal{X}$ by

$$\{x_n\} \sim \{y_n\} \text{ iff } \lim_{n\to\infty} \rho(x_n, y_n) = 0$$

It's easy to verify that $\sim$ is a equivalent relation. We see each equivalent class as an element and denote the set of all these equivalent classes by $\mathcal{X}_1$, then define metric on $\mathcal{X}_1$ by

$$\rho_1(\xi, \eta) = \lim_{n\to\infty} \rho(x_n, y_n)$$

where $x_n \in \xi$, $y_n \in \eta$.
We need to show that ρ_1 is well-defined.
Since for any $p \in \mathbb{Z}_+$,

$$\begin{aligned}
&|\rho(x_{n+p}, y_{n+p}) - \rho(x_n, y_n)| \\
&= |\rho(x_{n+p}, y_{n+p}) - \rho(x_{n+p}, y_n) + \rho(x_{n+p}, y_n) - \rho(x_n, y_n)| \\
&\le |\rho(x_{n+p}, y_{n+p}) - \rho(x_{n+p}, y_n)| + |\rho(x_{n+p}, y_n) - \rho(x_n, y_n)| \\
&\le |\rho(y_{n+p}, y_n)| + |\rho(x_{n+p}, x_n)| \to 0 \text{ as } n \to \infty
\end{aligned}$$

$\rho(x_n, y_n)$ is a Cauchy sequence in $\mathbb{R}$, which is a complete metric space. Hence the limit $\lim_{n\to\infty} \rho(x_n, y_n)$ does exist.
Note that if $x'_n \in \xi$, then

$$|\rho(x'_n, y_n) - \rho(x_n, y_n)| \le |\rho(x'_n, x_n)| \to 0$$

Hence the value of $\lim_{n\to\infty} \rho(x_n, y_n)$ doesn't depend on the selection of $\{x_n\}$ and $\{y_n\}$.
We also need to verify that the triangle inequality holds, i.e.

$$\rho_1(\xi, \zeta) \le \rho_1(\xi, \eta) + \rho_1(\eta, \zeta)$$

which can be obtained by taking the limit of both sides of the following inequality

$$\rho(x_n, z_n) \le \rho(x_n, y_n) + \rho(y_n, z_n)$$

Therefore, we showed that ρ_1 is a metric and $(\mathcal{X}_1, \rho_1)$ is a metric space.

Step 2. Show that $\mathcal{X}$ is dense in $\mathcal{X}_1$
$\forall x \in \mathcal{X}$, we denote by $\xi_x \in \mathcal{X}_1$ the equivalent class containing sequence $(x, x, \cdots, x, \cdots)$, and let $\mathcal{X}' = \{\xi_x : x \in \mathcal{X}\}$. Obviously, $\mathcal{X}' \subset \mathcal{X}_1$. Then we define a map

$$T : (\mathcal{X}, \rho) \to (\mathcal{X}', \rho_1), x \mapsto \xi_x$$

It's clear that T is onto. Also,

$$\rho(x_1, x_2) = \lim_{n\to\infty} \rho(x_1, x_2) = \rho_1(\xi_{x_1}, \xi_{x_2}) = \rho_1(Tx_1, Tx_2)$$

Thus T is an isometry and therefore $(\mathcal{X}, \rho) \subset (\mathcal{X}_1, \rho_1)$. By definition of $\mathcal{X}_1$, $\mathcal{X}$ is dense in $\mathcal{X}_1$.

Step 3. Show that $\mathcal{X}_1$ is complete
Let $\{\xi_n\} \subset \mathcal{X}_1$ be a Cauchy sequence. We show that $\exists \xi \in \mathcal{X}_1$ such that $\rho_1(\xi_n, \xi) \to 0$ as $n \to \infty$.
(1) Suppose $\{\xi_n\} \subset \mathcal{X}'$, let $x_n = T^{-1}\xi_n$, then $\{x_n\}$ is a Cauchy sequence. Suppose $\{x_n\} \in \xi$, then

$$\rho_1(\xi_n, \xi) = \lim_{m\to\infty} \rho(x_n, x_m) \to 0 \text{ as } n \to \infty$$

(2) Otherwise, note that $\mathcal{X}'$ is dense in $\mathcal{X}_1$, for each $\xi_n \in \mathcal{X}_1$, $\exists \hat{\xi}_n \in \mathcal{X}'$ such that $\rho_1(\xi_n, \hat{\xi}_n) < \frac{1}{n}$.
Since for each $p \in \mathbb{Z}_+$

$$\begin{aligned}
&\rho_1(\hat{\xi}_{n+p}, \hat{\xi}_n) \\
&\le \rho_1(\hat{\xi}_{n+p}, \xi_{n+p}) + \rho_1(\xi_{n+p}, \xi_n) + \rho_1(\xi_n, \hat{\xi}_n) \\
&< \frac{1}{n+p} + \frac{1}{n} + \rho_1(\xi_{n+p}, \xi_n) \to 0 \text{ as } n \to \infty
\end{aligned}$$

$\{\hat{\xi}_n\}$ is a Cauchy sequence in $\mathcal{X}'$ and by (1) we know that there exists ξ such that $\hat{\xi}_n \to \xi$. It follows that $\xi_n \to \xi$.
By *Proposition*2.1, $\mathcal{X}_1$ is the completion of $\mathcal{X}$. □

Example 2.2
The completion of $P[a, b]$ with metric

$$\rho(x, y) = \max_{a\le t\le b} |x(t) - y(t)|$$

is $C[a, b]$.

Example 2.3
The completion of $C[a, b]$ with metric

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

is $L^1[a, b]$.

3 Sequential Compactness

Definition 3.1 (bounded)
Let $(\mathcal{X}, \rho)$ be a metric space, $A \subset \mathcal{X}$, A is called **bounded** if there exists $x_0 \in \mathcal{X}$ and $r > 0$ such that $A \subset B(x_0, r)$ where

$$B(x_0, r) = \{x \in \mathcal{X} | \rho(x, x_0) < r\}$$

In finite-dimensional Euclidean space, infinite bounded set always contains a convergent subsequence (Bolzano-Weierstrass). However, the statement is not true for any metric space.

Example 3.1
Consider metric space $C[0, 1]$ and sequence

$$x_n(t) = \begin{cases} 0 & , t \geq \frac{1}{n} \\ 1 - nt & , t \leq \frac{1}{n} \end{cases}, n = 1, 2, \cdots$$

Obviously $\{x_n\} \subset B(0, 1)$ while $\{x_n\}$ doesn't contain a convergent subsequence.

Definition 3.2 (sequentially compact)
Let $(\mathcal{X}, \rho)$ be a metric space, $A \subset \mathcal{X}$. A is called a **sequentially compact set** if for any $\{x_n\} \subset A$, $\{x_n\}$ contains a convergent subsequence in $\mathcal{X}$. Moreover, if this subsequence converges to a point in A, then A is called a **self-sequentially compact set**. If $\mathcal{X}$ is sequentially compact, then $\mathcal{X}$ is called a **sequentially compact space**.

Proposition 3.1
Any bounded subset of $\mathbb{R}^n$ is sequentially compact. Any bounded closed subset of $\mathbb{R}^n$ is self-sequentially compact.

Proposition 3.2
In any sequentially compact space, any subset is sequentially compact, any closed subset is self-sequentially compact.

Proposition 3.3
Sequentially compact space is complete.
Proof.
Let $(\mathcal{X}, \rho)$ be a metric space, $\{x_n\} \subset \mathcal{X}$ is a Cauchy sequence. By sequentially compactness, there exists a subsequence converging to $x_0 \in \mathcal{X}$. Since $\{x_n\}$ is Cauchy we know that $x_n \to x_0$. $\square$

Definition 3.3 (ϵ-net)
Let $(\mathcal{X}, \rho)$ be a metric space, $M \subset \mathcal{X}$, $\epsilon > 0$, $N \subset M$.
If $\forall x \in M$, $\exists y \in N$ such that $\rho(x, y) < \epsilon$, then N is called an **ϵ-net** of M. Moreover, if N is a finite set, then N is called a **finite ϵ-net** of M.

