Notes on Real Analysis

Kai-Jyun Wang*

Fall 2024

Foreword

I took Real Analysis in the fall of 2024 with Professor Tien. This is my note on the course. I tried to include all the proofs and details that has or has not been covered in the class, in order to make this note as self-contained as possible. Some of the proofs might be taken from somewhere and some might be wrong. The following topics are covered in the lecture: measure theory, Lebesgue integration, Banach space, Hilbert space, and approximation theory.

Some funny things happened in the class. The professor had taught so fast that we had already reached the Banach space before our first midterm. Every student was wondering if the professor forgot that this is actually a one-year course. Time comes to the eleventh week, the professor walked into the class and said, “Few days ago, someone told me that we actually have two semesters for real analysis, and I didn’t know that before!” It turns out that our concern was right. The professor then said, “But that is also a good thing, because we can learn more advanced topics in the second semester, like the harmonic analysis, the Fourier analysis. . . ”

*National Taiwan University, Department of Economics.

Contents

1. Measure Theory and Integration

1.1. Lebesgue Measure

Definition 1.1
The ***length*** *of an open interval* $(a, b) = I$ *is* $b - a$ *in the extended sense, denoted by* $\ell(I)$.

Remark
We define $(a, a) = \varnothing$.

Definition 1.2
The ***Lebesgue outer measure*** *(or in brief,* ***outer measure****) of a set* $E \subset \mathbb{R}$ *is*

$$\mu^*(E) = \inf \left\{ \sum_n \ell(I_n) \;\middle|\; I_n \text{ are countable open intervals covering } E \right\}.$$

Proposition 1.3

(a) *Countable sets are of outer measure zero.*

(b) *If* $A \subset B$, *then* $\mu^*(A) \leq \mu^*(B)$.

(c) *For* $x \in \mathbb{R}$ *and* $A \subset \mathbb{R}$, $\mu^*(A + x) = \mu^*(A)$.

(d) *For countable* $A_n \subset \mathbb{R}$, $\mu^*(\bigcup_n A_n) \leq \sum_n \mu^*(A_n)$.

Proof. For (a), let x_n denumerate a countable set A. Then consider

$$I_n = (x_n - 2^{-n}\epsilon, x_n + 2^{-n}\epsilon)$$

for $n \in \mathbb{N}$. Then $A \subset \bigcup_n I_n$ and $\mu^*(A) \leq \sum_n 2 \cdot 2^{-n}\epsilon = 2\epsilon$. Since ϵ is arbitrary, $\mu^*(A) = 0$.

For (b), note that any cover of B must cover A. The result follows.

For (c), note that the translations of open intervals preserve their lengths.

For (d), let $\left\{I_j^n\right\}$ cover A_n for each n such that $\sum_j \ell(I_j^n) < \mu^*(A_n) + 2^{-n}\epsilon$. Then we have that $\bigcup_n \bigcup_j I_j^n$ covers $\bigcup_n A_n$ and

$$\sum_n \sum_j \ell(I_j^n) < \sum_n \mu^*(A_n) + 2^{-n}\epsilon = \epsilon + \sum_n \mu^*(A_n).$$

Since ϵ is arbitrary, it follows that $\mu^*(\bigcup_n A_n) \leq \sum_n \mu^*(A_n)$. ■

Definition 1.4
A family of sets $\mathcal{M}$ *is called a* σ***-algebra*** *if*

(a) $\varnothing \in \mathcal{M}$.

(b) $A \in \mathcal{M}$ *implies* $A^c \in \mathcal{M}$.

(c) *For countably many* $A_n \in \mathcal{M}$ *we have* $\bigcup_n A_n \in \mathcal{M}$.

*The space $(X, \mathcal{M})$ is called a **measurable space** and the sets in $\mathcal{M}$ are called **measurable sets**.*

Proposition 1.5
$\mathcal{M}$ is a σ-algebra if and only if the following hold:

(a) $X \in \mathcal{M}$.

(b) $A, B \in \mathcal{M}$ *implies* $A \bigcap B, A \bigcup B, A - B \in \mathcal{M}$.

(c) For countably many $A_n \in \mathcal{M}$ *we have* $\bigcap_n A_n \in \mathcal{M}$.

Proof. Omitted. ■

Proposition 1.6
Let $\mathcal{F}$ be a family of sets in X. Then there exists a unique smallest σ-algebra containing $\mathcal{F}$.

Proof. Let $\mathcal{M}$ be the intersection of all σ-algebras containing $\mathcal{F}$. Since $\mathcal{P}(X)$ must be such a σ-algebra, $\mathcal{M}$ is non-empty. Now we verify that $\mathcal{M}$ is a σ-algebra. First, $\varnothing \in \mathcal{M}$ since $\varnothing$ is in every σ-algebra. Second, if $A \in \mathcal{F}$ then A must belong to every σ-algebra containing $\mathcal{F}$ and so does A^c. Hence $A^c \in \mathcal{M}$. The closure under countable unions follows from a similar argument. We conclude that $\mathcal{M}$ is the desired σ-algebra. ■

Definition 1.7
For a family of sets $\mathcal{F}$, we denote the smallest σ-algebra containing $\mathcal{F}$ by $\sigma(\mathcal{F})$.

Definition 1.8
*Let $\mathcal{T}$ be the family of all open sets. The **Borel σ-algebra** is defined as $\mathcal{B} = \sigma(\mathcal{T})$. The sets in $\mathcal{B}$ are called **Borel sets**.*

Definition 1.9
*A set E is called **Lebesgue measurable** if for $\epsilon > 0$, there exists an open set V such that $E \subset V$ and $\mu^*(V - E) \leq \epsilon$.*

Remark
The Lebesgue measurable sets form a σ-algebra.

Remark
The Borel sets are Lebesgue measurable.

Remark
Not all subsets in $\mathbb{R}$ are Lebesgue measurable. Consider the Vitali set. For a Lebesgue measurable set that is not Borel, consider the preimage of a Vitali set of Cantor-Lebesgue function.

Definition 1.10
*A function $f : (X, \mathcal{M}) \to (\mathbb{R}, \mathcal{B})$ is called **$\mathcal{M}$-measurable** if $f^{-1}(B) \in \mathcal{M}$ for all $B \in \mathcal{B}$.*

Proposition 1.11
Let $f : X \to Y$ and A be an index set. Then

(a) $f^{-1}(B^c) = f^{-1}(B)^c$.

(b) $f^{-1}(\bigcup_{a\in A} B_a) = \bigcup_{a\in A} f^{-1}(B_a)$.

(c) $f^{-1}(\bigcap_{a\in A} B_a) = \bigcap_{a\in A} f^{-1}(B_a)$

Proof. Omitted. ■

Proposition 1.12
$f : (X, \mathcal{M}) \to (\mathbb{R}, \mathcal{B})$ is $\mathcal{M}$-measurable if $f^{-1}((a, \infty)) \in \mathcal{M}$.

Proof. Observe that $\{A \subset \mathbb{R} \mid f^{-1}(A) \in \mathcal{F}\}$ is a σ-algebra. By assumption, $[a, b]$, $(a, b]$, $[a, b)$ and (a, b) are in this σ-algebra for $a, b \in \overline{\mathbb{R}}$. ■

Proposition 1.13
f_n are measurable. Then $\sup_n f_n$, $\inf_n f_n$, $\limsup f_n$ and $\liminf f_n$ are measurable.

Proof. Note that $\{\sup_n f_n > a\} = \bigcup_n \{f_n > a\}$ and $\{\inf_n f_n < a\} = \bigcup_n \{f_n < a\}$ are measurable. $\limsup_n f_n = \inf_k \sup_{n\geq k} f_n$ and $\liminf_n f_n = \sup_k \inf_{n\geq k} f_n$ are measurable as well. ■

Remark
$\lim_n f_n = \limsup_n f_n = \liminf_n f_n$ *is measurable.*

Definition 1.14
*Let $(X, \mathcal{M})$ be a measurable space. A **measure** on X is a function $\mu : \mathcal{M} \to [0, \infty]$ satisfying*

(a) $\mu(\varnothing) = 0$.

(b) $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$ *for disjoint A_n.*

*The triple $(X, \mathcal{M}, \mu)$ is called a **measure space**.*

Proposition 1.15
Let $(X, \mathcal{M}, \mu)$ be a measure space and $A, B \in \mathcal{M}$. Then

(a) *$A \subset B$ implies $\mu(A) \leq \mu(B)$.*

(b) *$\mu(A - B) = \mu(A) - \mu(B)$ if $B \subset A$ and $\mu(B) < \infty$.*

Proof. Omitted. ■

Proposition 1.16
Let $(X, \mathcal{M}, \mu)$ be a measure space and E_n be a sequence of measurable sets. Then

(a) *If $E_n \nearrow E$, then $\mu(E_n) \to \mu(E)$ as $n \to \infty$.*

(b) *If $E_n \searrow E$ and $\mu(E_1) < \infty$, then $\mu(E_n) \to \mu(E)$ as $n \to \infty$.*

Proof. Suppose $\mu(E_n) < \infty$ for all n. Consider $S_n = E_n - E_{n-1}$ with $E_0 = \varnothing$. Then S_n are disjoint and $\bigcup_n S_n = E$. Then

$$\mu(E) = \mu(\bigcup_n S_n) = \sum_n \mu(S_n) = \sum_n \mu(E_n) - \mu(E_{n-1}) = \lim_n \mu(E_n).$$

If $\mu(E_n) = \infty$ for some n, then $\mu(E) = \infty$ and the result follows.

For the second part, note that $E_1 - E_n \nearrow E_1 - E$. Then

$$\mu(E_1) - \mu(E_n) = \mu(E_1 - E_n) \to \mu(E_1 - E) = \mu(E_1) - \mu(E).$$

Rearranging gives the desired result. ■

Theorem 1.17 (Egorov)
Let E be a measurable set with $\mu(E) < \infty$ and $f_n : E \to \mathbb{R}$ are measurable functions. If $f_n \to f$ a.e. on E, then for all $\epsilon > 0$, there exists a closed set $A_\epsilon \subset E$ such that $\mu(E - A_\epsilon) < \epsilon$ and $f_n \to f$ uniformly on A_ϵ.

Proof. Consider the case where $f_n \to f$ everywhere on E since $\{x \in E \mid f_n(x) \not\to f(x)\}$ is of measure zero. For each $n, k \in \mathbb{N}$, let $E^n_k = \left\{x \in E \,\middle|\, \left|f_j(x) - f(x)\right| < 1/n \text{ for all } j > k\right\}$. Then fix n and note that $E^n_k \nearrow E$ as $k \to \infty$. By proposition 1.16, there exists k_n such that $\mu(E - E^n_{k_n}) < 2^{-n}$. Then we have $\left|f_j(x) - f(x)\right| < 1/n$ for every $j > k_n$ and $x \in E^n_{k_n}$. Choose N such that $\sum_{n \geq N} 2^{-n} < \epsilon/2$ and let $\hat{A}_\epsilon = \bigcap_{n \geq N} E^n_{k_n}$. Then $\mu(E - \hat{A}_\epsilon) \leq \sum_{n \geq N} \mu(E - E^n_{k_n}) < \epsilon/2$. Also, for any $\delta > 0$, we may pick $n > N$ with $1/n < \delta$ and for $x \in \hat{A}_\epsilon$, $\left|f_j(x) - f(x)\right| < \delta$ whenever $j > k_n$. Hence $f_n \to f$ uniformly on $\hat{A}_\epsilon$. We may further find a closed $A_\epsilon \subset \hat{A}_\epsilon$ such that $\mu(\hat{A}_\epsilon - A_\epsilon) < \epsilon/2$. Then A_ϵ is the desired set. ■

Definition 1.18
*A sequence of measurable functions f_n is said to **converge almost uniformly** to a function f if for every $\epsilon > 0$, there exists a measurable set E_ϵ such that $\mu(E_\epsilon^c) < \epsilon$ and $f_n \to f$ uniformly on E_ϵ.*

Remark
The Egorov theorem states that if the space if of finite measure, then converging almost everywhere implies converging almost uniformly.

Definition 1.19
*A function $s : X \to Y$ is called **simple** if it only takes finitely many values.*

Lemma 1.20
$f : E \to [0, \infty]$ is measurable. Then there exists a sequence of simple functions $s_n \nearrow f$; furthermore, if f is bounded, then $s_n \to f$ uniformly.

Proof. Consider $s_n = \sum_{k=0}^{n2^n - 1} k2^{-n} \chi_{f^{-1}([k2^{-n}, (k+1)2^{-n}))} + n\chi_{f^{-1}([n,\infty])}$. Then s_n are simple and $s_n \nearrow f$. If f is bounded, then $f^{-1}([n, \infty]) = \varnothing$ for some n large enough and $s_n \to f$ uniformly. ■

Theorem 1.21 (Lusin)
Let $E \subset \mathbb{R}$ be a set of finite measure and $f : E \to \mathbb{R}$ be a measurable, finite-valued function. Then for all $\epsilon > 0$, there exists a closed set $F_\epsilon \subset E$ such that $\mu(E - F_\epsilon) < \epsilon$ and $f|_{F_\epsilon}$ is continuous.

Proof. First we may partition E into $E = \bigcup_{i\in\mathbb{N}} E_i$ where $E_i = E \bigcap [-i,i]$. We first prove the result for simple functions. Let $f = \sum_{j=1}^{N} c_j \chi_{A_j}$ be a simple function with the stated properties. Then for each j, we may find a closed set $F_j \subset A_j$ such that $\mu(A_j - F_j) < \epsilon/N$. Now since E_i are bounded, $F_j \bigcap E_i$ are compact and hence f being constant on each $F_j \bigcap E_i$ is continuous. Note that $F_\epsilon = \bigcup_{i,j=1}^{N} F_j \bigcap E_i$ satisfies the desired properties. Next, for a general measurable function f, we may find a sequence of simple functions $s_n \nearrow f$ by lemma 1.20. Now by Egorov's theorem, we may find a closed set $F_\epsilon \subset E$ such that $\mu(E - F_\epsilon) < \epsilon$ and $s_n \to f$ uniformly on F_ϵ. Since s_n are continuous on F_ϵ, f is continuous on F_ϵ. ■

Remark
By Tietze's extension theorem, f can be extended to a continuous function on all of $\mathbb{R}$.

Proposition 1.22
E is Lebesgue measurable if and only if $\mu(E\triangle B) = 0$ for some Borel set B.

Proof. Suppose E is Lebesgue measurable. Then for each n, there exists an open set V_n such that $E \subset V_n$ and $\mu(V_n - E) < 1/n$. Let $B = \bigcap_n V_n$. Then B is a Borel set and $\mu(E\triangle B) = 0$. Conversely, if $\mu(E\triangle B) = 0$ for some Borel set B, since B is measurable, there exists an open $V \supset B$ such that $\mu(V - B) < \epsilon$. Then $B = (E \bigcap B) \bigcup (B - E)$ and since the later set has outer measure zero, $E \bigcap B$ is measurable. And since $E - B$ is outer measure zero, $E \bigcap B = E$ is measurable. ■

Proposition 1.23
If f is Lebesgue measurable, then there exists a Borel measurable function g such that $f = g$ a.e.

Proof. Let $s_k \nearrow f$ be a sequence of simple functions with $s_k = \sum_{i=1}^{n_k} c_i \chi_{E_i}$ where E_i are measurable. Then for each E_i we may find a Borel set $B_i \subset E_i$ such that $\mu(E_i - B_i) = 0$ by the previous proposition. Then $t_k = \sum_{i=1}^{n_k} c_i \chi_{B_i}$ is a Borel measurable function. Let $g = \lim_{k\to\infty} t_k$. Then g is Borel measurable and $f = g$ a.e. since $\mu(E_i - B_i) = 0$ for countably many i. ■

1.2. Lebesgue Integration

Definition 1.24
*For a simple function $s = \sum_{i=1}^{n} c_i \chi_{E_i}$, its **Lebesgue integral** is defined as*

$$\int s\,d\mu = \sum_{i=1}^{n} c_i \mu(E_i).$$

Definition 1.25
*For a non-negative measurable function f, its **Lebesgue integral** is defined as*

$$\int f\,d\mu = \sup\left\{\int s\,d\mu \,\middle|\, s \text{ is simple and } 0 \le s \le f\right\}.$$

Definition 1.26
For a measurable function $f : X \to [-\infty, \infty]$, its ***Lebesgue integral*** *is defined as*

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu,$$

where $f^+ = \max\{f, 0\}$ and $f^- = \max\{-f, 0\}$ provided that

$$\int |f| \, d\mu = \int f^+ d\mu + \int f^- d\mu < \infty.$$

In such a case, we say that f is ***integrable****.*

Proposition 1.27
For f, g integrable and $c \in \mathbb{R}$,

(a) $\int cf + g d\mu = c \int f d\mu + \int g d\mu$.

(b) *If $f \le g$ a.e., then* $\int f d\mu \le \int g d\mu$.

Proof. Omitted. ■

Theorem 1.28 (Lebesgue Monotone Convergence Theorem)
Let $f_n : X \to [0, \infty]$ be a sequence of measurable functions with $f_n \nearrow f$ a.e. Then

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

Proof. By the monotonicity we have

$$\int f_n d\mu \le \int f d\mu$$

for all n and hence

$$\lim_{n\to\infty} \int f_n d\mu \le \int f d\mu.$$

To obtain the reverse inequality, note that for any $c \in (0, 1)$, there exists N such that $f_n \ge cf$ a.e. for all $n \ge N$. Then

$$\int f_n d\mu \ge c \int f d\mu$$

for all $n \ge N$. Letting $n \to \infty$,

$$\lim_{n\to\infty} \int f_n d\mu \ge c \int f d\mu.$$

Taking $c \to 1^-$ then

$$\lim_{n\to\infty} \int f_n d\mu \ge \int f d\mu \implies \lim_{n\to\infty} \int f_n d\mu = \int f d\mu.$$

■

Remark
As a consequence,

$$\int \sum_n f_n d\mu = \sum_n \int f_n d\mu.$$

Theorem 1.29 (Bounded Covergence Theorem)
Suppose $\mu(X) < \infty$. Let $f_n : X \to \mathbb{R}_+$ be a sequence measurable functions such that $f_n \leq M$ a.e. for some $M \in \mathbb{R}$. If $f_n \to f$ a.e., then f is integrable and

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

Proof. For any $\epsilon > 0$, by Egorov's theorem, there exists $F \subset X$ such that $\mu(X - F) < \epsilon$ and $f_n \to f$ uniformly on F. Then there exists N such that $|f_n - f| < \epsilon$ on F for all $n \geq N$. We have

$$\begin{aligned}\left|\int f_n d\mu - \int f d\mu\right| &\leq \int_X |f_n - f|\, d\mu \\ &= \int_F |f_n - f|\, d\mu + \int_{X-F} |f_n - f|\, d\mu \\ &\leq \epsilon\mu(F) + 2M\mu(X - F) = \epsilon(\mu(F) + 2M\epsilon).\end{aligned}$$

Since $\mu(X) < \infty$ and ϵ is arbitrary, we may conclude that

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

■

Lemma 1.30 (Fatou)
$f_n : X \to [0, \infty]$ are measurable. Then

$$\int \liminf_n f_n d\mu \leq \liminf_n \int f_n d\mu.$$

Proof. Let $g_n = \inf_{k\geq n} f_k$. Then $g_n \nearrow g = \liminf_n f_n$. By LMCT,

$$\int g_n d\mu \to \int g d\mu = \int \liminf_n f_n d\mu.$$

Note that $f_n \geq g_n$ and thus $\int f_n d\mu \geq \int g_n d\mu$. Hence

$$\liminf_n \int f_n d\mu \geq \liminf_n \int g_n d\mu = \int g d\mu = \int \liminf_n f_n d\mu.$$

■

Theorem 1.31 (Lebesgue Dominated Convergence Theorem)
Let $f_n : X \to [-\infty, \infty]$ be a sequence of measurable functions such that $f_n \to f$ a.e. and there

exists an integrable function g such that $|f_n| \leq g$ a.e. for all n. Then

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

Proof. Since $|f_n| \leq g$ a.e., $|f| \leq g$ a.e. Now $|f_n - f| \leq |f_n| + |f| \leq 2g$ a.e. Let $h_n = 2g - |f_n - f| \geq 0$ a.e. By Fatou's lemma,

$$\begin{aligned}\int 2g d\mu = \int \liminf_n h_n d\mu \leq \liminf_n \int h_n d\mu = \liminf_n \int 2g - |f_n - f| \, d\mu \\ = \int 2g d\mu - \limsup_n \int |f_n - f| \, d\mu.\end{aligned}$$

It follows that

$$0 \leq \liminf_n \int |f_n - f| \, d\mu \leq \limsup_n \int |f_n - f| \, d\mu \leq 0.$$

Hence

$$\lim_{n\to\infty} \int |f_n - f| \, d\mu = 0.$$

By the triangle inequality,

$$\left| \int f d\mu - \int f_n d\mu \right| \leq \int |f - f_n| \, d\mu \to 0.$$

So

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

■

Remark

If $\operatorname{supp}(f)$ has finite measure and f is bounded, then

$$\int f = \inf_{s \geq f} \int s d\mu,$$

where s is simple.

Definition 1.32

*$\mathcal{L}^1 = \{f : X \to \mathbb{R} \mid f$ is integrable$\}$ with the norm $\|f\|_{\mathcal{L}^1} = \int |f| \, d\mu$ is called the $\mathcal{L}^1$ **space**.*

Remark

The elements in $\mathcal{L}^1$ are in fact equivalence classes of functions that are equal a.e.

Proposition 1.33

Let $f \in \mathcal{L}^1$ be a nonnegative function. Then for every $\epsilon > 0$, there is some $\delta > 0$ such that for any measurable E with $\mu(E) \leq \delta$,

$$\int_E f d\mu \leq \epsilon.$$

Proof. Let $E_n = \{x \in X \mid f(x) > n\}$. Then by Lebesgue dominated convergence theorem, since $f\chi_{E_n} \leq f$,

$$\int_{E_n} f d\mu \to 0.$$

For any $\epsilon > 0$, there exists n such that

$$\int_{E_n} f d\mu \leq \frac{\epsilon}{2}.$$

Pick $\delta \leq \epsilon/(2n)$. Then for any measurable E with $\mu(E) \leq \delta$,

$$\int_E f d\mu = \int_{E\cap E_n} f d\mu + \int_{E\cap E_n^c} f d\mu \leq \int_{E_n} f d\mu + n\mu(E) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

since $f \leq n$ on E_n^c. This completes the proof. ■

Theorem 1.34 (Lebesgue-Vitali)
$f : X \to \mathbb{R}$ is Riemann integrable if and only if the discontinuity set of f has Lebesgue measure zero. Furthermore, if f is Riemann integrable, then the Riemann integral and the Lebesgue integral agrees.

Proof. Define the oscillation of f at x as

$$\operatorname{osc}(f, x) = \inf_{U:x\in U} \operatorname{diam}(f(U)),$$

where U is open.

We first claim that f is continuous at x if and only if $\operatorname{osc}(f, x) = 0$. Indeed, if f is continuous at x, then $\forall\, \epsilon > 0$, $\exists\, \delta > 0$ such that $|f(x) - f(y)| < \epsilon$ for all $y \in B_\delta(x)$. Then $\operatorname{diam}(f(B_\delta(x))) \leq 2\epsilon$. Since ϵ is arbitrary, $\operatorname{osc}(f, x) = 0$. Conversely, if $\operatorname{osc}(f, x) = 0$, then $\forall \epsilon > 0$, $\exists$ open U containing x such that $\operatorname{diam}(f(U)) < \epsilon$. This implies that $|f(x) - f(y)| < \epsilon$ for all $y \in U$ and hence f is continuous at x.

Next, let D_ϵ collect all points x such that $\operatorname{osc}(f, x) \geq \epsilon > 0$. We claim that D_ϵ is closed. For any convergent sequence $x_k \in D_\epsilon$, let $x_k \to x$. For any open U containing x, $\exists\, N$ such that $x_k \in U$ for all $k \geq N$. Then $\exists$ an open neighborhood of x_N, U', such that $U' \subset U$ and $\operatorname{diam}(f(U')) \geq \epsilon$. Hence $\operatorname{osc}(f, x) \geq \epsilon$ and $x \in D_\epsilon$, showing that D_ϵ is closed. Observe that $D = \bigcup_{n=1}^\infty D_{1/n}$.

Now suppose that f is Riemann integrable. Then for any $\epsilon > 0$, $\exists\, \mathcal{P}$ such that $\mathrm{U}(f, \mathcal{P}) - \mathrm{L}(f, \mathcal{P}) < \frac{1}{n}$ and $\|\mathcal{P}\| < \frac{1}{n}$. Then

$$\begin{aligned}
&\sum_{\substack{Q\in\mathcal{P},\\ Q\bigcap D_{\frac{1}{n}}\neq\varnothing}} (\sup_Q f - \inf_Q f)\,|Q| + \sum_{\substack{Q\in\mathcal{P},\\ Q\bigcap D_{\frac{1}{n}}=\varnothing}} (\sup_Q f - \inf_Q f)\,|Q| \\
&= \sum_{Q\in\mathcal{P}} (\sup_Q f - \inf_Q f)\,|Q| = \mathrm{U}(f, \mathcal{P}) - \mathrm{L}(f, \mathcal{P}) < \epsilon.
\end{aligned}$$

Note that $\sup_Q f - \inf_Q f = \operatorname{diam}(f(Q))$. This gives that $2M\mu^*(D_{\frac{1}{n}}) < \epsilon$ for every n. Since ϵ is arbitrary, we conclude that $\mu^*(D_{\frac{1}{n}}) = 0$ for each n. Thus D is an union of sets of measure zero and hence also has measure zero.

For the converse, suppose that $m(D) = 0$. Then D_ϵ also has measure zero. Let $\mathcal{P}$ be a partition on E with $\|\mathcal{P}\| < \delta$ for some $\delta > 0$, which will be determined later. Then

$$\begin{aligned}
\mathrm{U}(f,\mathcal{P}) - \mathrm{L}(f,\mathcal{P}) &= \sum_{Q\in\mathcal{P}} (\sup_Q f - \inf_Q f)\,|Q| \\
&= \sum_{\substack{Q\in\mathcal{P},\\ Q\bigcap D_\epsilon=\varnothing}} (\sup_Q f - \inf_Q f)\,|Q| + \sum_{\substack{Q\in\mathcal{P},\\ Q\bigcap D_\epsilon\neq\varnothing}} (\sup_Q f - \inf_Q f)\,|Q|
\end{aligned}$$

For the first term, $\sup_Q f - \inf_Q f < \epsilon$ for $\|\mathcal{P}\| < \delta_1$ for some $\delta_1 > 0$. And thus the first term is bounded by $\epsilon m(E)$. For the second term, $\sup_Q f - \inf_Q f < 2M$ and since D_ϵ has measure zero, $\exists Q_k$ cubic cover of D_ϵ such that $\sum_k |Q_k| < \epsilon$. Now if $\operatorname{diam}(Q) < \delta_2$ for some $\delta_2 > 0$, then those Q intersecting D_ϵ nonempty are subset of $\bigcup_k Q_k$. Thus the second term is bounded by $2M\epsilon$. Choosing $\delta = \min\{\delta_1, \delta_2\}$ yields that

$$\mathrm{U}(f,\mathcal{P}) - \mathrm{L}(f,\mathcal{P}) < \epsilon m(E) + 2M\epsilon$$

whenever $\|\mathcal{P}\| < \delta$. Since ϵ is arbitrary, f is Riemann integrable. ■

Proposition 1.35

(a) Step functions are dense in $\mathcal{L}^1$.

(b) Continuous functions with compact support are dense in $\mathcal{L}^1$.

Proof. Let $f \in \mathcal{L}^1$. By lemma 1.20, we already know that simple functions are dense in $\mathcal{L}^1$. It now remains to show that step functions can approximate simple functions. Since simple functions are linear combinations of finitely many characteristic functions, it suffices to show that characteristic functions can be approximated by step functions. Now for any measurable E, there is a family of almost disjoint cubes Q_i such that $\mu(E \triangle \cup_{i=1}^M Q_i) \le 2\epsilon$, and thus we may set the step function to be $\phi = \sum_{i=1}^M \chi_{Q_i}$, with $\|\chi_E - \phi\|_{\mathcal{L}^1} \le 2\epsilon$.

For the second part, let it now suffices to show that continuous functions with compact support can approximate characteristic functions of a rectangle, say $[a,b]$. Then set

$$g(x) = \begin{cases} 0 & x \le a-\epsilon, \\ \frac{x-a+\epsilon}{\epsilon} & a-\epsilon \le x \le a, \\ 1 & a \le x \le b, \\ 1 - \frac{x-b}{\epsilon} & b \le x \le b+\epsilon, \\ 0 & x \ge b+\epsilon. \end{cases}$$

Then g is continuous with compact support and $\left\|\chi_{[a,b]} - g\right\|_{\mathcal{L}^1} \le \epsilon/2 + \epsilon/2 = \epsilon$. ■

1.3. Differentiation

Definition 1.36
Let $f \in \mathcal{L}^1(\mathbb{R}^d)$. The ***Hardy-Littlewood maximal function*** *is defined as*

$$f^*(x) = \sup_{B:x\in B} \frac{1}{\mu(B)} \int_B |f(y)|\, dy,$$

where the supremum is taken over all balls containing x.

Proposition 1.37
f^* is measurable.

Proof. Let $E_\alpha = \{x \mid f^*(x) > \alpha\}$. We claim that it is an open set. Indeed, if $p \in E_\alpha$, there exists a ball B containing p such that

$$\frac{1}{\mu(B)} \int_B |f(y)|\, dy > \alpha.$$

Now any x close enough to p will be contained in B and hence in E_α. Thus E_α is open. Hence f^* is measurable. ■

Lemma 1.38
[Vitali Covering Lemma] Suppose $\{B_1, \dots, B_N\}$ is a finite collection of open balls in $\mathbb{R}^d$. Then there exists a disjoint subcollection $\left\{B_{i_1}, \dots, B_{i_k}\right\}$ such that

$$\mu\left(\bigcup_{j=1}^{N} B_j\right) \le 3^d \sum_{j=1}^{k} \mu(B_{i_j}).$$

Proof. First we make an observation that if B and B' are balls intersecting with, say, the radius of B is greater than the radius of B', then B' is contained in the ball $\tilde{B}$ that is concentric with B but with 3 times the radius.

The construction of the subcollection is proceeded as follows. First, pick a ball B_{i_1} with the largest radius. Then remove all balls intersecting with $\tilde{B}_{i_1}$, the ball concentric with B_{i_1} but with 3 times the radius. Among the remaining balls, we repeat the process and pick B_{i_2}. The process terminates when no more balls can be picked, after at most N steps and we obtain a disjoint subcollection of balls $\left\{B_{i_1}, \dots, B_{i_k}\right\}$.

Lastly, we verify the inequality. By the construction, we know that $\cup_{j=1}^{N} B_j \subset \cup_{j=1}^{k} \tilde{B}_{i_j}$ and thus

$$\mu\left(\bigcup_{j=1}^{N} B_j\right) \le \mu\left(\bigcup_{j=1}^{k} \tilde{B}_{i_j}\right) \le \sum_{j=1}^{k} \mu(\tilde{B}_{i_j}) = \sum_{j=1}^{k} 3^d \mu(B_{i_j}).$$

■

Theorem 1.39 (Weak-Type Inequality)

Let $f \in \mathcal{L}^1(\mathbb{R}^d)$. Then for all $\alpha > 0$,

$$\mu\Big(\big\{x \in \mathbb{R}^d \mid f^*(x) > \alpha\big\}\Big) \le \frac{A}{\alpha} \|f\|_{\mathcal{L}^1(\mathbb{R}^d)},$$

where $A = 3^d$.

Proof. Let $E_\alpha = \{x \mid f^*(x) > \alpha\}$. For each $x \in E_\alpha$ there exists a ball B_x containing x such that

$$\frac{1}{\mu(B_x)} \int_{B_x} |f(y)|\, dy > \alpha \quad \Rightarrow \quad \mu(B_x) < \frac{1}{\alpha} \int_{B_x} |f(y)|\, dy.$$

Now for any fixed compact $K \subset E_\alpha$, K is covered by $\bigcup_{x \in E_\alpha} B_x$, and hence there exists a finite subcover $\{B_1, \ldots, B_N\}$ of K. By the Vitali covering lemma, there exists a disjoint subcollection $\{B_{i_1}, \ldots, B_{i_k}\}$ with

$$\mu\left(\bigcup_{j=1}^N B_j\right) \le 3^d \sum_{j=1}^k \mu(B_{i_j}).$$

As a result,

$$\begin{aligned}\mu(K) \le \mu\left(\bigcup_{j=1}^N B_j\right) &\le 3^d \sum_{j=1}^k \mu(B_{i_j}) \le \frac{3^d}{\alpha} \sum_{j=1}^k \int_{B_{i_j}} |f(y)|\, dy \\ &\le \frac{3^d}{\alpha} \int_{\cup_{j=1}^k B_{i_j}} |f(y)|\, dy \le \frac{3^d}{\alpha} \int_{\mathbb{R}^d} |f(y)|\, dy.\end{aligned}$$

Since the inequality holds for all compact subset K of E_α, the proof is complete. ■

Remark

Note that $\{x \mid f^*(x) = \infty\} \subset \{x \mid f^*(x) > \alpha\}$ for every $\alpha > 0$. Taking $\alpha \to \infty$ yields

$$\mu(\{x \mid f^*(x) = \infty\}) = 0.$$

Hence $f^*(x) < \infty$ a.e.

Theorem 1.40 (Lebesgue Differentiation Theorem)
Let $f \in \mathcal{L}^1(\mathbb{R}^d)$. Then for almost every $x \in \mathbb{R}^d$,

$$\lim_{m(B) \to 0, x \in B} \frac{1}{m(B)} \int_B f(y) dy = f(x).$$

Proof. Since continuous functions are dense in $\mathcal{L}^1$, we may find a continuous g such that $\|f - g\|_{\mathcal{L}^1} < \epsilon$. For such g, by the continuity, there exists a ball such that $|g(y) - g(x)| < \epsilon$

for all $x, y \in B$. Thus

$$\begin{aligned}
\left|\frac{1}{m(B)}\int_B f(y)dy - f(x)\right| &= \left|\frac{1}{m(B)}\int_B (f(y)-g(y))dy + \frac{1}{m(B)}\int_B g(y)-g(x)dy + g(x)-f(x)\right| \\
&\leq \frac{1}{m(B)}\int_B |(f(y)-g(y))|\,dy + \frac{1}{m(B)}\int_B |g(y)-g(x)|\,dy + |g(x)-f(x)| \\
&\leq (f-g)^*(x) + \epsilon + |g(x)-f(x)|\,.
\end{aligned}$$

Since ϵ can be arbitrary small, we have

$$\left|\frac{1}{m(B)}\int_B f(y)dy - f(x)\right| \leq (f-g)^*(x) + |g(x)-f(x)|\,.$$

Now we let

$$E_\alpha = \left\{x \;\middle|\; \limsup_{m(B)\to 0, x\in B}\left|\frac{1}{m(B)}\int_B f(y)dy - f(x)\right| > 2\alpha\right\}.$$

We claim that E_α has measure zero. Set

$$F_\alpha = \{x \mid (f-g)^*(x) > \alpha\} \quad \text{and} \quad G_\alpha = \{x \mid |g(x)-f(x)| > \alpha\}\,.$$

Then we have $E_\alpha \subset F_\alpha \cup G_\alpha$. By the weak-type inequality and Tchebyshev's inequality,

$$\mu(F_\alpha) \leq \frac{A}{\alpha}\|f-g\|_{\mathcal{L}^1} < \frac{A}{\alpha}\epsilon \quad \text{and} \quad \mu(G_\alpha) \leq \frac{1}{\alpha}\|f-g\|_{\mathcal{L}^1} < \frac{1}{\alpha}\epsilon.$$

Thus $\mu(E_\alpha) \leq \mu(F_\alpha \cup G_\alpha) < \frac{A+1}{\alpha}\epsilon$. Since ϵ is arbitrary, we have $\mu(E_\alpha) = 0$ and the proof is complete. ■

Remark

For $f \in \mathcal{L}^1(\mathbb{R})$, and $F(x) = \int_{-\infty}^x f(y)dy$, we have $F'(x) = f(x)$ a.e. Indeed,

$$\begin{aligned}
\left|\frac{F(x+h)-F(x)}{h} - f(x)\right| &= \frac{1}{h}\left|\int_x^{x+h} f(y)-f(x)dy\right| \leq \frac{1}{h}\int_x^{x+h} |f(y)-f(x)|\,dy \\
&\leq \frac{1}{h}\int_{x-h}^{x+h} |f(y)-f(x)|\,dy \leq 2\frac{1}{2h}\int_{x-h}^{x+h} |f(y)-f(x)|\,dy \to 0
\end{aligned}$$

as $h \to 0$ a.e. x.

Remark

In fact, the requirement that $f \in \mathcal{L}^1$ can be relaxed to $f \in \mathcal{L}^1_{loc}$, which is defined as the set of all locally integrable functions, i.e., $f\chi_B \in \mathcal{L}^1$ for all finite balls B since the proof only requires B to be a ball near x.

1.4. Radon-Nikodym Theorem

Definition 1.41
Let $(X, \mathcal{A})$ be a measurable space. A ***signed measure*** *is a function $\mu : \mathcal{A} \to [-\infty, \infty]$ such that $\mu(\varnothing) = 0$ and for any countable disjoint collection $\{A_i\}_{i\in\mathbb{N}}$,*

$$\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mu(A_i).$$

Remark
The range of μ can only include one of $\pm\infty$.

Definition 1.42
Let $(X, \mathcal{A}, \mu)$ be a measure space. μ is called ***σ-finite*** *if X can be covered by countably many $A_n \in \mathcal{A}$ such that $\mu(A_n) < \infty$ for all n. In this case, we also call X* ***σ-finite****.*

Definition 1.43
Let ν, λ be two measures defined on a measurable space. ν is said to be ***absolutely continuous*** *with respect to λ if $\lambda(A) = 0$ implies that $\nu(A) = 0$ for all measurable A, denoted as $\nu \ll \lambda$.*

Example
Let

$$\nu(A) = \int_A f \, d\lambda$$

where $f \geq 0$ is measurable. Then $\lambda(A) = 0$ implies $\nu(A) = 0$. $\nu \ll \lambda$.