Note.
By definition we have

$$M \subset \bigcup_{y \in N} B(y, \epsilon)$$

Definition 3.4 (totally bounded)
A set M is called **totally bounded**, if for any $\epsilon > 0$, there exists a finite ϵ-net of M.

Theorem 3.1 (Hausdorff)
Let $(\mathcal{X}, \rho)$ be a complete metric space, $M \subset \mathcal{X}$, then M is sequentially compact if and only if M is totally bounded.

Proof.
$\Rightarrow$:
Assume that M is not totally bounded. Then there exists ϵ_0 such that there is no finite ϵ_0-net of M.
Pick any $x_1 \in M$, then for each $n \in \mathbb{N}$, choose x_{n+1} inductively by

$$x_{n+1} \in M \setminus \bigcup_{k=1}^{n} B(x_k, \epsilon_0)$$

Then we obtain a infinite sequence $\{x_n\}$.
Note that for any $n \neq m$, $\rho(x_n, x_m) \geq \epsilon_0$. Hence it can't contain a convergent subsequence, leading a contradiction.
$\Leftarrow$:
Suppose $\{x_n\}$ is a infinite sequence in M, we want to find a convergent subsequence.
Note that, given any $\epsilon > 0$, the ϵ-net of M is finite, hence there must exists $y \in M$ such that $B(y, \epsilon)$ contains infinitely many terms of $\{x_n\}$.
Thus for 1-net, $\exists y_1 \in M$ and a subsequence $\{x_n^{(1)}\} \subset B(y_1, 1)$.
For $\frac{1}{2}$-net, $\exists y_2 \in M$ and a subsequence $\{x_n^{(2)}\} \subset B(y_2, \frac{1}{2})$ of $\{x_n^{(1)}\}$.

$$\cdots\cdots$$

For $\frac{1}{k}$-net, $\exists y_k \in M$ and a subsequence $\{x_n^{(k)}\} \subset B(y_k, \frac{1}{k})$ of $\{x_n^{(k-1)}\}$.

$$\cdots\cdots$$

Then we obtain a diagonal subsequence $\{x_n^{(n)}\}$, it's a Cauchy sequence.
In fact, $\forall \epsilon > 0$, when $n > \frac{2}{\epsilon}$, $\forall p \in \mathbb{N}$

$$\begin{aligned} \rho(x_{n+p}^{(n+p)}, x_n^{(n)}) &\leq \rho(x_{n+p}^{(n+p)}, y_n) + \rho(x_n^{(n)}, y_n) \\ &\leq \frac{2}{n} < \epsilon \end{aligned}$$

Since $\mathcal{X}$ is complete, $\{x_n^{(n)}\}$ is convergent. $\square$

Definition 3.5 (separable)
A metric space is called **separable** if it has countable dense subset.

Theorem 3.2
If a metric space is totally bounded, then it's separable.
Proof.
Let N_n denote the finite $\frac{1}{n}$-net, then $\bigcup_{n=1}^{\infty} N_n$ is a countable dense subset. $\square$

Definition 3.6 (compact)
Let $\mathcal{X}$ be a topological space. $M \subset \mathcal{X}$ is called **compact** if every open cover of M in $\mathcal{X}$ has a finite subcover.

Theorem 3.3
Let $(\mathcal{X}, \rho)$ be a metric space, $M \subset \mathcal{X}$. Then M is compact if and only if M is self-sequentially compact.
Proof.
$\Rightarrow$:
Let M be a compact set. First we show that M is closed.
(Actually, all metric spaces are Hausdorff and any compact set in Hausdorff space is closed)
$\forall x_0 \in \mathcal{X} \setminus M$, since

$$M \subset \bigcup_{x \in M} B(x, \frac{1}{2}\rho(x, x_0))$$

By compactness of M, $\exists x_k \in M, k = 1, 2, \cdots, n$ such that

$$M \subset \bigcup_{k=1}^{n} B(x_k, r_k)$$

where $r_k = \frac{1}{2}\rho(x_k, x_0)$
Take

$$\delta = \min_{1\leq k\leq n} r_k$$

Then $\forall x \in M$, suppose that $x \in B(x_k, r_k)$, we have

$$\rho(x, x_0) \geq \rho(x_k, x_0) - \rho(x, x_k) = 2r_k - \rho(x, x_k) > r_k \geq \delta$$

Thus $B(x_0, \delta) \cap M = \varnothing$ and therefore M is closed.
Next, assume that M is not self-sequentially compact, then there exists $\{x_n\} \subset M$ that doesn't have any convergent subsequence. Without loss of generality, we can assume that all x_n's are distinct.
Then for each $n \in \mathbb{N}$, let S_n denote $\{x_1, x_2, \cdots, x_{n-1}, x_{n+1}, \cdots\}$. Since S_n doesn't have a convergent subsequence, S_n is a closed set and therefore each $\mathcal{X} \setminus S_n$ is open. However,

$$\bigcup_{n=1}^{\infty}(\mathcal{X} \setminus S_n) = \mathcal{X} \setminus \bigcap_{n=1}^{\infty} S_n = \mathcal{X} \setminus \varnothing = \mathcal{X} \supset M$$

By compactness of M, there is a finite subcover

$$\bigcup_{n=1}^{N}(\mathcal{X} \setminus S_{k_n}) \supset M$$

This is impossible since for any x_m, $m \neq k_1, k_2, \cdots, k_N$,
$x_m \in M$ but $x_m \notin \bigcup_{n=1}^{N}(\mathcal{X} \setminus S_{k_n})$.
Hence M is self-sequentially compact.
$\Leftarrow$:
Since M is self-sequentially compact, M with metric ρ is complete. By Hausdorff theorem, M is totally bounded.
Assume that M is not compact, then there exists an open cover

$$\bigcup_{\lambda\in\Lambda} G_\lambda \supset M$$

that doesn't have a finite subcover.
For each $n \in \mathbb{N}$, there is a finite $\frac{1}{n}$-net

$$N_n = \{x_{k_1}^{(n)}, x_{k_2}^{(n)}, \cdots, x_{k_n}^{(n)}\}$$

Obviously

$$\bigcup_{y\in N_n} B(y, \frac{1}{n}) \supset M$$

Thus, $\forall n \in \mathbb{N}$, $\exists y_n \in N_n$ such that $B(y_n, \frac{1}{n})$ can't be covered by finitely many G_λ (Otherwise, there exists n such that $\bigcup_{y\in N_n} B(y, \frac{1}{n})$ can be covered by finitely many G_λ and therefore there exists a finite subcover of M).
Then we obtain a sequence $\{y_n\}$, since M is self-sequentially compact, there exists a convergent subsequence $\{y_{n_k}\}$, say, converging to $y_0 \in G_{\lambda_0}$.
Since G_{λ_0} is open and $\{y_{n_k}\}$ converges to $y_0 \in G_{\lambda_0}$, when k is large enough, $B(y_{n_k}, \frac{1}{n_k}) \subset G_{\lambda_0}$, which is contradict to the fact that each $B(y_n, \frac{1}{n})$ can be covered by finitely many G_λ. $\square$

Proposition 3.1
Let (M, ρ) be a compact metric space. Let $C(M)$ denote the set of all continuous mapping from M to $\mathbb{R}$. Define

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

Then $(C(M), d)$ is a metric space.
Proof.
It suffices to show that $d(u, v)$ is well-defined, i.e. we shall show that for each $u \in C(M)$, $\max_{x\in M} |u(x)|$ exists.

Since M is compact and u is continuous, $u(M)$ is also compact and therefore $u(M)$ is a bounded closed set. It follows that $\max_{x \in M} |u(x)|$ exists.

Proposition 3.2
$(C(M), d)$ is complete.
Proof.
Let $\{u_n(t)\}$ be a Cauchy sequence in $C(M)$.
Fix $t_0 \in M$, $\{u_n(t_0)\}$ is also a Cauchy sequence in $\mathbb{R}$. Let $u(t_0)$ denote the limit of $\{u_n(t_0)\}$.
$\forall \epsilon > 0$, $\exists N \in \mathbb{Z}_+$ such that $\forall m, n > N$, $d(u_m, u_n) < \epsilon$, let $n \to \infty$, we have $d(u_m, u) < \epsilon$ if $m > N$. Thus $u_n(t)$ converges to $u(t)$ uniformly. It follows that $u(t) \in C(M)$. $\square$

Definition 3.7 (uniformly bounded)
Let F be a subset of $C(M)$. F is called **uniformly bounded**, if $\exists M_1 > 0$ such that $|\phi(x)| \leq M_1, \forall x \in M, \forall \phi \in F$.

Definition 3.8 (equicontinuous)
Let F be a subset of $C(M)$. F is called **equicontinuous**,
if $\forall \epsilon > 0$, $\exists \delta = \delta(\epsilon) > 0$ such that

$$|\phi(x_1) - \phi(x_2)| \leq \epsilon, \quad \forall x_1, x_2 \in M, \rho(x_1, x_2) < \delta, \forall \phi \in F$$

Theorem 3.4 (Arzela-Ascoli)
Let F be a subset of $C(M)$.
Then F is sequentially compact if and only if F is uniformly bounded and equicontinuous.
Proof.
$\Rightarrow$:
Since $C(M)$ is complete, by Hausdorff theorem, F is totally bounded. Thus F is bounded and therefore uniformly bounded. Then we shall show that F is equicontinuous.
$\forall \epsilon > 0$, there exists a finite $\frac{\epsilon}{3}$-net N of M.
Suppose $N = \{f_1, f_2, \cdots, f_n\}$. Since M is compact, each f_k is uniformly continuous. Hence, there exists $\delta = \delta(\epsilon)$ such that for each f_k,

$$|f_k(x_1) - f_k(x_2)| < \frac{\epsilon}{3}, \quad \forall x_1, x_2 : \rho(x_1, x_2) < \delta$$

So for any $\phi \in F$, there exists some $f_j \in N$ such that

$$\begin{aligned} |\phi(x_1) - \phi(x_2)| &\leq |\phi(x_1) - f_j(x_1)| + |f_j(x_1) - f_j(x_2)| + |\phi(x_2) - f_j(x_2)| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon, , \quad \forall x_1, x_2 : \rho(x_1, x_2) < \delta \end{aligned}$$

$\Leftarrow$:
Suppose F is uniformly bounded and equicontinuous, we show that F is totally bounded.
$\forall \epsilon > 0$, $\exists \delta = \delta(\epsilon) > 0$ such that when $\rho(y_1, y_2) < \delta$, $\forall \phi \in F$, $|\phi(y_1) - \phi(y_2)| < \frac{\epsilon}{3}$.
For this δ, there exists a finite δ-net N of M.
Suppose $N = \{x_1, x_2, \cdots, x_n\}$, define a map $T : F \to \mathbb{R}^n$ by

$$T\phi = (\phi(x_1), \phi(x_2), \cdots, \phi(x_n)), \quad \forall \phi \in F$$

Let $\hat{F} = T(F)$, then $\hat{F}$ is a bounded set in $\mathbb{R}^n$. (Because F is uniformly bounded)
By Bolzano-Weierstrass, any sequence of $\hat{F}$ has a convergent subsequence, it follows that $\hat{F}$ is sequentially compact. By Hausdorff theorem, $\hat{F}$ is totally bounded. For given ϵ, there exists a finite $\frac{\epsilon}{3}$-net of $\hat{F}$

$$\hat{N} = \{T\phi_1, T\phi_2, \cdots, T\phi_m\}$$

Thus for any $\phi \in F$, there exists ϕ_i such that $\rho_n(T\phi, T\phi_i) < \frac{\epsilon}{3}$. Then pick $x_r \in N$ such that $\rho(x, x_r) < \delta$, and

$$\begin{aligned} |\phi(x) - \phi_i(x)| &\leq |\phi(x) - \phi(x_r)| + |\phi(x_r) - \phi_i(x_r)| + |\phi_i(x_r) - \phi_i(x)| \\ &< \frac{1}{3}\epsilon + \rho_n(T\phi, T\phi_i) + \frac{1}{3}\epsilon < \epsilon \end{aligned}$$

where ρ_n denote the metric on $\mathbb{R}^n$. $\square$

Example 3.2
Let $\Omega \subset \mathbb{R}^n$ be an open bounded convex set. If M_1, M_2 are two given positive numbers, then

$$F = \{\phi \in C^{(1)}(\bar{\Omega}) : |\phi(x)| \leq M_1, |grad(\phi(x))| \leq M_2, \forall x \in \Omega\}$$

is a sequentially compact set in $C(\bar{\Omega})$.
Proof.
$\forall \phi \in F$, $\forall x_1, x_2 \in \bar{\Omega}$, $\exists \theta \in (0,1)$ such that

$$\phi(x_1) - \phi(x_2) = grad(\phi(\theta x_1 + (1-\theta)x_2))(x_1 - x_2)$$

So $|\phi(x_1) - \phi(x_2)| \leq M_2 \rho_n(x_1, x_2), \quad \forall \phi \in F$
Hence F is equicontinuous. Obviously F is uniformly bounded. $\square$

4 Normed Vector Space

Definition 4.1 (vector space)
Let $\mathcal{X}$ be a non-empty set, $\mathbb{K}$ is a field ($\mathbb{R}$ or $\mathbb{C}$).
$\mathcal{X}$ is called a **vector space** if
(1) $\mathcal{X}$ is an additive abelian group.
(2) $\mathcal{X}$ is equipped with scalar multiplication $F \times \mathcal{X} \to \mathcal{X}$

Definition 4.2 (linear isomorphism)
Let $\mathcal{X}$ and $\mathcal{Y}$ be vector spaces. $T : \mathcal{X} \to \mathcal{Y}$ be a **linear isomorphism** if
(1) T is a bijection.
(2) $T(\alpha x + \beta y) = \alpha Tx + \beta Ty, \forall x, y \in \mathcal{X}, \forall \alpha, \beta \in \mathbb{K}$

Definition 4.3 (vector subspace)
Let E be a subset of $\mathcal{X}$. If E equipped with the same addition and scalar multiplication as $\mathcal{X}$ is also a vector space, then E is called a **vector subspace** of $\mathcal{X}$.

Definition 4.4 (norm)
A **norm** on vector space $\mathcal{X}$ is a function $\|\cdot\| : \mathcal{X} \to \mathbb{R}$, satisfying
(1) $\|x\| \geq 0$, $\forall x \in \mathcal{X}$. $\|x\| = 0$ iff $x = 0$.
(2) $\|\alpha x\| = |\alpha|\|x\|$, $\forall \alpha \in \mathbb{K}$, $\forall x \in \mathcal{X}$
(3) $\|x + y\| \leq \|x\| + \|y\|$, $\forall x, y \in \mathcal{X}$.

Definition 4.5 (normed vector space)
A **normed vector space** is a vector space $\mathcal{X}$ equipped with a norm.
It is also called a B^* space.

Definition 4.6 (Banach space)
A complete normed vector space is called a **Banach space**.

Definition 4.7 (equivalence of norm)
Let $\mathcal{X}$ be a vector space, $\|\cdot\|_1$ and $\|\cdot\|_2$ are two norms on $\mathcal{X}$.
$\|\cdot\|_2$ is **stronger** than $\|\cdot\|_1$, if

$$\|x_n\|_2 \to 0 \Rightarrow \|x_n\|_1 \to 0, \text{ as } n \to \infty$$

If $\|\cdot\|_2$ is stronger than $\|\cdot\|_1$ and $\|\cdot\|_1$ is stronger than $\|\cdot\|_2$, then $\|\cdot\|_1$ and $\|\cdot\|_2$ are **equivalent**.