Definition 1.44
Let ν, λ be two measures defined on a measurable space. ν is said to be ***singular*** *with respect to λ if there exists a measurable set A such that $\lambda(A) = 0$ and $\nu(A^c) = 0$, denoted as $\nu \perp \lambda$.*

Example
Let λ be the Lebesgue measure on $[0, 1]$ and

$$\nu(A) = \sum_i c_i \delta_{q_i}(A), \quad \text{with} \quad \sum_i c_i < \infty, \quad \delta_{q_i}(A) = \mathbf{1}\{q_i \in A\},$$

where q_i enumerates the rationals in $[0, 1]$ and $\mathbf{1}$ is the indicator function. Then $\nu \perp \lambda$.

Definition 1.45
ν and λ are said to be ***equivalent*** *if $\nu \ll \lambda$ and $\lambda \ll \nu$.*

Definition 1.46
Let $(X, \mathcal{A}, \mu)$ be a measure space. A set $P \in \mathcal{A}$ is said to be ***positive*** *if $\mu(A) \geq 0$ for all measurable $A \subset P$; a set $N \in \mathcal{A}$ is said to be* ***negative*** *if $\mu(A) \leq 0$ for all measurable $A \subset N$.*

Theorem 1.47 (Hahn Decomposition)
Let μ be a signed measure on a measurable space $(X, \mathcal{A})$. Then X can be partitioned into a positive set P and a negative set N. Furthermore, if P', N' form another such partition, then $P \triangle P'$ and $N \triangle N'$ are measure zero.

Proof. We may consider the case where $\mu(A) \neq -\infty$ for all $A \in \mathcal{A}$. The other case is similar. We first claim that every measurable set A contains a postive set P such that $\mu(P) \geq \mu(A)$.

To prove the claim, we first show that for every $\epsilon > 0$, there exists $A_\epsilon \subset A$ such that $\mu(A_\epsilon) \geq \mu(A)$ and $B \subset A_\epsilon$ implies $\mu(B) > -\epsilon$. Otherwise, we can pick a sequence of set B_k inductively, such that $B_1 \subset A, \ldots, B_k \subset A - (B_1 \cup \cdots \cup B_{k-1}), \ldots$ with $\mu(B_k) \leq -\epsilon$. Put $B = \cup_k B_k$. Since B_k are disjoint, $\mu(B) = -\infty$. Also, $\mu(A - B) = \mu(A) - \mu(B) = \infty$, contradicting to the remark that μ cannot take both $\pm\infty$. Now choose $\epsilon_n \to 0$ and let $P = \cap_n A_{\epsilon_n}$. $A_{\epsilon_n} \searrow P$ and then $\mu(A_{\epsilon_n}) \to \mu(P)$ by proposition 1.16. Thus $\mu(P) \geq \mu(A)$.

Next, let $s = \sup\{\mu(A) \mid A \in \mathcal{A}\}$. There is a sequence P_n such that $\mu(P_n) \to s$. Note that $s \geq 0$ since $\varnothing \in \mathcal{A}$. By the claim, we may assume that P_n are positive. Putting $P = \cup_n P_n$, we have $\mu(P) = s$ and P is positive. Now let $N = X - P$. N is negative; otherwise if $E \subset N$ and $\mu(E) > 0$, then $\mu(P \cup E) = \mu(P) + \mu(E) > s$, which contradicts to the definition of s.

Finally, suppose P' and N' are another such partition. Then $P \cap N'$ and $N \cap P'$ are both negative and positive, implying that they are measure zero. $\mu(P \triangle P') = \mu(P \cap N') + \mu(N \cap P') = 0$. This furnishes the proof. ■

Corollary 1.48 (Hahn-Jordan Decomposition)
If ν is a signed measure on a measurable space $(X, \mathcal{A})$, then there exists a unique pair of positive measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$.

Proof. By the Hahn decomposition, X can be partitioned into a positive set P and a negative set N. Define $\nu^+(A) = \nu(A \cap P)$ and $\nu^-(A) = -\nu(A \cap N)$. Then ν^+ and ν^- are positive measures and $\nu = \nu^+ - \nu^-$. The uniqueness follows from the uniqueness of the Hahn decomposition. ■

Theorem 1.49 (Radon-Nikodym)
Let $(X, \mathcal{A})$ be a measurable space and ν, λ are σ-finite measures on $(X, \mathcal{A})$. If $\nu \ll \lambda$, then there exists an $\mathcal{A}$-measurable function $f : X \to [0, \infty)$ such that for every $A \in \mathcal{A}$,

$$\nu(A) = \int_A f \, d\lambda.$$

Furthermore, if f and f' are two such functions, then $f = f'$ a.e.

Proof. We first consider the case where ν and λ are finite. Let

$$F = \left\{ f : X \to [0, \infty] \;\middle|\; \int_A f \, d\lambda \leq \nu(A) \text{ for all } A \in \mathcal{A} \right\}.$$

$F \neq \varnothing$ since $f = 0$ is in F. Now let $f_1, f_2 \in F$ and $A \in \mathcal{A}$ and define

$$A_1 = \{x \in A \mid f_1(x) > f_2(x)\}, \quad A_2 = \{x \in A \mid f_1(x) \leq f_2(x)\}.$$

Then
$$\int_A \max\{f_1, f_2\}\, d\lambda = \int_{A_1} f_1 d\lambda + \int_{A_2} f_2 d\lambda \le \nu(A_1) + \nu(A_2) = \nu(A).$$
Thus $\max\{f_1, f_2\} \in F$. Next, for any sequence of functions $f_n \in F$ such that

$$\lim_{n\to\infty} \int_X f_n d\lambda = \sup_{f\in F} \int_X f d\lambda,$$

we may assume that $f_n \nearrow$ by replacing f_n with the maximum among $f_1, \dots, f_n$. Let g be the pointwise limit of f_n. By Lebesgue's monotone convergence theorem,

$$\int_A g d\lambda = \lim_{n\to\infty} \int_A f_n d\lambda \le \nu(A),$$

so $g \in F$. Also, by construction,

$$\int_X g d\lambda = \sup_{f\in F} \int_X f d\lambda.$$

Now define
$$\nu_0(A) = \nu(A) - \int_A g d\lambda.$$
Since $g \in F$, ν_0 is a nonnegative measure. To prove the equality, we need to show that $\nu_0(A) = 0$ for all $A \in \mathcal{A}$. Suppose $\nu_0 > 0$. Then there exists $\epsilon > 0$ such that $\nu_0(X) > \epsilon\lambda(X)$. By the Hahn decomposition theorem, we can find a positive set P such that $\nu_0(A) \ge \epsilon\lambda(A)$ for each $A \subset P$. Thus

$$\nu(A) = \int_A g d\lambda + \nu_0(A) \ge \int_A g d\lambda + \nu_0(P \cap A) \ge \int_A g d\lambda + \epsilon\lambda(P \cap A) = \int_A (g + \epsilon\chi_P) d\lambda.$$

Note that $\lambda(P) > 0$, for otherwise $\lambda(P) = 0$ and $\nu_0(P) \le \nu(P) = 0 \implies \nu(P) = 0$ by the absolute continuity and hence

$$\nu_0(X) - \epsilon\lambda(X) = (\nu_0 - \epsilon\lambda)(N) \le 0,$$

posing a contradiction. Meanwhile,

$$\int_X (g + \epsilon\chi_P) d\lambda \le \nu(X) < \infty \implies g + \epsilon\chi_P \in F,$$

and
$$\int_X (g + \epsilon\chi_P) d\lambda > \int_X g d\lambda = \sup_{f\in F} \int_X f d\lambda.$$
This violates the definition of the supremum. Thus $\nu_0 = 0$ and we obtain that

$$\nu(A) = \int_A g d\lambda.$$

Finally, if we define

$$f(x) = \begin{cases} g(x) & \text{if } g(x) < \infty, \\ 0 & \text{if } g(x) = \infty, \end{cases}$$

since g is λ-integrable, $f = g$ λ-a.e. and f is the desired function.

For the uniqueness, suppose f and f' are two such functions. Then

$$\nu(A) = \int_A f d\lambda = \int_A f' d\lambda \implies \int_A (f - f') d\lambda = 0$$

for every A. In particular, letting $A = \{x \in X \mid f(x) \leq f'(x)\}$ or $A = \{x \in X \mid f(x) \geq f'(x)\}$ gives

$$\int_X (f - f')^+ d\lambda = \int_X (f - f')^- d\lambda = 0.$$

Thus $f = f'$ λ-a.e.

For the general case where ν and λ are σ-finite, we can write $X = \cup_n X_n$ such that $\lambda(X_n) < \infty$ and X_n are disjoint. For each n we can find f_n such that

$$\nu(A) = \int_A f_n d\lambda.$$

for every $\mathcal{A}$-measurable $A \subset X_n$. Let $f = \sum_n f_n \chi_{X_n}$.

$$\int_A f d\lambda = \sum_n \int_{A \cap X_n} f_n d\lambda = \sum_n \nu(A \cap X_n) = \nu(A),$$

for every $A \in \mathcal{A}$. The uniqueness follows from the uniqueness of f_n. ■

Remark

The function f can be chosen in $\mathcal{L}^1(X, \lambda)$ if ν is finite.

Definition 1.50

*The function f in the Radon-Nikodym theorem is called the **Radon-Nikodym derivative** of ν with respect to λ, denoted as $f = \frac{d\nu}{d\lambda}$.*

Proposition 1.51

Let ν, μ and λ be σ-finite measures defined on measurable space $(X, \mathcal{A})$. If $\nu \ll \lambda$ and $\mu \ll \lambda$, then

(a) $\frac{d(\nu+\mu)}{d\lambda} = \frac{d\nu}{d\lambda} + \frac{d\mu}{d\lambda}$ *λ-a.e.*

(b) If $\nu \ll \mu \ll \lambda$, then $\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}$ λ-a.e.

(c) If ν and μ are equivalent, then $\frac{d\nu}{d\mu} = \left(\frac{d\mu}{d\nu}\right)^{-1}$ μ-a.e.

(d) If g is ν-integrable, then

$$\int_X g d\nu = \int_X g \frac{d\nu}{d\lambda} d\lambda.$$

Proof. For (a), note that $\nu + \mu \ll \lambda$ as well. Let $f = \frac{d\nu}{d\lambda}$ and $g = \frac{d\mu}{d\lambda}$. Then

$$\int_A (f+g)d\lambda = \int_A f d\lambda + \int_A g d\lambda = \nu(A) + \mu(A) = (\nu+\mu)(A) = \int_A \frac{d(\nu+\mu)}{d\lambda} d\lambda \quad \text{for all } A \in \mathcal{A}.$$

Thus $\frac{d\nu}{d\lambda} + \frac{d\mu}{d\lambda} = f + g = \frac{d(\nu+\mu)}{d\lambda}$ λ-a.e.

Next, we jump to (d). We start by considering the case where $g = \chi_A$ with $A \in \mathcal{A}$. By the Radon-Nikodym theorem,

$$\int_X g d\nu = \int_X \chi_A d\nu = \nu(A) = \int_A \frac{d\nu}{d\lambda} d\lambda = \int_X \chi_A \frac{d\nu}{d\lambda} d\lambda = \int_X g \frac{d\nu}{d\lambda} d\lambda.$$

By linearity, the result holds for simple functions. For a nonnegative $g \in \mathcal{L}^1(\nu)$, we can find a sequence of simple functions $g_n \nearrow g$ so that

$$\int_X g d\nu = \lim_{n\to\infty} \int_X g_n d\nu = \lim_{n\to\infty} \int_X g_n \frac{d\nu}{d\lambda} d\lambda = \int_X g \frac{d\nu}{d\lambda} d\lambda$$

by Lebesgue's monotone convergence theorem. For general $g \in \mathcal{L}^1(\nu)$, we can write $g = g^+ - g^-$ and apply the result to g^+ and g^-.

$$\int_X g d\nu = \int_X g^+ d\nu - \int_X g^- d\nu = \int_X g^+ \frac{d\nu}{d\lambda} d\lambda - \int_X g^- \frac{d\nu}{d\lambda} d\lambda = \int_X g d\nu.$$

With (d) established, we can now prove (b). By the Radon-Nikodym theorem,

$$\int_A \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda} d\lambda = \int_A \frac{d\nu}{d\mu} d\mu = \int_A d\nu = \nu(A) = \int_A \frac{d\nu}{d\lambda} d\lambda.$$

Finally, for (c), letting $\lambda = \nu$ and applying (b) gives $1 = \frac{d\nu}{d\nu} = \frac{d\nu}{d\mu}\frac{d\mu}{d\nu}$ ν-a.e. and thus μ-a.e. by the equivalence of ν and μ. Hence $\frac{d\nu}{d\mu} = \left(\frac{d\mu}{d\nu}\right)^{-1}$ μ-a.e. ■

Theorem 1.52 (Lebesgue Decomposition)
Let ν, λ be two σ-finite measures defined on a measurable space $(X, \mathcal{A})$. Then ν can be decomposed uniquely into $\nu = \nu_a + \nu_s$ where $\nu_a \ll \lambda$ and $\nu_s \perp \lambda$.

Proof. We first assume that ν, λ are finite measures. Let $\mu = \nu + \lambda$. Then clearly $\lambda \ll \mu$ and μ is σ-finite. By the Radon-Nikodym theorem, there exists a Radon-Nikodym derivative f such that

$$\lambda(A) = \int_A f d\mu.$$

Denote $\{x \in X \mid f(x) = 0\}$ by E. Define

$$\nu_a(A) = \nu(A \cap E^c), \quad \nu_s(A) = \nu(A \cap E)$$

for each $A \in \mathcal{A}$. Then clearly $\nu_a(A) + \nu_s(A) = \nu(A \cap E^c) + \nu(A \cap E) = \nu(A)$ for all $A \in \mathcal{A}$.

Also, suppose $\lambda(A) = 0$. Then by proposition 1.51,

$$0 = \lambda(A) = \int_A f d\mu = \int_A f d\lambda + \int_A f d\nu = \int_A f d\nu.$$

Hence $f(x) = 0$ ν-a.e. on A. This implies that $\nu(A) = \nu(A \cap E)$ and thus $\nu_a(A) = \nu(A \cap E^c) = \nu(A) - \nu(A \cap E) = 0$, so $\nu_a \ll \lambda$. Also, since $\lambda(E) = 0$ and $\nu_s(E^c) = \nu(\varnothing) = 0$, $\nu_s \perp \lambda$. For the uniqueness, suppose $\nu = \nu_a + \nu_s = \nu_a' + \nu_s'$ both satisfy the conditions. Since $\nu_a \ll \lambda$ and $\nu_a' \ll \lambda$, by the uniqueness of the Radon-Nikodym derivative, $\nu_a = \nu_a'$ and hence $\nu_s = \nu_s'$ as well.

Finally, for the general case where ν, λ are σ-finite, write $X = \cup_n X_n$ where $\lambda(X_n) < \infty$ and X_n are disjoint. For each n we can find the corresponding decomposition ν_a^n and ν_s^n. Let $\nu_a = \sum_n \nu_a^n$ and $\nu_s = \sum_n \nu_s^n$. Then $\nu_a \ll \lambda$ and $\nu_s \perp \lambda$. The uniqueness follows from the uniqueness of the decompositions in each X_n. This establishes the proof. ∎

Corollary 1.53

Let ν be a signed measure and λ be a measure defined on a measurable space $(X, \mathcal{A})$. Suppose both ν and λ are finite and $\nu \ll \lambda$. Then there exists a unique $f \in \mathcal{L}^1(X, \lambda)$ such that

$$\nu(A) = \int_A f d\lambda.$$

Proof. By Hahn decomposition, there exists a positive set P and a negative set N such that $P \cup N = X$. Define

$$\nu_P(A) = \nu(A \cap P), \quad \nu_N(A) = -\nu(A \cap N).$$

Then clearly $\nu_P - \nu_N = \nu$ and $|\nu| = \nu_P + \nu_N$. Note that ν_P and ν_N are both positive measures. Also, by assumption, if $\lambda(A) = 0$ then $\nu(A) = 0$ and hence so are ν_P and ν_N. Thus $\nu_P \ll \lambda$ and $\nu_N \ll \lambda$. By the Radon-Nikodym theorem, there exists $f_P, f_N \in \mathcal{L}^1(X, \lambda)$ such that

$$\nu_P(A) = \int_A f_P d\lambda, \quad \nu_N(A) = \int_A f_N d\lambda.$$

Hence

$$\nu(A) = \nu_P(A) - \nu_N(A) = \int_A f_P d\lambda - \int_A f_N d\lambda = \int_A (f_P - f_N) d\lambda.$$

By setting $f = f_P - f_N$, we obtain the desired function. Uniqueness follows from the uniqueness of the Radon-Nikodym derivative. ∎

1.5. Product Measure

Definition 1.54

Let $\mathcal{S}, \mathcal{T}$ be two σ-algebra on X and Y respectively. The smallest σ-algebra on $X \times Y$ containting the collection $\{S \times T \mid S \in \mathcal{S}, T \in \mathcal{T}\}$ is called the ***product σ-algebra*** *of $\mathcal{S}$ and $\mathcal{T}$, denoted by $\mathcal{S} \otimes \mathcal{T}$.*

Definition 1.55
Suppose X is an arbitrary set and $\mathcal{M}$ is a collection of subsets of X. We say that $\mathcal{M}$ is a ***monotone class*** *if*

(a) *If $E_i \subset E_{i+1}$ for countably many $E_i \in \mathcal{M}$, then $\bigcup_{i=1}^{\infty} E_i \in \mathcal{M}$.*

(b) *If $E_i \supset E_{i+1}$ for countably many $E_i \in \mathcal{M}$, then $\bigcap_{i=1}^{\infty} E_i \in \mathcal{M}$.*

Definition 1.56
A collection $\mathcal{A}$ of subsets in X is called an ***algebra*** *on X if*

(a) *$\varnothing \in \mathcal{A}$.*

(b) *If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.*

(c) *If $A_1, A_2 \in \mathcal{A}$, then $A_1 \cup A_2 \in \mathcal{A}$.*

Remark
The condition (c) implies that for finitly many $A_i \in \mathcal{A}$, $\cup_{i=1}^{n} A_i \in \mathcal{A}$.

Theorem 1.57 (Monotone Class Theorem)
Suppose $\mathcal{A}$ is an algebra on X. Then the smallest σ-algebra containing $\mathcal{A}$ is the smallest monotone class containing $\mathcal{A}$.

Proof. Let $\mathcal{M}$ be the smallest monotone class containing $\mathcal{A}$. The theorem can be written as $\sigma(\mathcal{A}) = \mathcal{M}$. First we show that $\mathcal{M} \subset \sigma(\mathcal{A})$. To see this, we claim first that a σ-algebra is automatically a monotone class. Indeed, let $\mathcal{S}$ be a σ-algebra. Then for any countably many $E_i \in \mathcal{S}$ with $E_i \nearrow E$, we have $E = \bigcup_{i=1}^{\infty} E_i \in \mathcal{S}$. Also, for any countably many $E_i \in \mathcal{S}$ with $E_i \searrow E$, we have $E_i^c \nearrow E^c$. Thus $E^c \in \mathcal{S}$ and $E \in \mathcal{S}$. Therefore $\mathcal{S}$ is a monotone class. It follows that $\sigma(\mathcal{A})$ is a monotone class and hence $\mathcal{M} \subset \sigma(\mathcal{A})$ by the minimality of $\mathcal{M}$.

Next, we claim that $\mathcal{M}$ is a σ-algebra. By definition, we already have $\varnothing \in \mathcal{M}$. Let $E \in \mathcal{M}$. Then there is a sequence of sets $E_i \in \mathcal{A}$ such that either $E_i \nearrow E$ or $E_i \searrow E$. In the former case, we have $E^c = \bigcap_{i=1}^{\infty} E_i^c \in \mathcal{M}$; in the latter case, we have $E^c = \bigcup_{i=1}^{\infty} E_i^c \in \mathcal{M}$. Thus $E^c \in \mathcal{M}$. Lastly, we need to show that $\mathcal{M}$ is closed under countable unions. We start by showing that it is closed under finite unions. Consider $A \in \mathcal{A}$. Define $\mathcal{D}_1 = \{D \in \mathcal{M} \mid D \cup A \in \mathcal{M}\}$. It is clear that $\mathcal{D}_1$ is a monotone class and $\mathcal{A} \subset \mathcal{D}_1$. Consider also $\mathcal{D}_2 = \{D \in \mathcal{M} \mid D \cup E \in \mathcal{M} \text{ for all } E \in \mathcal{M}\}$. Then $\mathcal{D}_2$ is also a monotone class and $\mathcal{A} \subset \mathcal{D}_2$. By the minimality of $\mathcal{M}$, we have $\mathcal{M} \subset \mathcal{D}_1 \cap \mathcal{D}_2$ and hence $\mathcal{M}$ is closed under finite unions. Now let $E_i \in \mathcal{M}$ be countably many sets. Put $F_n = \cup_{i=1}^{n} E_i$. Then $F_n \nearrow E = \bigcup_i E_i$. By the closure of $\mathcal{M}$ under countable unions, $F_n \in \mathcal{M}$; by the definition of $\mathcal{M}$, $E \in \mathcal{M}$. We conclude that $\mathcal{M}$ is closed under countable unions. Thus $\mathcal{M}$ forms a σ-algebra. It now follows by the minimality of $\sigma(\mathcal{A})$ that $\sigma(\mathcal{A}) \subset \mathcal{M}$. We conclude that $\sigma(\mathcal{A}) = \mathcal{M}$. ∎

Lemma 1.58
Suppose $(X, \mathcal{S}, \mu)$ and $(Y, \mathcal{T}, \nu)$ are two finite measure spaces. Let

$$\mathcal{F} = \left\{ E \subset X \times Y \,\middle|\, \int \int \chi_E(x,y) d\nu(y) d\mu(x) = \int \int \chi_E(x,y) d\mu(x) d\nu(y) \right\}.$$

Then $\mathcal{S} \otimes \mathcal{T} \subset \mathcal{F}$.

Proof. Since $\varnothing \in \mathcal{F}$, $\mathcal{F}$ is non-empty. Let $E = A \times B$ for some $A \in \mathcal{S}$ and $B \in \mathcal{T}$. Then

$$\begin{aligned} \int\int \chi_E(x,y)d\nu(y)d\mu(x) &= \int_A\int_B d\nu(y)d\mu(x) = \nu(B)\int_A d\mu(x) \\ &= \nu(B)\mu(A) = \int_B \mu(A)d\nu(y) \\ &= \int_B\int_A d\mu(x)d\nu(y) = \int\int \chi_E(x,y)d\mu(x)d\nu(y). \end{aligned}$$

Now let $\mathcal{R}$ be the collection of all rectangles on $X \times Y$, i.e., $\mathcal{R} = \{A \times B \mid A \in \mathcal{S}, B \in \mathcal{T}\}$. For $R_1, R_2 \in \mathcal{R}$, $R_1 \cap R_2 = \varnothing$ implies $\chi_{R_1 \cup R_2} = \chi_{R_1} + \chi_{R_2}$. By the above calculation, we know that $\mathcal{R} \subset \mathcal{F}$. Consider a sequence of sets $E_i \in \mathcal{F}$. If $E_i \nearrow E$, then

$$\begin{aligned} \int\int \chi_E(x,y)d\nu(y)d\mu(x) &= \lim_{i\to\infty}\int\int \chi_{E_i}(x,y)d\nu(y)d\mu(x) \\ &= \lim_{i\to\infty}\int\int \chi_{E_i}(x,y)d\mu(x)d\nu(y) = \int\int \chi_E(x,y)d\mu(x)d\nu(y). \end{aligned}$$

Also, if $E_i \searrow E$, then

$$\begin{aligned} \int\int \chi_E(x,y)d\nu(y)d\mu(x) &= \lim_{i\to\infty}\int\int \chi_{E_i}(x,y)d\nu(y)d\mu(x) \\ &= \lim_{i\to\infty}\int\int \chi_{E_i}(x,y)d\mu(x)d\nu(y) = \int\int \chi_E(x,y)d\mu(x)d\nu(y). \end{aligned}$$

Hence $\mathcal{F}$ is a monotone class containing $\mathcal{R}$. By the monotone class theorem, $\mathcal{S} \otimes \mathcal{T} \subset \mathcal{F}$. ■

Theorem 1.59 (Existence and Uniqueness of Product Measure)
Let $(X, \mathcal{S}, \mu)$ and $(Y, \mathcal{T}, \nu)$ be two σ-finite measure spaces. Let ω be a set function on $\mathcal{S} \otimes \mathcal{T}$. For $A \in \mathcal{S}$ and $B \in \mathcal{T}$, define

$$\omega(A \times B) = \mu(A)\nu(B).$$

Then, ω extends uniquely to a measure on $(X \times Y, \mathcal{S} \otimes \mathcal{T})$ such that for every $E \in \mathcal{S} \otimes \mathcal{T}$,

$$\omega(E) = (\mu \times \nu)(E) = \int\int \chi_E(x,y)d\nu(y)d\mu(x) = \int\int \chi_E(x,y)d\mu(x)d\nu(y).$$

Proof. If we consider μ and ν to be σ-finite measures, lemma 1.58 gives us that $\omega(A \times B) = \mu(A)\nu(B)$. We want to extend ω to a set function

$$\omega(E) = \int\int \chi_E(x,y)d\nu(y)d\mu(x) = \int\int \chi_E(x,y)d\mu(x)d\nu(y).$$

Since integrals are linear, ω is finitely additive. Applying the monotone class theorem, ω becomes σ-additive. Hence ω becomes a measure on $(X \times Y, \mathcal{S} \otimes \mathcal{T})$. To see the uniqueness, let ρ be another measure on $(X \times Y, \mathcal{S} \otimes \mathcal{T})$ such that $\rho(A \times B) = \mu(A)\nu(B)$. Let $\mathcal{M} =$

$\{E \subset X \times Y \mid \omega(E) = \rho(E)\}$. For countably many $E_i \in \mathcal{M}$ with $E_i \nearrow E$, we can write $E = \cup_{i=1}^{\infty} D_i$ where $D_i = E_{i+1} - E_i$ and $E_0 = \varnothing$ are disjoint. The σ-additivity gives $\omega(E) = \rho(E)$. Thus $E \in \mathcal{M}$. A similar argument gives us that $E_i \searrow E$ implies $E \in \mathcal{M}$. Hence $\mathcal{M}$ is a monotone class. By the monotone class theorem, $\mathcal{S} \otimes \mathcal{T} \subset \mathcal{M}$. Thus $\omega = \rho$.

For the case μ, ν being σ-finite, consider $\{A_i\}$ and $\{B_i\}$ to be two disjoint partitions of X and Y respectively with $\mu(A_i) < \infty$ and $\nu(B_i) < \infty$ for all i. Let $E_{ij} = E \cap (A_i \times B_j)$. By the established result for finite measures,

$$\int \int \chi_{E_{ij}} d\mu d\nu = \int \int \chi_{E_{ij}} d\nu d\mu.$$

Taking the sum over i, j and applying Lebesgue monotone convergence theorem gives us

$$\omega(E) = \int \int \chi_E d\nu d\mu = \int \int \chi_E d\mu d\nu$$

for any $E \in \mathcal{S} \otimes \mathcal{T}$. Applying Lebesgue monotone convergence theorem again results in that ω is σ-additive. Hence ω is a measure on $(X \times Y, \mathcal{S} \otimes \mathcal{T})$. To see the uniqueness, let ρ be another measure on $(X \times Y, \mathcal{S} \otimes \mathcal{T})$ such that $\rho(A \times B) = \mu(A)\nu(B)$. By the σ-additivity and the uniqueness of the finite measure case,

$$\omega(E) = \sum_{i,j} \omega(E_{ij}) = \sum_{i,j} \rho(E_{ij}) = \rho(E)$$

for all $E \in \mathcal{S} \otimes \mathcal{T}$. Thus $\omega = \rho$. ■

Theorem 1.60 (Fubini-Tonelli)
Let $(X, \mathcal{S}, \mu)$ and $(Y, \mathcal{T}, \nu)$ be two σ-finite measure spaces. Let $F : X \times Y \to \mathbb{R}$ be a $\mathcal{S} \otimes \mathcal{T}$-measurable function such that one of the following conditions holds:

(a) $F \geq 0$ a.e. (Tonelli);

(b) F is integrable (Fubini).

Then

$$\int F(x, y) d(\mu \times \nu) = \int \int F d\mu d\nu = \int \int F d\nu d\mu$$

and furthermore,

$$\begin{cases} y \mapsto \int F(x, y) d\mu(x) & \text{is } \mathcal{T}\text{-measurable}, \\ x \mapsto \int F(x, y) d\nu(y) & \text{is } \mathcal{S}\text{-measurable}. \end{cases}$$

Proof. By theorem 1.59, the statement holds for indicator functions and hence for simple functions. By Lebesgue monotone convergence theorem, the non-negative case (Tonelli) is proved. For the integrable case (Fubini), write $F = F^+ - F^-$. We also have that $y \mapsto \int F^\pm(x, y) d\mu(x)$ and $x \mapsto \int F^\pm(x, y) d\nu(y)$ are $\mathcal{S}$-measurable and $\mathcal{T}$-measurable by the theorem 1.59. Furthermore, $y \mapsto \int F^\pm(x, y) d\mu(x)$ and $x \mapsto \int F^\pm(x, y) d\nu(y)$ are integrable a.e. or the condition (b) is violated. Thus

$$\begin{aligned}\int F d(\mu\times\nu) &= \int\int F^+ d(\mu\times\nu) - \int\int F^- d(\mu\times\nu)\\ &= \int\int F^+ d\mu d\nu - \int\int F^- d\mu d\nu\\ &= \int\int F^+ d\nu d\mu - \int\int F^- d\nu d\mu.\end{aligned}$$

The proof is complete. ∎

Remark

By induction, one can extend the Lebesgue measure to any $\mathbb{R}^d$, $d \in \mathbb{N}$. $\mathcal{B}_m \otimes \mathcal{B}_n = \mathcal{B}_{m+n}$, where $\mathcal{B}_n$ is the Borel σ-algebra on $\mathbb{R}^n$. Extension to $\mathbb{R}^\infty$ is also possible.

1.6. Convergence in Measure

Definition 1.61

*Let $(\Omega, \mathcal{S}, \mu)$ be a measure space. We say that a seuqnce of function f_n on Ω **converges to a function f in measure** if for every $\epsilon > 0$,*

$$\mu(\{x \in \Omega \mid |f_n(x) - f(x)| \geq \epsilon\}) \to 0$$

as $n \to \infty$. We write $f_n \xrightarrow{m} f$.

Theorem 1.62 (Markov Inequality)

Let $(\Omega, \mathcal{S}, \mu)$ be a measure space. For any non-negative measurable function f on Ω,

$$\mu(\{x \in \Omega \mid f \geq t\}) \leq \frac{1}{t}\int_\Omega f d\mu.$$

Proof. Let $E_t = \{x \in \Omega \mid f(x) \geq t\}$. Then

$$\mu(E_t) = \int \chi_{E_t} d\mu \leq \int \frac{f}{t} d\mu = \frac{1}{t}\int f d\mu.$$

∎

Corollary 1.63 (Chebyshev Inequality)

Let $(\Omega, \mathcal{S}, \mu)$ be a measure space. For any measurable function f on Ω, and $\alpha \in \mathbb{R}$,

$$\mu(\{x \in \Omega \mid |f(x) - \alpha| \geq t\}) \leq \frac{1}{t^2}\int_\Omega (f-\alpha)^2 d\mu.$$

Proof. Let $g = |f - \alpha|^2$. Apply Markov inequality,

$$\mu(\{x \in \Omega \mid |f(x) - \alpha| \geq t\}) = \mu\Big(\Big\{x \in \Omega \;\Big|\; g \geq t^2\Big\}\Big) \leq \frac{1}{t^2}\int_\omega g d\mu = \frac{1}{t^2}\int_\Omega (f-\alpha)^2 d\mu.$$

■

Corollary 1.64 (Chernoff Bound)
Let $(\Omega, \mathcal{S}, \mu)$ *be a measure space. For any measurable function* f *on* Ω*, and* $\eta \in \mathbb{R}$*,*

$$\mu(\{x \in \Omega \mid f(x) \geq t\}) \leq e^{-\eta t} \int_\Omega e^{\eta f} d\mu$$

for all $t \in \mathbb{R}$*.*

Proof. Let $g = e^{\eta f}$. Then by Markov inequality,

$$\mu(\{x \in \Omega \mid f(x) \geq t\}) = \mu\left(\left\{x \in \Omega \;\middle|\; g \geq e^{\eta t}\right\}\right) \leq \frac{1}{e^{\eta t}} \int_\Omega g d\mu = e^{-\eta t} \int_\Omega e^{\eta f} d\mu.$$

■

Corollary 1.65
If $f_n \to f$ *in* $\mathcal{L}^1$*, then* $f_n \xrightarrow{m} f$*.*

Proof. Let $\epsilon > 0$. By Markov inequaltiy, we have

$$\mu(\{x \in \Omega \mid |f_n(x) - f(x)| \geq \epsilon\}) \leq \frac{1}{\epsilon} \int_\Omega |f_n - f| d\mu \to 0$$

as $n \to \infty$. Thus $f_n \xrightarrow{m} f$. ■

Remark
The converse is not true. Simply find a sequence of functions converging in L^1 *but not almost everywhere will do. However, even stronger, we can actually find a sequence converging in measure but neither in* $\mathcal{L}^1$ *nor almost everywhere. For example, let* $\Omega = [0, 1]$ *with usual measure. Then let* $f_{k,j} = k^2 \chi_{[\frac{j}{k}, \frac{j+1}{k}]}$ *for* $j = 0, 1, \ldots, k-1$ *and* $k \in \mathbb{N}$*. Reindex the sequence recursively by letting* $g_0 = f_{1,0}$ *and*

$$g_{n+1} = \begin{cases} f_{k,j+1} & \text{if } g_n = f_{k,j} \text{ with } j \neq k-1, \\ f_{k+1,0} & \text{if } g_n = f_{k,j} \text{ with } j = k-1. \end{cases}$$

This also defines a injective function $\phi : n \mapsto (k_n, j_n)$*. Then* $g_n \to 0$ *in measure because for any* $\epsilon > 0$*,*

$$\mu(\{x \mid |g_n| \geq \epsilon\}) = \frac{1}{k_n} \to 0.$$

But

$$\int_0^1 |g_n| \, d\mu = k_n \to \infty$$

and since $[\frac{j_n}{k_n}, \frac{j_n+1}{k_n}]$ *includes* x *infinitely many times for any* $x \in [0, 1]$*,* g_n *converges nowhere in* $[0, 1]$*.*

Theorem 1.66
Let $(\Omega, \mathcal{S}, \mu)$ be a σ-finite measure space. If $f_n \xrightarrow{m} f$, then there exists a subsequence f_{n_k} such that $f_{n_k} \to f$ almost everywhere.

Proof. Since $f_n \xrightarrow{m} f$, we can choose n_k such that

$$\mu\left(\left\{x \in \Omega \,\middle|\, |f_n(x) - f(x)| \geq \frac{1}{k}\right\}\right) \leq \frac{1}{2^k} \quad \text{for all } n \geq n_k.$$

Let $E_k = \left\{x \in \Omega \,\middle|\, |f_n(x) - f(x)| \geq \frac{1}{k} \text{ for all } n \geq n_k\right\}$. Then $\mu(E_k) \leq 2^{-k}$. Put $H_m = \bigcup_{k=m}^{\infty} E_k$. We have

$$\mu(H_m) \leq \sum_{k \geq m} \mu(E_k) \leq \sum_{k \geq m} 2^{-k} = 2^{-m+1}.$$

Put $H = \bigcap_{m=1}^{\infty} H_m$, $H_m \searrow H$. Then

$$\mu(H) = \lim_{m \to \infty} \mu(H_m) = 0.$$

If $x \notin H$, then $x \notin H_m$ for some m. Then

$$\left|f_{n_k}(x) - f(x)\right| < \frac{1}{k} \text{ for all } k \geq m.$$

Thus $f_{n_k}(x) \to f(x)$ almost everywhere as $k \to \infty$. ■

Definition 1.67
*Let f_n be a sequence of measurable functions on $(\Omega, \mathcal{S}, \mu)$. We say that f_n is **Cauchy in measure** if for every $\epsilon > 0$,*

$$\mu(\{x \in \Omega \mid |f_n(x) - f_m(x)| \geq \epsilon\}) \to 0$$

as $n, m \to \infty$.

Theorem 1.68 (Cauchy Criterion for Convergence in Measure)
Let $(\Omega, \mathcal{S}, \mu)$ be a measure space. A sequence of measurable functions f_n on Ω converges in measure if and only if it is Cauchy in measure.

Proof. Suppose that $f_n \xrightarrow{m} f$. Let $\epsilon > 0$ be given. We have

$$\mu(\{|f_n - f| \geq \epsilon\}) \to 0$$

as $n \to \infty$. Then since $\{|f_n - f_m| \geq \epsilon\} \subset \{|f_n - f| \geq \epsilon/2\} \cup \{|f_m - f| \geq \epsilon/2\}$,

$$\mu(\{|f_n - f_m| \geq \epsilon\}) \leq \mu(\{|f_n - f| \geq \epsilon/2\}) + \mu(\{|f_m - f| \geq \epsilon/2\}) \to 0$$

as $n, m \to \infty$. Thus f_n is Cauchy in measure.

Conversely, suppose that f_n is Cauchy in measure. We can take a subsequence f_{n_j} such that

$$\mu(E_j) = \mu\left(\left\{\left|f_{n_j} - f_{n_{j+1}}\right| \geq 2^{-j}\right\}\right) \leq 2^{-j}.$$

Put $F_k = \cup_{j=k}^{\infty} E_j$. Then $\mu(F_k) \le \sum_{j=k}^{\infty} \mu(E_j) \le 2^{-k+1}$. For $x \notin F_k$, $i > j$,

$$\left|f_{n_i} - f_{n_j}\right| \le \sum_{l=j}^{i-1} \left|f_{n_{l+1}} - f_{n_l}\right| \le \sum_{l=j}^{i-1} 2^{-l} \le 2^{-j+1}.$$

Hence f_{n_j} is Cauchy on F_k^c. By the completeness of $\mathbb{R}$, f_{n_j} converges pointwise on F_k^c for each k. Put $F = \cap_{k=1}^{\infty} F_k$. Then $\mu(F) = 0$. Let

$$f(x) = \begin{cases} \lim_{j\to\infty} f_{n_j}(x) & \text{if } x \notin F, \\ 0 & \text{if } x \in F. \end{cases}$$

Since f_{n_j} are measurable, f is measurable. Also, $f_{n_j} \to f$ pointwisely almost everywhere. Thus $f_{n_j} \xrightarrow{m} f$. Observe that $\{|f_n - f| \ge \epsilon\} \subset \left\{\left|f_n - f_{n_j}\right| \ge \epsilon/2\right\} \cup \left\{\left|f_{n_j} - f\right| \ge \epsilon/2\right\}$. Hence

$$\mu(\{|f_n - f| \ge \epsilon\}) \le \mu\Big(\left\{\left|f_n - f_{n_j}\right| \ge \epsilon/2\right\}\Big) + \mu\Big(\left\{\left|f_{n_j} - f\right| \ge \epsilon/2\right\}\Big) \to 0$$

as $n \to \infty$. Thus $f_n \xrightarrow{m} f$. ■

Definition 1.69
*A function $\phi : (a,b) \to \mathbb{R}$, where $-\infty \le a < b \le \infty$, is **convex** if for any $x, y \in (a,b)$ and $\lambda \in [0,1]$,*

$$\phi(\lambda x + (1-\lambda)y) \le \lambda\phi(x) + (1-\lambda)\phi(y).$$

Remark
Every convex function is continuous.

Remark
The definition of convexity can also be written as

$$\frac{\phi(t) - \phi(s)}{t - s} \le \frac{\phi(u) - \phi(t)}{u - t},$$

whenever $a < s < t < u < b$.

Theorem 1.70 (Jensen's Inequality)
Let $(\Omega, \mathcal{S}, \mu)$ be a measure space with $\mu(\Omega) = 1$. Suppose that $f : \Omega \to I$, $f \in \mathcal{L}^1(\Omega)$ and $\phi : I \to \mathbb{R}$ is a convex function on an interval I. Then

$$\phi\left(\int_\Omega f\,d\mu\right) \le \int_\Omega \phi(f)\,d\mu.$$

Proof. Put $t = \int_\Omega f\,d\mu$. Then $a < t < b$. Let

$$\beta = \sup_{s\in(a,t)} \frac{\phi(t) - \phi(s)}{t - s}.$$

By the convexity,

$$\beta \leq \frac{\phi(u) - \phi(t)}{u - t}$$

for any $u \in (t, b)$. Thus

$$\phi(y) \geq \phi(t) + \beta(y - t)$$

for all $y \in (a, b)$. Hence

$$\phi(f(x)) - \phi(t) - \beta(f(x) - t) \geq 0$$

for every $x \in \Omega$. Since ϕ is continuous, $\phi \circ f$ is measurable. Thus

$$\int_\Omega \phi(f)d\mu - \phi(t) = \int_\Omega \phi(f)d\mu - \phi(t) - \beta\left(\int_\Omega f d\mu - t\right) = \int_\Omega \phi(f)d\mu - \phi(t) - \beta \int_\Omega (f - t)d\mu \geq 0.$$

Since $t = \int_\Omega f d\mu$, we have

$$\phi\left(\int_\Omega f d\mu\right) \leq \int_\Omega \phi(f)d\mu.$$

■

Definition 1.71
A family of measure $\{\nu_\alpha\}$ is said to be ***equicontinuous at*** $\varnothing$ *if for any $\epsilon > 0$ and $B_k \searrow \varnothing$, there exists k_0 such that*

$$\sup_\alpha \nu_\alpha(B_k) < \epsilon$$

for all $k \geq k_0$.

Definition 1.72
A family of measure $\{\nu_\alpha\}$ is said to be ***uniformly absolutely continuous*** *with respect to μ if for any $\epsilon > 0$, there exists $\delta > 0$ such that for any B with $\mu(B) < \delta$,*

$$\sup_\alpha \nu_\alpha(B) < \epsilon.$$

Lemma 1.73
If $\{\nu_\alpha\}$ is equicontinuous at $\varnothing$ and $\nu_\alpha \ll \mu$ for all α, then $\{\nu_\alpha\}$ is uniformly absolutely continuous with respect to μ.

Proof. Suppose that $\{\nu_\alpha\}$ is not uniformly absolutely continuous with respect to μ. Then there exists $\epsilon > 0$ such that for any n, we can find B_n with $\mu(B_n) \leq 2^{-n}$ and some α_n with $\nu_{\alpha_n}(B_n) \geq \epsilon$. Put $A_k = \cup_{n=k}^\infty B_n$. Then $\mu(A_k) \leq 2^{-k+1}$. Set $A = \cap_{k=1}^\infty A_k$. Then $A_k \searrow A$ and $\mu(A) = 0$. This implies $\nu_\alpha(A) = 0$ for all α since $\nu\alpha \ll \mu$. Observe now that

$$\nu_{\alpha_n}(A_k - A) = \nu_{\alpha_n}(A_k) \geq \nu_{\alpha_n}(B_n) \geq \epsilon$$

for all $n \geq k$. But $\nu_{\alpha_n}(A_k - A) \to 0$ as $k \to \infty$, a contradiction. Thus $\{\nu_\alpha\}$ is uniformly absolutely continuous with respect to μ. ■

Theorem 1.74
Let $(\Omega, \mathcal{S}, \mu)$ be a σ-finite measure space. Suppose $f_n \in \mathcal{L}^p(\Omega)$. Consider a family of measures ν_n defined by

$$\nu_n(A) = \int_A |f_n|^p \, d\mu.$$

If ν_n is equicontinuous at $\varnothing$ and $f_n \xrightarrow{m} f$, then $f_n \to f$ in $\mathcal{L}^p(\Omega)$.

Proof. Since $(\Omega, \mathcal{S}, \mu)$ is σ-finite, we can write $\Omega = \cup_k E_k$ with $\mu(E_k) < \infty$ for all k. Then $E_k^c \searrow \varnothing$ and $\nu_n(E_k^c) \to 0$ as $k \to \infty$. Also, since ν_n is equicontinuous at $\varnothing$, for any $\epsilon > 0$, there exists k_0 such that

$$\sup_n \nu_n(E_k^c) < \epsilon$$

for all $k \geq k_0$.

We claim that f_n is Cauchy in $\mathcal{L}^p$. Indeed,

$$\int |f_n - f_m|^p \, d\mu = \int_{E_{k_0}^c} |f_n - f_m|^p \, d\mu + \int_{E_{k_0} \cap \{|f_n - f_m| \leq \epsilon/\mu(E_{k_0})\}} |f_n - f_m|^p \, d\mu$$
$$+ \int_{E_{k_0} \cap \{|f_n - f_m| > \epsilon/\mu(E_{k_0})\}} |f_n - f_m|^p \, d\mu.$$

Estimate from Jensen's inequality,

$$\int_{E_{k_0}^c} |f_n - f_m|^p \, d\mu \leq 2^p \int_{E_{k_0}^c} |f_n|^p \, d\mu + 2^p \int_{E_{k_0}^c} |f_m|^p \, d\mu = 2^p \nu_n(E_{k_0}^c) + 2^p \nu_m(E_{k_0}^c) \to 0,$$

$$\int_{E_{k_0} \cap \{|f_n - f_m| \leq \epsilon/\mu(E_{k_0})\}} |f_n - f_m|^p \, d\mu \leq \frac{\epsilon}{\mu(E_{k_0})} \mu(E_{k_0}) \to 0.$$

For the last term, since $\nu_n \ll \mu$ for all μ, lemma 1.73 gives that ν_n is uniformly absolutely continuous with respect to μ. Given any $\epsilon > 0$, there is $\delta > 0$ such that for all B with $\mu(B) \leq \delta$, $\nu_n(B) \leq \epsilon$ for all n. Thus

$$\mu\left(\left\{ \left| f_j - f \right| \geq \frac{\epsilon}{\mu(E_{k_0})} \right\}\right) \to 0$$

as $j \to \infty$. Hence we obtain that f_n is Cauchy in $\mathcal{L}^p$. It follows from the Riesz-Fischer thoerem that $f_n \to g$ in $\mathcal{L}^p(\Omega)$ for some $g \in \mathcal{L}^p(\Omega)$. Since $f_n \xrightarrow{m} f$, $f = g$ almost everywhere. Thus $f_n \to f$ in $\mathcal{L}^p(\Omega)$. ∎

2. Banach Space

2.1. Banach Space and Bounded Linear Functional

Definition 2.1
A space X is called a ***Banach space*** *if it is a complete normed vector space.*

Remark
$\mathcal{L}^1$ is a Banach space with the norm

$$\|f\|_{\mathcal{L}^1} = \int |f| \, d\mu.$$

We treat $f = g$ a.e. as the same element in $\mathcal{L}^1$.

Definition 2.2
Let V, W be vector spaces. A map $T : V \to W$ is ***linear*** *if for every $c \in \mathbb{R}$, $f, g \in V$, $T(cf + g) = cT(f) + T(g)$.*

Definition 2.3
A linear map $T : V \to W$ has ***operator norm*** *defined by*

$$\|T\| = \sup_{\|f\|_V = 1} \|T(f)\|_W .$$

T is ***bounded*** *if $\|T\| < \infty$. We denote the set of all bounded linear operators from V to W by $B(V, W)$.*

Proposition 2.4
Suppose W is a Banach space. Then $B(V, W)$ is a Banach space with the operator norm.

Proof. It suffices to show that $B(V, W)$ is complete. Let $\{T_i\} \subset B(V, W)$ be a Cauchy sequence. Then for $f \in V$,

$$\left\|T_i(f) - T_j(f)\right\|_W \le \left\|T_i - T_j\right\| \|f\|_V .$$

Hence $\{T_i(f)\}$ is a Cauchy sequence in W. By the completeness of W, we may define Tf as the limit of $T_i(f)$ as $i \to \infty$. Now,

$$\|Tf\| \le \sup_i \|T_i(f)\| \le \sup_i \|T_i\| \|f\| .$$

Since Cauchy sequences are bounded, $\|Tf\| < \infty$ for all $f \in V$ and $T \in B(V, W)$. It remains to show that T_i converges to T in the operator norm. For any $f \in V$, pick N such that $\left\|T_i(f) - T_j(f)\right\| \le \epsilon$ for all $i, j \ge N$. Then for fixed i,

$$\left\|(T_i - T_j)f\right\| \le \left\|T_i - T_j\right\| \|f\| \le \epsilon \|f\|$$

for every $f \in V$ and $j \ge N$. Hence $\|T_i - T\| \le \epsilon$ for all $i \ge N$ and the proof is complete. ■

Remark
Consider X, Y are two normed vector space. $\overline{X}$, $\overline{Y}$ are the completion of X and Y, respectively.

$$\overline{X} = \{\{x_n\} \subset X \mid \{x_n\} \text{ is Cauchy}\}.$$

Define the equivalence relation $\{x_n\} \sim \{y_n\}$ if $\lim_{n\to\infty} \|x_n - y_n\| = 0$. It is easy to see that $\overline{X}$ is a Banach space with $\|\{x_n\}\| = \lim_{n\to\infty} \|x_n\|$.

For $L : X \to Y$, a bounded linear operator, its counterpart $\overline{L} : \overline{X} \to \overline{Y}$ is also a bounded linear operator.

Definition 2.5
T is continuous if $f_i \to f$ in V implies that $T(f_i) \to T(f)$ in W.

Proposition 2.6
Suppose $T : V \to W$ is linear. Then T is continuous if and only if T is bounded.

Proof. Suppose T is not bounded. Then there exists $f_i \in V$ with $\|f_i\| \leq 1$ for all i and $\|Tf_i\| \to \infty$. Thus

$$\frac{f_i}{\|Tf_i\|} \to 0, \quad \text{but} \quad T\left(\frac{f_i}{\|Tf_i\|}\right) = \frac{Tf_i}{\|Tf_i\|} \not\to 0 \quad \text{as} \quad \frac{\|Tf_i\|}{\|Tf_i\|} = 1.$$

Hence T is not continuous.

Conversely, suppose T is bounded. Let $f_i \to f$ in V. Then

$$\|Tf_i - Tf\| = \|T(f_i - f)\| \leq \|T\| \, \|f_i - f\| \to 0.$$

Hence T is continuous. ■

Definition 2.7
*A **linear functional** T is a linear map $T : V \to \mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or C is the scalar field of V.*

Definition 2.8
*Let V, W be vector spaces. $T : V \to W$ is linear. The **kernel** of T is defined as*

$$\ker(T) = \{f \in V \mid T(f) = 0\}.$$

Proposition 2.9
Let X be a normed vector space and $T \in X'$. Then

(a) *$\ker(T)$ is a closed subspace of X.*

(b) *If $T \neq 0$, there exists $x \in X$ such that $T(x) \neq 0$. Then for any $y \in X$, there exists $c \in \mathbb{R}$ and $z \in \ker(T)$ such that $y = cx + z$.*

Proof. For (a), let $x, y \in \ker(T)$ and $c \in \mathbb{R}$.

$$T(cx + y) = cT(x) + T(y) = 0. \implies cx + y \in \ker(T).$$

Also, let $x_i \to x$ in X. Then since T is continuous,

$$T(x) = \lim_{n\to\infty} T(x_n) = 0. \implies x \in \ker(T).$$

Hence $\ker(T)$ is a closed subspace of X.

For the rest part, fix $x \in X$ and $f(x) \neq 0$. For each $y \in X$, let $\alpha = T(y)/T(x)$ and $z = y - T(y)x/T(x)$. Then

$$\alpha x + z = \frac{T(y)}{T(x)}x + y - \frac{T(y)}{T(x)}x = y - \frac{T(y)}{T(x)}x = y.$$

■

Definition 2.10
*The **dual space** of V is defined as $V' = B(V, \mathbb{F})$, where $\mathbb{F} = \mathbb{R}$ or C.*

Remark
The dual space is a Banach space.

Remark
$T : X \to Y$ is bounded and linear. Then

$$\|T\| = \inf\{c \in [0,\infty) \mid \|Tx\|_Y \leq c\,\|x\|_X \textit{ for all } x \in X\}.$$

Example
Let $X = C([0,1])$ with the supremum norm and $Y = \mathbb{R}$ with the usual norm. For $g \in X$, $g(t) \neq 0$ on $[0,1]$, define $Tg : X \to \mathbb{R}$ by

$$Tg(f) = \int_0^1 f(t)g(t)dt.$$

Now for $\|f\|_\infty \leq 1$,

$$\begin{aligned}|Tg(f)| &= \left|\int_0^1 f(t)g(t)dt\right| \leq \int_0^1 |f(t)g(t)|\,dt \leq \int_0^1 |g(t)| \sup_{[0,1]} |f(t)|\,dt \\ &= \|f\|_\infty \int_0^1 |g(t)|\,dt \leq \int_0^1 |g(t)|\,dt.\end{aligned}$$

Take $f = g/|g|$,

$$|Tgf| = \left|\int_0^1 \frac{g^2(t)}{|g(t)|}dt\right| = \int_0^1 |g(t)|\,dt. \implies \|Tg\| = \int_0^1 |g(t)|\,dt.$$

Example
Consider $X = Y = C([0,1])$ with the supremum norm. Define $T : C^1([0,1]) \to Y$ by $Tf = f'$. Then consider the sequnce $f_n(x) = e^{-n(x-1/2)^2}$, $f_n'(x) = e^{-n(x-1/2)^2}(-2n(x-1/2))$. Hence $\|Tf_n\| / \|f_n\| = \sqrt{2n}e^{-1/2} \to \infty$ as $n \to \infty$. Thus T is not bounded.

2.2. ℓ^p Space

Definition 2.11
*$\ell^p = \left\{ \{x_i\}_{i\in\mathcal{I}} \;\middle|\; \|x\|_p < \infty \right\}$, where $\mathcal{I}$ is an countable index set and $\|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$, $1 \le p < \infty$, is called the ℓ^p **space**. For $p = \infty$, the norm is defined as $\|x\|_\infty = \sup_i |x_i|$.*

Definition 2.12
*$f : X \to Y$ is called a **homomophism** if it preserves the algebraic structure. In particular, for X, Y being vector spaces, f is a homomorphism if $f(cx + y) = cf(x) + f(y)$.*

Definition 2.13
*$f : X \to Y$ is called an **isomorphism** if it is a bijective homomorphism.*

Definition 2.14
*$f : X \to Y$ is called an **isometry** if $\|f(x)\|_Y = \|x\|_X$ for all $x \in X$.*

Example
A rightward shift operator $S_R : \ell^p(\mathbb{N}) \to \ell^p(\mathbb{N})$ is not an isomorphism, but $S_R : \ell^p(\mathbb{Z}) \to \ell^p(\mathbb{Z})$ is.

Lemma 2.15 (Young's Inequality)
Let $1 < p, p' < \infty$ with $\frac{1}{p} + \frac{1}{p'} = 1$. Then for all $a, b \ge 0$,

$$ab \le \frac{a^p}{p} + \frac{b^{p'}}{p'}.$$

Furthermore, the equality holds if and only if $a^p = b^{p'}$.

Proof. If $a = 0$ or $b = 0$, the inequality is trivial. Suppose $a, b > 0$. Let $t = 1/p$ and we can write

$$\log(ab) = \log(a) + \log(b) = t\log(a^p) + (1-t)\log(b^{p'}) \le \log\Big(ta^p + (1-t)b^{p'}\Big)$$

by the concavity of logarithm and Jensen's inequality. Exponentiating both sides yields the desired inequality. The equality holds if and only if $a^p = b^{p'}$ by the Jensen's inequality. ∎

Theorem 2.16 (Hölder's Inequality in ℓ^p)
Let $1 \le p, p' \le \infty$ with $\frac{1}{p} + \frac{1}{p'} = 1$. Then for all $f \in \ell^p$ and $g \in \ell^{p'}$,

$$\|fg\|_1 \le \|f\|_p \|g\|_{p'}.$$

Moreover, the equality holds if and only if $f = cg$ for some constant c.

Proof. If one of f or g is zero, the inequality is trivial. If $p = 1$ and $p' = \infty$, $|f_i g_i| \le \|g\|_\infty |f_i|$. Summing over i yields the desired inequality. For the case $p = \infty$ and $p' = 1$ the proof is similar. Now suppose $1 < p < \infty$ and $1 < p' < \infty$. Without loss of generality, we may assume that $\|f\|_p = \|g\|_{p'} = 1$. By Young's inequality,

$$|f_i g_i| \le \frac{|f_i|^p}{p} + \frac{|g_i|^{p'}}{p'}.$$

Thus

$$\|fg\|_1 = \sum_i |f_i g_i| \le \sum_i \frac{|f_i|^p}{p} + \sum_i \frac{|g_i|^{p'}}{p'} = \frac{1}{p}\|f\|_p^p + \frac{1}{p'}\|g\|_{p'}^{p'} = 1.$$

Hence we obtain the desired inequality. The equality holds if and only if $|f_i|^p = |g_i|^{p'}$ for all i by the Young's inequality. In general, the equality holds if and only if $f = cg$ for some constant c after scaling the both sides of the inequality by c. ∎

Remark
We call p' the ***conjugate exponent*** *of p for $1/p + 1/p' = 1$.*

Theorem 2.17 (Minkowski's Inequality in ℓ^p)
Let $1 \le p \le \infty$. Then for all $f, g \in \ell^p$,

$$\|f + g\|_p \le \|f\|_p + \|g\|_p .$$

Proof. If $p = 1$, the inequality comes from the triangle inequality. For $1 < p < \infty$,

$$\begin{aligned}
\|f+g\|_p^p &= \sum_i |f_i + g_i|\,|f_i + g_i|^{p-1} \\
&\le \sum_i |f_i|\,|f_i + g_i|^{p-1} + \sum_i |g_i|\,|f_i + g_i|^{p-1} \\
&\le \|f\|_p \left(\sum_i |f_i + g_i|^{(p-1)p'}\right)^{1/p'} + \|g\|_p \left(\sum_i |f_i + g_i|^{(p-1)p'}\right)^{1/p'} \\
&= \|f\|_p \|f+g\|_p^{p/p'} + \|g\|_p \|f+g\|_p^{p/p'}
\end{aligned}$$

by the Hölder's inequality. Rearranging the inequality yields

$$\|f + g\|_p = \|f + g\|_p^{p-p/p'} \le \|f\|_p + \|g\|_p .$$

For $p = \infty$,

$$\|f + g\|_\infty = \sup_i |f_i + g_i| \le \sup_i |f_i| + \sup_i |g_i| = \|f\|_\infty + \|g\|_\infty .$$

The proof is complete. ∎

Remark
The Minkowski's inequality is exactly the triangle inequality in ℓ^p spaces. We can thus confirm that ℓ^p norms are indeed norms.

2.3. $\mathcal{L}^p$ Space

Definition 2.18
Let $(X, \mathcal{A}, \mu)$ be a measure space and $1 \leq p < \infty$. The space $\mathcal{L}^p(X)$ consists of all equivalence classes of measurable functions $f : X \to \mathbb{R}$ such that

$$\|f\|_{\mathcal{L}^p} = \left(\int_X |f|^p \, d\mu \right)^{1/p} < \infty,$$

where $f \sim g$ if $f = g$ a.e. and the norm is defined on a representative of the equivalence class.

Definition 2.19
*$f : X \to \mathbb{R}$ is measurable. The **essential supremum** of f on X is defined as*

$$\operatorname{ess\,sup}_X f = \inf \left\{ \sup_X g \;\middle|\; g = f \mu\text{-a.e.} \right\} = \inf \left\{ c \in \mathbb{R} \mid \mu(\{x \mid f(x) > c\}) = 0 \right\}.$$

*We called f **essentially bounded** if $\operatorname{ess\,sup}_X f < \infty$. The space $\mathcal{L}^\infty(X)$ consists of all equivalence classes of essentially bounded measurable functions with the norm*

$$\|f\|_{\mathcal{L}^\infty} = \operatorname{ess\,sup}_X |f|.$$

Theorem 2.20 (Hölder's Inequality in $\mathcal{L}^p$)
Let $1 \leq p, p' \leq \infty$ with $\frac{1}{p} + \frac{1}{p'} = 1$. Then for all $f \in \mathcal{L}^p$ and $g \in \mathcal{L}^{p'}$,

$$\|fg\|_1 \leq \|f\|_p \|g\|_{p'}.$$

Moreover, the equality holds if and only if $f = cg$ for some constant c.

Proof. For the case $p = 1$ and $p' = \infty$, notice that

$$|fg| \leq |f| \operatorname{ess\,sup} |g| \implies \|fg\|_1 = \int |fg| \, d\mu \leq \int |f| \operatorname{ess\,sup} |g| \, d\mu = \|f\|_1 \|g\|_\infty.$$

For the case $p = \infty$ and $p' = 1$, the proof is similar. Now suppose $1 < p < \infty$ and $1 < p' < \infty$. If one of f or g is zero, the inequality is trivial. Without loss of generality, we may assume that $\|f\|_p = \|g\|_{p'} = 1$. By the Young's inequality,

$$|fg| \leq \frac{|f|^p}{p} + \frac{|g|^{p'}}{p'}.$$

Integrating both sides yields

$$\|fg\|_1 = \int |fg| \, d\mu \leq \int \frac{|f|^p}{p} d\mu + \int \frac{|g|^{p'}}{p'} d\mu = \frac{1}{p} + \frac{1}{p'} = 1.$$

Hence we obtain the desired inequality. The equality holds if and only if $|f|^p = |g|^{p'}$ a.e.

by the Young's inequality. In general, the equality holds if and only if $f = cg$ a.e. for some constant c after scaling the both sides of the inequality by c. ■

Theorem 2.21 (Minkowski's Inequality in $\mathcal{L}^p$)
Let $1 \leq p \leq \infty$. Then for all $f, g \in \mathcal{L}^p$,

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p .$$

Proof. If $p = 1$, the inequality comes from the triangle inequality. For $1 < p < \infty$,

$$\begin{aligned}
\|f + g\|_p^p &= \int |f + g|^p \, d\mu = \int |f + g| \, |f + g|^{p-1} \, d\mu \\
&\leq \int |f| \, |f + g|^{p-1} \, d\mu + \int |g| \, |f + g|^{p-1} \, d\mu \\
&\leq \|f\|_p \left(\int |f + g|^{(p-1)p'} \, d\mu \right)^{1/p'} + \|g\|_p \left(\int |f + g|^{(p-1)p'} \, d\mu \right)^{1/p'} \\
&= \|f\|_p \, \|f + g\|_p^{p/p'} + \|g\|_p \, \|f + g\|_p^{p/p'} .
\end{aligned}$$

Rearranging the inequality yields

$$\|f + g\|_p = \|f + g\|_p^{p - p/p'} \leq \|f\|_p + \|g\|_p .$$

For $p = \infty$,

$$\|f + g\|_\infty = \operatorname{ess\,sup} |f + g| \leq \operatorname{ess\,sup} |f| + \operatorname{ess\,sup} |g| = \|f\|_\infty + \|g\|_\infty .$$

The proof is complete. ■

Theorem 2.22
$1 \leq p \leq \infty$. *Simple functions are dense in $\mathcal{L}^p$.*

Proof. For $p < \infty$, consider $f \geq 0$ and $f \in \mathcal{L}^1$. There exists a sequence of simple functions $f_n \nearrow f$ a.e. Note that $|f - f_n|^p \leq |f|^p \in \mathcal{L}^1$. By Lebesgue's dominated convergence theorem, $\|f_n - f\|_p \to 0$ as $n \to \infty$. For $p = \infty$, pick an f in the f-equivalent class such that f is bounded. Then since the approximation of simple functions can be done uniformly, the result follows. ■

Remark
A simple function $s = \sum_{i=1}^n c_i \chi_{A_i} \in \mathcal{L}^p$ must have $\mu(A_i) < \infty$ for every i such that $c_i > 0$. Since contiuous functions can approximate simple functions, they are dense in $\mathcal{L}^p$ as well.

Remark
Step functions and continuous functions with compact supports are dense in $\mathcal{L}^p$ for $1 \leq p < \infty$. This can be seen by a slight modification of the proof of proposition 1.35. Let $\epsilon > 0$ be

given. First, for $f \in \mathcal{L}^p$, we can find some $M > 0$ such that

$$\int_{|x|>M} |f|^p \, d\mu < \epsilon.$$

Next, since $\mathcal{L}^p([-M, M]) \subset \mathcal{L}^1([-M, M])$, the result from proposition 1.35 applies, and we can find a step function s such that $\|f - s\|_\infty < \epsilon$ on $[-M, M]$. This implies

$$\int_{|x|\leq M} |f - s|^p \, d\mu \leq \int_{|x|\leq M} \epsilon^p \, d\mu = \epsilon^p \mu([-M, M]).$$

Thus

$$\|f - s\|_p^p = \int_{|x|>M} |f|^p \, d\mu + \int_{|x|\leq M} |f - s|^p \, d\mu \leq \epsilon + \epsilon^p \mu([-M, M]).$$

Hence step functions with compact supports are dense in $\mathcal{L}^p$. Using the same approximation technique in proposition 1.35, we can find a continuous function g such that $\|f - g\|_p < \epsilon$ as well.

Lemma 2.23
$1 \leq p < \infty$. $g_k \in \mathcal{L}^p$ and $\sum_k \|g_k\|_p < \infty$. Then there exists $f \in \mathcal{L}^p$ such that $\sum_k g_k = f$ pointwise a.e. and in $\mathcal{L}^p$.

Proof. Define h_n and h by $h_n = \sum_{k=1}^n |g_k|$ and $h = \sum_k |g_k|$. Then $h_n \nearrow h$. By Lebesgue's monotone convergence theorem,

$$\lim_{n\to\infty} \int h_n^p \, d\mu = \int h^p \, d\mu.$$

By Minkowski's inequality,

$$\left(\int h_n^p \, d\mu\right)^{1/p} = \left(\int \left(\sum_{k=1}^n |g_k|\right)^p d\mu\right)^{1/p} \leq \sum_{k=1}^n \left(\int |g_k|^p \, d\mu\right)^{1/p} \leq \sum_{k=1}^n \|g_k\|_p < \infty$$

for every n, so $h \in \mathcal{L}^p$ and $\|h\|_p \leq M$ for some M bounding $\sum_k \|g_k\|_p$. Now since $\sum_k g_k$ converges absolutely to some f pointwisely a.e. and $|f| \leq h$,

$$\left|f - \sum_{k=1}^n g_k\right|^p \leq \left(|f| + \sum_{k=1}^n |g_k|\right)^p \leq (2h)^p \in \mathcal{L}^1.$$

By Lebesgue's dominated convergence theorem, $\left\|f - \sum_{k=1}^n g_k\right\|_p \to 0$ as $n \to \infty$. Thus the proof is complete. ■

Theorem 2.24 (Riesz-Fischer)
$\mathcal{L}^p$ spaces are complete.

Proof. First, we focus on the case where $1 \leq p < \infty$. Let f_k be a Cauchy sequence in $\mathcal{L}^p$. Take a subsequence f_{k_j} such that $\left\|f_{k_{j+1}} - f_{k_j}\right\| \leq 2^{-j}$. Let $g_j = f_{k_{j+1}} - f_{k_j} \in \mathcal{L}^p$ and we have

$\sum_j \left\|g_j\right\|_p < \infty$. By the lemma 2.23, there exists $f \in \mathcal{L}^p$ such that $f = \sum_j g_j$ a.e. and

$$\lim_{j\to\infty} f_{k_j} = \lim_{j\to\infty} f_{k_1} + \sum_{i=1}^{j-1} g_i = f_{k_1} + f \in \mathcal{L}^p.$$

Since f_k is Cauchy and a subsequence converges, the original sequence f_k converges to $f_{k_1} + f \in \mathcal{L}^p$ as well. We now consider the case where $p = \infty$. Let f_k be a Cauchy sequence in $\mathcal{L}^\infty$. Then for almost every x, $\{f_k(x)\}$ is a Cauchy sequence in $\mathbb{R}$. Thus we can define $f(x)$ as the limit of $f_k(x)$ as $k \to \infty$. On the set where $f_k(x)$ does not converge, we let $f(x)$ be zero. Then $f \in \mathcal{L}^\infty$ since $\{f_k\}$ is Cauchy and has an uniform bound except on a measure zero set. Also, for any $\epsilon > 0$, we can find N such that $\left\|f_k - f_j\right\|_\infty < \epsilon$ for all $k, j \geq N$. Hence $\|f_k - f\|_\infty < \epsilon$ for all $k \geq N$. Thus $f_k \to f$ in $\mathcal{L}^\infty$. We conclude that $\mathcal{L}^p$ spaces are complete. ■

Theorem 2.25
Let $1 \leq p < \infty$. Let $f_n \in \mathcal{L}^p$ be a sequence of measurable funcitons on a σ-finite measure space X. If $f_n \to f$ in $\mathcal{L}^p$, then there exists a subsequence f_{n_k} such that $f_{n_k} \to f$ a.e. on X.

Proof. Using the Markov inequality,

$$\begin{aligned}\mu(\{x \in X \mid |f_n(x) - f(x)| \geq \epsilon\}) &= \mu(\{x \in X \mid |f_n(x) - f(x)|^p \geq \epsilon^p\}) \\ &\leq \frac{1}{\epsilon^p}\int |f_n(x) - f(x)|^p \, d\mu = \frac{1}{\epsilon^p}\|f_n - f\|_p^p \to 0\end{aligned}$$

as $n \to \infty$. Thus $f_n \xrightarrow{m} f$. By theorem 1.66, there is a subsequence f_{n_k} such that $f_{n_k} \to f$ a.e. on X. ■

Definition 2.26
A metric space (X, d) is ***separable*** *if there exists a countable dense subset.*

Theorem 2.27
Let $1 \leq p < \infty$. $\mathcal{L}^p(\mathbb{R})$ is separable.

Proof. Consider the collection of sets $\mathcal{I} = \{(q, r) \mid q < r \in \mathbb{Q}\}$. Then the family of functions $F = \left\{\sum_{i=1}^n c_i \chi_{I_i} \,\middle|\, I_i \in \mathcal{I}, c_i \in \mathbb{Q}, n \in \mathbb{N}\right\}$ is countable. We claim that F is dense in $\mathcal{L}^p(\mathbb{R})$. Indeed, since the continuous functions with compact supports are dense in $\mathcal{L}^p(\mathbb{R})$, it suffices to show that any such function can be approximated by functions in F. Let $f \in \mathcal{L}^p(\mathbb{R})$ be a continuous function with compact support. By the uniform continuity, there exists $\delta > 0$ such that for all $x, y \in \mathbb{R}$ with $|x - y| < \delta$, $|f(x) - f(y)| < \epsilon$.

Consider $\mathcal{I}' = \{I \in \mathcal{I} \mid I \cap \operatorname{supp}(f) \neq \varnothing, \mu I < \delta\}$, an open cover of $\operatorname{supp}(f)$. By the compactness of $\operatorname{supp}(f)$, we can find a finite subcover $\mathcal{I}'' = \{I_i \mid i = 1, \ldots, n\}$ such that $\operatorname{supp}(f) \subset \cup_{i=1}^n I_i$. By the density of $\mathbb{Q}$ in $\mathbb{R}$, we can find $c_i \in \mathbb{Q}$ such that $|f(x) - c_i| < \epsilon$ for all $x \in I_i$, for $i = 1, \ldots, n$. Let $g = \sum_{i=1}^n c_i \chi_{I_i} \in F$. Then $\|f - g\|_\infty < \epsilon$.

$$\|f - g\|_p^p = \int |f - g|^p \, d\mu \leq \int_{\operatorname{supp}(f-g)} \epsilon^p \, d\mu = \epsilon^p \mu(\operatorname{supp}(f - g)).$$

Since ϵ is arbitrary, we conclude that F is dense in $\mathcal{L}^p(\mathbb{R})$. Thus $\mathcal{L}^p(\mathbb{R})$ is separable. ■

Remark

$\mathcal{L}^\infty(\Omega,\mu)$ is not separable in general. For example, let $\Omega = [a,b]$. Suppose that $\{f_n\}$ is a countable dense subset of $\mathcal{L}^\infty(\Omega)$. Define $\eta : [a,b] \to \mathbb{N}$ such that $\left\|\chi_{[a,b]} - f_{\eta(x)}\right\| < \frac{1}{2}$. Then if $x_1 \neq x_2$, $\left\|\chi_{[a,x_1]} - \chi_{[a,x_2]}\right\|_\infty = 1$. This implies that $f_{\eta(x_1)} \neq f_{\eta(x_2)}$ and $\eta(x_1) \neq \eta(x_2)$. Thus η is injective. But $[a,b]$ is uncountable, a contradiction. Hence $\mathcal{L}^\infty(\Omega)$ is not separable.

2.4. Dual Space

Theorem 2.28 (Dualities of ℓ^p Spaces)
Let $1 < p < \infty$. Then $(\ell^p)' \cong \ell^{p'}$, where p' is the conjugate exponent of p.

Proof. We need to prove that there exists an isometric isomorphism $\psi : \ell^{p'} \to (\ell^p)'$ such that $\psi g f = \sum_i f_i g_i$ for all $g \in \ell^{p'}$ and $f \in \ell^p$. We show that ψ is well-defined, linear, bounded, bijective, and isometric.

First, we show that ψ is well-defined. For $f \in \ell^p$ and $g \in \ell^{p'}$,

$$|\psi g f| \leq \sum_i |f_i g_i| \leq \|f\|_p \|g\|_{p'} < \infty$$

by the Hölder's inequality. Thus $\psi g \in (\ell^p)'$ is well-defined.

Next, ψ is linear since for $g_1, g_2 \in \ell^{p'}$ and $c \in \mathbb{R}$,

$$\psi(cg_1 + g_2)(f) = \sum_i f_i(cg_{1i} + g_{2i}) = c\sum_i f_i g_{1i} + \sum_i f_i g_{2i} = c\psi g_1(f) + \psi g_2(f)$$

for all $f \in \ell^p$. Hence $\psi(cg_1 + g_2) = c\psi g_1 + \psi g_2$.

Now, to show that ψ is bounded,

$$\begin{aligned} \|\psi g\| &= \sup\left\{|\psi g f| \;\middle|\; \|f\|_p = 1\right\} = \sup\left\{\left|\sum_i f_i g_i\right| \;\middle|\; \|f\|_p = 1\right\} \\ &\leq \sup_{\|f\|_p = 1}\left\{\|g\|_{p'}\right\} \leq \|g\|_{p'}. \end{aligned}$$

We see that $\|\psi\| \leq 1$. Next, let $h \in (\ell^p)'$ and define g by $g_i = h(e_i)$. Then

$$\|g\|_{p'} = \left(\sum_i |g_i|^{p'}\right)^{1/p'} = \left(\sum_i |h(e_i)|^{p'}\right)^{1/p'} \leq \left(\sum_i \|h\|^{p'}\right)^{1/p'} = \|h\|.$$

Then $g \in \ell^{p'}$. Furthermore, for such g,

$$\psi g(f) = \sum_i f_i g_i = \sum_i f_i h(e_i) = h\left(\sum_i f_i e_i\right) = h(f)$$

for every $f \in \ell^p$. Hence ψ is surjective and $\|\psi g\| = \|h\|$. The isometry of ψ is immediate from that

$$\|\psi g\| \leq \|g\|_{p'} \leq \|h\| = \|\psi g\| .$$

Finally, ψ is injective since otherwise there exists $g \neq 0$ such that $\psi g = 0$. Then $\|g\|_{p'} = 0$ by the isometry of ψ, which implies that $g = 0$, a contradiction. We conclude that ψ is an isometric isomorphism and the proof is complete. ■

Proposition 2.29
$1 \leq p < \infty$. $1/p + 1/p' = 1$. *Let* $g \in \mathcal{L}^{p'}(X, \mu)$. *Then the mapping* $Tg : \mathcal{L}^p(X, \mu) \to \mathbb{R}$ *defined by*

$$Tg(f) = \int_X fg d\mu$$

is a bounded linear functional. Furthermore, $\|Tg\|_{\mathcal{L}^p \to \mathbb{R}} = \|g\|_{p'}$.