Proposition 4.1

$\|\cdot\|_2$ is stronger than $\|\cdot\|_1$ if and only if there exists a constant C such that

$$\|\cdot\|_1 \leq C\|\cdot\|_2, \forall x \in \mathcal{X}$$

Proof.
$\Rightarrow$:
Suppose not, then for each $n \in \mathbb{Z}_+$, there exists $x_n \in \mathcal{X}$ such that $\|x_n\|_1 \geq n\|x_n\|_2$, let $y_n = \frac{x_n}{\|x_n\|_1}$, then $\|y_n\|_1 = 1$. On the other hand,

$$0 \leq \|y_n\|_2 < \frac{1}{n}, \forall n \in \mathbb{N}$$

So $\|y_n\|_2 \to 0$ as $n \to \infty$ and therefore $\|y_n\|_1 \to 0$, leading a contradiction.
$\Leftarrow$:
Trivial. $\square$

Corollary 4.1
$\|\cdot\|_1$ is equivalent to $\|\cdot\|_2$ if and only if there exists constants $C_1, C_2 > 0$ such that

$$C_1\|x\|_1 \leq \|x\|_2 \leq C_2\|x\|_1$$

Let $\mathcal{X}$ be a normed vector space, $dim\mathcal{X} = n$, then there is a basis of $\mathcal{X}$: $e_1, e_2, \cdots, e_n$. And any element $x \in \mathcal{X}$ has a unique representation:

$$x = \xi_1 e_1 + \xi_2 e_2 + \cdots + \xi_n e_n$$

Therefore, any point $x \in \mathcal{X}$ corresponds to a unique point $\xi = Tx = (\xi_1, \xi_2, \cdots, \xi_n)$ in $\mathbb{R}^n$.
We show that, the norm in $\mathcal{X}$ is equivalent to the norm in $\mathbb{R}^n$.
Consider function

$$p(\xi) = \|x\| = \|\sum_{j=1}^{n} \xi_j e_j\|, \quad \forall \xi \in \mathbb{R}^n$$

First note that p is uniformly continuous with respect to ξ:
$\forall \xi = (\xi_1, \xi_2, \cdots, \xi_n)$, $\forall \eta = (\eta_1, \eta_2, \cdots, \eta_n) \in \mathbb{R}^n$
By Triangle inequality and Cauchy-Schwarz inequality,

$$\begin{aligned}
&|p(\xi) - p(\eta)| \leq p(\xi - \eta) \leq \sum_{i=1}^{n} |\xi_i - \eta_i| \|e_i\| \\
&\leq (\sum_{i=1}^{n} |\xi_i - \eta_i|^2)^{\frac{1}{2}} (\sum_{i=1}^{n} \|e_i\|^2)^{\frac{1}{2}} \\
&= |\xi - \eta| (\sum_{i=1}^{n} \|e_i\|^2)^{\frac{1}{2}}
\end{aligned}$$

Then, $\forall \xi \in \mathbb{R}^n \setminus \{0\}$,

$$p(\xi) = \|\sum_{j=1}^{n} \xi_j e_j\| = |\xi| \|\sum_{j=1}^{n} \frac{\xi_j}{|\xi|} e_j\| = |\xi| p(\frac{\xi}{|\xi|})$$

Note that the unit sphere of $\mathbb{R}^n$, denoted by $S_1 = \{\xi \in \mathbb{R}^n | |\xi| = 1\}$ is compact. Hence $p(\xi)$ obtains its minimum C_1 and maximum C_2 on S_1, i.e.

$$C_1 \leq p(\xi) \leq C_2, \quad \forall \xi \in S_1$$

It follows that

$$C_1|\xi| \leq p(\xi) \leq C_2|\xi|, \quad \forall \xi \in \mathbb{R}^n$$

It remains to show that $C_1 > 0$.
Assume that $C_1 = 0$, then $\exists \xi^* \in S_1$ such that $p(\xi^*) = 0$.
Suppose $\xi^* = (\xi_1^*, \xi_2^*, \cdots, \xi_n^*)$,i.e.

$$\xi_1^* e_1 + \xi_2^* e_2 + \cdots + \xi_n^* e_n = 0$$

Since $\{e_i\}$ is a basis, it follows that $\xi^* = 0$, which is contradict to the fact that $\xi^* \in S_1$.
Thus we have

$$C_1|Tx| \leq \|x\| \leq C_2|Tx|, \quad \forall x \in X$$

If we regard $|Tx|$ as another norm, denoted by $\|x\|_T$, then it shows that $\|\cdot\|$ and $\|\cdot\|_T$ are equivalent. Therefore, the norm of n-dimensional normed vector space is equivalent to the norm of $\mathbb{R}^n$.

Theorem 4.1
Let $\mathcal{X}$ be a finite dimensional normed vector space, if $\|\cdot\|_1$ and $\|\cdot\|_2$ are both norm on $\mathcal{X}$, then there exists positive constants C_1, C_2 such that

$$C_1\|x\|_1 \leq \|x\|_2 \leq C_2\|x\|_1, \quad \forall x \in \mathcal{X}$$

Proof.
Suppose $dim\mathcal{X} = n$, since $\|\cdot\|_1$ and $\|\cdot\|_2$ are both equivalent to the norm of $\mathbb{R}^n$, $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent.

Note.
This theorem shows that any two n-dimensional normed vector space are isomorphic and homeomorphic.

Corollary 4.2
Any finite-dimensional normed vector space is a Banach space.

Corollary 4.3
Any finite-dimensional subspace of a normed vector space is closed.

Definition 4.8 (sublinear functional)
Let $P : \mathcal{X} \to \mathbb{R}$ be a function on vector space $\mathcal{X}$.If
(1) $P(x + y) \leq P(x) + P(y), \forall x, y \in \mathcal{X}$.
(2) $P(\lambda x) = \lambda P(x), \forall \lambda > 0, \forall x \in \mathcal{X}$.
Then P is called a sublinear functional on $\mathcal{X}$.

Theorem 4.2
Let $\mathcal{X}$ be a normed vector space. If $e_1, e_2, \cdots, e_n \in \mathcal{X}$ are given vectors, then $\forall x \in \mathcal{X}$, there exists best approximation coefficients $\lambda_1, \lambda_2, \cdots, \lambda_n$.
Proof.
Given any vector $x \in \mathcal{X}$, we want to find $(\lambda_1, \lambda_2, \cdots, \lambda_n) \in \mathbb{K}^n$ such that

$$\|x - \sum_{i=1}^{n} \lambda_i e_i\| = \min_{a \in \mathbb{K}^n} \|x - \sum_{i=1}^{n} a_i e_i\|$$

where $a = (a_1, a_2, \cdots, a_n)$.
Consider function

$$F(a) = \|x - \sum_{i=1}^{n} a_i e_i\|, a \in \mathbb{K}^n$$

We want to find its minimum. It's easy to see that F is a continuous function on $\mathbb{K}^n$.
Also,

$$F(a) \geq \|\sum_{i=1}^{n} a_i e_i\| - \|x\|, \forall a \in \mathbb{K}^n$$

Let $P(a) = \|\sum_{i=1}^{n} a_i e_i\|$, then $P(\cdot)$ is a norm on $\mathbb{K}^n$. Since $\mathbb{K}^n$ is a finite-dimensional space, by theorem 4.1, there exists $C_1 > 0$ such that

$$P(a) \geq C_1|a|, \forall a \in \mathbb{K}^n$$

where $|a| = (|a_1|^2 + |a_2|^2 + \cdots + |a_n|^2)^{\frac{1}{2}}$.
Thus $F(a) \to \infty$ as $|a| \to \infty$, therefore the minimum of F exists. □

Note.
If we write $M = span\{e_1, e_2, \cdots, e_n\}$, $\rho(x, M) = \inf_{y \in M} \|x - y\|$, $x_0 = \sum_{i=1}^{n} \lambda_i e_i$. Then $\rho(x, x_0) = \rho(x, M)$.