Proof. We start by checking that Tg is well-defined. For $f \in \mathcal{L}^p$,

$$|Tg(f)| = \left| \int fg d\mu \right| \leq \int |fg| \, d\mu \leq \|f\|_p \, \|g\|_{p'}$$

by Hölder's inequality. Thus $Tg(f) \in \mathbb{R}$. Also, we obtain that $\|Tg\|_{\mathcal{L}^p \to \mathbb{R}} \leq \|g\|_{p'}$. For the linearity, let $c \in \mathbb{R}$ and $f_1, f_2 \in \mathcal{L}^p$.

$$Tg(cf_1 + f_2) = \int (cf_1 + f_2) g d\mu = c \int f_1 g d\mu + \int f_2 g d\mu = cTg(f_1) + Tg(f_2).$$

Lastly, to furnish the isometry, let $g \neq 0$ and define

$$f = \operatorname{sgn}(g) \left(\frac{|g|}{\|g\|_{p'}} \right)^{p'/p} \implies \int |f|^p \, d\mu = \int \left(\frac{|g|}{\|g\|_{p'}} \right)^{p'} d\mu < \infty.$$

Then $f \in \mathcal{L}^p$ and $\|f\|_p = 1$. Also,

$$Tg(f) = \int \operatorname{sgn}(g) \left(\frac{|g|}{\|g\|_{p'}} \right)^{p'/p} g d\mu = \|g\|_{p'} .$$

It follows that $\|Tg\|_{\mathcal{L}^p \to \mathbb{R}} = \|g\|_{p'}$. ■

Theorem 2.30 (Riesz Representation)
Let $(X, \mathcal{A}, \mu)$ *be a* σ*-finite measure space and* $1 \leq p < \infty$. *Then the mapping* $T : \mathcal{L}^{p'}(X, \mu) \to (\mathcal{L}^p(X, \mu))'$ *defined by* $Tg \in \mathcal{L}^p(X, \mu)$,

$$Tg(f) = \int fg d\mu,$$

is an isometric isomorphism.

Proof. By proposition 2.29, Tg is a bounded linear functional. Besides, let $c \in \mathbb{R}$ and $g_1, g_2 \in \mathcal{L}^{p'}$,

$$T(cg_1+g_2)(f) = \int (cg_1+g_2)f d\mu = c\int g_1 f d\mu + \int g_2 f d\mu = cTg_1(f)+Tg_2(f) = (cTg_1+Tg_2)(f)$$

for all $f \in \mathcal{L}^p$. Thus T is linear. It remains to show that T is a bijection. We first verify that T is surjective.

Consider the case where $p > 1$ and $\mu(X) < \infty$. Let $h \in (\mathcal{L}^p)'$. Define $\nu : \mathcal{A} \to \mathbb{R}$ by $\nu(A) = h(\chi_A)$. We claim that ν is a finite measure and $\nu \ll \mu$. Since

$$|\nu(A)| = |h(\chi(A))| \leq \|h\|_{\mathcal{L}^p \to \mathbb{R}} \|\chi_A\|_p = \|h\|_{\mathcal{L}^p \to \mathbb{R}} \mu(A)^{1/p},$$

we see that ν is finite since so is μ. Also, if $\mu(A) = 0$, then $|\nu(A)| = 0$ and hence $\nu(A) = 0$. Thus $\nu \ll \mu$. For finite additivity, let $A_1, A_2 \in \mathcal{A}$ be disjoint.

$$\nu(A_1 \cup A_2) = h(\chi_{A_1 \cup A_2}) = h(\chi_{A_1} + \chi_{A_2}) = h(\chi_{A_1}) + h(\chi_{A_2}) = \nu(A_1) + \nu(A_2).$$

To show the σ-additivity, let $A_j \in \mathcal{A}$ be countably many disjoint sets. Put $A = \cup_j A_j$, $A = B_n + C_n$ where $B_n = \cup_{j=1}^n A_j$ and $C_n = \cup_{j=n+1}^\infty A_j$. Then since $B_n \cap C_n = \varnothing$,

$$\nu(A) = \nu(B_n + C_n) = \nu(B_n) + \nu(C_n) = \sum_{j=1}^n \nu(A_j) + \nu(C_n)$$

for all n. Since $\mu(X) < \infty$, $\sum_j \mu(A_j) < \infty$ and $\mu(C_n) \to 0$ as $n \to \infty$. Thus

$$|\nu(C_n)| = |h(C_n)| \leq \|h\|_{\mathcal{L}^p \to \mathbb{R}} \mu(C_n)^{1/p} \to \infty.$$

We conclude that $\nu(A) = \sum_j \nu(A_j)$ and ν is a measure.

Next, since $\nu \ll \lambda$, by the Radon-Nikodym theorem, there exists a unique $g \in \mathcal{L}^1(X, \mu)$ such that

$$h(\chi_A) = \nu(A) = \int_A g d\mu = \int_X \chi_A g d\mu = Tg(\chi_A).$$

for arbitrary $A \in \mathcal{A}$. Extend by linearity to p-integrable simple functions, say $s = \sum_{i=1}^n c_i \chi_{A_i}$.

$$h(s) = \sum_{i=1}^n c_i h(\chi_{A_i}) = \sum_{i=1}^n c_i \int_X \chi_{A_i} g d\mu = \int_X \sum_{i=1}^n c_i \chi_{A_i} g d\mu = \int_X s g d\mu = Tg(s).$$

For a general $f \in \mathcal{L}^p$, by separating $f = f^+ - f^-$ if necessary, we may assume that $f \geq 0$. By lemma 1.20, there exists a sequence of simple functions $s_n \nearrow f$. Then by Lebesgue's monotone convergence theorem, $\|f - s_n\|_p \to 0$. Since h is a bounded linear functional, it is continuous, and hence $h(s_n) \to h(f)$ as $n \to \infty$. We obtain that

$$h(f) = \lim_{n\to\infty} h(s_n) = \lim_{n\to\infty} \int_X s_n g d\mu = \int_X f g d\mu = Tg(f)$$

for all $f \in \mathcal{L}^p$. Thus $Tg = h$. It remains to check that $g \in \mathcal{L}^{p'}$. Let

$$f_n = \begin{cases} |g|^{p'-1} \operatorname{sgn}(g) & \text{if } |g(x)|^{p'-1} \le n, \\ n \operatorname{sgn}(g) & \text{otherwise.} \end{cases}$$

Then $f_n \in \mathcal{L}^p$ and $f_n g \nearrow |g|^{p'}$.

$$|Tg(f_n)| = \left| \int f_n g d\mu \right| \le \|Tg\|_{\mathcal{L}^p \to \mathbb{R}} \|f_n\|_p .$$

Also, $f_n g = |f_n|\,|g| \ge |f_n|\,|f_n|^{1/(p'-1)} = |f_n|^p$ and

$$\|f_n\|_p^p = \int |f_n|^p \, d\mu \le \int f_n g d\mu \le \|Tg\|_{\mathcal{L}^p \to \mathbb{R}} \|f_n\|_p .$$

As a result,

$$\|g\|_{p'}^{p'} = \int |g|^{p'} \, d\mu = \lim_{n\to\infty} \int f_n g d\mu \le \|Tg\|_{\mathcal{L}^p \to \mathbb{R}} \|f_n\|_p < \infty.$$

Hence $g \in \mathcal{L}^{p'}$ and T is indeed surjective. Furthermore, such g is unique by the uniqueness of the Radon-Nikodym derivative. We also conclude that T is injective.

For the case where $p = 1$ and $\mu(X) < \infty$, $p' = \infty$. We consider the same mapping T with $Tg(f) = \int fg d\mu$. We claim that $g \in \mathcal{L}^\infty$. Suppose $g \notin \mathcal{L}^\infty$. Then for every K, the set $A_K = \{x \in X \mid |g(x)| > K\}$ has positive measure. Define $f_K = \operatorname{sgn}(g)\chi_{A_K}/\mu(A_K)$. Note that $\|f_K\|_1 = 1$. If $g \ge 0$, then

$$|Tg(f_K)| = \int f_K g d\mu > K$$

for all K. But Tg is a bounded linear functional, which is a contradiction. Thus $g \in \mathcal{L}^\infty$.

Finally, we prove the case where X is σ-finite. Write $X = \cup_n X_n$ where $\mu(X_n) < \infty$ and $X_n \subset X_{n+1}$. For every $f \in \mathcal{L}^p(X_k, \mu)$, consider $\hat{f} \in \mathcal{L}^p(X, \mu)$ defined by $\hat{f} = f$ on X_k and $\hat{f} = 0$ on $X - X_k$. Then $\|f\|_{\mathcal{L}^p(X_k)} = \|f\|_{\mathcal{L}^p(X)}$. Let $h \in (\mathcal{L}^p(X))'$ and consider $h_k \in (\mathcal{L}^p(X_k))'$ by $h_k(f) = h(\hat{f})$. Then $\|h_k\| \le \|h\|$. By the previous result, we can find a unique $g_k \in \mathcal{L}^{p'}(X_k, \mu)$ such that

$$h_k(f) = \int f g_k d\mu, \|g_k\|_{\mathcal{L}^{p'}(X_k)} \le \|h_k\| \le \|h\| .$$

Since $X_n \subset X_{n+1}$, for $f \in \mathcal{L}^p(X_k)$, we have $h_k(f) = h(\hat{f}) = h_{k+1}(f)$ and $g_k = g_{k+1}$ μ-a.e. in X_k. Define $g = g_k$ on X_k with $\|g\|_{\mathcal{L}^{p'}(X)} \le \|h\|$. Let $f \in \mathcal{L}^p(X, \mu)$. Hölder's inequality implies that $fg \in \mathcal{L}^1(X, \mu)$ and

$$h(f\chi_{X_k}) = h_k(f) = \int f \chi_{X_k} g_k d\mu$$

Since $f\chi_{X_k} \le |f|$, $f\chi_k \to f \in \mathcal{L}^p(X,\mu)$ by Lebesgue's dominated convergence theorem. Also,

$$h_k(f) = \int f\chi_{X_k} g_k d\mu \to \int fgd\mu = Tg(f)$$

by Lebesgue's dominated convergence theorem. Thus T is indeed the desired isometric isomorphism. ∎

Remark
$(\mathcal{L}^\infty)' \not\cong \mathcal{L}^1$. Consider $C^\infty([-1,1])$, a subspace of $\mathcal{L}^\infty$. Define a linear functional $\delta : \mathbb{C}^\infty([-1,1]) \to \mathbb{R}$ by $\delta(f) = f(0)$. Clearly $\delta \in (\mathcal{L}^\infty)'$. Now suppose there exists $g \in \mathcal{L}^1$ such that $\delta(f) = \int_{-1}^1 fgdx$. Let $f = \chi_A$ where A is measurable. Then $f \in \mathcal{L}^\infty$ and by definition,

$$0 = f(0) = \delta(f) = \int_{-1}^1 fgdx = \int_A gdx.$$

Thus $g = 0$ a.e. and $\delta = 0$, a contradiction.

Definition 2.31
$M(X)$ is a space consisting of all finite signed measures. For $\nu \in M(X)$, the total variation norm of ν is defined by $\|\nu\| = \nu^+(X) + \nu^-(X)$, where ν^+ and ν^- are the Hahn-Jordan decompositions of ν.

Proposition 2.32
$M(X)$ with the total variation norm forms a Banach space.

Proof. Clearly, $M(X)$ forms a vector space. We check that $\|\cdot\|$ is indeed a norm. For $\nu \in M(X)$, clearly $\|\nu\| \ge 0$. If $\|\nu\| = 0$, then $\nu^+(X) = \nu^-(X) = 0$, $\nu^+(A)$ and $\nu^-(A)$ are zero for all $A \in \mathcal{A}$, and hence $\nu = 0$. Conversely, if $\nu = 0$, then so are ν^+ and ν^- and hence $\|\nu\| = 0$. For $c \in \mathbb{R}$,

$$\|c\nu\| = |c|\,\nu^+(X) + |c|\,\nu^-(X) = |c|\,(\nu^+(X) + \nu^-(X)) = |c|\,\|\nu\|.$$

Lastly, let $\nu, \mu \in M(X)$. Notice that $(\nu+\mu)^+ \le \nu^+ + \mu^+$ and $(\nu+\mu)^- \le \nu^- + \mu^-$. Thus

$$\|\nu+\mu\| = (\nu+\mu)^+(X) + (\nu+\mu)^-(X) \le \nu^+(X) + \mu^+(X) + \nu^-(X) + \mu^-(X) = \|\nu\| + \|\mu\|,$$

proving that $\|\cdot\|$ is indeed a norm.

For the completeness, let ν_n be a Cauchy sequence in $M(X)$. We define a measure ν by $\nu(A) = \lim_{n\to\infty} \nu_n(A)$ for all $A \in \mathcal{A}$. We claim that the limit exists and ν is indeed a finite signed measure. Since the sequence is Cauchy, for every $\epsilon > 0$, there exists N such that

$$(\nu_m - \nu_n)^+(X) + (\nu_m - \nu_n)^-(X) = \|\nu_m - \nu_n\| \le \epsilon$$

for all $m, n \ge N$. Since both $(\nu_m - \nu_n)^+$ and $(\nu_m - \nu_n)^-$ are positive measures, we have

$$(\nu_m - \nu_n)^+(A) \le (\nu_m - \nu_n)^+(X) \le \epsilon, \quad \text{and} \quad (\nu_m - \nu_n)^-(A) \le (\nu_m - \nu_n)^-(X) \le \epsilon$$

for every $A \in \mathcal{A}$. Thus

$$|\nu_m(A) - \nu_n(A)| = \left|(\nu_m - \nu_n)^+(A) - (\nu_m - \nu_n)^-(A)\right| \leq \epsilon.$$

It follows that for any fixed $A \in \mathcal{A}$, $\nu_n(A)$ is a Cauchy sequence in $\mathbb{R}$ and hence the limit exists. Also, taking $A = X$, we see that $\nu(X)$ is finite. To show that ν is a measure, first note that $\nu(\varnothing) = 0$. For finite additivity, let $A_1, A_2 \in \mathcal{A}$ be disjoint. Then

$$\nu(A_1 \cup A_2) = \lim_{n\to\infty} \nu_n(A_1 \cup A_2) = \lim_{n\to\infty} \nu_n(A_1) + \nu_n(A_2) = \nu(A_1) + \nu(A_2).$$

For the σ-additivity, let $A_n \in \mathcal{A}$ be countably many disjoint sets. Put $A = \cup_n A_n$, $A = B_n \cup C_n$ where $B_n = \cup_{j=1}^n A_j$ and $C_n = \cup_{j=n+1}^\infty A_j$. Since $\nu(X) < \infty$, $\sum_j \nu(A_j) < \infty$ and hence $\nu(C_n) \to 0$ as $n \to \infty$. Thus

$$\nu(A) = \nu(B_n) + \nu(C_n) = \sum_{j=1}^n \nu(A_j) + \nu(C_n)$$

for every n and by letting $n \to \infty$, we obtain $\nu(A) = \sum_j \nu(A_j)$. Finally, fix n and let $m \to \infty$,

$$\|\nu - \nu_n\| = \lim_{m\to\infty} \|\nu_m - \nu_n\| = \lim_{m\to\infty} |\nu_m(X) - \nu_n(X)| = |\nu(X) - \nu_n(X)| \leq \epsilon$$

for all $n \geq N$. Thus $\nu_n \to \nu$ in norm and $M(X)$ is complete. ∎

Definition 2.33
*Let $f : [a,b] \to \mathbb{R}$. The **variation** of f is defined by*

$$V_{\mathcal{P}}(f) = \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|,$$

*where $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_n = b\}$ is a partition of $[a,b]$. The **total variation** of f on $[a,b]$ is defined by*

$$V(f) = \sup_{\mathcal{P}} V_{\mathcal{P}}(f).$$

Definition 2.34
*The **bounded variation space** $BV([a,b])$ consists of all functions $f : [a,b] \to \mathbb{R}$ such that $V(f) < \infty$. For $f \in BV([a,b])$, the **total variation norm** is defined by $\|f\|_{TV} = |f(a)| + V(f)$.*

Proposition 2.35
$BV([a,b])$ with the total variation norm forms a Banach space.

Proof. It clearly forms a vector space. We check that $\|\cdot\|_{TV}$ is indeed a norm. First, clearly $\|f\|_{TV} \geq 0$. If $\|f\|_{TV} = 0$, then $f(a) = 0$ and $f(t) = f(t')$ for all $t, t' \in [a,b]$. Hence $f = 0$; if $f = 0$, then $V(f) = 0$ and $f(a) = 0$ and $\|f\|_{TV} = 0$. Next, for $c \in \mathbb{R}$,

$$\|cf\|_{TV} = |cf(a)| + \sum_{i=0}^{n-1} |cf(t_{i+1}) - cf(t_i)| = |c| \left(|f(a)| + \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)| \right) = |c| \, \|f\|_{TV}.$$

Lastly, let $f, g \in BV([a,b])$. Then

$$\begin{aligned}
\|f+g\|_{TV} &= \sup_{\mathcal{P}} |(f+g)(a)| + \sum_{i=0}^{n-1} |(f+g)(t_{i+1}) - (f+g)(t_i)| \\
&\le \sup_{\mathcal{P}} |f(a)| + |g(a)| + \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)| + \sum_{i=0}^{n-1} |g(t_{i+1}) - g(t_i)| \\
&\le \sup_{\mathcal{P}} |f(a)| + \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)| + \sup_{\mathcal{P}} |g(a)| + \sum_{i=0}^{n-1} |g(t_{i+1}) - g(t_i)| = \|f\|_{TV} + \|g\|_{TV} .
\end{aligned}$$

Thus $\|\cdot\|_{TV}$ is indeed a norm.

For the completeness, let f_n be a Cauchy sequence in $BV([a,b])$. For $\epsilon > 0$, there exists N such that $\|f_m - f_n\|_{TV} < \epsilon$ for all $m, n \ge N$. Given any $x \in [a,b]$, consider the partition $\mathcal{P} = \{a < x < b\}$.

$$\begin{aligned}
|f_m(x) - f_n(x)| &= |f_m(x) - f_m(a) + f_m(a) - f_n(a) + f_n(a) - f_n(x)| \\
&\le |((f_m(x) - f_n(x))) - (f_m(a) - f_n(a))| + |f_m(a) - f_n(a)| \\
&\le V(f_m - f_n) + |f_m(a) - f_n(a)| = \epsilon.
\end{aligned}$$

Thus $\{f_n(x)\}$ is a Cauchy sequence in $\mathbb{R}$ and hence converges pointwisely to, say $f(x)$. Furthermore, observe that the choice of N does not depend on x, and thus the convergence is uniform. We claim that $f \in BV([a,b])$. Indeed, for any partition $\mathcal{P} = \{a = t_0 < \cdots < t_n = b\}$,

$$\sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)| \le \sum_{i=0}^{n-1} |f(t_{i+1}) - f_N(t_{i+1})| + \sum_{i=0}^{n-1} |f(t_i) - f_N(t_i)| + V(f_N).$$

Since the convergence is uniform, we can choose N such that $|f(t) - f_N(t)| \le \epsilon/(2n)$. Thus

$$\sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)| \le \epsilon + V(f_N).$$

Since f_N is of bounded variation, we see that $f \in BV([a,b])$ as well. Lastly, to show that $\|f - f_n\|_{TV} \to 0$, first note that by definition we have $|f_n(a) - f(a)| \to 0$. It remains to show that $V(f_n - f) \to 0$. For any $\epsilon > 0$, there exists N such that $V_{\mathcal{P}}(f_m - f_n) < \epsilon$ for all $m, n \ge N$ and some partition $\mathcal{P}$. Taking $m \to \infty$, we obtain $V_{\mathcal{P}}(f - f_n) < \epsilon$ for all $n \ge N$. Since the partition is arbitrary, we have $V(f - f_n) < \epsilon$ for all $n \ge N$. Thus $f_n \to f$ in $BV([a,b])$ and $BV([a,b])$ is complete. ■

Theorem 2.36

$M([a,b])$ is isometrically isomorphic to $BV([a,b])$.

Proof. We define the mapping $\phi : M([a,b]) \to BV([a,b])$ by

$$\rho(t) = \phi\nu(t) = \nu([a,t]).$$

First, we show that $\rho \in BV([a,b])$. For any partition $\mathcal{P} = \{a = t_0 < \cdots < t_n = b\}$,

$$\begin{aligned}\sum_{i=0}^{n-1} |\rho(t_{i+1}) - \rho(t_i)| + |\rho(a)| &= \sum_{i=0}^{n-1} |\nu([a, t_{i+1}]) - \nu([a, t_i])| + |\nu(\{a\})| \\ &= \sum_{i=0}^{n-1} |\nu((t_i, t_{i+1}])| + |\nu(\{a\})| \\ &= \sum_{i=0}^{n-1} |\nu| \left((t_i, t_{i+1}]\right) + |\nu| \left(\{a\}\right) = |\nu| \left([a,b]\right) = \|\nu\| .\end{aligned}$$

Since ν is a finite signed measure, $\rho \in BV([a,b])$. Furthermore, taking supremum over all partitions, we obtain that $\|\rho\|_{TV} = \|\nu\|$. It remains to show that ϕ is an isomorphism. Suppose $\nu, \mu \in M([a,b])$ and $\phi\nu = \phi\mu$. Then $\nu([a,t]) = \mu([a,t])$ for all $t \in [a,b]$. Since $[a,t]$ generates the Borel σ-algebra on $[a,b]$, we have $\nu = \mu$. Thus ϕ is injective. For surjectivity, let $\rho \in BV([a,b])$. Consider the signed measure ν defined by $\nu([a,t]) = \rho(t)$ and $\nu(\varnothing) = 0$. Then ν is a finite signed measure and $\phi\nu = \rho$. The proof is complete. ■

Lemma 2.37

Let X be a normed vector space and $M \subset X$ be a proper subspace. Suppose $S : M \to \mathbb{R}$ is a bounded linear functional. Then for every $x \in X \setminus M$, there exists a linear $U : M' \to \mathbb{R}$ such that $\|U\|_{M' \to \mathbb{R}} = \|S\|_{M \to \mathbb{R}}$, where $M' = M + \mathbb{R}x$.

Proof. Clearly M' is a subspace; furthermore, $M' = M \bigoplus \mathbb{R}x$ since if $v = w + cx = w' + c'x$ for some $w, w' \in M$ and $c, c' \in \mathbb{R}$, then $(c - c')x = w - w' \in M$. Since $x \notin M$, this implies that $c = c'$, $w = w'$ and hence the representation is unique.

Now we can define U on M' by $U(w + cx) = Sw + c\lambda$ for any $w + cx \in M'$ and some $\lambda \in \mathbb{R}$ to be determined. To make U have the same norm as U, we need to find λ such that $|Sw + c\lambda| \leq \|S\| \|w + cx\|$ holds for all $w \in M$ and $c \in \mathbb{R}$. Clearly if $c = 0$, the inequality is already satisfied. For $c \neq 0$, by deviding both sides by $|c|$, we see that the condition is equivalent to $|Sw + \lambda| \leq \|S\| \|w + x\|$ for all $w \in M$. Now for any $w, v \in M$,

$$Sw - Sv = S(w - v) \leq |S(w - v)| \leq \|S\| \|w - v\| = \|S\| \|w + x - (v + x)\| \leq \|S\| \left(\|w + x\| + \|v + x\|\right).$$

Thus

$$Sw - \|S\| \|w + x\| \leq Sv + \|S\| \|v + x\| .$$

Fix v and taking supremum over all $w \in M$ on the left,

$$\sup_{w \in M} Sw - \|S\| \|w + x\| \leq Sv + \|S\| \|v + x\| .$$

Taking infimum over all $v \in M$ on the right,

$$\sup_{w \in M} Sw - \|S\| \|w + x\| \leq \inf_{v \in M} Sv + \|S\| \|v + x\| .$$

Hence there exists $\lambda \in \mathbb{R}$ such that

$$S(w) - \|S\|\,\|w + x\| \le -\lambda \le S(w) + \|S\|\,\|w + x\|$$

for all $w \in M$. Picking this λ, we see that

$$|Sw + \lambda| \le \|S\|\,\|w + x\|$$

as desired. Thus U is a bounded linear functional on M' with $\|U\|_{M' \to \mathbb{R}} = \|S\|_{M \to \mathbb{R}}$. Also, on M, $U = S$ and hence U is an extension of S. ■

Theorem 2.38 (Hahn-Banach)
Let X be a normed vector space and $M \subset X$ be a subspace. Suppose $S : M \to \mathbb{R}$ is a bounded linear functional on M. Then there exists a bounded linear functional $T : X \to \mathbb{R}$ such that $T|_M = S$ and $\|T\|_{X \to \mathbb{R}} = \|S\|_{M \to \mathbb{R}}$.

Proof. The proof relies on Zorn's lemma.[1] We start by constructing a partial order space. Let $(P, \preceq)$ be a partial order space with

$$P = \{(U, Y) \mid M \subset Y \subset X, Y \text{ is a subspace of } X, U \text{ is a bounded extension of } S \text{ on } V\}$$

and the partial order: $(U_1, Y_1) \preceq (U_2, Y_2)$ if $Y_1 \subset Y_2$ and U_2 is a bounded extension of U_1 on Y_2. Clearly the pair indeed forms a partial order space. We now check the assumptions of Zorn's lemma. Let $C = \{(U_\alpha, Y_\alpha) \mid \alpha \in A\}$ with an arbitrary index set A be a chain in P. Put $Y = \cup_{\alpha \in A} Y_\alpha$. We claim that Y is a subspace of X. Indeed, for $y_1, y_2 \in Y$ and $c_1, c_2 \in \mathbb{R}$, there exist $\alpha_1, \alpha_2 \in A$ such that $y_1 \in Y_{\alpha_1}$ and $y_2 \in Y_{\alpha_2}$. Since Y is a chain, one of them is a subspace of the other, say Y_{α_1} is a subspace of Y_{α_2}. Then $y_1, y_2 \in Y_{\alpha_2}$ and hence $c_1 y_1 + c_2 y_2 \in Y_2 \subset Y$. Thus Y is a subspace.

Next we need to define a bounded linear functional U on Y so that U is a bounded extension of S on Y. For $y \in Y$, we can find an $\alpha \in A$ such that $y \in Y_\alpha$ and set $U(y) = U_\alpha(y)$. Such U is well-defined since if α_1 and α_2 are two indices satisfying $y \in Y_{\alpha_1} \cap Y_{\alpha_2}$, then $U_{\alpha_1}(y) = U_{\alpha_2}(y)$ since one of them is an extension of the other. Also, U is linear since U_α is linear for every $\alpha \in A$. Lastly, U is a bounded extension of U_α on Y for any $\alpha \in A$ because every $U_{\alpha'}$ with $(U_\alpha, Y_\alpha) \preceq (U_{\alpha'}, Y_{\alpha'})$ is a bounded extension of U_α. We conclude that $(U, Y) \in P$ is an upper bound of C.

By Zorn's lemma, there exists a maximal element $(T, Z) \in P$. We claim that $Z = X$. Suppose $Z \subsetneq X$. Then there exists $x \in X \setminus Z$ and also a bounded extension T' of T on $Z + \mathbb{R}x \supsetneq Z$ by lemma 2.37. But then $(T', Z + \mathbb{R}x) \in P$ and $(T, Z) \preceq (T', Z + \mathbb{R}x)$, contradicting the maximality of (T, Z). Thus $Z = X$ and T is a bounded extension of S on X. ■

[1] Zorn's lemma states that if every chain in a partially ordered set has an upper bound, then the set has a maximal element. It is a direct consequence of the axiom of choice.

Theorem 2.39 (Riesz Representation of $C([a,b])$)
$C([a,b])' \cong BV([a,b]) \cong M([a,b])$ isometrically.

Proof. In theorem 2.36, we have shown that $M([a,b]) \cong BV([a,b])$. We are going to show this by constructing an isometric isomorphism between $C([a,b])'$ and $BV([a,b])$.

Let $X = C([a,b])$ and $\ell \in X'$. $\ell : X \to \mathbb{R}$ is a bounded linear functional. We need to find a $\nu \in M([a,b])$ such that

$$\ell(f) = \int_{[a,b]} f d\nu$$

for $f \in C([a,b])$. Let $Y = B([a,b]) = \{f : [a,b] \to \mathbb{R} \mid f \text{ is bounded}\}$. By Hahn-Banach theorem, there exists a bounded linear extension $L : Y \to \mathbb{R}$ of ℓ. Now if $f = \chi_{[a,t]} \in Y$, then

$$L(f) = \int_{[a,b]} \chi_{[a,t]} d\nu = \nu([a,t]) = \rho(t).$$

We claim that $\rho \in BV([a,b])$. For any partition $\mathcal{P} = \{a = t_0 < \cdots < t_n = b\}$,

$$\begin{aligned} V_{\mathcal{P}}(\rho) &= \sum_{i=0}^{n-1} |\rho(t_{i+1}) - \rho(t_i)| = \sum_{i=0}^{n-1} \left| L(\chi_{[a,t_{i+1}]}) - L(\chi_{[a,t_i]}) \right| \\ &= \sum_{i=0}^{n-1} L(\chi_{(t_i,t_{i+1}]}) s_i = L\left(\sum_{i=0}^{n-1} \chi_{(t_i,t_{i+1}]} s_i \right) \le \|L\| \left\| \sum_{i=0}^{n-1} \chi_{(t_i,t_{i+1}]} s_i \right\|_\infty \le \|L\| \end{aligned}$$

by letting $s_i = \operatorname{sgn}(\rho(t_{i+1}) - \rho(t_i))$. Thus $\rho \in BV([a,b])$ and $\|\rho\|_{TV} \le \|L\| = \|\ell\|$. To extend to $f \in C([a,b])$ so that

$$\ell(f) = L(f) = \int_{[a,b]} f d\nu,$$

we first note that by our established result, $f = \chi_{[a,t]} \in Y$ holds. By linearity so does simple functions. For $f \in C([a,b])$, consider

$$h_{\mathcal{P}}(t) = f(a) + \sum_{i=0}^{n-1} f(t_i) \chi_{(t_i,t_{i+1}]}(t).$$

Since L is continuous and $h_{\mathcal{P}} \to f$ uniformly as $\|\mathcal{P}\| \to 0$, we have

$$L(f) = \lim_{\|\mathcal{P}\| \to 0} L(h_{\mathcal{P}}) = \int_a^b f d\rho.$$

L is an extension of ℓ and hence

$$\ell(f) = \int_a^b f d\rho = f(a)\rho(a) + \int_a^b f d\rho.$$

Finally, we claim that $\|\ell\| \leq \|\rho\|_{TV} \leq \|L\| = \|\ell\|$. Take $f \in X$.

$$|\ell(f)| = \left|\int_a^b f d\rho\right| \leq \|f\|_\infty \|\rho\|_{TV} \leq \|f\|_\infty \|L\| = \|\ell\| \|f\|_\infty .$$

Hence $\|\ell\| \leq \|\rho\|_{TV} \leq \|L\| = \|\ell\|$. It follows that the mapping $\ell \mapsto \rho$ is isometric. Conversely, if $\rho \in BV([a,b])$, define

$$\ell_\rho(f) = f(a)\rho(a) + \int_a^b f d\rho.$$

We need to check that ℓ_ρ is linear and $\|\rho\|_{TV} \leq \|\ell\| \leq \|\rho\|_{TV}$. ℓ_ρ has an extension $L_\rho : Y \to \mathbb{R}$. Define $\lambda(t) = L_\rho(\chi_{[a,t]})$. Then $\|\rho\|_{TV} = \|\lambda\| \leq \|L_\rho\| = \|\ell_\rho\|$. ■

Remark
If $\ell \in C([a,b])'$, there exists $\rho \in BV([a,b])$ such that

$$\ell(f) = \int_a^b f d\rho;$$

if $\rho \in BV([a,b])$,

$$\ell_\rho(f) = f(a)\rho(a) + \int_a^b f d\rho$$

and $\|\ell_\rho\| = \|\rho\|_{TV}$.

Definition 2.40
Let X be a Banach space and $J : X \to X''$, the canonical mapping defined by $J : x \mapsto (T \mapsto Tx)$ for $T \in X'$. X is called ***reflexive*** *if J is surjective.*

Remark
Intuitively, X is reflexive meaning that $X \cong X''$. Such canonical mapping is well-defined since $\hat{x} = T \mapsto T(x)$ is indeed a bounded linear functional on X', which we verify here. This is linear because

$$\hat{x}(cT + S) = (cT + S)(x) = cT(x) + S(x) = c\hat{x}(T) + \hat{x}(S)$$

for $c \in \mathbb{R}$ and $S \in X'$. To show that $\hat{x}$ is bounded, we have

$$|\hat{x}(T)| = |T(x)| \leq \|T\| \|x\| = \|T\|_{X'} \|x\|_X$$

Definition 2.41
A Banach space X is said to be ***uniformly convex*** *if for all $\epsilon > 0$, $x, y \in X$ with $\|x - y\| \geq \epsilon > 0$ and $\|x\|, \|y\| \leq 1$, we have*

$$\left\|\frac{x+y}{2}\right\| \leq 1 - \delta$$

for some $\delta > 0$ depending on ϵ.

Remark
An equibalent definition of the uniform convexity is that for $\|x\|, \|y\| \leq 1$ with $\left\|\frac{x+y}{2}\right\| > 1 - \delta$

for some $\delta = \delta(\epsilon) > 0$, then $\|x - y\| < \epsilon$. Indeed, if $\|x - y\| \geq \epsilon$, then from the definition of uniform convexity, we have

$$\left\|\frac{x+y}{2}\right\| \leq 1 - \delta.$$

By contrapositive, if $\left\|\frac{x+y}{2}\right\| > 1 - \delta$, then $\|x - y\| < \epsilon$.

Theorem 2.42 (Clarkson)
$\mathcal{L}^p(\Omega, \mu)$ is uniformly convex for $1 < p < \infty$.

Proof. Consider the function $\alpha : \mathbb{R}^2 \to [0, \infty)$ defined by

$$\alpha(x) = \frac{|x_1|^p + |x_2|^p}{2} - \left|\frac{x_1 + x_2}{2}\right|^p$$

for $x = (x_1, x_2) \in \mathbb{R}^2$. Observe that $\alpha(x) \geq 0$ for all $x \in \mathbb{R}^2$ and $\alpha(x) = 0$ if and only if $x_1 = x_2$ by the strict convexity of $|\cdot|^p$. Now given $\epsilon > 0$, choose $\eta < \epsilon^p/2$. Consider the set $D = \left\{x \in \mathbb{R}^2 \;\middle|\; |x_1|^p + |x_2|^p = 1,\ |x_1 - x_2|^p \geq \eta\right\}$. Then D is a compact set. By the compactness, we can set $\theta = \inf_{x \in D} \alpha(x) > 0$. We now claim that if $x \in \mathbb{R}^2$ satisfies $|x_1 - x_2|^p \geq \eta(|x_1|^p + |x_2|^p)$, then

$$|x_1|^p + |x_2|^p \leq \frac{\alpha(x)}{\theta}.$$

By the assumption, we may assume that $x \neq (0, 0)$. Set $t = (|x_1|^p + |x_2|^p)^{1/p}$. Then

$$\left|\frac{x_1}{t}\right|^p + \left|\frac{x_2}{t}\right|^p = 1, \quad \text{and} \quad \left|\frac{x_1}{t} - \frac{x_2}{t}\right|^p \geq \eta.$$

Thus

$$\theta \leq \alpha\left(\frac{x}{t}\right) = \frac{\alpha(x)}{t^p} \Rightarrow |x_1|^p + |x_2|^p = t^p \leq \frac{\alpha(x)}{\theta}.$$

The claim follows.

Now let $f, g \in \mathcal{L}^p(\Omega, \mu)$ with $\|f\|_p, \|g\|_p \leq 1$ and $\|f - g\|_p \geq \epsilon$. Put

$$E = \{x \in \Omega \mid |f(x) - g(x)|^p \geq \eta(|f(x)|^p + |g(x)|^p)\}.$$

Using the claim and $\alpha(x) \geq 0$,

$$\begin{aligned}
\epsilon^p &\leq \int_\Omega |f - g|^p \, d\mu = \int_E |f - g|^p \, d\mu + \int_{E^c} |f - g|^p \, d\mu \\
&\leq 2^p \int_E \left|\frac{f - g}{2}\right|^p d\mu + \eta \int_{E^c} |f|^p \, d\mu + \eta \int_{E^c} |g|^p \, d\mu \\
&\leq 2^p \int_E \frac{|f|^p + |g|^p}{2} d\mu + 2\eta \\
&\leq 2^{p-1} \int_E \frac{\alpha(f, g)}{\theta} d\mu + 2\eta \\
&\leq \frac{2^{p-1}}{\theta} \int_\Omega \frac{|f|^p + |g|^p}{2} - \left|\frac{f + g}{2}\right|^p d\mu + 2\eta \leq \frac{2^{p-1}}{\theta} + 2\eta - \frac{2^{p-1}}{\theta} \left\|\frac{f + g}{2}\right\|_p^p.
\end{aligned}$$

Hence

$$\left\|\frac{f+g}{2}\right\|_p^p \le 1-\theta\frac{\epsilon^p-2\eta}{2^{p-1}} \Rightarrow \left\|\frac{f+g}{2}\right\|_p \le 1-\delta$$

for some $\delta > 0$. We conclude that $\mathcal{L}^p(\Omega,\mu)$ is uniformly convex for $1 < p < \infty$. ■

Example

Fix $x \in X$ and define the functional $L_x : X' \to \mathbb{R}$ by $L_x(\ell) = \ell(x)$. Then L_x is a bounded linear functional. To see this, let $c \in \mathbb{R}$ and $\ell_1, \ell_2 \in X'$.

$$L_x(c\ell_1+\ell_2) = (c\ell_1+\ell_2)(x) = c\ell_1(x)+\ell_2(x) = cL_x(\ell_1)+L_x(\ell_2).$$

And also

$$\|L_x\| = \sup_{\|\ell\|=1}|L_x(\ell)| = \sup_{\|\ell\|=1}|\ell(x)| \le \sup_{\|\ell\|=1}\|\ell\|\,\|x\| = \|x\|\,.$$

In fact, we have $\|L_x\| = \|x\|$. To see this, consider the one-dimensional subspace $Y = \mathbb{R}x$ and the functional $s : Y \to \mathbb{R}$ defined by $s(\lambda x) = \lambda\|x\|$ for some $\lambda \neq 0$. Then by the Hahn-Banach theorem, there exists a bounded linear functional $s' : X \to \mathbb{R}$ such that $s'|_Y = s$ with $\|s'\|_X = \|s\|_Y$. Then $s' \in X'$ and

$$\|L_x\| \ge |L_x(s')| = |s'(x)| = \|x\|\,.$$

Hence $\|L_x\| = \|x\|$.

Remark

Another important observation from the above example is that $\|x\| = \sup_{\|\ell\|=1}|\ell(x)|$.

Proposition 2.43

If X is a finite-dimensional Banach space, then X is reflexive.

Proof. Let X be a finite-dimensional Banach space. Then X is isomorphic to $\mathbb{R}^n$ for some $n \in \mathbb{N}$. Since the dual of $\mathbb{R}^n$ is also $\mathbb{R}^n$, we have $X' \cong X$. Thus $X'' \cong X' \cong X$. Hence X is reflexive. ■

Example

Consider the space $C[-1,1]$ with $\|f\| = \sup_{x\in[-1,1]}|f(x)|$. Then $C[-1,1]$ is a Banach space, but not reflexive. Suppose $C[-1,1]$ is reflexive. Write $(C[-1,1])'' \cong C[-1,1]$. Then the mapping $L \mapsto L_x$ is an isomorphism, where $L : (C[-1,1])' \to \mathbb{R}$, $L \in (C[-1,1])''$. Then, for all $\ell \in C[-1,1]'$, there exists an $x \in C[-1,1]$ and $L_x \in C[-1,1]''$ such that

$$\begin{cases} \|L_x\| = 1, & \textit{and } \|\ell\| = L_x(\ell), \\ \|x\| = 1, & \textit{and } \|\ell\| = \ell(x). \end{cases}$$

Now consider the functional

$$\ell(g) = \int_{-1}^{0} g(x)dx - \int_{0}^{1} g(x)dx.$$

Clearly ℓ is a bounded linear functional on $C[-1,1]$.

$$|\ell(g)| \leq \int_{-1}^{0} |g(x)|\, dx + \int_{0}^{1} |g(x)|\, dx \leq 2 \sup_{x\in[-1,1]} |g(x)| = 2\, \|g\|\,.$$

Thus $\|\ell\| \leq 2$. In fact, we have $\|\ell\| = 2$ by considering the continuous functions

$$g_\epsilon(x) = \begin{cases} 1, & x \in [-1,-\epsilon], \\ -\frac{x}{\epsilon}, & x \in [-\epsilon,\epsilon], \\ -1, & x \in [\epsilon, 1]. \end{cases}$$

Then $\|g_\epsilon\| = 1$ and $\ell(g_\epsilon) = 2 - \epsilon \to 2$ as $\epsilon \to 0$.

However, if $f \in C[-1,1]$ has $\|f\| = 1$, then there exists an interval $I \subset [-1,1]$ with $\mu(I) > 0$ such that $\sup_I |f| < 1 - \delta$, where $\delta > 0$. Then

$$\begin{aligned} |\ell(f)| &= \left| \int_{[-1,0]-I} f(x)dx + \int_{[-1,0]\cap I} f(x)dx - \int_{[0,1]\cap I} f(x)dx - \int_{[0,1]-I} f(x)dx \right| \\ &\leq \mu([-1,0]-I) + \mu([-1,0]\cap I)(1-\delta) + \mu([0,1]\cap I)(1-\delta) + \mu([0,1]-I) \\ &= 2 - \delta\mu(I). \end{aligned}$$

This contradicts the fact that there is an $x \in C[-1,1]$ such that $\|x\| = 1$ and $\|\ell\| = \ell(x)$. Thus $C[-1,1]$ is not reflexive.

Theorem 2.44
Let X be a Banach space. If X' is separable, then X is also separable.

Proof. By the assumption, let $\{\ell_n\} \subset X'$ be a countable dense subset. By definition, since $\|\ell_n\| = \sup_{\|z\|=1} |\ell_n(z)|$, for each $n \in \mathbb{N}$, we can find a $z_n \in X$ such that $\|z_n\| = 1$ and $|\ell_n(z_n)| \geq \frac{1}{2}\|\ell_n\|$.

Now we claim that $Y = \operatorname{span}\{z_n\}$ is dense in X. Suppose not. Then there is an $x \in X \setminus Y$. Consider the space $W = \{cx + y \mid c \in \mathbb{R},\ y \in Y\}$. Define a linear functional $\ell(v) = c\ell(x) \neq 0$ for $v = cx + y$ with $c \neq 0$ on W. By the Hahn-Banach theorem, we can extend ℓ to X. For such ℓ on X, we have $\ell(z_n) = 0$ for all n since $z_n \in Y$. Assume without loss of generality that $\|\ell\| = 1$ and $\|\ell_n - \ell\| \leq \epsilon$ by the density of $\{\ell_n\}$ in X'. Hence

$$\|\ell_n\| \geq \|\ell\| - \|\ell_n - \ell\| \geq 1 - \epsilon.$$

On the other hand,

$$\|\ell_n\| \leq 2\,|\ell_n(z_n)| = 2\,|\ell_n(z_n) - \ell(z_n)| \leq 2\,\|\ell_n - \ell\|\,\|z_n\| \leq 2\epsilon.$$

This implies that $1 \leq 3\epsilon$. Picking $\epsilon < 1/3$ leads to a contradiction. Hence Y is dense in X.

Finally, write $X = \overline{Y} = \overline{\left\{\sum_{k=1}^{M} \alpha_k z_k \;\middle|\; M \geq 1, \alpha_k \in \mathbb{R}\right\}} = \overline{\left\{\sum_{k=1}^{M} \alpha_k z_k \;\middle|\; M \geq 1, \alpha_k \in \mathbb{Q}\right\}}$. Thus

X is separable. ■

Theorem 2.45

Let X be a reflexive Banach space. If $Y \subset X$ is a closed subspace, then Y is reflexive.

Proof. Fix a bounded lineal functional $L : Y' \to \mathbb{R}$. We want to show that there exists a unique $z \in Y$ such that $L(\ell) = L_z(\ell) = \ell(z)$ for all $\ell \in Y'$. Suppose $\ell : X \to \mathbb{R}$ is a bounded linear functional on X. Consider its restriction on Y, $\ell|_Y$. Note that $\|\ell|_Y\| \leq \|\ell\|$.

Now for L, we cna extend by Hahn-Banach theorem to $L_0 : X' \to \mathbb{R}$. For $m \in X'$, $m : X \to \mathbb{R}$, consider its restriction on Y, $m|_Y$. Then $L_0(m) = L(m|_Y)$. We check that L_0 is linear and bounded. For $c \in \mathbb{R}$ and $m, \ell \in X'$,

$$L_0(cm + \ell) = L((cm + \ell)|_Y) = L(cm|_Y + \ell|_Y) = cL(m|_Y) + L(\ell|_Y) = cL_0(m) + L_0(\ell).$$

And also

$$|L_0(m)| = |L(m|_Y)| \leq \|L\| \, \|m|_Y\| \leq \|L\| \, \|m\| \, . \Rightarrow \|L_0\| \leq \|L\| \, .$$

Thus L_0 is a bounded linear functional on X'. We now use the reflexivity of X. Since $X'' \cong X$, there exists a $z \in X$ such that $L_0(m) = L_z(m)$ for all $m \in X'$.

We claim that $z \in Y$. Suppose not. Then there exists a bounded linear functional $m : \{cz + y \mid c \in \mathbb{R},\ y \in Y\} \to \mathbb{R}$ such that $m(z) \neq 0$ and $m(y) = 0$ for all $y \in Y$. Extend m to $m_0 : X \to \mathbb{R}$ by Hahn-Banach theorem. Then

$$L_0(m_0) = L(m_0|_Y) = L(0) = 0 \neq m_0(z) = L_0(m_0),$$

which is absurd. Hence $z \in Y$ and we see that $L(m) = m(z)$ for all $m \in Y'$. Take $m \in Y'$ and its extension $m_0 \in X'$. If $m_0, m_0' \in X'$ are two extensions of m, then $L_0(m_0) = L_0(m_0')$ and hence $L(m) = L_0(m_0) = m_0(z) = m(z)$. Thus the extension is unique. We conclude that Y is a reflexive Banach space. ■

Definition 2.46

*Let X be a Banach space and $Y \subset X$ be a closed subset. Then the **quotient space** X/Y is defined as*

$$X/Y = \{x + Y \mid x \in X\}$$

with the norm $\|x + Y\| = \inf_{y \in Y} \|x + y\|$.

Remark

In the quotient space X/Y, two elements $x_1 + Y$ and $x_2 + Y$ are equal if $x_1 - x_2 \in Y$.

Remark

For any $T \in B(X, Y)$, consider its kernel $\ker(T) \subset X$. By proposition 2.9, $\ker(T)$ is closed, and thus $X/\ker(T)$ is well-defined.

Proposition 2.47
Let $T \in B(X, Y)$ be a bounded linear operator. Define $T_0 : X/\ker(T) \to Y$ by $T_0 : x + \ker(T) \mapsto Tx$. Then T_0 is a bounded linear operator with $\|T_0\| = \|T\|$.

Proof. We first check that T_0 is well-defined. Suppose $x_1 + Y = x_2 + Y$. Then $x_1 - x_2 \in \ker(T)$ and hence $T(x_1 - x_2) = 0$. Thus $Tx_1 = Tx_2$ and

$$T_0(x_1 + \ker(T)) = Tx_1 = Tx_2 = T_0(x_2 + \ker(T)).$$

Next, T_0 is clearly linear. For $x + \ker(T) \in X/\ker(T)$ and any $\epsilon > 0$, there exists $x_0 \in X$ such that $\|x + x_0\| \leq \|x + \ker(T)\| + \epsilon$ by the definition of the norm on the quotient space. Then

$$\|T_0(x + \ker(T))\| = \|Tx\| = \|T(x + x_0)\| \leq \|T\| \, \|x + x_0\| \leq \|T\| \, (\|x + \ker(T)\| + \epsilon).$$

Since ϵ is arbitrary, we have $\|T_0(x + \ker(T))\| \leq \|T\| \, \|x + \ker(T)\|$. This shows that T_0 is bounded and $\|T_0\| \leq \|T\|$. Conversely, notice that $0 \in \ker(T)$. Thus

$$\|Tx\| = \|T_0(x + \ker(T))\| \leq \|T_0\| \, \|x + \ker(T)\| \leq \|T_0\| \, \|x + 0\| = \|T_0\| \, \|x\| \, .$$

Hence $\|T\| \leq \|T_0\|$. We conclude that $\|T_0\| = \|T\|$. ■

Remark
T_0 is injective and $T_0(X/\ker(T)) = T(X)$.

Definition 2.48
*Let X, Y be two Banach spaces and $T \in B(X, Y)$. The **transpose** of T is defined as $T' : Y' \to X'$ by $T' : \ell \mapsto \ell T \in X'$.*

Remark
$T'\ell = \ell T$.

Proposition 2.49
Suppose $T \in B(X, Y)$. Then $T' : Y' \to X'$ is a bounded linear operator with $\|T'\| = \|T\|$.

Proof. The linearity of T' is trivial. By definition,

$$\|T'\| = \sup_{\|\ell\|=1} \|T'\ell\| = \sup_{\|\ell\|=1} \|\ell T\| = \sup_{\|\ell\|=1} \sup_{\|x\|=1} |\ell(Tx)| \leq \sup_{\|\ell\|=1} \sup_{\|x\|=1} \|\ell\| \, \|T\| \, \|x\| = \|T\| \, .$$

Conversely,

$$\|T\| = \sup_{\|x\|=1} \|Tx\| = \sup_{\|x\|=1} \sup_{\|\ell\|=1} |\ell(Tx)| \leq \sup_{\|\ell\|=1} \sup_{\|x\|=1} \|\ell T\| \, \|x\| = \sup_{\|\ell\|=1} \|T'\ell\| = \|T'\| \, .$$

Hence $\|T'\| = \|T\|$. ■

Definition 2.50
*Let $T \in B(X, Y)$. The **orthogonal complement** of $T(X)$ is defined as*

$$T(X)^{\perp} = \{\ell \in Y' \mid \ell(Tx) = 0 \text{ for all } x \in X\}.$$

Proposition 2.51
Let $T \in B(X, Y)$. Then $\ker(T') = T(X)^{\perp}$.

Proof. Let $\ell \in T(X)^{\perp}$. Then for all $x \in X$,

$$T'\ell x = \ell(Tx) = 0.$$

Hence $\ell \in \ker(T')$ and $T(X)^{\perp} \subset \ker(T')$. Conversely, if $\ell \in \ker(T')$, then $T'\ell = 0$ and

$$\ell(Tx) = T'\ell(x) = 0$$

for all $x \in X$. Thus $\ell \in T(X)^{\perp}$ and $\ker(T') \subset T(X)^{\perp}$. We conclude that $\ker(T') = T(X)^{\perp}$. ∎

2.5. Hahn-Banach Separation Theorem

Definition 2.52
*An **affine hyperplane** in a vector space X is a set of the form*

$$H = \{x \in X \mid f(x) = \alpha\}$$

where f is a linear functional on X and $\alpha \in \mathbb{R}$. We denote the affine hyperplane by $H(f, \alpha)$.

Remark
The linear functional f need not be continuous.

Proposition 2.53
The hyperplane $H(f, \alpha)$ is closed if and only if f is continuous.

Proof. Suppose first that f is continuous. Clearly $\{\alpha\} \subset \mathbb{R}$ is closed. It follows that $f^{-1}(\{\alpha\}) = H(f, \alpha)$ is closed.

Conversely, assume that $H(f, \alpha)$ is closed. If $H(f, \alpha) = X$, then $f = 0$ and is continuous. If not, then $H(f, \alpha)^c \neq \varnothing$. Let $x_0 \in H(f, \alpha)^c$ and $f(x_0) \neq \alpha$. Without loss of generality, assume that $f(x_0) < \alpha$.

Fix $r > 0$ such that $B_r(x_0) \subset H(f, \alpha)^c$. We claim that $f(x) < \alpha$ for all $x \in B_r(x_0)$. Suppose not. Then there is $x_1 \in B_r(x_0)$ such that $f(x_1) > \alpha$. We have that the segment

$$\{x_t \in X \mid x_t = (1 - t)x_0 + tx_1, t \in [0, 1]\}$$

lies in $B_r(x_0)$ and $f(x_t) \neq \alpha$ for all $t \in [0,1]$. However, it is clear that

$$t = \frac{f(x_1) - \alpha}{f(x_1) - f(x_0)} \in [0,1] \quad \text{and} \quad f(x_t) = (1-t)f(x_0) + tf(x_1) = \alpha,$$

a contradiction. Thus $f(x) < \alpha$ for all $x \in B_r(x_0)$. It follows that $f(x_0 + rz) < \alpha$ for all $\|z\| < 1$. Then

$$\|f\| = \sup_{\|z\| \leq 1} |f(z)| \leq \frac{1}{r}(\alpha - f(x_0)) < \infty.$$

Hence f is continuous. ■

Definition 2.54
*Let $A, B \subset X$ be two subsets of X. We say that a hyperplane $F(f,\alpha)$ **weakly separates** A and B if*

$$\sup_{x \in A} f(x) \leq \alpha \leq \inf_{x \in B} f(x).$$

Definition 2.55
*Let $A, B \subset X$ be two subsets of X. We say that a hyperplane $F(f,\alpha)$ **strictly separates** A and B if*

$$\sup_{x \in A} f(x) \leq \alpha - \epsilon < \alpha + \epsilon \leq \inf_{x \in B} f(x)$$

for some $\epsilon > 0$.

Lemma 2.56
Let $C \subset X$ be an open convex set containing 0. For every $x \in X$, set

$$p(x) = \inf \left\{ \alpha > 0 \;\middle|\; \frac{1}{\alpha} x \in C \right\}.$$

Then

(a) $p(\lambda x) = \lambda p(x)$ for all $\lambda > 0$ and $x \in X$,

(b) $p(x+y) \leq p(x) + p(y)$ for all $x, y \in X$,

(c) there is $M < \infty$ such that $0 \leq p(x) \leq M\|x\|$ for all $x \in X$,

(d) $C = \{x \in X \mid p(x) < 1\}$.

Proof. For (c), let $r > 0$ be such that $B_r(0) \subset C$. $x \in B_{\|x\|}(0)$ implies that $rx/\|x\| \in B_r(0) \subset C$. Thus

$$p(x) \leq \frac{1}{r}\|x\|$$

for all $x \in X$.

For (d), let $x \in C$. Since C is open, there is $\delta > 0$ such that $(1+\delta)x \in C$. Thus

$$p(x) \leq \frac{1}{1+\delta} < 1.$$

Conversely, suppose $p(x) < 1$. There is $\alpha \in (0,1)$ such that $\frac{1}{\alpha}x \in C$. Then $x = \alpha(x/\alpha) + (1-\alpha)\cdot 0 \in C$ by convexity of C. We conclude that $C = \{x \in X \mid p(x) < 1\}$.

(a) is obvious. For (b), let $x, y \in X$ be given. For $\epsilon > 0$, from the definition of p, $\frac{x}{p(x)+\epsilon} \in C$ and $\frac{y}{p(y)+\epsilon} \in C$. Now for $t \in [0,1]$,

$$t\frac{x}{p(x)+\epsilon} + (1-t)\frac{y}{p(y)+\epsilon} \in C$$

by the convexity of C. Thus

$$t = \frac{p(x)+\epsilon}{p(x)+p(y)+2\epsilon} \in [0,1] \quad \Rightarrow \quad \frac{x+y}{p(x)+p(y)+2\epsilon} \in C.$$

Hence

$$p(x+y) \leq p(x) + p(y) + 2\epsilon$$

for all $\epsilon > 0$. Thus $p(x+y) \leq p(x) + p(y)$. ■

Theorem 2.57 (Hahn-Banach Separation Theorem I)
Let $A, B \subset X$ be two non-empty convex sets such that $A \cap B = \varnothing$. If one of the sets is open, there is a closed hyperplane $H(f, \alpha)$ separating A and B.

Proof. We first prove the case where $A = \{x_0\}$ is a singleton and B is open. By translation we may assume without loss of generality that B contains 0. Consider the set $G = \operatorname{span}(\{x_0\})$. Define the functional g on G by

$$g(tx_0) = t$$

for $t \in R$. Apply lemma 2.56 to the open convex set B to obtain the corresponding p. We claim that $g(x) \leq p(x)$ for all $x \in G$.

Indeed, let $x = tx_0$. If $t > 0$, then $g(x) = t$ and

$$p(x) = p(tx_0) = tp(x_0) \geq t = g(x).$$

If $t \leq 0$, then $g(x) = t \leq 0$ and by definition $p(x) \geq 0$. We conclude that $g(x) \leq p(x)$ for all $x \in G$. ■

2.6. Weak and Weak* Convergence

Definition 2.58
Let $(X, \|\cdot\|)$ be a normed space. A sequence $\{x_n\}$ in X is said to ***converge weakly*** *to $x \in X$, denoted by $x_n \xrightarrow{w} x$, if for every $L \in X'$, $L(x_n) \to L(x)$ as $n \to \infty$.*

Remark
Strong convergence implies weak convergence. If $x_n \to x$,

$$|L(x_n) - L(x)| = |L(x_n - x)| \leq \|L\|\,\|x_n - x\| \to 0$$

as $n \to \infty$. Thus $x_n \xrightarrow{w} x$. However, the converse is not true in general.

Example

Consider ℓ^2. Note that $(\ell^2)' \cong \ell^2$. For all $L \in (\ell^2)'$, there exists $y \in \ell^2$ such that $L(x) = \sum_{n=1}^{\infty} x_n y_n$. Let $x_n = e^n$ be the sequence with 1 at the n-th position and 0 elsewhere. Then $x_n \xrightarrow{w} 0$ since for every $L \in (\ell^2)'$,

$$L(x_n) = \sum_i e_i^n y_i = y_n \to 0$$

for $y \in \ell^2$. However, $\|x_n\|_{\ell^2} = 1$ for every n and thus $x_n \not\to 0$.

Example

Consider $X = C([0,1])$ with the supremum norm. Let

$$x_n(t) = \begin{cases} nt & \text{if } 0 \le t \le 1/n, \\ 2 - nt & \text{if } 1/n \le t \le 2/n, \\ 0 & \text{if } 2/n \le t \le 1. \end{cases}$$

Then $\|x_n\|_\infty = 1$ and thus $x_n \not\to 0$. Instead, we have $x_n \xrightarrow{w} 0$. Assume not, then we can find $T \in X'$ and a subsequence $\{x_{n_k}\}$ such that $|T(x_{n_k})| \ge \delta > 0$. For simplicity, we consider the case $T(x_{n_k}) \ge \delta$, but the other case is similar. Since $T \in X'$, $|T(x_{n_k})| \le \|T\|_{X\to\mathbb{R}} \|x_{n_k}\|_\infty$. Let $y_K = \sum_{k=1}^K x_{n_k}$. Then $T(y_K) = \sum_{k=1}^K T(x_{n_k}) \ge K\delta$ and $T(y_K) \le \|T\|_{X\to\mathbb{R}} \|y_K\|_\infty$. This implies that y_K cannot be bounded. Now consider x_{n_k} with $n_{k+1} \ge 2n_k$. For $t \in [0, 1/n_K]$, $x_{n_k}(t) = n_k t$.

$$y_K(t) = \sum_{k=1}^K n_k t \le \sum_{k=1}^K n_k/n_K \le 1 + \sum_{k=1}^K 2^{K-k} \le 1 + \sum_k 2^{-k} = 2.$$

For $t \in [1/n_K, 1/n_{K-1}]$,

$$y_K(t) = \sum_{k=1}^K x_{n_k}(t) \le 1 + \sum_{k=1}^{K-1} n_k t \le 1 + \frac{1}{n_{K-1}} \sum_{k=1}^{K-1} n_k \le 1 + 1 + \sum_k 2^{-k} = 3.$$

On $[1/n_K, 1/n_{K-1}]$, we have $\|y_K\| \le 3$. Thus $\delta K \le \|T\|_{X\to\mathbb{R}} \|y_K\|_\infty \le 3\|T\|_{X\to\mathbb{R}}$, which is impossible for sufficiently large K. Hence $x_n \xrightarrow{w} 0$.

Proposition 2.59

$(X, \|\cdot\|_X)$ is a normed space and $x_n \in X$. If $\|x_n\|_X \le C$ for all $n \in \mathbb{N}$ and $L(x_n) \to L(x)$ for all $L \in A \subset X'$, where A is dense in X', then $x_n \xrightarrow{w} x$ in X.

Proof. Let $\epsilon > 0$ be given. A is dense in X'. For $T \in X'$, there is an $L \in A$ such that $\|T - L\|_{X'\to\mathbb{R}} \le \epsilon$. Also, there exists N such that $|L(x_n) - L(x)| \le \epsilon$ for all $n \ge N$. Then

$$\begin{aligned} |T(x_n) - T(x)| &\le |T(x_n) - L(x_n)| + |L(x_n) - L(x)| + |L(x) - T(x)| \\ &\le \|T - L\|_{X'\to\mathbb{R}} (\|x_n\|_X + \|x\|_X) + |L(x_n) - L(x)| \le 2C\epsilon + \epsilon \end{aligned}$$

for all $n \le N$. Since ϵ is arbitrary, $x_n \xrightarrow{w} x$. ■

Definition 2.60

*A space X is called a **Baire space** if for any sequence of open dense subsets $\{E_n\}$, $\cap_n E_n$ is dense in X.*

Theorem 2.61 (Baire Category Theorem)

A complete metric space is a Baire space.

Proof. Let X be a complete metric space and $\{E_n\}$ be a sequence of open dense subsets in X. Put $E = \cap_n E_n$. We want to show that any nonempty open set $G \subset X$ intersects E.

E_1 is dense in X so $G \cap E_1$ is nonempty. Then there exists $x_1 \in E_1 \cap G$. Note that $E_1 \cap G$ is open; there exists $1 > \delta_1 > 0$ such that $B_{\delta_1}(x_1) \subset E_1 \cap G$. By shrinking δ_1, we can have $\overline{B_{\delta_1}(x_1)} \subset E_1 \cap G$. Now since E_2 is dense in X, there exists $x_2 \in E_2 \cap B_{\delta_1}(x_1)$ and also a $1/2 > \delta_2 > 0$ such that $\overline{B_{\delta_2}(x_2)} \subset E_2 \cap B_{\delta_1}(x_1)$. Continue this process, we obtain a sequence $\{x_n\}$ and $\delta_n \le 1/n$ such that $\overline{B_{\delta_n}(x_n)} \subset E_n \cap B_{\delta_{n-1}}(x_{n-1})$.

For every $m, n \ge N$, we have $x_n \in B_{\delta_n}(x_n) \subset \cdots \subset B_{\delta_N}(x_N)$ and $x_m \in B_{\delta_m}(x_m) \subset \cdots \subset B_{\delta_N}(x_N)$ by construction. Hence $d(x_n, x_m) \le 2\delta_N \le 2/N$ and $\{x_n\}$ is a Cauchy sequence. Since X is complete, $\{x_n\}$ converges to some $x \in X$. We claim that $x \in E \cap G$. Clearly $x \in G$. By construction $x_m \in \overline{B_{\delta_n}(x_n)}$ for all $m \ge n$. Thus $x \in B_{\delta_n}(x_m) \subset E_N$ for $m \ge n \ge N$. We see that $x \in \cap_n E_n$. Notice that G is arbitrary, so E is dense in X, proving that X is a Baire space. ■

Theorem 2.62 (Uniform Boundedness Principle I)

X is a complete metric space. $f_\alpha : X \to \mathbb{R}$ is continuous for every $\alpha \in A$, where A is an index set. If for every $x \in X$, there exists $M(x) < \infty$ such that

$$\sup_{\alpha \in A} |f_\alpha(x)| \le M(x),$$

then there exists an open G and a constant $C < \infty$ such that

$$\sup_{\alpha \in A} |f_\alpha(x)| \le C$$

for all $x \in G$.

Proof. By Baire Category Theorem, X is a Baire space. For each n, let

$$F_n = \left\{ x \in X \;\middle|\; \sup_{\alpha \in A} |f_\alpha(x)| \le n \right\}.$$

We claim that F_n is closed and $X = \cup_n F_n$. Indeed, set $x_k \to x \in X$, where $x_k \in F_n$ for all k. For any $\alpha \in A$, $|f_\alpha(x_k)| \le n$ for all k and by continuity of f_α,

$$|f_\alpha(x)| = \lim_{k \to \infty} |f_\alpha(x_k)| \le n.$$

Hence $x \in F_n$ and F_n is closed. Next, for any $x \in X$, take $N \geq M(x)$. Then $x \in F_N \subset \cup_n F_n$. This shows that $X = \cup_n F_n$.

Finally, observe that F_n cannot have empty interiors for all n. Otherwise, $\emptyset = X^c = (\cup_n F_n)^c = \cap F_n^c \neq \emptyset$ since F_n^c are open dense subsets of X, which is absurd. Hence there is some n such that F_n has nonempty interior, say $G \subset F_n$. Then $\sup_{\alpha \in A} |f_\alpha(x)| \leq n$ for all $x \in G$ as desired. ■

Definition 2.63

*A function $f : X \to \mathbb{R}$ is said to be **sub-additive** if $f(x+y) \leq f(x) + f(y)$ for all $x, y \in X$.*

Theorem 2.64 (Uniform Boundedness Principle II)

X is a Banach space. $\alpha \in A$ is an arbitrary index set. $f_\alpha : X \to \mathbb{R}$ are continuous, sub-additive and satisfy $f_\alpha(cx) = |c|\, f_\alpha(x)$ for all $x \in X$ and $c \in \mathbb{R}$. If for every $x \in X$, there exists $M(x) < \infty$ such that

$$\sup_{\alpha \in A} |f_\alpha(x)| \leq M(x),$$

then there exists a constant $C < \infty$ such that

$$\sup_{\alpha \in A} |f_\alpha(x)| \leq C \, \|x\|_X$$

for all $x \in X$.

Proof. By theorem 2.62, there exists an open G and a constant $C < \infty$ such that

$$\sup_{\alpha \in A} |f_\alpha(x)| \leq C$$

for all $x \in G$. The proof will be complete if we can extend G to X. Since G is open, there exists $r > 0$ such that $B_r(z) \subset G$ for all $z \in G$. For any $x \in B_r(z)$, $\sup_{\alpha \in A} |f_\alpha(x)| \leq C$ and hence $\sup_{\alpha \in A} |f_\alpha(z+y)| \leq C$ for all $y \in B_r(0)$. Take y with $\|y\| \leq r/2$. Then

$$-2C \leq f_\alpha(y+z) - f_\alpha(z) \leq f_\alpha(y) \leq f_\alpha(y+z) + f_\alpha(-z) = f_\alpha(y+z) + f_\alpha(z) \leq 2C.$$

Hence $|f_\alpha(y)| \leq 2C$ for all y with $\|y\| \leq r/2$. Take $x \in X$.

$$|f_\alpha(x)| = \left| f_\alpha\left(\frac{x}{\|x\|} \frac{r}{2} \frac{2}{r} \|x\| \right) \right| = \frac{2}{r} \|x\| \, |f_\alpha(y)| \leq \frac{4C}{r} \|x\| .$$

Thus

$$\sup_{\alpha \in A} |f_\alpha(x)| \leq \frac{4C}{r} \|x\|$$

for all $x \in X$. ■

Corollary 2.65

X is a Banach space. $L_\alpha \in X'$ and $\alpha \in A$. If for every $x \in X$, there exists $M(x) < \infty$ such that $\sup_{\alpha \in A} |L_\alpha(x)| \leq M(x)$, then there exists a constant $C < \infty$ such that $\sup_{\alpha \in A} \|L_\alpha\| \leq C$.

Proof. Apply theorem 2.64 to $f_\alpha(x) = |L_\alpha(x)|$. First, L_α is linear and the sub-linearity follows from the triangle inequality. Next, $|L_\alpha(cx)| = |c|\,|L_\alpha(x)|$ for all $c \in \mathbb{R}$. Also, $L_\alpha \in X'$ implies that f_α is continuous. The conclusion follows from theorem 2.64. ■

Corollary 2.66

X is a normed space. $x_\alpha \in X$ for all $\alpha \in A$ with the property that for every $L \in X'$, there is $M(L) < \infty$ such that $\sup_\alpha |L(x_\alpha)| \leq M(L)$ and $(X', \|\cdot\|_{X \to \mathbb{R}})$ is a Banach space. Then there exists $C < \infty$ such that $\|x_\alpha\|_X \leq C$ for all $\alpha \in A$.

Proof. Apply the theorem 2.64 to $f_\alpha(L) = |L(x_\alpha)|$. First, for $L, T \in X'$,

$$f_\alpha(L+T) = |L(x_\alpha) + T(x_\alpha)| \leq |L(x_\alpha)| + |T(x_\alpha)| = f_\alpha(L) + f_\alpha(T).$$

Next, for $c \in \mathbb{R}$,

$$f_\alpha(cL) = |cL(x_\alpha)| = |c|\,|L(x_\alpha)| = |c|\, f_\alpha(L).$$

Finally, to verify that f_α is continuous, note that for $L_n \to L$ in X',

$$|f_\alpha(L_n) - f_\alpha(L)| = |L_n(x_\alpha) - L(x_\alpha)| \leq \|L_n - L\|_{X' \to \mathbb{R}} \|x_\alpha\|_X \to 0$$

for each $\alpha \in A$. The conclusion follows from theorem 2.64. ■

Corollary 2.67

X is a normed space and $x_n \in X$ with $x_n \xrightarrow{w} x$ in X. Then there exists $C < \infty$ such that $\|x_n\|_X \leq C$ for all n.

Proof. This is a direct consequence of corollary 2.66 with $A = \mathbb{N}$. ■

Proposition 2.68

Let $f_n \in \mathcal{L}^p(X, \mu)$ and $1 \leq p < \infty$. Then $f_n \xrightarrow{w} f \in \mathcal{L}^p$ if

$$\lim_{n \to \infty} \int f_n g \, d\mu = \int f g \, d\mu$$

for all $g \in \mathcal{L}^{p'}(X, \mu)$ and some f in $\mathcal{L}^p$ where p' is the conjugate exponent of p.

Proof. By the assumption and Riesz representation theorem, for every $T \in (\mathcal{L}^p)'$, there exists a unique $g \in \mathcal{L}^{p'}$ such that

$$T(f_n) = \int f_n g \, d\mu \to \int f g \, d\mu = T(f).$$

Hence $f_n \xrightarrow{w} f$. ■

Proposition 2.69

$f_n \in \mathcal{L}^p(X, \mu)$ and $1 \leq p < \infty$. If $f_n \xrightarrow{w} f$ in $\mathcal{L}^p$, then f_n is bounded and

$$\|f_n\|_p \leq \liminf_{n \to \infty} \|f_n\|_p \, .$$

Proof. Consider the function

$$g = \frac{|f|^{p/p'}}{\|f\|_p^{p/p'}}.$$

Note that

$$\|g\|_{p'}^{p'} = \int |g|^{p'}\, d\mu = \int \frac{|f|^p}{\|f\|_p^p} d\mu = 1.$$

Hence $g \in \mathcal{L}^{p'}$ with $\|g\|_{p'} = 1$. Also notice that $|g| = |f|^{p/p'}/\|f\|_p^{p/p'} = |f|^{p-1}/\|f\|_p^{p-1}$. By the weak convergence and Riesz representation theorem,

$$\|f\|_p = \int \frac{|f|^p}{\|f\|_p^{p-1}} d\mu = \int |fg|\, d\mu = \lim_{n\to\infty} \int |f_n g|\, d\mu \le \liminf_{n\to\infty} \|f_n\|_p \|g\|_{p'} = \liminf_{n\to\infty} \|f_n\|_p$$

by the Hölder inequality. Note that by corollary 2.67, f_n is bounded uniformly in n. ∎

Proposition 2.70

$1 \le p < \infty$ and $1/p + 1/p' = 1$. Suppose $f_n \to f$ in $\mathcal{L}^p$ and $g_n \to g$ in $\mathcal{L}^{p'}$. Then

$$\lim_{n\to\infty} \int f_n g_n d\mu = \int f g d\mu.$$

Proof. By the Hölder inequality,

$$\begin{aligned}\left|\int f_n g_n d\mu - \int f g d\mu\right| &\le \left|\int f_n(g_n - g) d\mu\right| + \left|\int (f_n - f) g d\mu\right| \\ &\le \|f_n\|_p \|g_n - g\|_{p'} + \|f_n - f\|_p \|g\|_{p'}.\end{aligned}$$

Note that by proposition 2.69, f_n converges to f strongly and hence weakly. It follows that $\|f_n\|$ is bounded by some $C < \infty$. Since g_n converges to g and f_n converges to f in their respective norms, the right hand side of the inequality converges to 0 as $n \to \infty$. ∎

Remark

If we loosen the condition to $f_n \xrightarrow{w} f$ in $\mathcal{L}^p$ and $g_n \xrightarrow{w} g$ in $\mathcal{L}^{p'}$, then the conclusion fails.

Example

Suppose $p = p' = 2$ and $f_n(x) = \sqrt{2/\pi}\sin(nx)$ for $x \in [0,\pi]$. Then $f_n \in \mathcal{L}^2([0,\pi])$ and

$$\int_0^\pi f_n^2 dx = \frac{2}{\pi}\int_0^\pi \sin^2(nx) dx = 1.$$

To see that $f_n \xrightarrow{w} 0$, let $g \in \mathcal{L}^2([0,\pi])$. For every $\epsilon > 0$, there is a step function ϕ such that $\|g - \phi\|_2 < \epsilon$. Note that every step function is a finite linear combination of characteristic functions of intervals. Hence it suffices to show that $f_n \chi_I$ can be arbitrary small for n sufficiently large. On every interval,

$$\left|\int_I \sin(nx) dx\right| \le \int_0^{\pi/n} \sin(nx) dx = \frac{2}{n} \to 0$$

as $n \to \infty$. Thus $f_n \xrightarrow{w} 0$ in $\mathcal{L}^2([0,\pi])$. However, f_n does not converge to 0 strongly in $\mathcal{L}^2([0,\pi])$ since $\|f_n\|_2 = 1 \neq 0$ for all n.

Proposition 2.71

$1 \leq p < \infty$. Let $f_n \in \mathcal{L}^p(X,\mu)$ be a bounded sequence of functions. Then $f_n \xrightarrow{w} f$ in $\mathcal{L}^p$ if and only if

$$\lim_{n\to\infty} \int_A f_n d\mu = \int_A f d\mu$$

for all $A \in \mathcal{A}$ when $p = 1$ and for A with finite measure when $p > 1$.

Proof.

$$\begin{aligned} f_n \xrightarrow{w} f &\iff \int f_n g d\mu \to \int f g d\mu \text{ for all } g \in \mathcal{L}^{p'} \\ &\iff \int_A f_n s d\mu \to \int_A f s d\mu \text{ for all simple } s \in \mathcal{L}^{p'} \\ &\iff \int_A f_n d\mu = \int f_n \chi_A d\mu \to \int f \chi_A d\mu = \int_A f d\mu \end{aligned}$$

for $A \in \mathcal{A}$ such that $\chi_A \in \mathcal{L}^{p'}$. If $p = 1$, then A can be taken to be any $A \in \mathcal{A}$; if $p > 1$, then A must have finite measure. ∎

Proposition 2.72

$1 < p < \infty$. Let $f_n \in \mathcal{L}^p(X,\mu)$ be a sequence with $\|f_n\|_p \leq M$ and $f_n \to f$ pointwise a.e. Then $f_n \xrightarrow{w} f$ in $\mathcal{L}^p$.

Proof. Since $\|f_n\|_p \leq M$,

$$\int |f|^p \, d\mu = \int \liminf_{n\to\infty} |f_n|^p \, d\mu \leq \liminf_{n\to\infty} \int |f_n|^p \, d\mu = M^p$$

by Fatou's lemma. Hence $f \in \mathcal{L}^p$. It remains to show that the convergence is weak. By proposition 2.71, it is equivalent to show that

$$\lim_{n\to\infty} \int_A f_n d\mu = \int_A f d\mu$$

for all $A \in \mathcal{A}$ with $\mu(A) < \infty$. Indeed, by Egorov's theorem, for every $\epsilon > 0$, there exists $F_\epsilon \subset A$ with $\mu(A - F_\epsilon) \leq \epsilon$ and $f_n \to f$ uniformly on F_ϵ. Furthermore, by proposition 1.33, we can choose F_ϵ so that

$$\int_{A-F_\epsilon} |f_n - f|^p \, d\mu \leq \epsilon$$

since $f_n, f \in \mathcal{L}^p$ and so does $|f_n - f|^p$. Also, let $E = \{x \in A - F_\epsilon \mid |f_n - f| > 1\}$. Then for n

sufficiently large,

$$\begin{aligned}\int_A |f_n - f|\, d\mu &\le \int_{F_\epsilon} |f_n - f|\, d\mu + \int_{A-F_\epsilon} |f_n - f|\, d\mu \\ &\le \int_A \epsilon d\mu + \int_{A-F_\epsilon-E} |f_n - f|\, d\mu + \int_E |f_n - f|\, d\mu \\ &\le \epsilon\mu(A) + \mu(A - F_\epsilon) + \int_{A-F_\epsilon} |f_n - f|^p\, d\mu \le \epsilon\mu(A) + \epsilon + \epsilon.\end{aligned}$$

Hence $f_n \xrightarrow{w} f$. ■

Remark

The proposition fails for $p = 1$*. Consider* $f_n = n\chi_{[0,1/n]}$*. Then* $\|f_n\|_1 = 1$ *and* $f_n \to 0$ *pointwise a.e. However,*

$$\int_0^1 f_n(x)dx = 1 \neq 0 = \int_0^1 0dx.$$

Thus f_n *does not converge weakly to* 0 *in* $\mathcal{L}^1$*.*

Theorem 2.73 (Radon-Riesz)

$1 < p < \infty$*. Then* $f_n \to f$ *in* $\mathcal{L}^p$ *if and only if* $\lim_{n\to\infty} \|f_n\|_p = \|f\|_p$ *and* $f_n \xrightarrow{w} f$ *in* $\mathcal{L}^p$*.*

Proof. Suppose $f_n \to f$ in $\mathcal{L}^p$. Then the strong convergence immediately implies the weak convergence. Also, note that $\|f_n\|_p \le \|f_n - f\|_p + \|f\|_p$ and thus

$$\left|\|f_n\|_p - \|f\|_p\right| \le \|f_n - f\|_p \to 0$$

by the strong convergence. Conversely, suppose that $\|f_n\|_p \to \|f\|_p$ and $f_n \xrightarrow{w} f$ in $\mathcal{L}^p$.