Definition 4.9 (strictly convex)
Let $(\mathcal{X}, \|\cdot\|)$ be a normed vector space.
If $\forall x \neq y \in \mathcal{X}, \|x\| = \|y\| = 1$, then

$$\|\alpha x + \beta y\| < 1, \forall \alpha, \beta > 0, \alpha + \beta = 1$$

Theorem 4.3
Let $\mathcal{X}$ be a normed vector space which is strictly convex. $\{e_1, e_2, \cdots, e_n\} \subset \mathcal{X}$ are linear independent, then $\forall x \in \mathcal{X}$, there exists a unique set of best approximation $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$.
Proof.
Suppose $d = \rho(x, M)$, $\exists y_1, y_2 \in M$ such that $\|x - y_1\| = \|x - y_2\| = d$, then $\forall \alpha, \beta > 0$, $\alpha + \beta = 1$, since M is strictly convex,

$$\begin{aligned}\frac{\|x - (\alpha y_1 + \beta y_2)\|}{d} &= \frac{\|\alpha(x - y_1) + \beta(x - y_2)\|}{d} \\ &= \|\alpha(\frac{x - y_1}{d}) + \beta(\frac{x - y_2}{d})\| < 1\end{aligned}$$

i.e. $\|x - (\alpha y_1 + \beta y_2)\| < d$, which is contradict to the definition of d.
If $d = 0$, the best approximation of x is y, then $x = y$. □

Theorem 4.4
Let $\mathcal{X}$ be a normed vector space, then $\mathcal{X}$ is finite-dimensional if and only if the unit sphere of $\mathcal{X}$ is sequentially compact.
Proof.
$\Rightarrow$:
If $\mathcal{X}$ is finite-dimensional, then $\mathcal{X}$ is homeomorphic to $\mathbb{R}^n$. Since the unit sphere of $\mathbb{R}^n$ is compact, the unit sphere of $\mathcal{X}$ is also compact and therefore sequentially compact.
$\Leftarrow$:
Assume that $\mathcal{X}$ is infinite-dimensional, let S_1 be the unit surface of $\mathcal{X}$.
Pick any $x_1 \in S_1$. Let M_n denote the span of $x_1, x_2, \cdots, x_n$, then we can always find $x_{n+1} \notin M_n$ such that $\|x_{n+1} - x_i\| \geq 1, \forall i = 1, 2, \cdots, n$.
It is because $\forall y \in \mathcal{X} \setminus M_n$, by theorem 4.3, $\exists x \in M_n$ such that

$$\|y - x\| = d = \rho(y, M_n)$$

Let $x_{n+1} = \frac{y-x}{d}$, then $x_{n+1} \in S_1$ and

$$\|x_{n+1} - x_i\| = \frac{\|y - (x + dx_i)\|}{d} \geq \frac{d}{d} = 1, \forall i = 1, 2, \cdots, n$$

So that we can obtain a sequence $\{x_n\}$ satisfying $\|x_n - x_m\| \geq 1, \forall n \neq m \in \mathbb{N}$, which doesn't have any convergent subsequence and therefore S_1 is not sequentially compact. □

Definition 4.10 (bounded)
Let $\mathcal{X}$ be a normed vector space, $A \subset \mathcal{X}$ is called bounded if there exists a constant $c > 0$ such that $\|x\| \leq c$, $\forall x \in A$.

Corollary 4.2
Let $\mathcal{X}$ be a normed vector space, $\mathcal{X}$ is finite dimensional if and only if any bounded subset of $\mathcal{X}$ is sequentially compact.

Lemma 4.1 (F.Riesz)
Let $\mathcal{X}$ be a normed vector space, $\mathcal{X}_0$ be a proper closed subspace, then for $\forall 0 < \epsilon < 1$, $\exists y \in \mathcal{X}$, such that $\|y\| = 1$ and $\|y - x\| \geq 1 - \epsilon$, $\forall x \in \mathcal{X}_0$
Proof.
By theorem 4.3, there exists $y_0 \in \mathcal{X} \setminus \mathcal{X}_0$ such that

$$\inf_{x \in \mathcal{X}_0} \|y_0 - x\| = 1$$

$\forall \epsilon > 0$, $\exists x_0 \in \mathcal{X}_0$ such that

$$1 \leq \|y_0 - x_0\| < 1 + \frac{\epsilon}{1-\epsilon} = \frac{1}{1-\epsilon}$$

Let $y = \frac{y_0 - x_0}{\|y_0 - x_0\|}$, then $\|y\| = 1$, and $\forall x \in \mathcal{X}_0$

$$\|y - x\| = \frac{y_0 - x'}{\|y_0 - x_0\|} > \frac{1}{\frac{1}{1-\epsilon}} = 1 - \epsilon \quad \square$$

5 Convex Set And Fixed Point

Definition 5.1 (convex)
Let $\mathcal{X}$ be a vector space, $E \subset \mathcal{X}$, E is called a **convex** set, if

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

Proposition 5.1
If $\{E_\lambda | \lambda \in \Lambda\}$ is a family of convex set in vector space $\mathcal{X}$, then $\bigcap_{\lambda \in \Lambda} E_\lambda$ is also a convex set.

Definition 5.2 (convex hull)
Let $\mathcal{X}$ be a vector space, $A \subset \mathcal{X}$. If $\{E_\lambda \subset \mathcal{X} | \lambda \in \Lambda\}$ is the family of all convex sets that contain A, then $\bigcap_{\lambda \in \Lambda} E_\lambda$ is called the **convex hull** of A, denoted by $conv(A)$.
$\forall n \in \mathbb{N}$, $x_1, x_2, \cdots, x_n \in A$,

$$\sum_{i=1}^{n} \lambda_i x_i, \text{ where } \lambda_i \geq 0, \sum_{i=1}^{n} \lambda_i = 1$$

is called the **convex combination** of $x_1, x_2, \cdots, x_n$

Proposition 5.2
Let $\mathcal{X}$ be a vector space, $A \subset \mathcal{X}$, then the convex hull of A is the set of all convex combination of elements of A, i.e.

$$conv(A) = \{\sum_{i=1}^{n} \lambda_i x_i | \sum_{i=1}^{n} \lambda_i = 1, x_i \in A, i = 1, 2, \cdots n, \forall n \in \mathbb{N}\}$$

Proof.
Let

$$S = \{\sum_{i=1}^{n} \lambda_i x_i | \sum_{i=1}^{n} \lambda_i = 1, x_i \in A, i = 1, 2, \cdots n, \forall n \in \mathbb{N}\}$$

Then $S \supset A$ and S by definition is a convex set. So it suffices to show that $S \subset conv(A)$.
Let C be any convex set that contains A. Pick any point in S, say

$$y = \sum_{i=1}^{n} \lambda_i x_i$$

Note that

$$y = \lambda_1 x_1 + (1-\lambda_1) x_1'$$

where

$$x_1' = \sum_{i=2}^{n} \frac{\lambda_i}{1-\lambda_1} x_i$$

$$\sum_{i=2}^{n} \frac{\lambda_i}{1-\lambda_1} = \frac{1-\lambda_1}{1-\lambda_1} = 1$$

It implies that y can be obtained inductively, hence $y \in C$. Since C is arbitrary, $y \in conv(A)$. Therefore $S = conv(A)$. $\square$

Definition 5.3 (Minkowski functional)
Let $\mathcal{X}$ be a vector space, $C \subset \mathcal{X}$ is convex, containing origin.
Define a function $P : \mathcal{X} \to [0, \infty]$ by

$$P(x) = \inf\{\lambda > 0 | \frac{x}{\lambda} \in C\}, \quad \forall x \in \mathcal{X}$$

Function P is called the **Minkowski functional** of C.
We can also write

$$P(x) = \inf\{\lambda > 0 | x \in \lambda C\}, \quad \forall x \in \mathcal{X}$$

where $\lambda C = \{\lambda x | x \in C\}$

Proposition 5.3
Let $\mathcal{X}$ be a vector space, $C \subset \mathcal{X}$ is convex, containing origin. If P is the Minkowski functional of C, then P has the following properties:
(1) $P(x) \in [0, \infty]$, $P(0) = 0$.
(2) $P(\lambda x) = \lambda P(x)$, $\forall x \in \mathcal{X}$, $\forall \lambda > 0$. (positive homogeneity)
(3) $\forall \epsilon > 0$, $\forall x \in \mathcal{X}$, there exists $\lambda = p(x) + \epsilon$ such that $x \in \lambda C$.
(4) $P(x + y) \leq P(x) + P(y)$, $\forall x, y \in \mathcal{X}$. (subadditivity)
(5) $\forall x \in C$, $p(x) \leq 1$.
Proof.