Assume $p \ge 2$. For any $y \in \mathbb{R}$, notice that $|1 + y|^p \ge 1 + py + c\,|y|^p$ for some $c \in (0,1)$. Let $E = \{x \in X \mid f(x) = 0\}$ and apply $y = (f_n - f)/f$ on E^c. Then on E^c,

$$\left|\frac{f_n}{f}\right|^p \ge 1 + p\left(\frac{f_n - f}{f}\right) + c\left|\frac{f_n - f}{f}\right|^p$$

Thus

$$|f_n|^p \ge |f|^p + p(f_n - f)\,|f|^{p-1}\operatorname{sgn}(f) + c\,|f_n - f|^p.$$

Rearranging the inequality and integrating both sides on E^c gives

$$c\int_{E^c} |f_n - f|^p\, d\mu \le \int_{E^c} |f_n|^p - |f|^p\, d\mu - p\int_{E^c} |f|^{p-1}\operatorname{sgn}(f)(f_n - f)d\mu$$

Note that as shown in the proof of proposition 2.69, $|f|^{p-1}\operatorname{sgn}(f) \in \mathcal{L}^{p'}$. By the assumptions we see that

$$\int_{E^c} |f_n - f|^p\, d\mu \to 0$$

as $n \to \infty$. On E, we have $f = 0$ and

$$\int_E |f_n - f|^p \, d\mu = \int_E |f_n|^p \, d\mu \to 0$$

as $n \to \infty$. Hence $f_n \to f$ in $\mathcal{L}^p$.

Assume $1 < p < 2$. Then we have the same inequality for $|z| \geq 1$, i.e.,

$$|1+z|^p \geq 1 + p\,|z| + c\,|z|^p$$

Also, for $|z| \leq 1$,

$$\frac{|1+z|^p - 1 - pz}{z^2}$$

is strictly positive. Now let $E_n = \{x \in X \mid |f_n(x) - f(x)| \leq |f(x)|\}$. Then by applying the same argument above on E_n^c, we have

$$\int_{E_n^c} |f_n - f|^p \, d\mu \leq \frac{1}{c} \int_{E_n^c} |f_n|^p - |f|^p \, d\mu - \frac{p}{c} \int_{E_n^c} |f|^{p-1} \operatorname{sgn}(f)(f_n - f) d\mu$$

as $n \to \infty$. On E_n, replacing z by $(f_n - f)/f$,

$$\left|\frac{f_n}{f}\right|^p \geq 1 + p\frac{f_n - f}{f} + c'\left(\frac{f_n - f}{f}\right)^2 \implies |f_n|^p \geq |f|^p + p(f_n - f)\,|f|^{p-1} \operatorname{sgn}(f) + c'\,|f_n - f|^2\,|f|^{p-2}$$

for some $c' > 0$. Thus

$$\int_{E_n} |f_n - f|^2\,|f|^{p-2} \, d\mu \leq \frac{1}{c'} \int_{E_n} |f_n|^p - |f|^p \, d\mu - \frac{p}{c'} \int_{E_n} |f|^{p-1} \operatorname{sgn}(f)(f_n - f) d\mu.$$

Adding up the two inequalities, we have

$$\int_{E_n^c} |f_n - f|^p \, d\mu + \int_{E_n} |f_n - f|^2\,|f|^{p-2} \, d\mu \to 0$$

as $n \to \infty$ by the assumptions. Note that on E_n, $|f| \geq |f_n - f|$ and

$$\begin{aligned}
\int_{E_n} |f_n - f|^p \, d\mu \leq \int_{E_n} |f_n - f|\,|f|^{p-1} \, d\mu &\leq \left(\int_{E_n} |f_n - f|^2\,|f|^{p-2} \, d\mu\right)^{1/2} \left(\int_{E_n} |f|^p \, d\mu\right)^{1/2} \\
&\leq \left(\int_{E_n} |f_n - f|^2\,|f|^{p-2} \, d\mu\right)^{1/2} \|f\|_p^{p/2} \to 0.
\end{aligned}$$

Hence $f_n \to f$ in $\mathcal{L}^p$. We conclude that $f_n \to f$ strongly in $\mathcal{L}^p$ if and only if $f_n \xrightarrow{w} f$ in $\mathcal{L}^p$ and $\|f_n\|_p \to \|f\|_p$. ■

Remark

Radon-Riesz theorem fails for $p = 1$. Consider $f_n(x) = 1 + \sin(nx)$ on $X = [-\pi, \pi]$. Then for *every* $g \in \mathcal{L}^\infty$,

$$\int (f_n - 1) g d\mu \leq \int \sin(nx) g d\mu \to 0$$

by the step function approximation argument. Also, $\|f_n\|_1 = 2\pi$ *for all* n *and hence converges to* $\|1\|_1 = 2\pi$*. However,* f_n *does not converge to* 1 *in* $\mathcal{L}^1$ *since*

$$\int_{-\pi}^{\pi} |f_n - 1| \, d\mu = \int_{-\pi}^{\pi} |\sin(nx)| \, d\mu = 2n \int_0^{\frac{\pi}{2n}} \sin(nx) dx = 2$$

for all n*.*

Definition 2.74
Let X *be a Banach space. A subset* $K \subset X$ *is* ***weakly sequentially compact*** *if every sequence* $\{x_n\} \subset K$ *has a subsequence* $\{x_{n_k}\}$ *such that* $x_{n_k} \xrightarrow{w} x \in K$*.*

Proposition 2.75
Let X *be a Banach space. If* $K \subset X$ *is weakly sequentially compact, then* K *is bounded.*

Proof. Suppose K is not bounded. Then we can choose an unbounded sequence $\{x_n\} \subset K$ such that $\|x_n\| \geq n$ for all n. By the weakly sequential compactness of K, there exists a weaky convergent subsequence $\{x_{n_k}\}$ such that $x_{n_k} \xrightarrow{w} x \in K$ and also $\|x_{n_k}\| \geq n_k$. However, by corollary 2.67, $\|x_{n_k}\| \leq C$ for some $C < \infty$, which is absurd. Hence K is bounded. ∎

Theorem 2.76 (Kakutani)
Let X *be a reflexive Banach space. Then the closed unit ball*

$$B = \{x \in X \mid \|x\| \leq 1\}$$

is weakly sequentially compact.

Proof. We consider first the case when X is separable. Reflexivity gives that $X'' \cong X$ and hence X'' is separable. By theorem 2.44, X' is separable, and there exists a countable dense subset $\{m_j\} \subset X'$. Given $x_n \in X$ with $\|x_n\| \leq 1$, we need to show that there exists a subsequence $x_{n_k} \xrightarrow{w} x \in B$. Since $m_j(x_n)$ is a bounded sequence for each j, we can extract a subsequence x_{n_k} such that $m_j(x_{n_{k_j}}) \to A(m_{k_j})$ as $j \to \infty$, where

$$A(m_j) = \lim_{j \to \infty} m_j(x_{n_{k_j}}).$$

We claim that for all $m \in X'$, $m(x_{n_k}) \to A(m)$ as $k \to \infty$. Indeed, for any $m \in X'$, we can find a sequence $\{m_j\}$ such that $m_j \to m$ as $j \to \infty$. Then

$$\begin{aligned}
\left|m(x_{n_k}) - m(x_{n_l})\right| &\leq \left|m(x_{n_k}) - m_j(x_{n_k})\right| + \left|m_j(x_{n_k}) - m_j(x_{n_l})\right| + \left|m_j(x_{n_l}) - m(x_{n_l})\right| \\
&\leq \left\|m - m_j\right\| \left\|x_{n_k}\right\| + \left|m_j(x_{n_k}) - m_j(x_{n_l})\right| + \left\|m - m_j\right\| \left\|x_{n_l}\right\| \\
&\leq 2\left\|m - m_j\right\| + \left|m_j(x_{n_k}) - m_j(x_{n_l})\right| \to 0
\end{aligned}$$

as $k, l \to \infty$. Hence the sequence $\{m(x_{n_k})\}$ is Cauchy and A is well-defined. Notice that A is also bounded:

$$|A(m)| = \lim_{k\to\infty} \left|m(x_{n_k})\right| \le \lim_{k\to\infty} \|m\| \left\|x_{n_k}\right\| = \|m\| .$$

We see that $\|A\| \le 1$. Because A is bounded, it is continuous and thus $m(x_{n_k}) \to m(x)$ for some $x \in X$ by the reflexivity of X. Such x belongs to B since $\|x\| = \|A\| \le 1$. Thus B is weakly sequentially compact.

For the general case where X is not separable, consider the sequence $\{x_n\} \subset B$. Let $Y = \left\{\sum_{n=1}^{N} \alpha_n x_n \;\middle|\; N \in \mathbb{N}, \alpha_n \in \mathbb{R}\right\}$ be the closed subspace of X spanned by $\{x_n\}$. Since X is reflexive, Y is also reflexive by theorem 2.45. Note that Y is also separable. The established results above show that there exists a subsequence $\{x_{n_k}\} \subset Y$ and $x \in Y$ such that $x_{n_k} \xrightarrow{w} x$ in Y, i.e., for every $m \in Y'$, $m(x_{n_k}) \to m(x)$. Extend the functionals $m \in Y'$ to $\ell \in X'$ by Hahn-Banach theorem. Then $\ell|_Y = m \in Y'$ implies that $\ell(x_{n_k}) = m(x_{n_k}) \to m(x) = \ell(x)$. We conclude that $x_{n_k} \xrightarrow{w} x \in B$. Thus B is weakly sequentially compact. ■

Example

Let $p \in (1, \infty)$. Then $\mathcal{L}^p(\Omega, \mu)$ is reflexive. Then for all $\{f_n\}$ with $\|f_n\|_p \le 1$, there exists a subsequence $f_{n_k} \xrightarrow{w} f$ in $\mathcal{L}^p$ for some f with $\|f\|_p \le 1$. By Riesz representation theorem, this is equivalent to saying that for every $g \in \mathcal{L}^q(\Omega, \mu)$,

$$\lim_{k\to\infty} \int f_{n_k} g d\mu = \int f g d\mu,$$

where q is the conjugate exponent of p.

Definition 2.77

*Let M be a Banach space. A sequence of bounded linear functionals $\{x_n\} \subset M'$ **converges weakly*** to x if for all $m \in M$, $x_n(m) \to x(m)$ as $n \to \infty$. We denote the convergence by $x_n \xrightarrow{w^*} x$.*

Remark

Since the canonical mapping $M \to M''$ is always injective, w^* convergence is weaker than weak convergence. Allowing for the abuse of notation, we can write $M \subset M''$. Consider now a sequence $x_n \in M'$ with $x_n \xrightarrow{w} x$ in M'. Then $\ell(x_n) \to \ell(x)$ for any $\ell \in M''$. This implies that $x_n(m) \to x(m)$ for all $m \in M$ and hence $x_n \xrightarrow{w^*} x$ in M'. Thus weak convergence implies w^* convergence.

The converse is true if M is reflexive. However, once we remove the reflexivity condition, the converse fails. Let X be the space of finite signed measures on $[-1, 1]$. We have already seen in theorem 2.39 that $C([-1, 1])' \cong X$. Consider the measures $\nu_n(A) = n\mu(A \cap [-1/n, 1/n])/2$. We claim that $\nu_n \xrightarrow{w^*} \delta_0$, where δ_0 is the Dirac measure at 0. Indeed, for any $f \in C([-1, 1])$, the corresponding functional ℓ_n for ν_n is given by

$$\ell_n(f) = \int_{-1}^{1} f d\nu_n = \frac{n}{2} \int_{-1/n}^{1/n} f(x) dx \to f(0) = \ell_0(f),$$

where ℓ_0 is the functional defined as $\ell_0 : f \mapsto f(0)$. Thus $\ell_n \xrightarrow{w^*} \ell_0$.

However, ℓ_n is not weakly convergent to ℓ_0. To see this, consider the evaluation functional $L_{\{0\}} : \ell \mapsto (\phi\ell)(\{0\})$, where ϕ is the isometric isomorphism from $C([-1,1])'$ to X. Then $L_{\{0\}} \in X' = M''$. However,

$$L_{\{0\}}(\ell_n) = \nu_n(\{0\}) = 0 \not\to 1 = \delta_0(\{0\}) = L_{\{0\}}(\ell_0).$$

Thus ν_n does not converge weakly to δ_0.

Definition 2.78
*Let M be a Banach space. A subset $K \subset M'$ is **weakly* sequentially compact** if every sequence $\{x_n\} \subset K$ has a subsequence $\{x_{n_k}\}$ such that $x_{n_k} \xrightarrow{w^*} x \in K$.*

Theorem 2.79 (Banach-Alaoglu)
Let M be a separable Banach space. Then the closed unit ball

$$B = \{x \in M' \mid \|x\| \leq 1\}$$

is weakly sequentially compact.*

Proof. This proof is similar to the proof of Kakutani theorem. Let $\{x_n\} \subset B$ be a sequence. By the separability of M, for every $m \in M$, there is a sequence $\{m_j\} \subset M$ such that $m_j \to m$ as $j \to \infty$. For any fixed j, $|x_n(m_j)| \leq \|m_j\|$ is a bounded sequence. Hence we can extract a subsequence x_{n_k} such that

$$x_{n_k}(m_j) \to A(m_j) \text{ as } k \to \infty$$

for some $A(m_j) \in \mathbb{R}$, with $A(m) = \lim_{k\to\infty} x_{n_k}(m)$. A is a bounded linear functional on M since it is clearly linear and

$$|A(m)| = \lim_{k\to\infty} |x_{n_k}(m)| \leq \lim_{k\to\infty} \|x_{n_k}\| \, \|m\| = \|m\|,$$

showing that $\|A\| \leq 1$. We claim that A is well-defined, i.e., the limit exists. Indeed, for any $m \in M$, we can find a sequence $\{m_j\}$ such that $m_j \to m$ as $j \to \infty$. Then

$$\begin{aligned}
|x_{n_k}(m) - x_{n_l}(m)| &\leq |x_{n_k}(m) - x_{n_k}(m_j)| + |x_{n_k}(m_j) - x_{n_l}(m_j)| + |x_{n_l}(m_j) - x_{n_l}(m)| \\
&\leq \|x_{n_k} - x_{n_l}\| \, \|m\| + |x_{n_k}(m_j) - x_{n_l}(m_j)| + \|x_{n_k} - x_{n_l}\| \, \|m\| \\
&\leq 2 \|x_{n_k} - x_{n_l}\| + |x_{n_k}(m_j) - x_{n_l}(m_j)| \to 0
\end{aligned}$$

as $k, l \to \infty$. Hence the sequence $\{x_{n_k}(m)\}$ is Cauchy and A is well-defined. Because of the boundedness of A, it is continuous and thus $x_{n_k}(m) \to x(m)$ for some $x \in M'$. Since $\|x_{n_k}\| \leq 1$, we have $\|x\| \leq 1$. Thus $x \in B$ and $x_{n_k} \xrightarrow{w^*} x$ in M'. We conclude that B is weakly* sequentially compact. ■

Definition 2.80
Let X be a normed space. $A \subset B \subset X''$ is ***weakly* dense*** *in B if for every $f \in B$, there exists a sequence $\{f_n\} \subset A$ such that $f_n \xrightarrow{w^*} f$ in X'. We also say that B is the* ***weak* closure*** *of A.*

Theorem 2.81 (Goldstine)
Let X be a Banach space and B be the closed unit ball in X. Consider the canonical mapping $J : X \to X''$ given by $J : x \mapsto (f \mapsto f(x))$. Then $J(B)$ is weakly dense in the closed unit ball in X''.*

Proof. We begin by showing the following claim: for all $\xi \in B''$, the closed unit ball in X'', $f_1, \ldots, f_n \in X'$, and $\delta > 0$, there exists an $x \in (1+\delta)B$ such that $f_i(x) = \xi(f_i)$ for all $i = 1, \ldots, n$. To show this, consider the mapping $\Phi : X \to \mathbb{R}^n$ given by

$$\Phi(x) = (f_1(x), \ldots, f_n(x)).$$

Then Φ is a surjective bounded linear mapping. Hence we can find $x \in X$ such that $f_i(x) = \xi(f_i)$ for all $i = 1, \ldots, n$. Now, define $Y = \bigcap_{i=1}^n \ker(f_i) = \ker(\Phi)$. Every $z \in (x+Y) \cap (1+\delta)B$ satisfies that $z \in (1+\delta)B$ and $f_i(z) = f_i(x)$ for all $i = 1, \ldots, n$. The claim follows once we show that $(x+Y) \cap (1+\delta)B \neq \varnothing$.

Suppose not. Then $d(x, Y) \geq 1+\delta$. Clearly, Y is closed by proposition 2.9 and $\{x\}$ is compact. By Hahn-Banach seperation theorem, we can find $f \in X'$ such that $f|_Y = 0$, $\|f\| = 1$, and $f(x) \geq 1+\delta$. $f \in \operatorname{span}\{f_1, \ldots, f_n\}$ and $1+\delta \leq f(x) = \xi(f) \leq \|f\|\,\|\xi\| \leq 1$, which is absurd.

Now, fix $\xi \in B''$, $f_1, \ldots, f_n \in X'$, and $\epsilon > 0$. Consider the weak* neighborhood of ξ given by

$$U = \{\zeta \in X'' \mid |\zeta(f_i) - \xi(f_i)| < \epsilon, i = 1, \ldots, n\}.$$

This is the base of the weak* topology on x''. The density of $J(B)$ in B'' follows once we show that $U \cap J(B) \neq \varnothing$. Our claim above asserts that since $J(B) \subset B''$, for any $\delta > 0$, there exists $x \in (1+\delta)B$ such that $J(x) \in (1+\delta)J(B) \cap U$. Rescaling gives $(1+\delta)^{-1}J(x) \in J(B)$. We proceed to show that for sufficiently small δ, we also have $(1+\delta)^{-1}U$.

$$\left|\xi(f_i) - \frac{1}{1+\delta}J(x)(f_i)\right| = \left|f_i(x) - \frac{1}{1+\delta}f_i(x)\right| = \frac{\delta}{1+\delta}\,|f_i(x)|.$$

Now pick δ such that $\delta \max_{1\leq i\leq n} \|f_i\| < \epsilon$. Since $\|x\| \leq 1+\delta$,

$$\frac{\delta}{1+\delta}\,|f_i(x)| \leq \frac{\delta}{1+\delta}\,\|f_i\|\,\|x\| \leq \delta \max_{1\leq i\leq n} \|f_i\| < \epsilon.$$

Thus $(1+\delta)^{-1}J(x) \in U$ and we conclude that $J(B) \cap U \neq \varnothing$. This shows that $J(B)$ is weakly* dense in B''. ∎

Theorem 2.82 (Milman-Pettis)
Every uniformly convex Banach space is reflexive.

Proof. Let X be a uniformly convex Banach space and $\xi \in X''$. We need to show that there exists a unique $x \in X$ such that $\xi = J(x)$, where J is the canonical mapping from X to X''. Without loss of generality, we can assume that $\|\xi\| = 1$. The injectivity of J gives the uniqueness of x. It remains to show the existence of x.

Consider the closed unit ball B in X. We first show that $J(B)$ is closed in X''. Indeed, if $\zeta_n \in J(B)$ is a sequence converging to $\zeta \in X''$, then there exists a sequence $\{z_n\} \subset B$ such that $\|\zeta_m - \zeta_n\| = \|J(z_m) - J(z_n)\| = \|z_m - z_n\|$ and z_n is Cauchy. By the completeness of X, $z_n \to z \in X$. Take $\zeta = J(z)$ and using the fact that J is isometric, we deduce that $J(B)$ is closed in X''. It now suffices to show that for any $\epsilon > 0$, there exists $x \in B$ such that $\|\xi - J(x)\| \leq \epsilon$.

Now, fix $\epsilon > 0$ and by the uniform convexity of X, there is $\delta > 0$ such that $\|(x+y)/2\| \leq 1 - \delta$ for all $x, y \in B$ with $\|x - y\| \geq \epsilon$. Choose $f \in X'$ such that $\|f\| = 1$ and $\xi(f) \geq 1 - \delta/2$. Set

$$V = \{\eta \in X'' \mid |\eta(f) - \xi(f)| < \delta/2\}.$$

By the Goldstine theorem, $J(B)$ is weakly* dense in the closed unit ball in X''. Hence $V \cap J(B) \neq \emptyset$ and there is $x \in B$ such that $J(x) \in V$. We claim that this x is the desired choice.

Suppose not, i.e., $\|\xi - J(x)\| > \epsilon$. Then $\xi \in (J(x) + \epsilon B'')^c$, where B'' is the closed unit ball in X''. $(J(x) + \epsilon B'')^c$ is also a neighborhood of ξ in weakly* topology. Using the Goldstine theorem again, we have that $V \cap (J(x) + \epsilon B'')^c \cap J(B) \neq \emptyset$. This means that there is some $y \in B$ such that $J(y) \in V \cap (J(x) + \epsilon B'')^c$. Then we obtain that

$$|J(y)(f) - \xi(f)| < \delta/2 \text{ and } |J(x)(f) - \xi(f)| < \delta/2.$$

Hence

$$2\xi(f) < J(x)(f) + J(y)(f) + \delta \leq \|x + y\| + \delta.$$

Recall that $\xi(f) \geq 1 - \delta/2$. Thus

$$2 - \delta < \|x + y\| + \delta \quad \Rightarrow \quad \left\|\frac{x+y}{2}\right\| > 1 - \delta.$$

This implies that $\|x - y\| \leq \epsilon$ by the uniform convexity of X. This is contradicting to our assumption. Hence we conclude that ξ lies in $J(B)$ and that X is reflexive. ■

Corollary 2.83
$\mathcal{L}^p(\Omega, \mu)$ is reflexive for any $1 < p < \infty$.

Proof. This is a direct consequence of the Clarkson theorem and Milman-Pettis theorem. Since for every $1 < p < \infty$, $\mathcal{L}^p(\Omega, \mu)$ is uniformly convex, and every uniformly convex Banach space is reflexive, we conclude that $\mathcal{L}^p(\Omega, \mu)$ is reflexive. ■

Remark
This corollary can also be inferred from the Riesz representation theorem twice.

Remark
Let X, Y be Banach spaces and $B(X,Y)$ be the space of bounded linear operators from X to Y. Consider the following topologies on $B(X,Y)$:

- *The uniform topology on $B(X,Y)$ is the topology induced by the uniform norm:*

$$\|m\|_{B(X,Y)} = \sup_{\|x\|_X \le 1} \|m(x)\|_Y .$$

  This is the coarest among the three topologies.

- *The strong topology on $B(X,Y)$ is the topology generated by the collection of sets:*

$$\left\{ B_{x,\epsilon}(T) = \{S \in B(X,Y) \mid \|Sx - Tx\| < \epsilon\} \,\middle|\, \epsilon > 0, x \in X, T \in B(X,Y) \right\} .$$

  This is the coarest topology that makes the evaluation map $m \mapsto m(x)$ continuous for all $x \in X$.

- *The weak topology on $B(X,Y)$ is the topology generated by the collection*

$$\left\{ B_{y',x,\epsilon}(T) = \{S \in B(X,Y) \mid |y'Sx - y'Tx| < \epsilon\} \,\middle|\, \epsilon > 0, y' \in Y', x \in X, T \in B(X,Y) \right\} .$$

2.7. Open Mapping Theorem and Closed Graph Theorem

Proposition 2.84
If X is a Baire space and F_n is a sequence of closed sets in X such that $\bigcup_{n=1}^{\infty} F_n = X$, then there exists some n and a nonempty open set G such that $G \subsetneq F_n$.

Proof. Let $G_n = F_n^c$ be open sets in X. Then $\bigcap_{n=1}^{\infty} G_n = \bigcap_{n=1}^{\infty} F_n^c = \left(\bigcup_{n=1}^{\infty} F_n\right)^c = \varnothing$. By the Baire category theorem, at least one of the G_n is not dense in X. Thus there is some $x \in G^c$ and an open neighborhood U of x such that $U \bigcap G_n = \varnothing$. This implies $U \subsetneq F_n$. ■

Theorem 2.85 (Open Mapping Theorem)
Let X and Y be Banach spaces and $T : X \to Y$ be a bounded surjective linear map. Then for any open set $U \subset X$, $T(U)$ is open in Y.

Proof. We first claim that for any open ball B centered at 0 in X, $\overline{T(B)}$ contains an open neighborhood of zero in Y. By the surjectivity, $Y \subset T(X) = T(\bigcup_n nB) = \bigcup_n T(nB) \subset \bigcup_n \overline{T(nB)}$. By proposition 2.84, there is some n such that $\overline{T(nB)}$ contains an interior point, say y, and some open ball $B_r(y) \subset \overline{T(nB)}$. Then for every $z \in Y$ with $\|z\| < r$, $z - y \in B_r(-y) \subset \overline{T(-nB)} = \overline{T(nB)}$ and

$$z = y + (z - y) \in y + B_r(-y) \subset \overline{T(nB)} + \overline{T(nB)} \subset \overline{T(2nB)}.$$

Deviding z by $2n$ gives that $z/2n \in \overline{T(B)}$ and $B_{r/2n}(0) \subset \overline{T(B)}$.

Next, let B be an unit ball. To shorten the notation, denote $r/2n$ by δ and $B_{r/2n}(0)$ by B_δ. Let $y \in B_\delta$ and $c_n > 0$ be a sequence. Since $B_\delta \subset \overline{T(B)}$, $\overline{B_\delta} \subset \overline{T(B)}$. Thus for every

$z \in Y$ and $\epsilon > 0$, we can find some $x \in X$ such that $\|x\| < \delta^{-1}\|z\|$ and $z \in B_\epsilon(T(x))$. Now taking $z = y$ and $\epsilon = c_1$, we can find an x_1 such that $\|x_1\| < \delta^{-1}\|y\|$ and $\|y - Tx_1\| < c_1$. Similarly, we can take $z = y - Tx_1$ and $\epsilon = c_2$ to find an x_2 such $\|x_2\| < \delta^{-1}\|y - Tx_1\| < \delta^{-1}c_1$ and $\|y - Tx_1 - Tx_2\| < c_2$. Iductively, we find a sequence $\{x_n\}$ such that $\|x_n\| < \delta^{-1}c_{n-1}$ and $\left\|y - T\left(\sum_{k=1}^n x_k\right)\right\| < c_n$. Now we choose $c_n = 2^{-n}c$ for arbitrary $c > 0$. Then

$$\left\|\sum_{k=1}^n x_k\right\| \le \sum_{k=1}^n \|x_k\| \le \frac{\|y\|}{\delta} + \sum_{k=2}^n \frac{c_{k-1}}{\delta} \le \frac{\|y\|}{\delta} + \frac{c}{\delta}\sum_{k=1}^\infty 2^{-k} = \frac{\|y\|}{\delta} + \frac{c}{\delta}.$$

Hence $\sum_n x_n$ converges in X to some x with $\|x\| < 1$ by making c arbitrarily small. Also,

$$\left\|y - T\left(\sum_{k=1}^n x_k\right)\right\| \le c_n = 2^{-n}c \to 0.$$

Thus $Tx = y$ and $y \in T(B)$, which implies $B_\delta \subset T(B)$.

Finally, let U be an open set in X. Then for any $y \in T(U)$, there is some $x \in U$ such that $y = Tx$. Since U is open, there is some $\epsilon > 0$ such that $B_\epsilon(x) \subset U$. By the previous claim, there is some $s > 0$ such that $B_s(0) \subset T(B_1(0))$. Multiplying both sides by ϵ gives $B_{s\epsilon}(0) \subset T(B_\epsilon(0))$. Then

$$B_{s\epsilon}(y) = y + B_{s\epsilon}(0) \subset y + T(B_\epsilon(0)) = Tx + T(B_\epsilon(0)) = T(x + B_\epsilon(0)) = T(B_\epsilon(x)) \subset T(U).$$

Thus $T(U)$ is open. This completes the proof. ∎

Theorem 2.86 (Bounded Inverse Theorem)
Let X and Y be Banach spaces and $T : X \to Y$ be a bounded linear map. If T is bijective, then T^{-1} is bounded.

Proof. By the open mapping theorem, there is $r > 0$ such that $B_r(0) \subset T(B_r(0))$. For any $y \in Y$ with $\|y\| = r/2$, there exists $x \in B_1(0)$ such that $y = Tx$. For $z \in Y$, write

$$z = \frac{rz}{2\|z\|}\frac{2}{r}\|z\|.$$

Then since $\left\|\frac{rz}{2\|z\|}\right\| = r/2$, there is some $x \in B_1(0)$ such that $\frac{rz}{2\|z\|} = Tx$. Thus $z = \frac{2}{r}\|z\|Tx$,

$$T^{-1}z = \frac{2}{r}\|z\|x \implies \left\|T^{-1}z\right\| \le \frac{2}{r}\|z\|\|x\|.$$

Note that $\|x\| \le 1$. We see that $\left\|T^{-1}\right\|$ is bounded by $2/r$. ∎

Remark
The completeness in the open mapping theorem is essential. For counterexample, consider X as the space of all sequences with finitely many nonzero terms equipped with the supremum

norm. Define $T : X \to X$ *by*

$$T(x_1, x_2, \ldots) = \left(x_1, \frac{x_2}{2}, \frac{x_3}{3}, \ldots\right).$$

Note that X *is not complete since the sequence* $x^{(n)} = (1, 1/2, \ldots, 1/n, 0, 0, \ldots)$ *converges to* $(1, 1/2, \ldots)$*, which does not belong to* X*. In this case* T^{-1} *exists but is not bounded.*

Definition 2.87
X, Y are Banach spaces. $T : X \to Y$ is a bounded linear map. The set

$$\Gamma(T) = \{(x, Tx) \in X \times Y \mid x \in X\}$$

is called the ***graph*** *of T. We define the norm of x on the graph by*

$$\|(x, Tx)\|_\Gamma = \|x\|_X + \|Tx\|_Y .$$

Note that $(\Gamma(T), \|\cdot\|_\Gamma)$ *forms a normed space.*

Definition 2.88
A linear map $T : X \to Y$ is called ***closed*** *if its graph is a closed, i.e., for any sequence $x_n \in X$, if $x_n \to x \in X$ and $Tx_n \to y \in Y$, then $Tx = y$ and $(x, y) \in \Gamma(T)$.*

Remark
If T is bounded, it is closed. To see this, note that if $x_n \to x \in X$, by the continuity we have $Tx_n \to Tx \in Y$.

Theorem 2.89 (Closed Graph Theorem)
Let X and Y be Banach spaces and $T : X \to Y$ be a linear map. If T is closed, then T is bounded.

Proof. Observe that $\Gamma(T)$ is a Banach space with the norm $\|\cdot\|$ on $\Gamma(T)$. This follows from the closedness of T. Now define $S : \Gamma(T) \to X$ by $S(x, Tx) = x$. We claim that S is bounded, linear and bijective. For linearity, let $(x_1, Tx_1), (x_2, Tx_2) \in \Gamma(T)$ and $c \in \mathbb{R}$.

$$S(c(x_1, Tx_1) + (x_2, Tx_2)) = S(cx_1 + x_2, cTx_1 + Tx_2) = cx_1 + x_2 = cS(x_1, Tx_1) + S(x_2, Tx_2).$$

For boundedness,

$$\|S(x, Tx)\|_X = \|x\|_X \le \|x\|_X + \|Tx\|_Y = \|(x, Tx)\|_\Gamma .$$

Thus $\|S\| \le 1$. For bijectivity, notice that

$$S(x_1, Tx_1) = S(x_2, Tx_2) \implies x_1 = S(x_1, Tx_1) = S(x_2, Tx_2) = x_2.$$

and for any $x \in X$, $(x, Tx) \in \Gamma(T)$ and $S(x, Tx) = x$. Thus S is bounded, linear and bijective.

By the bounded inverse theorem, $S^{-1} : X \to \Gamma(T)$ is bounded as well. For any $x \in X$,

$$\|Tx\|_Y = \|(x, Tx)\|_\Gamma - \|x\|_X = \left\|S^{-1}x\right\|_\Gamma - \|x\|_X \leq C\|x\|_X - \|x\|_X = (C-1)\|x\|_X$$

for some constant $C < \infty$. Thus T is bounded. ■

Remark

To apply the closed graph theorem, T must be closed in X. If T is only closed in $D(T)$, the domain of T, then the theorem does not hold. For example, let $X = Y = C[a,b]$ with sup norm and $T : C^1[a,b] \to C[a,b]$ be the differentiation operator $T(f) = f'$. Then T is closed in $C^1[a,b]$ while being unbounded. To see this, let $f_n(x) = \sin(nx)/n$. Then $T(f_n) = \cos(nx)$. $\|f_n\|_\infty \to 0$ while $\|T(f_n)\|_\infty = 1$. Thus T is not bounded. However, T is closed in $C^1[a,b]$. Let $u_n \in C^1[a,b]$ with $u_n \to u$ in $C[a,b]$ and $Tu_n = u_n' \to v \in C[a,b]$. Then $u \in C^1[a,b]$ and $Tu = v$. By definition, T is closed in $C^1[a,b]$.

Example

Let $X = \mathcal{L}^2(\mathbb{R})$ and $T : \left\{f \in \mathcal{L}^2(\mathbb{R}) \,\middle|\, xf(x) \in \mathcal{L}^2(\mathbb{R})\right\} \to X$ with $T(f) = xf(x)$. Consider $f_n = \frac{1}{n}\chi_{[n,n+1]}$. Then $\|f_n\|_2^2 = 1/n^2 \to 0$ and

$$\|T(f_n)\|_2^2 = \frac{1}{n}\int_n^{n+1} x dx = \frac{2n+1}{2n} \to 1.$$

Thus T is unbounded. If $u_n \to u$ in $\mathcal{L}^2(\mathbb{R})$ and $T(u_n) = xu_n(x) \to v$, then $T(u) = xu(x) = v$. Hence T is clsoed in $\left\{f \in \mathcal{L}^2(\mathbb{R}) \,\middle|\, xf(x) \in \mathcal{L}^2(\mathbb{R})\right\}$.

Definition 2.90

*Suppose X is a vector space with norms $\|\cdot\|_1$ and $\|\cdot\|_2$. The norms are said to be **compatible** if $x_n \to x$ in $\|\cdot\|_1$ and $x_n \to y$ in $\|\cdot\|_2$ implies $x = y$.*

Definition 2.91

*Let X be a vector space with norms $\|\cdot\|_1$ and $\|\cdot\|_2$. The norms are said to be **equivalent** if there are constants $c_1, c_2 > 0$ such*

$$c_1\|x\|_1 \leq \|x\|_2 \leq c_2\|x\|_1$$

for all $x \in X$.

Proposition 2.92

Suppose X is a vector space with norms $\|\cdot\|_1$ and $\|\cdot\|_2$. If the norms are equivalent, then they are compatible.

Proof. Suppose $\|x_n - x\|_1 \to 0$ and $\|x_n - y\|_2 \to 0$. By the equivalence, $\|x-y\|_1 \leq \|x - x_n\|_1 + \|x_n - y\|_1 \leq \|x - x_n\|_1 + c_2\|x_n - y\|_2 \to 0$ for some $c_2 > 0$. Similarly, $\|x-y\|_2 \leq c_1\|x - x_n\|_1 + \|x_n - y\|_2 \to 0$ for some $c_1 > 0$. Thus $x = y$. ■

Proposition 2.93

If $(X, \|\cdot\|_1)$ *and* $(X, \|\cdot\|_2)$ *are Banach spaces. Then the norms are equivalent.*

Proof. By the closed graph theorem, the identity map $I : (X, \|\cdot\|_1) \to (X, \|\cdot\|_2)$ is a closed linear map and thus bounded. Suppose $x_n \to x$ in $\|\cdot\|_1$. Then $x_n = Ix_n \to Ix = x$ in $\|\cdot\|_2$ by the continuity of I. Since I is bounded, $\|x\|_2 = \|Ix\|_2 \le c_1 \|x\|_1$ for some $c_1 > 0$. Applying the same argument exchanging the roles of $\|\cdot\|_1$ and $\|\cdot\|_2$ gives $\|x\|_1 \le c_2 \|x\|_2$ for some $c_2 > 0$. Hence

$$\frac{1}{c_2} \|x\|_1 \le \|x\|_2 \le c_1 \|x\|_1$$

and the norms are equivalent. ∎

3. Hilbert Space

3.1. Cauchy-Schwarz Inequality

Definition 3.1
Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{F}$ is an inner product on X if it satisfies

(a) $\langle cx + y, z \rangle = c \langle x, z \rangle + \langle y, z \rangle$ *for all $x, y, z \in X$ and $c \in \mathbb{F}$.*

(b) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ *for all $x, y \in X$.*

(c) $\langle x, x \rangle > 0$ *for all $x \neq 0$.*

Remark
An inner product automatically induces a norm on X by $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$.

Definition 3.2
*A **Hilbert space** is a complete vector space with an inner product inducing a norm that makes it a Banach space. We denote the Hilbert space by $\mathcal{H}$.*

Remark
*If X is a vector space with an inner product but not complete, then X is called a **pre-Hilbert space**.*

Proposition 3.3 (Cauchy-Schwarz Inequality)
For all $x, y \in \mathcal{H}$,

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

Furthermore, equality holds if and only if x and y are linearly dependent.

Proof. If $\langle x, y \rangle = 0$, then the inequality is trivial. Otherwise, let $t \in \mathbb{R}$. Then

$$\begin{aligned} 0 &\leq \left\langle t\frac{|\langle x, y\rangle|}{\langle x, y\rangle}x + y,\ t\frac{|\langle x, y\rangle|}{\langle x, y\rangle}x + y \right\rangle \\ &= t^2 \|x\|^2 + 2t\Re\left(\frac{|\langle x, y\rangle|}{\langle x, y\rangle} \langle x, y\rangle\right) + \|y\|^2 = t^2 \|x\|^2 + 2t |\langle x, y\rangle| + \|y\|^2 . \end{aligned}$$

Hence

$$4 |\langle x, y\rangle|^2 - 4 \|x\|^2 \|y\|^2 \leq 0 \implies |\langle x, y\rangle| \leq \|x\| \|y\| .$$

Note that if the equality holds, then

$$t^2 \|x\|^2 + 2t |\langle x, y\rangle| + \|y\|^2 = t^2 \|x\|^2 + 2t \|x\| \|y\| + \|y\|^2 = (t \|x\| + \|y\|)^2 = 0$$

by taking $t = -\|y\| / \|x\|$. But this implies that

$$t\frac{|\langle x, y\rangle|}{\langle x, y\rangle}x + y = 0$$

and so x and y are linearly dependent. Conversely, suppose $cx = y$. Then $|\langle x,\ y\rangle| = |c|\,\|y\|^2 = \|x\|\,\|y\|$. ■

Proposition 3.4 (Parallelogram Law)
For all $x, y \in \mathcal{H}$,

$$\|x + y\|^2 + \|x - y\|^2 = 2\,\|x\|^2 + 2\,\|y\|^2 .$$

Proof. Note that

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

Adding the two equations gives

$$\|x + y\|^2 + \|x - y\|^2 = 2\,\|x\|^2 + 2\,\|y\|^2 .$$

■

Proposition 3.5
For all $x \in \mathcal{H}$,

$$\|x\| = \sup_{\|y\|=1} |\langle x,\ y\rangle| .$$

Proof. By the Cauchy-Schwarz inequality,

$$|\langle x,\ y\rangle| \le \|x\|\,\|y\| \implies \left|\left\langle x,\ \frac{y}{\|y\|}\right\rangle\right| \le \|x\| \implies \sup_{\|y\|=1} |\langle x,\ y\rangle| \le \|x\| .$$

Taking $y = x/\|x\|$ gives the equality and $\|y\| = 1$. ■

Theorem 3.6 (Completion of Pre-Hilbert Space)
Let $(X, \langle\cdot,\ \cdot\rangle)$ be a pre-Hilbert space. Then there exists a Hilbert space $\mathcal{H}$ such that X is dense in $\mathcal{H}$ and $\langle\cdot,\ \cdot\rangle_*$ on $\mathcal{H}$ is an extension of $\langle\cdot,\ \cdot\rangle$.

Proof. Define $\langle x,\ y\rangle_* = \lim_{n\to\infty} \langle x_n,\ y_n\rangle$ for Cauchy sequences $\{x_n\}, \{y_n\} \subset X$ and $x, y \in \overline{X}$. We first check that $\langle\cdot,\ \cdot\rangle_*$ is well-defined. Note that

$$\begin{aligned} |\langle x_n,\ y_n\rangle - \langle x_m,\ y_m\rangle| &\le |\langle x_n,\ y_n\rangle - \langle x_n,\ y_m\rangle| + |\langle x_n,\ y_m\rangle - \langle x_m,\ y_m\rangle| \\ &\le \|x_n\|\,\|y_n - y_m\| + \|x_n - x_m\|\,\|y_m\| \longrightarrow 0 \end{aligned}$$

as $n, m \longrightarrow \infty$ by the Cauchy-Schwarz inequality. Since $\mathbb{F}$ is complete, the limit exists. To see that $\langle\cdot,\ \cdot\rangle_*$ is independent of the choice of sequences, suppose $\{x_n^1\}, \{y_n^1\}$ and $\{x_n^2\}, \{y_n^2\}$ are two pairs of Cauchy sequences converging to x and y respectively. Then

$$\begin{aligned} \left|\langle x_n^1,\ y_n^1\rangle - \langle x_n^2,\ y_n^2\rangle\right| &\le \left|\langle x_n^1,\ y_n^1\rangle - \langle x_n^1,\ y_n^2\rangle\right| + \left|\langle x_n^1,\ y_n^2\rangle - \langle x_n^2,\ y_n^2\rangle\right| \\ &\le \|x_n^1\|\,\|y_n^1 - y_n^2\| + \|x_n^1 - x_n^2\|\,\|y_n^2\| \longrightarrow 0. \end{aligned}$$

Hence $\langle x, y\rangle_*$ is well-defined. We now show that $\langle \cdot, \cdot\rangle_*$ is indeed an inner product on $\overline{X}$. For the linearity in the first argument, let $x, y, z \in \overline{X}$, $\{x_n\}, \{y_n\}, \{z_n\} \subset X$ be Cauchy sequences converging to x, y, z respectively and $c \in \mathbb{F}$. Then

$$\begin{aligned}\langle cx + y, z\rangle_* &= \lim_{n\to\infty} \langle cx_n + y_n, z_n\rangle = \lim_{n\to\infty} c\,\langle x_n, z_n\rangle + \langle y_n, z_n\rangle \\ &= c \lim_{n\to\infty} \langle x_n, z_n\rangle + \lim_{n\to\infty} \langle y_n, z_n\rangle = c\,\langle x, z\rangle_* + \langle y, z\rangle_* .\end{aligned}$$

For the conjugate symmetry, let $x, y \in \overline{X}$ and $\{x_n\}, \{y_n\} \subset X$ be Cauchy sequences converging to x and y respectively. Then

$$\overline{\langle x, y\rangle_*} = \overline{\lim_{n\to\infty} \langle x_n, y_n\rangle} = \lim_{n\to\infty} \overline{\langle x_n, y_n\rangle} = \lim_{n\to\infty} \langle y_n, x_n\rangle = \langle y, x\rangle_* .$$

For the positive definiteness, let $x \in \overline{X}$, $x \neq 0$ and $\{x_n\} \subset X$ be a Cauchy sequence converging to x. Then

$$\langle x, x\rangle_* = \lim_{n\to\infty} \langle x_n, x_n\rangle = \lim_{n\to\infty} \|x_n\|^2 > 0.$$

Hence $\langle \cdot, \cdot\rangle_*$ is an inner product on $\overline{X}$ and induces a norm on $\overline{X}$. Lastly, for every $x, y \in \overline{X}$, pick $x_n = x$ and $y_n = y$ to see that

$$\langle x, y\rangle_* = \lim_{n\to\infty} \langle x, y\rangle = \langle x, y\rangle ,$$

which shows that $\langle \cdot, \cdot\rangle_*$ is an extension of $\langle \cdot, \cdot\rangle$. We conclude that $\mathcal{H} = \overline{X}$ forms a Hilbert space. ■

Example
Let $X = C([0,1])$ with the inner product

$$\langle f, g\rangle = \int_0^1 f(x)g(x)dx.$$

Then X is a pre-Hilbert space. To see this, set $f_n(x) = x^n$.

$$\|f_m - f_n\|^2 = \int_0^1 (x^m - x^n)^2 dx = \frac{1}{2m+1} + \frac{2}{m+n+1} + \frac{1}{2n+1} \to 0$$

as $m, n \to \infty$. Hence $\{f_n\}$ is Cauchy in X. However, f_n converges to

$$f(x) = \begin{cases} 0 & x \in [0,1) \\ 1 & x = 1, \end{cases}$$

which is not in X. Hence X is not complete. But by the proposition 1.35, X is dense in $\mathcal{L}^2([0,1])$ and so X can be completed to a Hilbert space $\mathcal{H} = \mathcal{L}^2([0,1])$, which is complete by Riesz-

Fischer theorem.

Definition 3.7
A set X is called ***convex*** *if for all $x, y \in X$ and $t \in [0,1]$, $tx + (1-t)y \in X$.*

Theorem 3.8
Let $K \subset \mathcal{H}$ be a closed convex set. For $x \in \mathcal{H}$, define the distance from x to K as

$$d(x,K) = \inf_{y \in K} \|x - y\| .$$

Then there exists a unique $z \in K$ such that $d(x,K) = \|x - z\|$.

Proof. Let $\{y_n\} \subset K$ be a sequence such that $\|y_n - x\| \to d(x,K)$. We claim that $\{y_n\}$ is Cauchy. Let $\epsilon > 0$ be given. By the parallelogram law,

$$2\Big(\|x - y_n\|^2 + \|x - y_m\|^2\Big) = \|2x - y_n - y_m\|^2 + \|y_n - y_m\|^2$$

Rearranging gives

$$\begin{aligned}
\frac{1}{4}\|y_n - y_m\|^2 &= \frac{1}{2}\|x - y_n\|^2 + \frac{1}{2}\|x - y_m\|^2 - \left\|x - \frac{y_n + y_m}{2}\right\|^2 \\
&\le \frac{1}{2}(d(x,K) + \epsilon)^2 + \frac{1}{2}(d(x,K) + \epsilon)^2 - d(x,K)^2 \\
&= \epsilon^2 + 2\epsilon d(x,K)
\end{aligned}$$

for all $m, n \ge N$ for some $N \in \mathbb{N}$. The inequality follows from the fact that $(y_n + y_m)/2 \in K$ by the convexity of K. Since $\epsilon > 0$ is arbitrary, we conclude that $\{y_n\}$ is Cauchy. By the completeness of $\mathcal{H}$, $\{y_n\}$ converges to some $z \in \mathcal{H}$. Since K is closed, $z \in K$. To see the uniqueness, suppose $z_1, z_2 \in K$ are such that $\|x - z_1\| = \|x - z_2\| = d(x,K)$. Then by the parallelogram law,

$$\begin{aligned}
4d(x,K)^2 = 2\|x - z_1\|^2 + 2\|x - z_2\|^2 &= \|z_1 - z_2\|^2 + \|2x - z_1 - z_2\|^2 \\
&= \|z_1 - z_2\|^2 + 4\left\|x - \frac{z_1 + z_2}{2}\right\|^2 .
\end{aligned}$$

Hence

$$\|z_1 - z_2\|^2 = 4d(x,K)^2 - 4\left\|x - \frac{z_1 + z_2}{2}\right\|^2 \le 4d(x,K)^2 - 4d(x,K)^2 = 0$$

and so $z_1 = z_2$. ■

Definition 3.9
$Y \subset \mathcal{H}$ is a closed subspace. The ***orthogonal complement*** *of Y, denoted by $Y^\perp$, is defined as*

$$Y^\perp = \{x \in \mathcal{H} \mid \langle x,\ y\rangle = 0 \text{ for all } y \in Y\} .$$

Proposition 3.10
$Y \subset \mathcal{H}$ is a closed subspace. Then

(a) $Y^\perp$ *is a closed subspace.*

(b) $\mathcal{H} = Y \oplus Y^\perp$.

(c) $(Y^\perp)^\perp = Y$.

Proof. For (a), we first check that $Y^\perp$ is a subspace. First note that $0 \in Y^\perp$. Also, if $x, z \in Y^\perp$ and $c \in \mathbb{F}$, then

$$\langle cx + z,\, y\rangle = c\,\langle x,\, y\rangle + \langle z,\, y\rangle = 0$$

for all $y \in Y$. Hence $cx + z \in Y^\perp$. This shows that $Y^\perp$ is a subspace. To see that $Y^\perp$ is closed, let $\{x_n\} \subset Y^\perp$ be a sequnce converging to x. Then for all $y \in Y$,

$$|\langle x,\, y\rangle - \langle x_n,\, y\rangle| = |\langle x - x_n,\, y\rangle| \leq \|x - x_n\|\,\|y\| \to 0$$

by the Cauchy-Schwarz inequality. Thus $\langle \cdot,\, y\rangle$ is a continuous functional on $\mathcal{H}$. Therefore,

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

for all $y \in Y$ and so $x \in Y^\perp$. This shows that $Y^\perp$ is closed.

For (b), notice that Y as a closed subspace is convex. By theorem 3.8, for any $u \in \mathcal{H}$, there exists a unique $y \in Y$ such that $\|u - y\| \leq \|u - y'\|$ for all $y' \in Y$. Let $z = u - y$. We claim that $z \in Y^\perp$. To see this, let $y' \in Y$ and $t \in \mathbb{R}$. Then

$$\begin{aligned}\|z\|^2 = \|u - y\|^2 &\leq \|u - y - ty'\|^2 \\ &= \|u - y\|^2 - 2t\Re(\langle u - y,\, y'\rangle) + t^2\,\|y'\|^2 \\ &= \|z\|^2 - 2t\Re(\langle z,\, y'\rangle) + t^2\,\|y'\|^2.\end{aligned}$$

Rearranging gives

$$2t\Re(\langle z,\, y'\rangle) - t^2\,\|y'\|^2 \leq 0.$$

If $y' = 0$, we have $\langle z,\, y'\rangle = 0$; if $y' \neq 0$, then take $t = \Re\left(\langle z,\, y'\rangle / \|y'\|^2\right)$. Substituting this back gives

$$0 \geq 2\frac{(\Re(\langle z,\, y'\rangle))^2}{\|y'\|^2} - \frac{(\Re(\langle z,\, y'\rangle))^2}{\|y'\|^2} = \frac{(\Re(\langle z,\, y'\rangle))^2}{\|y'\|^2}.$$

Hence $\Re(\langle z,\, y'\rangle) = 0$ for all $y' \in Y$. Similarly, replacing t with it gives $\Im(\langle z,\, y'\rangle) = 0$ for all $y' \in Y$. Therefore, $\langle z,\, y'\rangle = 0$ for all $y' \in Y$ and so $z \in Y^\perp$. Since our choice of y is unique, we can write $u = y + z$ uniquely for $y \in Y$ and $z \in Y^\perp$. This shows that $\mathcal{H} = Y \oplus Y^\perp$.

For (c), note that we can apply (a) and (b) to $Y^\perp$ and obtain that $(Y^\perp)^\perp$ is a closed subspace and $\mathcal{H} = Y \oplus Y^\perp = (Y^\perp)^\perp \oplus Y^\perp$. It follows that for every $u \in \mathcal{H}$, we can write $u = y + z = x + z$ for $x \in (Y^\perp)^\perp$, $y \in Y$ and $z \in Y^\perp$ by the uniqueness of decomposition. This implies that $y = x$ and hence $(Y^\perp)^\perp = Y$. ∎

Remark

From the proposition, we can define the orthogonal projection P onto Y as $P(x) = P(y + y^\perp) = y$

for all $x \in \mathcal{H}$ where $y \in Y$ and $y^\perp \in Y^\perp$. Such decomposition $x = y + y^\perp$ is unique by (b) and hence P is well-defined.

3.2. Separability and Orthonormal Basis

Definition 3.11
*A Hilbert space $\mathcal{H}$ is said to be **separable** if there exists a countable dense subset in $\mathcal{H}$.*

Definition 3.12
*$\{x_\alpha \mid \alpha \in A\} \subset \mathcal{H}$, where A is an arbitrary index set. The **linear span** of $\{x_\alpha \mid \alpha \in A\}$ is defined as*

$$\operatorname{span}\{x_\alpha\} = \left\{ \sum_{\alpha \in A} c_\alpha x_\alpha \,\middle|\, c_\alpha \in \mathbb{F} = \mathbb{R} \text{ or } \mathbb{C} \right\},$$

where the sum is a finite sum.

Definition 3.13
*$\{x_\alpha \mid \alpha \in A\} \subset \mathcal{H}$, where A is an arbitrary index set. The **closed linear span** of $\{x_\alpha \mid \alpha \in A\}$ is defined as the smallest closed subspace of $\mathcal{H}$ containing $\{x_\alpha \mid \alpha \in A\}$.*

Proposition 3.14
Let $Y = \overline{\operatorname{span}(\{x_\alpha\})} \subset \mathcal{H}$ be a closed linear span of $\{x_\alpha\}$. Then for any $x \in \mathcal{H}$, $\langle x, x_\alpha \rangle = 0$ for all $\alpha \in A$ if and only if $\langle x, z \rangle = 0$ for all $z \in Y$.

Proof. Assume first that for $x \in \mathcal{H}$, $\langle x, x_\alpha \rangle = 0$ for all $\alpha \in A$. For each $z \in Y$, write $z = \sum_{\alpha_j \in A} c_{\alpha_j} x_{\alpha_j}$. Then

$$\langle x, z \rangle = \lim_{M \to \infty} \left\langle x, \sum_{j=1}^{M} c_{\alpha_j} x_{\alpha_j} \right\rangle = \lim_{M \to \infty} \sum_{j=1}^{M} \overline{c_{\alpha_j}} \left\langle x, x_{\alpha_j} \right\rangle = 0.$$

The converse is trivial since $x_\alpha \in Y$ for all $\alpha \in A$. ∎

Definition 3.15
*$\{x_\alpha \mid \alpha \in A\}$ is said to be **orthonormal** if $\left\langle x_\alpha, x_\beta \right\rangle = \delta_{\alpha\beta}$ for all $\alpha, \beta \in A$.*

Definition 3.16
*$\{x_\alpha \mid \alpha \in A\}$ forms a **othonormal basis** of $\mathcal{H}$ if it is orthonormal and $\overline{\operatorname{span}(\{x_\alpha\})} = \mathcal{H}$.*

Remark
This definition of basis is different from the definition of basis in linear algebra. In linear algebra, one can only express a vector as a finite linear combination of basis vectors; however, in Hilbert space, one can express a vector as a countable linear combination of basis vectors.

Lemma 3.17 (Bessel's Inequality)
Let $\{x_\alpha \mid \alpha \in A\}$ be an orthonormal set in $\mathcal{H}$. For any $x \in \mathcal{H}$, let $c_\alpha = \langle x, x_\alpha \rangle$. Then

(a) The set $\{\alpha \mid c_\alpha \neq 0\}$ *is at most countable.*

(b) $\sum_\alpha |c_\alpha|^2 \leq \|x\|^2$.

Proof. We assume that (a) is established and prove (b) first. Let $J \subset A$ be a countable subset with $J = \{\alpha_k \mid k \in \mathbb{N}\}$. For each $M \in \mathbb{N}$,

$$
\begin{aligned}
0 &\leq \left\| \sum_{k=1}^{M} c_{\alpha_k} x_{\alpha_k} - x \right\|^2 = \sum_{k=1}^{M} \left|c_{\alpha_k}\right|^2 - 2\Re\left(\left\langle x, \sum_{k=1}^{M} c_{\alpha_k} x_{\alpha_k} \right\rangle\right) + \|x\|^2 \\
&= \sum_{k=1}^{M} \left|c_{\alpha_k}\right|^2 - 2\Re\left(\sum_{k=1}^{M} \overline{c_{\alpha_k}} \left\langle x, x_{\alpha_k} \right\rangle\right) + \|x\|^2 = \sum_{k=1}^{M} \left|c_{\alpha_k}\right|^2 - 2\sum_{k=1}^{M} \left|c_{\alpha_k}\right|^2 + \|x\|^2 \\
&= \|x\|^2 - \sum_{k=1}^{M} \left|c_{\alpha_k}\right|^2 \implies \sum_{k=1}^{M} \left|c_{\alpha_k}\right|^2 \leq \|x\|^2 .
\end{aligned}
$$

Taking $M \to \infty$, we have $\sum_{k=1}^{\infty} \left|c_{\alpha_k}\right|^2 \leq \|x\|^2$.

Now we turn back to establish (a). For $m \in \mathbb{N}$, let $J_m = \{\alpha \in A \mid |c_\alpha| \geq 1/m\}$. Then J_m is finite or we can find infinitely many $\alpha \in J_m$ such that $|c_\alpha| \geq 1/m$. This implies that $x \in \mathcal{H}$ and

$$
\|x\|^2 = \sum_{\alpha \in A} |c_\alpha|^2 \geq \sum_{\alpha \in J_m} |c_\alpha|^2 \geq \sum_{\alpha \in J_m} \frac{1}{m^2} = \infty,
$$

which is absurd. Thus J_m is finite for all $m \in \mathbb{N}$. Observe that $\bigcup_{m \in \mathbb{N}} J_m = \{\alpha \in A \mid c_\alpha \neq 0\}$. It follows that as a countable union of finite sets, $\{\alpha \in A \mid c_\alpha \neq 0\}$ is at most countable. (b) follows from (a) and the previous argument. ■

Remark

There is a non-separable Hilbert space. Consider an uncountable set S. Let

$$
\mathcal{H} = \left\{ f : S \to \mathbb{R} \;\middle|\; \sum_{s \in S} f(s)^2 < \infty, f(S) \setminus \{0\} \text{ is at most countable} \right\}.
$$

Then $\mathcal{H}$ *is a Hilbert space with the inner product*

$$
\langle f, g \rangle = \sum_{s \in S} f(s) g(s).
$$

To see that $\mathcal{H}$ *is a Hilbert space, note that the countable union of countably many non-zero points is countable. Also, it is not separable since the set*

$$
\{e_s : S \to \mathbb{R} \mid e_s(t) = \delta_{st}\}
$$

forms an orthonormal set in $\mathcal{H}$ *and for each* $s \neq r$, $\|e_s - e_r\| = \sqrt{2}$. *This shows that it is nowhere dense in* $\mathcal{H}$. *Thus* $\mathcal{H}$ *is not separable.*

Proposition 3.18

Let $\{x_\alpha\}$ be an orthonormal set and Y be the closed linear span of $\{x_\alpha\}$. Then

$$Y = \left\{ \sum_j c_j x_{\alpha_j} \;\middle|\; \sum_j \left|c_j\right|^2 < \infty, \alpha_j \in A \right\}.$$

Proof. Let $S = \left\{ \sum_j c_j x_{\alpha_j} \;\middle|\; \sum_j \left|c_j\right|^2 < \infty, \alpha_j \in A \right\}$. For $x \in S$, $x = \sum_j c_j x_{\alpha_j}$ with $\sum_j \left|c_j\right|^2 < \infty$. Then $z_n = \sum_{j=1}^n c_j x_{\alpha_j} \to x$ as $n \to \infty$. Each $z_n \in Y$ and thus $x \in Y$. Hence $S \subset Y$.

Conversely, we claim that S is a closed subspace of $\mathcal{H}$. Clearly $0 \in S$. For $c \in \mathbb{F}$, $x = \sum_j c_j x_{\alpha_j} \in S$ and $y = \sum_j d_j x_{\alpha_j} \in S$, we have

$$cx + y = c \sum_j c_j x_{\alpha_j} + \sum_j d_j x_{\alpha_j} = \sum_j (cc_j + d_j) x_{\alpha_j} \in S,$$

where the summation is over all j such that either $c_j \neq 0$ or $d_j \neq 0$. To see that S is closed, let $z_n \in S$ where $z_n = \sum_j c_j^n x_{\alpha_j^n}$ with $\sum_j \left|c_j^n\right|^2 < \infty$ for all $n \in \mathbb{N}$. Let $J_n = \left\{ \alpha_j^n \;\middle|\; j \in \mathbb{N} \right\}$ and $J = \bigcup_n J_n \subset A$ is at most countable. Consider the transformation $T : S \to \ell^2$ defined by $\sum_j c_j x_{\alpha_j} \mapsto \left\{c_j\right\}$. Such definition is well-defined since if $\sum_j c_j x_{\alpha_j} = \sum_j d_j x_{\alpha_j}$, then $\sum_j (c_j - d_j) x_{\alpha_j} = 0$ and thus $c_j = d_j$ for all j since every x_{α_j} is orthogonal and thus linearly independent. Furthermore, T is clearly linear. Also, it is isometric since

$$\left\| \sum_j c_j x_{\alpha_j} \right\|^2 = \sum_j \left|c_j\right|^2 = \left\| \left\{c_j\right\} \right\|^2.$$

For $z_n \in S$, $z_n \to z$. Since z_n is Cauchy and T is isometric, $\left\{c_j^n\right\}$ is Cauchy in ℓ^2 and thus converges to some $\left\{c_j\right\} \in \ell^2$. Define $w = \sum_j c_j x_{\alpha_j} \in S$. It follows that $Tz_n \to Tw$. Hence $z = w \in S$ by the isometry of T. Thus S is closed. It follows that by the definition of Y, $Y \subset S$. We conclude that $Y = S$. ■

Lemma 3.19 (Gram-Schmidt)
Suppose $\{x_\alpha\}$ is an orthonormal set in $\mathcal{H}$ with $\overline{\operatorname{span}(\{x_\alpha\})} \neq \mathcal{H}$. Then there exists $y \in \mathcal{H}$ such that $\{x_\alpha\} \bigcup \{y\}$ is orthonormal.

Proof. Pick $z \in \mathcal{H}$ such that $z \notin \overline{\operatorname{span}(\{x_\alpha\})}$. By lemma 3.17, there are at most countably many α such that $\langle z, x_\alpha \rangle \neq 0$. Let α_j denumerate all α such that $\langle z, x_\alpha \rangle \neq 0$. Set $\hat{z} = \sum_j \left\langle z, x_{\alpha_j} \right\rangle x_{\alpha_j}$. For each x_{α_k},

$$\begin{aligned} \left\langle z - \hat{z}, x_{\alpha_k} \right\rangle &= \lim_{m \to \infty} \left\langle z - \sum_{j=1}^m \left\langle z, x_{\alpha_j} \right\rangle x_{\alpha_j}, x_{\alpha_k} \right\rangle \\ &= \lim_{m \to \infty} \left\langle z, x_{\alpha_k} \right\rangle - \sum_{j=1}^m \left\langle z, x_{\alpha_j} \right\rangle \delta_{jk} \\ &= \left\langle z, x_{\alpha_k} \right\rangle - \left\langle z, x_{\alpha_k} \right\rangle = 0. \end{aligned}$$

And for those x_α such that $\langle z, x_\alpha \rangle = 0$, $\langle z - \hat{z}, x_\alpha \rangle = \langle z, x_\alpha \rangle - \langle \hat{z}, x_\alpha \rangle = 0$ since x_{α_j} and x_α are orthogonal. Now set $y = (z - \hat{z})/\|z - \hat{z}\|$. Then $\{x_\alpha\} \bigcup \{y\}$ forms a orthonormal set. ■

Theorem 3.20

Every Hilbert space has an orthonormal basis.

Proof. We plan to use Zorn's lemma. Denote the space consisting of all orthonormal sets in $\mathcal{H}$ by $\mathcal{O}$. Define a partial order as the inclusion of sets. Let $\mathcal{C} \subset \mathcal{O}$ be a chain. We claim that $B = \bigcup_{\{x_\alpha\} \in \mathcal{C}} \{x_\alpha\}$ is an upper bound of $\mathcal{C}$. By construction we have $\{x_\alpha\} \subset B$ for all $\{x_\alpha\} \in \mathcal{C}$. We need to show that $B \in \mathcal{O}$. For distinct $x_\alpha, x_\beta \in B$, they belong to a common set $C \in \mathcal{C} \subset \mathcal{O}$. Hence C is orthonormal and $\langle x_\alpha, x_\beta \rangle = 0$. Also, it is clear that for every $x_\alpha \in B$, x_α belongs to some $C \in \mathcal{C}$ and thus $\|x_\alpha\| = 1$. It follows that B is also orthonormal. By Zorn's lemma, there exists a maximal element in $\mathcal{O}$, say B, such that if $C \in \mathcal{O}$ and $B \subset C$, then $B = C$. We claim that B is an orthonormal basis. It suffices to check that $\overline{\operatorname{span}(B)} = \mathcal{H}$. Suppose not, then by lemma 3.19, there exists $y \in \mathcal{H}$ such that $\{x_\alpha\} \bigcup \{y\}$ forms an orthonormal set. This contradicts the maximality of B. We conclude that B is an orthonormal basis. ■

Theorem 3.21

Let $\mathbb{F} = \mathbb{R}$ *or* $\mathbb{C}$. *Suppose that* $\mathcal{H}$ *is a Hilbert space. Then* $\mathcal{H}$ *is separable if and only if* $\mathcal{H}$ *has a countable orthonormal basis.*

Proof. Suppose that $\mathcal{H}$ has a countable orthonormal basis $\{x_n\}$. Then consider the sets

$$A_n = \left\{ \sum_{j=1}^{n} c_j x_j \;\middle|\; c_j \in S \right\},$$

where $S = \mathbb{Q}$ if $\mathbb{F} = \mathbb{R}$ and $S = \mathbb{Q} + \mathbb{Q}i$ if $\mathbb{F} = \mathbb{C}$. Since S is countable, each A_n being a finite union of countable sets is countable. Put $A = \bigcup_n A_n$ and let $\epsilon > 0$ be given. Since A is a countable union of countable sets, it is also countable. For every $x \in \mathcal{H}$, we can write $x = \sum_j \langle x, x_j \rangle x_j$ with

$$\left\| \sum_{j=N+1}^{\infty} \langle x, x_j \rangle x_j \right\| < \frac{\epsilon}{2}$$

for some $N \in \mathbb{N}$. Since S is dense in $\mathbb{F}$, we can pick some $c_j \in S$ with $\left| c_j - \langle x, x_j \rangle \right| < \epsilon/2^{j+1}$. Then

$$\begin{aligned}
\left\| x - \sum_{j=1}^{N} c_j x_j \right\| &= \left\| \sum_{j=1}^{\infty} \langle x, x_j \rangle x_j - \sum_{j=1}^{N} \langle x, x_j \rangle x_j + \sum_{j=1}^{N} \langle x, x_j \rangle x_j - \sum_{j=1}^{N} c_j x_j \right\| \\
&\leq \left\| \sum_{j=1}^{\infty} \langle x, x_j \rangle x_j - \sum_{j=1}^{N} \langle x, x_j \rangle x_j \right\| + \left\| \sum_{j=1}^{N} \langle x, x_j \rangle x_j - \sum_{j=1}^{N} c_j x_j \right\| \\
&\leq \left\| \sum_{j=N+1}^{\infty} \langle x, x_j \rangle x_j \right\| + \sum_{j=1}^{N} \left| \langle x, x_j \rangle - c_j \right| \|x_j\| \leq \frac{\epsilon}{2} + \sum_{j=1}^{N} \frac{\epsilon}{2^{j+1}} \leq \epsilon.
\end{aligned}$$

It follows that A is dense in $\mathcal{H}$ and hence $\mathcal{H}$ is separable.

Conversely, suppose that $\mathcal{H}$ is separable. Let $S \subset \mathcal{H}$ be a countable subset. Assume that every orthonormal basis of $\mathcal{H}$ is uncountable. Denote an orthonormal basis of $\mathcal{H}$ by $\{x_\alpha\}$. For each distinct $x_\alpha\ x_\beta \in S$, $\left\|x_\alpha - x_\beta\right\| = \sqrt{2}$. Consider the open balls $B_{1/2}(x_\alpha)$. They are clearly disjoint since if y lies in two such balls, then $\sqrt{2} = \left\|x_\alpha - x_\beta\right\| \leq \|x_\alpha - y\| + \left\|y - x_\beta\right\| < 1$, which is absurd. Now since S is dense in $\mathcal{H}$, for each α we can find some $s_\alpha \in S$ such that $s_\alpha \in B_{1/2}(x_\alpha)$. It follows that each s_α is distinct and thus S is uncountable. This contradicts to our assumption that S is countable. Thus $\mathcal{H}$ must have a countable orthonormal basis. ■

Proposition 3.22
Let $\mathcal{H}$ be a Hilbert space and $\{x_\alpha \mid \alpha \in A\}, \left\{y_\beta \mid \beta \in B\right\}$ be two orthonormal bases in $\mathcal{H}$. Then $\operatorname{card}(A) = \operatorname{card}(B)$.

Proof. Fixed an $\alpha \in A$, $B_\alpha = \left\{\beta \in B \,\middle|\, \left\langle y_\beta, x_\alpha \right\rangle \neq 0\right\}$ is at most countable by lemma 3.17 and $B_\alpha \subset B$. We claim that $B \subset \bigcup_{\alpha\in A} B_\alpha$. Take $\beta \in B$, we can write $y_\beta = \sum_k \left\langle y_\beta, x_{\alpha_k}\right\rangle x_{\alpha_k}$ with at least one $\left\langle y_\beta, x_{\alpha_k}\right\rangle \neq 0$. Hence $\beta \in B_{\alpha_k}$ for some $\alpha_k \in A$. It follows that $B \subset \bigcup_{\alpha\in A} B_\alpha$ and hence $\operatorname{card}(B) \leq \operatorname{card}(A)$. By symmetry, we have $\operatorname{card}(A) \leq \operatorname{card}(B)$ and thus $\operatorname{card}(A) = \operatorname{card}(B)$. ■

Remark
If $\mathcal{H}$ is separable, then $\mathcal{H}$ has a countable orthonormal basis and hence every orthonormal basis of $\mathcal{H}$ is countable.

Proposition 3.23 (Parseval's Identity)
Let $\{x_\alpha\}$ be an orthonormal basis of $\mathcal{H}$. Then

$$\|x\|^2 = \sum_j \left|\left\langle x, x_{\alpha_j}\right\rangle\right|^2.$$

Proof. Let $x \in \mathcal{H}$. Write $x = \sum_j c_j x_{\alpha_j}$ with $\sum_j \left|c_j\right|^2 < \infty$. Then

$$\left\langle x, x_{\alpha_k}\right\rangle = \lim_{M\to\infty} \left\langle \sum_{j=1}^M c_j x_{\alpha_j}, x_{\alpha_k}\right\rangle = \lim_{M\to\infty} \sum_{j=1}^M c_j \left\langle x_{\alpha_j}, x_{\alpha_k}\right\rangle = c_k.$$

It follows that

$$\|x\|^2 = \lim_{M\to\infty} \left\langle \sum_{j=1}^M c_j x_{\alpha_j}, \sum_{j=1}^M c_j x_{\alpha_j}\right\rangle = \lim_{M\to\infty} \sum_{j=1}^M \left|c_j\right|^2 = \sum_j \left|c_j\right|^2 = \sum_j \left|\left\langle x, x_{\alpha_j}\right\rangle\right|^2.$$

■

3.3. Riesz Representation and Bilinear Form

Proposition 3.24
$\mathbb{F} = \mathbb{R}$ *or* $\mathbb{C}$. *$T : \mathcal{H} \to \mathbb{F}$ is a nonzero bounded linear functional.*

(a) $\mathcal{H} = \operatorname{span}(\{w\}) \oplus \ker(T)$ *for* $w \notin \ker(T)$.

(b) *If* S, T *are bounded linear functionals and* $\ker(S) = \ker(T)$, *then there exists* $c \in \mathbb{F}$ *such that* $S = cT$.

(c) $\ker(T)$ *is closed.*

Proof. For (a), since T is nonzero, there is some w such that $Tw \neq 0$. For $x \in \mathcal{H}$, set $\alpha = Tx/Tw$ and $u = x - \alpha w$. Then $x = \alpha w + u$ and

$$Tu = Tx - \frac{Tx}{Tw}Tw = 0.$$

Hence $u \in \ker(T)$. Also, if $v \in \operatorname{span}(\{w\}) \bigcap \ker(T)$, $v = cw$ and $Tv = 0$. Then $cTw = Tv = 0$; $c = 0$ and thus $v = 0$. Therefore, $\operatorname{span}(\{w\}) \bigcap \ker(T) = \{0\}$ and $\mathcal{H} = \operatorname{span}(\{w\}) \oplus \ker(T)$.

To see (b), note that if $S = 0$, $\mathcal{H} = \ker(S) = \ker(T)$. Thus $T = 0$. If $S \neq 0$, by (a) we can write $\mathcal{H} = \operatorname{span}(\{w\}) \oplus \ker(S) = \operatorname{span}(\{w\}) \oplus \ker(T)$. Then for every $x \in \mathcal{H}$, $x = \alpha w + u$ for some $\alpha \in \mathbb{F}$ and $u \in \ker(T) = \ker(S)$. Then $Tw \neq 0$ and

$$Sx = S(\alpha w + u) = \alpha Sw = \alpha Tw\frac{Sw}{Tw} = \frac{Sw}{Tw}T(\alpha w + u) = \frac{Sw}{Tw}Tx.$$

Taking $c = Sw/Tw$ gives $S = cT$.

For (c), let $x_n \in \ker(T)$ be a sequence such that $x_n \to x \in \mathcal{H}$. Since T is continuous,

$$Tx = \lim_{n\to\infty} Tx_n = 0.$$

Hence $x \in \ker(T)$ and $\ker(T)$ is closed. ∎

Theorem 3.25 (Riesz Representation on $\mathcal{H}$)
$\mathbb{F} = \mathbb{R}$ *or* $\mathbb{C}$. $T : \mathcal{H} \to \mathbb{F}$ *is a bounded linear functional. Then there exists a unique* $x^* \in \mathcal{H}$ *such that* $Ty = \langle y, x^* \rangle$ *for all* $y \in \mathcal{H}$.

Proof. If $T = 0$, pick $x^* = 0$ then $Ty = 0 = \langle y, 0 \rangle$. If $T \neq 0$, there is some $w \in \mathcal{H}$ such that $Tw \neq 0$. By proposition 3.24, we can write $\mathcal{H} = \operatorname{span}(\{w\}) \oplus \ker(T)$ with $\ker(T)$ closed. Also, $\mathcal{H} = \ker(T) \oplus \ker(T)^\perp$ by proposition 3.10. We claim that $\ker(T)^\perp = \operatorname{span}(\{w\})$. First note that $\ker(T)^\perp \neq \{0\}$ or we would have $\mathcal{H} = \ker(T)$ and $T = 0$, contradicting to our assumption. Now if $z_1, z_2 \in \ker(T)^\perp$, write $z_1 = \alpha_1 w + u_1$ and $z_2 = \alpha_2 w + u_2$ for some $\alpha_1, \alpha_2 \in \mathbb{F}$ and $u_1, u_2 \in \ker(T)$. Then $\alpha_2 z_1 - \alpha_1 z_2 = \alpha_2 u_1 - \alpha_1 u_2 \in \ker(T)$ and $\alpha_2 z_1 - \alpha_1 z_2 \in \ker(T)^\perp$. Hence $\alpha_2 z_1 - \alpha_1 z_2 = 0$ and z_1, z_2 are linearly dependent. Now define $S : \mathcal{H} \to \mathbb{F}$ by $Sx = \langle x, w \rangle$. Then S is a bounded linear functional and $\ker(S) = \{x \in \mathcal{H} \mid \langle x, w \rangle = 0\} = (\ker(T)^\perp)^\perp = \ker(T)$ by proposition 3.10. Applying (b) of proposition 3.24 gives $cS = T$ for some $c \in \mathbb{F}$. Then $Tx = cSx = c\langle x, w \rangle = \langle x, \bar{c}w \rangle$. Set $x^* = \bar{c}w$ proves the existence of x^*.