(1) Trivial.

(2) Trivial.

(3) By definition, $\forall \epsilon > 0$, there exists $p(x) \leq \lambda' < p(x) + \epsilon$ such that $x \in \lambda' C$.
Since C is convex, $\forall t \in [0, 1]$, $tx + (1 - t)0 \in \lambda' C$, then $x \in \frac{\lambda'}{t} C$.
Take $t = \frac{\lambda'}{p(x)+\epsilon}$, then $\lambda = \frac{\lambda'}{t} = p(x) + \epsilon$ and $x \in \lambda C$

(4) Suppose $P(x)$ and $P(y)$ are finite. $\forall \epsilon > 0$, take $\lambda_1 = P(x) + \frac{\epsilon}{2}$, $\lambda_2 = P(y) + \frac{\epsilon}{2}$, then

$$\frac{x}{\lambda_1} \in C, \frac{x}{\lambda_2} \in C$$

Since C is convex, so

$$\frac{x + y}{\lambda_1 + \lambda_2} = \frac{\lambda_1}{\lambda_1 + \lambda_2} \cdot \frac{x}{\lambda_1} + \frac{\lambda_2}{\lambda_1 + \lambda_2} \cdot \frac{y}{\lambda_2} \in C$$

which implies

$$P(x + y) \leq \lambda_1 + \lambda_2 = P(x) + P(y) + \epsilon$$

Since $\epsilon > 0$ is arbitrary we obtain (4).

(5) $\forall x \in C$, $x \in 1 \cdot C$, by definition of $p(x)$, so $p(x) \leq 1$. $\square$

Definition 5.4 (absorbent)
Let $\mathcal{X}$ be a vector space, $C \subset \mathcal{X}$ is convex and it contains origin. C is called **absorbent** if $\forall x \in \mathcal{X}$, $\exists \lambda > 0$ such that $x \in \lambda C$.

Definition 5.5 (balanced)
Let $\mathcal{X}$ be a vector space, $C \subset \mathcal{X}$ is convex and it contains origin. C is called **balanced** if $\forall x \in C$, $\forall \lambda : |\lambda| = 1$ we have $\lambda x \in C$.

Proposition 5.4
Let $\mathcal{X}$ be a vector space on $\mathbb{C}$, then any absorbent balanced convex $C \subset \mathcal{X}$ decide a seminorm on $\mathcal{X}$.

Proposition 5.5
Let $\mathcal{X}$ be a normed vector space, C is a closed convex set containing origin. If $P(x)$ is the Minkowski functional of C, then $P(x)$ is lower semi-continuous, i.e.
$\forall x_0 \in \mathcal{X}$, $\forall \epsilon > 0$, $\exists$ a neighborhood U of x_0 such that $f(x) \geq f(x_0) - \epsilon$ for all $x \in U$.
And if C is bounded, then

$$P(x) = 0 \text{ iff } x = 0$$

Moreover, if 0 is an interior point of C then C is absorbent and $P(x)$ is uniformly continuous.
Proof.

(1) We show that $aC = \{x \in \mathcal{X} | P(x) \leq a\}$.
$\forall a > 0$, if $x \in aC$, then by definition $P(x) \leq a$.
On the other hand, if $P(x) \leq a$, then $\forall n \in \mathbb{N}$, we have

$$\frac{x}{a + \frac{1}{n}} \in C$$

Note that C is closed and

$$\frac{x}{a + \frac{1}{n}} \to \frac{x}{a} \text{ as } n \to \infty$$

So $x \in aC$. Therefore $aC = \{x \in \mathcal{X} | P(x) \leq a\}$.
Particularly, $C = \{x \in \mathcal{X} | P(x) \leq 1\}$.

(2) We show that $P(x)$ is lower semi-continuous.
$\forall x_0 \in \mathcal{X}$, $let \lambda_0 = P(x_0)$, $\forall \epsilon > 0$, $x_0 \notin (\lambda_0 - \epsilon)C$, which is a closed set. Thus there exists an open neighborhood U of x_0 such that $U \cap (\lambda_0 - \epsilon)C = \varnothing$, then for any point $x \in U$, $P(x) \geq \lambda_0 - \epsilon = P(x_0) - \epsilon$. Therefore $P(x)$ is lower semi-continuous.

(3) Obviously $P(0) = 0$. Since C is bounded, there $\exists r > 0$ such that $C \subset B(0, r)$, hence $\forall x \in \mathcal{X} \setminus \{0\}$, $2r\frac{x}{\|x\|} \notin C$, it follows that $P(x) \geq \frac{\|x\|}{2r}$. Hence if $P(x) = 0$, then $x = 0$

(4) If 0 is an interior point of C, then there exists $r > 0$ such that $B(0, r) \subset C$, then

$$\frac{rx}{2\|x\|} \in C, \quad \forall x \in \mathcal{X} \setminus \{0\}$$

Hence C is absorbent. Moreover, $P(x) \leq 2\frac{\|x\|}{r}, \forall x \in \mathcal{X}$. Thus

$$|P(x) - P(y)| \leq \max\{P(x - y), P(y - x)\} \leq \frac{2}{r}\|x - y\|, \forall x, y \in \mathcal{X}$$

Therefore $P(x)$ is uniformly continuous. $\square$

Theorem 5.1 (Brouwer)
Let B be the closed unit ball of $\mathbb{R}^n$, let $T : B \to B$ be a continuous mapping, then there exists a fixed point of T $x \in B$.

Theorem 5.2 (Schauder)
Let $\mathcal{X}$ be a normed vector space, $C \subset \mathcal{X}$ is a closed convex set, $T : C \to C$ is continuous and C is sequentially compact, then there exists a fixed point of T in C.

Definition 5.6 (compact mapping)
Let $\mathcal{X}$ be a normed vector space, E is a subset of $\mathcal{X}$, $T : E \to \mathcal{X}$ is called **compact** if T is continuous and for any bounded set $A \subset E$, $T(A)$ is sequentially compact.

Corollary 5.1
Let $\mathcal{X}$ be a normed vector space, C is a bounded closed convex subset, $T : C \to C$ is compact, then there exists fixed point of T in C.

6 Inner Product Space

Definition 6.1 (inner product)
Let $\mathcal{X}$ be a vector space on field $\mathbb{K}$, function $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{K}$ is called an **inner product** if
(1) $\forall x, y \in \mathcal{X}$, $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
(2) $\forall x, y \in \mathcal{X}$, $\forall k \in \mathbb{K}$, $\langle kx, y \rangle = k\langle x, y \rangle$.
(3) $\forall x, y, z \in \mathcal{X}$, $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$.
(4) $\forall x \in \mathcal{X}$, $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ iff $x = 0$.

Theorem 6.1 (Cauchy-Schwarz inequality)
Let $(\mathcal{X}, \langle \cdot, \cdot \rangle)$ be a inner product space.$\forall x, y \in \mathcal{X}$,

$$|\langle x, y \rangle| \leq \langle x, x \rangle^{\frac{1}{2}} \langle y, y \rangle^{\frac{1}{2}}$$

And the equality holds if and only if $x = \lambda y$, $\lambda \in \mathbb{K}$.

Proposition 6.1
Let $(\mathcal{X}, \|\cdot\|)$ be a normed vector space, the function $\langle \cdot, \cdot \rangle$ induced by $\langle x, x \rangle^{\frac{1}{2}} = \|x\|$ is an inner product if and only if $\|\cdot\|$ satisfies

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2), \forall x, y \in \mathcal{X}$$

Proof.
$\Rightarrow$:
Trivial.
$\Leftarrow$:
Define $\langle x, y \rangle$ on $\mathcal{X}$ by

$$\langle x, y \rangle = \begin{cases} \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2) & , \mathbb{K} = \mathbb{R} \\ \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2) & , \mathbb{K} = \mathbb{C} \end{cases}$$

It's easy to verify that $\langle x, y \rangle$ is an inner product. □

Definition 6.2 (Hilbert space)
A complete inner product space is called a **Hilbert space**.

Definition 6.3 (orthogonal)
Let $\mathcal{X}$ be an inner product space, $x, y \in \mathcal{X}$. x and y are called **orthogonal** if $\langle x, y \rangle = 0$. Written as $x \perp y$. If M is an unempty subset of $\mathcal{X}$, and $\forall y \in M$, $x \perp y$, then we call x and M are orthogonal, written as $x \perp M$. Moreover,

$$M^{\perp} = \{x \in \mathcal{X} | x \perp M\}$$

is called the **orthogonal complement** of M.

Definition 6.4 (complete)
Let $\mathcal{X}$ be an inner product space, $S \subset \mathcal{X}$ is an orthogonal set. If $S^{\perp} = \{0\}$, then S is called **complete**.

Proposition 6.2
(1) If $x \perp y_n$ for all $n \in \mathbb{N}$, $y_n \to y$, then $x \perp y$.
(2) If $x \perp M$, then $x \perp span\{M\}$.
(3) $M^{\perp}$ is a closed vector subspace of $\mathcal{X}$.

Definition 6.5 (basis, Fourier coefficients)
Let $\mathcal{X}$ be an inner product space, $S \subset \mathcal{X}$ is an orthonormal set, where

$$S = \{e_{\lambda} | \lambda \in \Lambda\}$$

S is called a **basis** if for any $x \in \mathcal{X}$,

$$x = \sum_{\lambda \in \Lambda} \langle x, e_{\lambda} \rangle e_{\lambda}$$

where $\{\langle x, e_\lambda \rangle | \lambda \in \Lambda\}$ is called the **Fourier coefficients** with respect to the basis S.

Theorem 6.2 (Bessel inequality)
Let $\mathcal{X}$ be an inner product space, $S = \{e_\lambda | \lambda \in \Lambda\}$ be an orthonormal set, then $\forall x \in \mathcal{X}$,

$$\|x\|^2 \geq \sum_{\lambda \in \Lambda} |\langle x, e_\lambda \rangle|^2$$

Proof.
First consider any finite subset of A, say $1, 2, \cdots, n$. Since

$$\begin{aligned} 0 &\leq \|x - \sum_{i=1}^{n} \langle x, e_i \rangle e_i\|^2 \\ &= \langle x - \sum_{i=1}^{n} \langle x, e_i \rangle e_i, x - \sum_{i=1}^{n} \langle x, e_i \rangle e_i \rangle \\ &= \|x\|^2 - \sum_{i=1}^{n} |\langle x, e_i \rangle|^2 \end{aligned}$$

Thus we have

$$\sum_{i=1}^{n} |\langle x, e_i \rangle|^2 \leq \|x\|^2$$

It also implies that $\forall n \in \mathbb{N}$, there is at most finitely many $\lambda \in \Lambda$ such that

$$|\langle x, e_\lambda \rangle|^2 > \frac{1}{n}$$

Otherwise, for given $M = \|x\|^2$, we can find $(n+1)M$ λ's, leading a contradiction.
Hence there is at most countably many $\lambda \in \Lambda$ such that

$$|\langle x, e_\lambda \rangle|^2 > 0$$

Let Λ_n denote the finite subset of Λ whose cardinality is n, then

$$\sum_{\lambda \in \Lambda} \langle x, e_\lambda \rangle e_\lambda = \lim_{n \to \infty} \sum_{\lambda \in \Lambda_n} \langle x, e_\lambda \rangle e_\lambda \leq \|x\|^2 \quad \square$$

Corollary 6.1
Let $\mathcal{X}$ be a Hilbert space, $\{e_\lambda | \lambda \in \Lambda\}$ is a orthonormal set in $\mathcal{X}$. Then $\forall x \in \mathcal{X}$, we have
(1)

$$\sum_{\lambda \in \Lambda} \langle x, e_\lambda \rangle e_\lambda \in \mathcal{X}$$

(2)

$$\|x - \sum_{\lambda \in \Lambda} \langle x, e_\lambda \rangle e_\lambda\|^2 = \|x\|^2 - \sum_{\lambda \in \Lambda} |\langle x, e_\lambda \rangle|^2$$

Proof.
Suppose the countably many $\lambda \in \Lambda$ such that $\langle x, e_\lambda \rangle \neq 0$ are $1, 2, \cdots, n, \cdots$, then

$$\sum_{\lambda \in \Lambda} \langle x, e_\lambda \rangle e_\lambda = \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n$$

Then by Bessel inequality, $\sum_{n=1}^{\infty} |\langle x, e_n \rangle|^2$ converges, thus for any $p \in \mathbb{N}$

$$\|\sum_{n=m}^{m+p} \langle x, e_n \rangle e_n\|^2 = \sum_{n=m}^{m+p} |\langle x, e_n \rangle|^2 \to 0 \text{ as } m \to \infty$$

Hence if we define $x_m = \sum_{n=1}^{m}\langle x, e_n\rangle e_n$, then $\{x_n\}$ is a Cauchy sequence. Since $\mathcal{X}$ is complete, we have

$$\sum_{\lambda\in\Lambda}\langle x, e_\lambda\rangle e_\lambda = \sum_{n=1}^{\infty}\langle x, e_n\rangle e_n = \lim_{n\to\infty} x_n \in \mathcal{X}$$

Moreover, note that

$$\begin{aligned}\langle x - \sum_{n=1}^{\infty}\langle x, e_n\rangle e_n, \sum_{n=1}^{\infty}\langle x, e_n\rangle e_n\rangle &= \langle\sum_{n=1}^{\infty}\langle x, e_n\rangle\langle x, e_n\rangle - \langle\sum_{n=1}^{\infty}\langle x, e_n\rangle e_n, \sum_{n=1}^{\infty}\langle x, e_n\rangle e_n\rangle \\ &= \sum_{n=1}^{\infty}|\langle x, e_n\rangle|^2 - \sum_{n=1}^{\infty}\langle\langle x, e_n\rangle e_n, \langle x, e_n\rangle e_n\rangle \\ &= \sum_{n=1}^{\infty}|\langle x, e_n\rangle|^2 - \sum_{n=1}^{\infty}|\langle x, e_n\rangle|^2 \\ &= 0\end{aligned}$$

□

Theorem 6.3 (Parseval)
Let $\mathcal{X}$ be a Hilbert space, if $S = \{e_\lambda|\lambda\in\Lambda\}$ is a orthonormal set in $\mathcal{X}$, then the following are equivalent:
(1) S is a basis.
(2) S is complete.
(3) Parseval equality holds. i.e.

$$\|x\|^2 = \sum_{\lambda\in\Lambda}|\langle x, e_\lambda\rangle|^2, \forall x\in\mathcal{X}$$

Proof.
(1)⇒(2):
If S is a basis, then for any $x\in\mathcal{X}\setminus\{0\}$, we have

$$x = \sum_{\lambda\in\Lambda}\langle x, e_\lambda\rangle e_\lambda \neq 0$$

So there exists some $\lambda_0\in\Lambda$ such that $\langle x, e_{\lambda_0}\rangle\neq 0$ Therefore $S^\perp = \{0\}$, i.e. S is complete.
(2)⇒(3):

$$\|x\|^2 - \sum_{\lambda\in\Lambda}|\langle x, e_\lambda\rangle|^2 = \|x - \sum_{\lambda\in\Lambda}\langle x, e_\lambda\rangle e_\lambda\|^2 = 0$$

Otherwise, $0\neq x - \sum_{\lambda\in\Lambda}\langle x, e_\lambda\rangle e_\lambda\in S^\perp$, which is contradict to the completeness of S.
(3)⇒(1):

$$\|x - \sum_{\lambda\in\Lambda}\langle x, e_\lambda\rangle e_\lambda\|^2 = \|x\|^2 - \sum_{\lambda\in\Lambda}|\langle x, e_\lambda\rangle|^2 = 0$$

Therefore

$$x = \sum_{\lambda\in\Lambda}\langle x, e_\lambda\rangle e_\lambda$$

□

Example 6.1
In $L^2[0, 2\pi]$,

$$e_n(t) = \frac{1}{\sqrt{2\pi}}e^{i\pi t}, \quad n = 0, \pm1, \pm2, \cdots$$

is an orthonormal basis.

Example 6.2
In l^2,

$$e_n = (0, \cdots, 0, 1, 0, \cdots), n = 1, 2, \cdots \text{ } (n-1 \text{ 0's before 1})$$

is an orthonormal basis.

Definition 6.6 (isomorphic)
Let $(\mathcal{X}_1, \langle\cdot,\cdot\rangle_1)$ and $(\mathcal{X}_2, \langle\cdot,\cdot\rangle_2)$ be two inner product space, if there exists a linear isomorphism $T : \mathcal{X}_1 \to \mathcal{X}_2$ such that

$$\langle x, y\rangle_1 = \langle Tx, Ty\rangle_2, \forall x, y \in \mathcal{X}_1$$

Then we say that $\mathcal{X}_1$ and $\mathcal{X}_2$ are **isomorphic**.

Theorem 6.4
Let $\mathcal{X}$ be a Hilbert space. X is separable if and only if it has a at most countbly many orthonormal basis S. Moreover, if the cardinality of S, say n, is finite, then $\mathcal{X}$ is isomorphic to $\mathbb{K}^n$. If $n = \infty$, then $\mathcal{X}$ is isomorphic to l^2.
Proof.
$\Rightarrow$:
Suppose X is separable, assume that $\{x_n\}$ is the countable dense subset of $\mathcal{X}$, then there must exists a linear independent subset $\{y_n\}_1^N (N < \infty$ or $N = \infty)$, such that

$$span\{y_n\}_1^N = span\{x_n\}$$

Then by applying Gram-Schmidt method on $\{y_n\}_1^N$ we can form a orthonormal set $\{e_n\}_1^N$.
Since

$$\overline{span\{e_n\}_1^N} = \overline{span\{y_n\}_1^N} = \mathcal{X}$$

For any $x \in \mathcal{X}$,

$$x = \sum_{n=1}^{N} c_n e_n, \quad c_n \in \mathbb{K}, n = 1, 2, \cdots, N, N < \infty \text{ or } N = \infty$$

Note that for each n,

$$\langle x, e_n\rangle = \langle \sum_{n=1}^{N} c_n e_n, e_n\rangle = c_n$$

Thus

$$x = \sum_{n=1}^{N} \langle x, e_n\rangle e_n$$

$\{e_n\}_1^N$ is an orthonormal basis.
$\Leftarrow$:
Suppose $\{e_n\}_1^N$ is the orthonormal basis of $\mathcal{X}$, then

$$\{x = \sum_{n=1}^{N} c_n e_n | Re\{c_n\}, Im\{c_n\} \in \mathbb{Q}\}$$

is a countable dense subset of $\mathcal{X}$, hence $\mathcal{X}$ is separable.
Moreover, given an orthonormal basis $\{e_n\}_1^N$ $(N < \infty$ or $N = \infty)$, define

$$T : x \mapsto \{\langle x, e_n\rangle\}_1^N, \forall x \in \mathcal{X}$$

Obviously T is a linear isomorphism and note that for all $x, y \in \mathcal{X}$

$$\begin{aligned}\langle x, y\rangle &= \langle \sum_{i=1}^{N} \langle x, e_i\rangle e_i, \sum_{j=1}^{N} \langle y, e_j\rangle e_j\rangle \\ &= \sum_{i=1}^{N} \langle x, e_i\rangle \overline{\langle y, e_i\rangle} \\ &= \langle Tx, Ty\rangle\end{aligned}$$

Hence $\mathcal{X}$ is isomorphic to $\mathbb{K}^n$(when $N < \infty$) or l^2(when $N = \infty$). $\square$

Theorem 6.5
Let C be a closed convex subset of Hilbert space $\mathcal{X}$, then there exists a unique $x_0 \in C$ such that

$$\|x_0\| = \inf_{x \in C} \|x\|$$

Proof.
If $0 \in C$, then $x_0 = 0$.
If $0 \notin C$, let $d = \inf_{x \in C} \|x\|$.
Given $n \in \mathbb{N}$, there exists $x_n \in C$ such that

$$\|x_n\| < d + \frac{1}{n}$$

$\{x_n\}$ is a Cauchy sequence since

$$\begin{aligned}\|x_m - x_n\|^2 &= 2(\|x_m\| + \|x_n\|)^2 - 4\|\frac{x_m + x_n}{2}\|^2 \\ &\leq 2\left[\left(d + \frac{1}{m}\right)^2 + \left(d + \frac{1}{n}\right)^2\right] - 4d^2 \to 0 \text{ as } n, m \to \infty\end{aligned}$$

Hence it converges to some point $x_0 \in \mathcal{X}$. Since C is closed, $x_0 \in C$ and we have

$$\|x_0\| = d$$

Assume that $\hat{x_0}$ is another point that satisfies $\|\hat{x_0}\| = d$, then

$$\begin{aligned}\|x_0 - \hat{x_0}\|^2 &= 2(\|x_0\|^2 + \|\hat{x_0}\|^2) - 4\|\frac{x_0 + \hat{x_0}}{2}\|^2 \\ &\leq 4d^2 - 4d^2 = 0\end{aligned}$$

Therefore $x_0 = \hat{x_0}$. □

Corollary 6.2
Let C be a closed convex subset of Hilbert space $\mathcal{X}$, then for any $y \in \mathcal{X}$, there exists a unique $x_0 \in C$ such that

$$\|y - x_0\| = \inf_{x \in C} \|x - y\|$$

Theorem 6.6
Let C be a closed convex subset of an inner product space $\mathcal{X}$, $\forall y \in \mathcal{X}$, x_0 is the best approximation of y in C if and only if

$$Re\{\langle y - x_0, x_0 - x\rangle\} \geq 0, \forall x \in C$$

Proof.
$\forall x \in C$, consider a function on $t \in [0, 1]$

$$\phi_x(t) = \|y - tx - (1-t)x_0\|^2$$

Note that, x_0 is the best approximation of y in C if and only if

$$\phi_x(t) \geq \phi_x(0), \forall x \in C, \forall t \in [0, 1]$$

Since

$$\begin{aligned}\phi_x(t) &= \|(y - x_0) + t(x_0 - x)\|^2 \\ &= \|y - x_0\|^2 + 2tRe\{\langle y - x_0, x_0 - x\rangle\} + t^2\|x_0 - x\|^2 \\ \phi_x'(0) &= 2Re\{\langle y - x_0, x_0 - x\rangle\}\end{aligned}$$

Thus

$$Re\{\langle y - x_0, x_0 - x\rangle\} \geq 0 \Leftrightarrow \phi_x'(0) \geq 0$$

Also,

$$\phi_x(t) - \phi_x(0) = \phi_x'(0)t + \|x_0 - x\|^2 t^2$$

Therefore,

$$Re\{\langle y - x_0, x_0 - x\rangle\} \geq 0 \Leftrightarrow \phi_x(t) \geq \phi_x(0) \quad \square$$

Corollary 6.3
Let M be a closed linear submanifold of a Hilbert space $\mathcal{X}$. Then $\forall x \in \mathcal{X}$, y is the best approximation of x in M if and only if

$$x - y \perp M - \{y\}$$

Corollary 6.4 (orthogonal decomposition)
Let M be a closed subspace of a Hilbert space $\mathcal{X}$, then for $\forall x \in \mathcal{X}$, there exists a unique orthogonal decomposition

$$x = y + z, \quad y \in M, z \in M^\perp$$