To see uniqueness, suppose $x_1^*, x_2^* \in \mathcal{H}$ are such that $Ty = \langle y, x_1^* \rangle = \langle y, x_2^* \rangle$ for all $y \in \mathcal{H}$. Then $\langle y, x_1^* - x_2^* \rangle = 0$ for all $y \in \mathcal{H}$. Hence $x_1^* - x_2^* = 0$ and $x_1^* = x_2^*$. Such x^* is unique. ∎

Remark
From the Riesz representation, we can see that $\mathcal{H}' \cong \mathcal{H}$ and applying the Riesz representation theorem again gives $\mathcal{H}'' \cong \mathcal{H}$. Thus $\mathcal{H}$ is reflexive.

Definition 3.26
*The **adjoint operator** of $T : \mathcal{H} \to \mathcal{H}$ is the operator $T^* : \mathcal{H} \to \mathcal{H}$ such that $\langle Tx,\ y\rangle = \langle x,\ T^*y\rangle$ for all $x, y \in \mathcal{H}$.*

Remark
$T : \mathcal{H} \to \mathcal{H}$ is a bounded linear operator. By the Riesz representation, $\mathcal{H}' = \mathcal{H}$. Thus $T' : \mathcal{H}' \to \mathcal{H}'$ is defined by $T' : \ell \mapsto T'\ell = \ell T$. $T^* : \mathcal{H} \to \mathcal{H}' = \mathcal{H}$ is defined by $T^* : x \mapsto T^*x = T'\ell y$. For $x, y \in \mathcal{H}$,

$$\langle Tx,\ y\rangle = \langle x,\ T^*y\rangle = \ell_{T^*y}(x) = T'\ell_y(x) = \ell_y(Tx).$$

Definition 3.27
*Let X, Y be vector spaces. $T : X \to Y$ is called **skew-linear** if $T(cx + y) = \overline{c}Tx + Ty$ for all $x, y \in X$ and $c \in \mathbb{F}$.*

Definition 3.28
*$\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. $B : \mathcal{H} \times \mathcal{H} \to \mathbb{F}$ is called a **bilinear form** if*

(a) $B(\cdot, x)$ is linear for all $x \in \mathcal{H}$.

(b) $B(x, \cdot)$ is skew-linear for all $x \in \mathcal{H}$.

Definition 3.29
*A bilinear form $B : \mathcal{H} \times \mathcal{H} \to \mathbb{F}$ is called **bounded** if there exists $C < \infty$ such that $|B(x, y)| \leq C \|x\| \|y\|$ for all $x, y \in \mathcal{H}$.*

Definition 3.30
*A bilinear form $B : \mathcal{H} \times \mathcal{H} \to \mathbb{F}$ is called **coercive** if there exists $\delta > 0$ such that $B(x, x) \geq \delta \|x\|^2$ for all $x \in \mathcal{H}$.*

Theorem 3.31 (Lax-Milgram I)
Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. $B : \mathcal{H} \times \mathcal{H} \to \mathbb{F}$ is a bounded coercive bilinear form. Then for every $L \in \mathcal{H}'$, there exists $x \in \mathcal{H}$ such that $Ly = B(y, x)$ for all $y \in \mathcal{H}$.

Proof. Fixed $x \in \mathcal{H}$. Then $B(\cdot, x)$ is a bounded linear functional defined on $\mathcal{H}$. By Riesz representation, there exists a unique $x^* \in \mathcal{H}$ such that $B(y, x) = \langle y,\ x^*\rangle$ for all $y \in \mathcal{H}$. Define $T : \mathcal{H} \to \mathcal{H}$ by $Tx = x^*$. Such definition is well-defined because x^* is unique. We claim that T is bounded and linear. For linearity, let $x, y, z \in \mathcal{H}$ and $c \in \mathbb{F}$. Then

$$\langle y,\ T(cx + z)\rangle = B(y, cx + z) = \overline{c}B(y, x) + B(y, z) = \overline{c}\,\langle y,\ Tx\rangle + \langle y,\ Tz\rangle = \langle y,\ cTx + Tz\rangle\,.$$

Hence $T(cx + z) = cTx + Tz$. For boundedness, by proposition 3.5,

$$\|Tx\| = \sup_{\|y\|=1} |\langle y,\ Tx\rangle| = \sup_{\|y\|=1} |B(y, x)| \leq \sup_{\|y\|=1} C \|x\| \|y\| = C \|x\|\,.$$

Hence T is bounded.

Next, let $A = T(\mathcal{H})$. We claim that A is closed. Let $y_n \to y$ and $y_n \in A$. By the boundedness of T we have $\|Tx\| \le C\|x\|$. Also, by the coerciveness and proposition 3.5,

$$\delta\|x\|^2 \le B(x,x) \le |B(x,x)| = |\langle x, Tx\rangle| = \|x\| \left|\left\langle \frac{x}{\|x\|}, Tx \right\rangle\right| \le \|x\| \sup_{\|y\|=1} |\langle y, Tx\rangle| = \|x\| \|Tx\|.$$

So $\|x\| \le \frac{1}{\delta}\|Tx\|$. Then we see that the norms $\|\cdot\|$ and $\|T(\cdot)\|$ are equivalent. For $y_n \in A$, we can find $x_n \in \mathcal{H}$ such that $Tx_n = y_n$. Since y_n is Cauchy, x_n is also Cauchy by the equivalence of norms. By the completeness of $\mathcal{H}$, $x_n \to x \in \mathcal{H}$. Then the boundedness of T implies the continuity and $Tx_n \to Tx$. It follows that $y = Tx$ by the uniqueness of the limit. Hence $y \in A$ and A is closed.

Finally, we claim that $A = \mathcal{H}$. Assume not. Then because A is closed, $\mathcal{H} = A \oplus A^\perp$ with $A^\perp \ne \{0\}$ by proposition 3.10. There is some $z \in A$ such that $\langle z, y\rangle = 0$ for all $y \in A^\perp$. This implies $B(z,x) = \langle z, Tx\rangle = 0$ for all $x \in \mathcal{H}$. Taking $x = z$ gives $0 = B(z,z) \ge \delta\|z\|^2$ by the coerciveness of B. Hence $z = 0$, $A^\perp = \{0\}$, and $A = \mathcal{H}$, a contradiction. We conclude that $A = \mathcal{H}$.

For any bounded linear functional L, there is $x^* \in \mathcal{H}$ such that $Ly = \langle y, x^*\rangle$ for all $y \in \mathcal{H}$ by Riesz representation. Then there is $x \in \mathcal{H}$ such that $Tx = x^*$. Then for all $y \in \mathcal{H}$,

$$Ly = \langle y, x^*\rangle = \langle y, Tx\rangle = B(y,x).$$

This completes the proof. ■

Remark

Lax-Milgram theorem ensures the existence of weak solutions to linear PDEs. For example, consider the Poisson equation

$$-\Delta u = f \quad on\ \Omega \subset \mathbb{R}^d, \quad u|_{\partial\Omega} = 0$$

for $f \in \mathcal{L}^2(\Omega)$. $\Delta = \sum_{i=1}^d \partial_i^2$ *is the Laplacian. Then for all* $\varphi \in C_c^\infty$*,*

$$L(\varphi) = B(u,\varphi) = \int_\Omega \nabla u \cdot \nabla\varphi = \int_\Omega f\varphi.$$

Definition 3.32

A bilinear form $B : \mathcal{H} \times \mathcal{H} \to \mathbb{F}$ *is called* ***symmetric*** *if* $B(x,y) = B(y,x)$ *for all* $x, y \in \mathcal{H}$.

Theorem 3.33 (Lax-Milgram II)

Let $\mathbb{F} = \mathbb{R}$ *or* $\mathbb{C}$. *Suppose that* $B : \mathcal{H} \times \mathcal{H} \to \mathbb{F}$ *is a bounded symmetric coercive bilinear form and* $L \in \mathcal{H}'$ *is a bounded linear functional. Then*

$$\inf_{x\in\mathcal{H}} \frac{1}{2}B(x,x) - Lx$$

is attained at a unique $x \in \mathcal{H}$.

Proof. Set $F(x) = \frac{1}{2}B(x,x) - Lx$ and

$$\alpha = \inf_{x\in\mathcal{H}} F(x).$$

We check that α is finite. Notice that for all $x \in \mathcal{H}$,

$$-M < \frac{1}{2}\delta \|x\|^2 - \|L\| \|x\| \le \frac{1}{2}B(x,x) - |Lx| \le F(x) \le \left|\frac{1}{2}B(x,x) - Lx\right| \le \frac{1}{2}C \|x\|^2 + \|L\| \|x\| < \infty$$

for some finite $M, C > 0$ and $\delta > 0$. The first inequality is due to $\delta > 0$ and the quadratic function is bounded below. Taking infimum of $F(x)$ gives that α is finite.

Now by definition we can find $u_n \in \mathcal{H}$ such that $F(u_n) \to \alpha$. We claim that u_n is Cauchy. For all $n, m \in \mathbb{N}$, we have

$$\begin{aligned}
F(u_m) + F(u_n) &= \frac{1}{2}B(u_m, u_m) - Lu_m + \frac{1}{2}B(u_n, u_n) - Lu_n \\
&= 2\Big(B\Big(\frac{u_m}{2}, \frac{u_m}{2}\Big) + B\Big(\frac{u_n}{2}, \frac{u_n}{2}\Big)\Big) - 2L\Big(\frac{u_m + u_n}{2}\Big) \\
&= 2\left[\frac{1}{2}\Big(B\Big(\frac{u_m + u_n}{2}, \frac{u_m + u_n}{2}\Big) + B\Big(\frac{u_m - u_n}{2}, \frac{u_m - u_n}{2}\Big)\Big) - L\Big(\frac{u_m + u_n}{2}\Big)\right] \\
&= 2F\Big(\frac{u_m + u_n}{2}\Big) + B\Big(\frac{u_m - u_n}{2}, \frac{u_m - u_n}{2}\Big) \ge 2\alpha + \frac{\delta}{4} \|u_m - u_n\|^2 .
\end{aligned}$$

For arbitrary $\epsilon > 0$, there is $N \in \mathbb{N}$ such that for all $m, n \ge N$, $F(u_m), F(u_n) \le \alpha + \epsilon$. Then

$$2\alpha + \frac{\delta}{4} \|u_m - u_n\|^2 \le F(u_m) + F(u_n) \le 2\alpha + 2\epsilon \implies \|u_m - u_n\|^2 \le \frac{8\epsilon}{\delta}.$$

Since ϵ is arbitrary, we obtain that u_n is Cauchy. By the completeness, $u_n \to u$ for some $u \in \mathcal{H}$. We check that u is the minimizer of F, i.e. $F(u) = \alpha$. Observe that if for $u_n \to u$ and $v_n \to v$, we have $B(u_n, v_n) \to B(u,v)$ and $Lu_n \to Lu$. Indeed, $Lu_n \to u$ by the boundedness and hence the continuity of L. Also,

$$|B(u_n, v_n) - B(u,v)| \le |B(u_n - u, v_n)| + |B(u, v_n - v)| \le C \|u_n - u\| \|v_n\| + C \|u\| \|v_n - v\| \to 0$$

since $u_n \to u$ as a Cauchy sequence must be bounded. This implies

$$F(u_n) = \frac{1}{2}B(u_n, u_n) - Lu_n \to \frac{1}{2}B(u,u) - Lu = F(u).$$

The uniqueness of the limit of u_n ensures that the minimizer is unique. ∎

Theorem 3.34
Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Suppose that $B : \mathcal{H} \times \mathcal{H} \to \mathbb{F}$ is a bounded symmetric coercive bilinear form and $L \in \mathcal{H}'$ is a bounded linear functional. $x_0 \in \mathcal{H}$ is a minimizer of $F(x) = \frac{1}{2}B(x,x) - Lx$ if and only if $B(x_0, y) = L(y)$ for all $y \in \mathcal{H}$.

Proof. Suppose that x_0 is a solution to $B(x_0, y) = L(y)$ for all $y \in \mathcal{H}$. Then for all $x \in \mathcal{H}$,

$$\begin{aligned} F(x_0 + x) - F(x_0) &= \frac{1}{2}B(x_0 + x, x_0 + x) - L(x_0 + x) - \frac{1}{2}B(x_0, x_0) + L(x_0) \\ &= B(x_0, x) - L(x) + \frac{1}{2}B(x, x) \geq \frac{\delta}{2}\|x\|^2 \geq 0 \end{aligned}$$

by the coerciveness of B. Hence $F(x_0) \leq F(x_0 + x)$ for all $x \in \mathcal{H}$ and x_0 is a minimizer.

Conversely, suppose that x_0 minimizes F. For all $t \in \mathbb{R}$ and $y \in \mathcal{H}$, consider the function $\phi(t) = F(x_0 + ty)$. Then since x_0 minimizes F, $\phi'(t)|_{t=0} = 0$. We compute that

$$\begin{aligned} \phi(t) = F(x_0 + ty) &= \frac{1}{2}B(x_0 + ty, x_0 + ty) - L(x_0 + ty) \\ &= \frac{1}{2}B(x_0, x_0) + tB(x_0, y) + \frac{1}{2}t^2B(y, y) - L(x_0) - tL(y). \end{aligned}$$

Differentiating gives

$$0 = \phi'(t)|_{t=0} = B(x_0, y) - L(y) + tB(y, y)|_{t=0} = B(x_0, y) - L(y)$$

for each given $y \in \mathcal{H}$. Hence x_0 satisfies $B(x_0, y) = L(y)$ for all $y \in \mathcal{H}$. ■

3.4. Symmetric and Compact Operators

Definition 3.35

*$D(A) \subset \mathcal{H}$ is dense in $\mathcal{H}$. A linear operator $A : D(A) \to \mathcal{H}$ is said to be **symmetric** if $\langle Ax, y\rangle = \langle x, Ay\rangle$ for all $x, y \in D(A)$.*

Remark

Note that the domain $D(A)$ is dense in $\mathcal{H}$. It follows that by density, the domain can often be extended to $\mathcal{H}$. For simplicity, we consider the domain to be $\mathcal{H}$, but the domain can be any dense subset of $\mathcal{H}$.

Definition 3.36

*$\lambda \in \mathbb{F}$ is an **eigenvalue** of a linear operator $A : \mathcal{H} \to \mathcal{H}$ if there exists a non-zero vector $x \in \mathcal{H}$ such that $Ax = \lambda x$. The vector x is called the **eigenvector** corresponding to the eigenvalue λ.*

Proposition 3.37

Let $A : \mathcal{H} \to \mathcal{H}$ be a symmetric operator. The followings are true.

(a) $\langle Ax, x\rangle \in \mathbb{R}$ *for all* $x \in \mathcal{H}$.

(b) If $\lambda \in \mathbb{F}$ is an eigenvalue of A, then $\lambda \in \mathbb{R}$.

(c) If $\lambda_1, \lambda_2 \in \mathbb{F}$ are two distinct eigenvalues with respect to eigenvectors $x_1, x_2 \in \mathcal{H}$, then $\langle x_1, x_2\rangle = 0$.

(d) *Suppose $\{x_\alpha\}$ is an orthonormal basis of $\mathcal{H}$ with the property that each x_α is an eigenvector of A corresponding to the eigenvalue λ_α. Then if $\mu \in \mathbb{F}$ is also an eigenvalue of A, then $\mu = \lambda_\alpha$ for some α.*

Proof. For (a), $\langle Ax, x\rangle = \langle x, Ax\rangle = \overline{\langle Ax, x\rangle}$. Then $\Im(\langle Ax, x\rangle) = 0$ and $\langle Ax, x\rangle \in \mathbb{R}$.

For (b), let $x \in \mathcal{H}$ be the corresponding eigenvector to λ. Then

$$\langle Ax, x\rangle = \langle \lambda x, x\rangle = \lambda \langle x, x\rangle = \lambda \|x\|^2 \implies \lambda = \frac{\langle Ax, x\rangle}{\|x\|^2} \in \mathbb{R}.$$

For (c), by symmetry and (b), we have

$$\lambda_1 \langle x_1, x_2\rangle = \langle Ax_1, x_2\rangle = \langle x_1, Ax_2\rangle = \overline{\lambda_2} \langle x_1, x_2\rangle = \lambda_2 \langle x_1, x_2\rangle .$$

Since $\lambda_1 \neq \lambda_2$, $\langle x_1, x_2\rangle = 0$.

For (d), let $\mu \in \mathbb{F}$ be an eigenvalue of A with eigenvector $y \in \mathcal{H}$, $y \neq 0$. We claim that $\mu = \lambda_\alpha$ for some α. Suppose not. Then write $y = \sum_j c_j x_{\alpha_j}$, where $c_j \in \mathbb{F}$. We see that

$$\|y\|^2 = \lim_{M\to\infty} \left\langle y, \sum_{j=1}^{M} c_j x_{\alpha_j} \right\rangle = \lim_{M\to\infty} \overline{c_j} \left\langle y, x_{\alpha_j} \right\rangle = 0$$

by (c), but this is a contradiction since $y \neq 0$. Thus $\mu = \lambda_\alpha$ for some α. ■

Definition 3.38

*A linear operator $A : \mathcal{H} \to \mathcal{H}$ is called **bounded** if*

$$\|A\| = \sup_{\|x\|=1} \|Ax\| < \infty.$$

Proposition 3.39

$A : \mathcal{H} \to \mathcal{H}$ is a symmetric bounded linear operator. Then

$$\|A\| = \sup_{\|x\|=1} |\langle Ax, x\rangle| .$$

Proof. Assume $\|x\| = 1$. By Cauchy-Schwarz inequality, $|\langle Ax, x\rangle| \le \|Ax\| \|x\| = \|Ax\|$. Taking supremum,

$$\sup_{\|x\|=1} |\langle Ax, x\rangle| \le \sup_{\|x\|=1} \|Ax\| = \|A\| .$$

To see the reverse inequality, note that $\|Ax\|^2 = \langle Ax, Ax\rangle = \left\langle A^2 x, x\right\rangle$. For any nonzero $\lambda \in \mathbb{R}$, define $x^+ = \lambda x + \frac{1}{\lambda}Ax$ and $x^- = \lambda x - \frac{1}{\lambda}Ax$. Then $x = \frac{1}{2\lambda}(x^+ + x^-)$ and $Ax = \frac{\lambda}{2}(x^+ - x^-)$. Now,

$$\begin{aligned}
\left\langle A^2x,\, x\right\rangle &= \left\langle A\left(\frac{\lambda}{2}(x^+ - x^-)\right),\, \frac{1}{2\lambda}(x^+ + x^-)\right\rangle \\
&= \frac{1}{4}\left\langle Ax^+ - Ax^-,\, x^+ + x^-\right\rangle \\
&= \frac{1}{4}(\left\langle Ax^+,\, x^+\right\rangle + \left\langle Ax^+,\, x^-\right\rangle - \left\langle Ax^-,\, x^+\right\rangle - \langle Ax^-,\, x^-\rangle)
\end{aligned}$$

Notice that $\left\langle A^2x,\, x\right\rangle$, $\langle Ax^+,\, x^+\rangle$ and $\langle Ax^-,\, x^-\rangle$ are real numbers by proposition 3.37; hence $\Im(\langle Ax^+,\, x^-\rangle - \langle Ax^-,\, x^+\rangle) = 0$. Also, $\langle Ax^-,\, x^+\rangle = \overline{\langle x^+,\, Ax^-\rangle} = \overline{\langle Ax^+,\, x^-\rangle}$. We have $\Im(\langle Ax^+,\, x^-\rangle) = 0$ and $\langle Ax^+,\, x^-\rangle - \langle Ax^-,\, x^+\rangle = 0$. Thus, letting $C = \sup_{\|x\|=1} |\langle Ax,\, x\rangle|$,

$$\begin{aligned}
\left\langle A^2x,\, x\right\rangle &= \frac{1}{4}(\left\langle Ax^+,\, x^+\right\rangle + \langle Ax^-,\, x^-\rangle) \\
&\leq \frac{1}{4}C\left(\left\|x^+\right\|^2 + \|x^-\|^2\right) \\
&= \frac{1}{4}C(\left\langle x^+,\, x^+\right\rangle + \langle x^-,\, x^-\rangle) \\
&= \frac{1}{4}C\left(2\lambda^2\left\|x^2\right\| + \frac{2}{\lambda^2}\|Ax\|^2\right) = \frac{1}{2}C\left(\lambda^2\|x\|^2 + \frac{1}{\lambda^2}\|Ax\|^2\right).
\end{aligned}$$

Notice that for $a, b \in \mathbb{R}$, $(a-b)^2 \geq 0$ and thus $a^2 + b^2 \geq 2ab$. Hence

$$\lambda^2\|x\|^2 + \frac{1}{\lambda^2}\|Ax\|^2 \geq 2\lambda\|x\|\frac{1}{\lambda}\|Ax\| = 2\|Ax\|\|x\|.$$

We see that

$$\|Ax\|^2 = \left\langle A^2x,\, x\right\rangle \leq \frac{1}{2}C\inf_{\lambda\neq 0}\lambda^2\|x\|^2 + \frac{1}{\lambda^2}\|Ax\|^2 \leq C\|Ax\|\|x\|.$$

Clearly if $Ax = 0$ the inequality holds. Suppose $\|Ax\| \neq 0$. Then deviding both sides by $\|Ax\|$ and taking supremum gives

$$\|A\| = \sup_{\|x\|=1}\|Ax\| \leq C\sup_{\|x\|=1}\|x\| = C = \sup_{\|x\|=1}|\langle Ax,\, x\rangle|.$$

We conclude that $\|A\| = \sup_{\|x\|=1}|\langle Ax,\, x\rangle|$. ■

Definition 3.40
*X and Y are normed spaces. $M \subset X$. $A : M \to Y$ is an operator. We say that A is **compact** if A is continuous and for every bounded sequence $x_n \in M$, the sequence $Ax_n \in Y$ has a convergent subsequence.*

Remark
A compact operator A transfers bounded sets in X to relatively compact sets in Y.

Example

Consider the integral operator $A : C([0,1]) \to C([0,1])$ equipped with the supremum norms. Define

$$Au(x) = \int_0^1 K(x,y)f(y)dy,$$

where $K \in C([0,1]^2)$. We verify that A is well-defined, i.e., $Au \in C([0,1])$. Let $x_n \to x \in [0,1]$.

$$\begin{aligned} |Au(x_n) - Au(x)| &= \left| \int_0^1 K(x_n,y)u(y)dy - \int_0^1 K(x,y)u(y)dy \right| \\ &\leq \int_0^1 |K(x_n,y) - K(x,y)|\,|u(y)|\,dy \leq \|K(x_n,\cdot) - K(x,\cdot)\|_\infty \|u\|_\infty . \end{aligned}$$

Since K is continuous, $(x_n,y) \to (x,y)$ implies $K(x_n,y) \to K(x,y)$. It follows that $Au(x_n) \to Au(x)$. Hence $Au \in C([0,1])$.

We claim that A is compact. Let $\{u_n\}$ be a bounded sequence in $C([0,1])$. By Arzelà-Ascoli theorem, it suffices to show that $\{Au_n\}$ is bounded and uniformly equicontinuous. To see the boundedness, note that by assumption we have $\|u_n\|_\infty \leq M$ for all n. Also, since K is continuous on a compact set, we have $K(x,y) \leq C$ for all $x,y \in [0,1]$. Then

$$\begin{aligned} \|Au_n\|_\infty &= \sup_{x\in[0,1]} \left| \int_0^1 K(x,y)u_n(y)dy \right| \\ &\leq \sup_{x\in[0,1]} \int_0^1 |K(x,y)|\,|u_n(y)|\,dy \\ &\leq \sup_{x\in[0,1]} \int_0^1 C\,\|u_n\|_\infty\,dy = C\,\|u_n\|_\infty \leq CM. \end{aligned}$$

Thus $\{Au_n\}$ is bounded. To see the uniform equicontinuity, let $\epsilon > 0$ be given. By continuity of K, we can find $\delta > 0$ such that whenever $|x - z| < \delta$, $|K(x,y) - K(z,y)| < \epsilon/M$ for all $y \in [0,1]$. Then for each $n \in \mathbb{N}$,

$$\begin{aligned} |Au_n(x) - Au_n(z)| &= \left| \int_0^1 K(x,y)u_n(y)dy - \int_0^1 K(z,y)u_n(y)dy \right| \\ &\leq \int_0^1 |K(x,y) - K(z,y)|\,|u_n(y)|\,dy \leq \frac{\epsilon}{M}\,\|u_n\|_\infty \leq \epsilon. \end{aligned}$$

Thus $\{Au_n\}$ is uniformly equicontinuous. By Arzelà-Ascoli theorem, $\{Au_n\}$ has a convergent subsequence, i.e., A is compact.

Definition 3.41
Let $A : \mathcal{H} \to \mathcal{H}$ be a linear operator. If $\lambda \in \mathbb{F}$ is an eigenvalue of A, then the corresponding ***eigenspace*** *is defined as*

$$E_\lambda = \{x \in \mathcal{H} : Ax = \lambda x\}.$$

Remark
Clearly E_λ is a subspace of $\mathcal{H}$.

Theorem 3.42 (Spectral Theorem for Compact Symmetric Operators)
Let $\mathcal{H}$ be a separable Hilbert space. Suppose that $A : \mathcal{H} \to \mathcal{H}$ is a symmetric compact linear operator. Then the followings are true.

(a) There exists an at most countable orthobormal basis $\{x_j\}$, in which each x_j is an eigenvector of A corresponding to an eigenvalue λ_j.

(b) If $\lambda_i \neq \lambda_j$, then $\langle x_i, x_j \rangle = 0$.

(c) For any $\lambda \neq 0$, $\dim(E_\lambda) < \infty$.

(d) If $\dim(\mathcal{H}) = \infty$, then either $\lambda_j \to 0$ or there is only finitely many $\lambda_j \neq 0$.

Proof. First note that if $\mathcal{H} = \{0\}$, then the statements are vacuously true. We assume that $\mathcal{H} \neq \{0\}$. Assume first that $\dim(\mathcal{H}) = \infty$ and $\ker(A) = \{0\}$. By the assumptions we can find $x \in \mathcal{H}$ such that $\|Ax\| > 0$ and thus $\|A\| > 0$. By proposition 3.39, $\|A\| = \sup_{\|x\|=1} |\langle Ax, x\rangle|$. Hence there exists a sequence $z_n \in \mathcal{H}$ such that $|\langle Az_n, z_n\rangle| \to \|A\|$ with $\|z_n\| = 1$. Let $\lambda_1 \in \mathbb{R}$ satisfying that $\lambda_1 = \operatorname{sgn}(\langle Az_n, z_n\rangle) \|A\|$ for n greater than some N so that the sign of $\langle Az_n, z_n\rangle$ does not alternate. Now notice that

$$\begin{aligned} 0 &\leq \|\lambda_1 z_n - Az_n\|^2 \\ &= |\lambda_1|^2 \|z_n\|^2 + \|Az_n\|^2 - 2|\lambda_1| \langle Az_n, z_n\rangle \\ &\leq 2|\lambda_1|^2 - 2\lambda_1 \langle Az_n, z_n\rangle \leq 2|\lambda_1|^2 - 2|\lambda_1| |\langle Az_n, z_n\rangle| \to 0. \end{aligned}$$

Hence $\lambda_1 z_n - Az_n \to 0$. Since A is compact, $\{Az_n\}$ has a convergent subsequence, say $A(z_{n_k}) \to y$. Then $\lambda_1 z_{n_k} \to \lambda_1 x_1$ for some $x_1 \in \mathcal{H}$ with $Ax_1 = y$ and then $z_{n_k} \to x_1$. Since A is continuous, $A(z_{n_k}) \to Ax_1$ implies that

$$Ax_1 = \lim_{k\to\infty} A(z_{n_k}) = \lim_{k\to\infty} \lambda_1 z_{n_k} = \lambda_1 x_1.$$

Note that $\|z_{n_k}\| = 1$ and thus $\|x_1\| = 1$. We have shown that there exists an eigenvector x_1 corresponding to an eigenvalue λ_1, with $\|x_1\| = 1$ and $|\lambda_1| = \|A\|$.

Next, define $W_1 = \operatorname{span}(\{x_1\})$ and $W_1^\perp = \{y \in \mathcal{H} \mid \langle y, x_1\rangle = 0\}$. Consider $A_1 = A|_{W_1^\perp}$. We verify that $A_1 : W_1^\perp \to W_1^\perp$ is well-defined. For any $y \in W_1^\perp$,

$$\langle A_1 y, x_1\rangle = \langle Ay, x_1\rangle = \langle y, Ax_1\rangle = \langle y, \lambda_1 x_1\rangle = \lambda_1 \langle y, x_1\rangle = 0.$$

Hence $A_1 y \in W_1^\perp$. Observe that A_1 is also symmetric since for every $y_1, y_2 \in W_1^\perp$,

$$\langle A_1 y_1, y_2\rangle = \langle Ay_1, y_2\rangle = \langle y_1, Ay_2\rangle = \langle y_1, A_1 y_2\rangle.$$

We show that A_1 is compact. Suppose $y_n \in w_1^\perp$ and $y_n \to y \in W_1^\perp$. Then $A_1 y_n = Ay_n \to Ay = A_1 y$ by the continuity of A. Also, if $\{y_n\}$ is a bounded sequence in $W_1^\perp$, then $\{A_1 y_n\} = \{Ay_n\} \subset W_1^\perp$ has a convergent subsequence. Since W_1 is finite-dimensional, W_1 is itself closed and thus so does $W_1^\perp$ by proposition 3.10. It follows that the subsequence converges in $W_1^\perp$.

Hence A_1 is compact.

Now by similar argument as above, we can find $x_2 \in W_1^\perp$ such that $\|x_2\| = 1$, $|\lambda_2| = \|A_1\| \le \|A\| = |\lambda_1|$ and $Ax_2 = A_1 x_2 = \lambda_2 x_2$ for some $\lambda_2 \in \mathbb{R}$. Continue the process. We obtain a sequence $\{x_j\}$ such that each x_j is an eigenvector of A corresponding to an eigenvalue λ_j with $|\lambda_j| \le |\lambda_{j-1}|$ and $\|x_j\| = 1$. Furthermore, observe that $\langle x_i, x_j \rangle = 0$ for all $i \ne j$ and $|\lambda_{i+1}| = \|A_i\|$.

We verify (d) first. Notice that $|\lambda_j|$ decreases and bounded below by 0. Hence $|\lambda_j| \to \alpha$. By the compactness of A, there exists a subsequence x_{n_i} such that Ax_{n_i} converges. Thus Ax_{n_i} is Cauchy and

$$\left\|Ax_{n_i} - Ax_{n_j}\right\|^2 = \left\|\lambda_{n_i} x_{n_i} - \lambda_{n_j} x_{n_j}\right\|^2 = \left|\lambda_{n_i}\right|^2 + \left|\lambda_{n_j}\right|^2.$$

Note that the left hand side converges to 0 and the right hand side converges to $2\alpha^2$. Hence $\alpha = 0$.

Next, we show that $\dim(E_\lambda) < \infty$ for $\lambda \ne 0$. Suppose not. Then we can find a countable orthonormal basis $\{x_i\}$ of E_λ with each x_i corresponding to λ. By the compactness of A, there exists a subsequence x_{n_i} such that Ax_{n_i} converges and thus Cauchy.

$$2|\lambda|^2 = |\lambda|^2 \left\|x_{n_i}\right\|^2 + |\lambda|^2 \left\|x_{n_j}\right\|^2 = \left\|\lambda x_{n_i} - \lambda x_{n_j}\right\|^2 = \left\|Ax_{n_i} - Ax_{n_j}\right\|^2 \to 0.$$

Hence $|\lambda| = 0$, a contradiction. Thus $\dim(E_\lambda) < \infty$.

Lastly, we show that $\overline{\operatorname{span}(\{x_j\})} = \mathcal{H}$. For every $x \in \mathcal{H}$, consider the partial sum $z_n = \sum_{i=1}^n \langle x, x_i \rangle x_i$. We want to show that $z_n \to x$. $Az_n = \sum_{i=1}^n \lambda_i \langle x, x_i \rangle x_i$. Notice that

$$\langle x - z_n, x_j \rangle = \left\langle x - \sum_{i=1}^n \langle x, x_i \rangle x_i, x_j \right\rangle = \langle x, x_j \rangle - \langle x, x_j \rangle = 0.$$

Hence $x - z_n \in W_n^\perp$. Thus since $\|x - z_n\|$ is bounded,

$$\|Ax - Az_n\| = \|A(x - z_n)\| = \|A_n(x - z_n)\| \le \|A_n\| \|x - z_n\| = |\lambda_{n+1}| \|x - z_n\| \to 0$$

Hence $Az_n \to Ax$ and thus we can write $Ax = \sum_j \lambda_j \langle x, x_j \rangle x_j$. If $y = \sum_j \langle x, x_j \rangle x_j$, then $Ay = \sum_j \lambda_j \langle x, x_j \rangle x_j = Ax$. Because A has zero kernel, $x = y = \sum_j \langle x, x_j \rangle x_j$. Thus we have $\overline{\operatorname{span}(\{x_j\})} = \mathcal{H}$.

Now we drop the assumption that $\ker(A) = \{0\}$. Note that $\ker(A)$ is a closed subspace of $\mathcal{H}$ since if $x_n \in \ker(A)$ and $x_n \to x \in \mathcal{H}$, by continuity of A,

$$Ax = \lim_{n \to \infty} Ax_n = 0 \implies x \in \ker(A).$$

It follows that by proposition 3.10, $\mathcal{H} = \ker(A) \oplus \ker(A)^\perp$. For $\ker(A)$, we apply theorem 3.20 to find an orthonormal basis $\{w_k\}$ of $\ker(A)$. Note that since $\{w_k\} \subset \ker(A)$, each w_k is an eigenvector of A corresponding to the eigenvalue 0. Also, define $A^\perp : \ker(A)^\perp \to \ker(A)^\perp$ by $A^\perp y = Ay$. We verify that such definition is well-defined, i.e., $A^\perp y \in \ker(A)^\perp$. For each

$y \in \ker(A)^\perp$,

$$\left\langle A^\perp y,\, w\right\rangle = \langle Ay,\, w\rangle = \langle y,\, Aw\rangle = 0$$

for all $w \in \ker(A)$. Hence $A^\perp y \in \ker(A)^\perp$. Also, $A^\perp$ inherits the compactness and symmetry of A. By the previous argument, we can find an orthonormal basis $\{x_j\}$ of $\ker(A)^\perp$ such that each x_j is an eigenvector of $A^\perp$ and thus A, corresponding to an eigenvalue λ_j. Notice that $\left\langle w_k,\, x_j\right\rangle = 0$. Thus $\{w_k\} \cup \{x_j\}$ forms the desired orthonormal basis of $\mathcal{H}$.

Finally, if $\dim(\mathcal{H}) = \infty$, then (c) and (d) are vacuously true. (a) follows by applying the above construction with the process being terminated in finite steps. Once (a) is established, (b) follows by observing that x_i and x_j are distinct eigenvectors in the orthonormal basis and thus must be orthogonal. ∎

Definition 3.43
Let $\mathcal{H}$ be a separable Hilbert space and $A : \mathcal{H} \to \mathcal{H}$ be a symmetric compact linear operator. Then for each $x \in \mathcal{H}$, we can find $\{x_j\}$ and $\{w_k\}$ are orthonormal bases for $\ker(A)^\perp$ *and* $\ker(A)$ *respectively such that*

$$x = \sum_j \left\langle x,\, x_j\right\rangle x_j + \sum_k \langle x,\, w_k\rangle\, w_k.$$

Such decomposition is called the ***spectral decomposition*** *of A.*

Theorem 3.44 (Fredholm Alternative)
Let $\mathcal{H}$ be separable. Suppose $A : \mathcal{H} \to \mathcal{H}$ is a symmetric compact linear operator, $\lambda \neq 0$. Let $N_\lambda = \{x \in \mathcal{H} \mid Ax = \lambda x\}$. Then the equation

$$\lambda x - Ax = z$$

has a solution if and only if $z \in N_\lambda^\perp$. Furthermore, if λ is not an eigenvalue of A, then the solution is unique.

Proof. Consider the orthonormal eigenbasis $\{x_j\} \bigcup \{w_k\}$ of A with nonzero eigenvalues λ_j for x_j and zeros for w_k. Suppose first that $\lambda \neq \lambda_j$ for all j. This is equivalent to that $N_\lambda = \{0\}$ and $N_\lambda^\perp = \mathcal{H}$. For every $z \in \mathcal{H}$, by setting

$$x = \sum_j \frac{1}{\lambda - \lambda_j} \left\langle z,\, x_j\right\rangle x_j + \sum_k \frac{1}{\lambda} \langle z,\, w_k\rangle\, w_k,$$

we see that

$$\begin{aligned}\lambda x - Ax &= \lambda \sum_j \frac{1}{\lambda - \lambda_j} \left\langle z,\, x_j\right\rangle x_j + \lambda \sum_k \frac{1}{\lambda} \langle z,\, w_k\rangle\, w_k - \sum_j \frac{\lambda_j}{\lambda - \lambda_j} \left\langle z,\, x_j\right\rangle x_j \\ &= \sum_j \left\langle z,\, x_j\right\rangle x_j + \sum_k \langle z,\, w_k\rangle\, w_k = z.\end{aligned}$$

We verify that such x indeed belongs to $\mathcal{H}$.

$$\|x\|^2 = \sum_j \left| \frac{1}{\lambda - \lambda_j} \langle z, x_j \rangle \right|^2 + \sum_k \left| \frac{1}{\lambda} \langle z, w_k \rangle \right|^2 \le \sup_j \frac{1}{|\lambda - \lambda_j|^2} \|z\|^2 + \sup_k \frac{1}{|\lambda|^2} \|z\|^2 < C_\lambda \|z\|^2$$

for some C_λ by the Parseval's identity. Since $\lambda \neq 0$, C_λ is finite and thus $x \in \mathcal{H}$. To check the uniqueness, it suffices to show that the homogeneous equation $\lambda x - Ax = 0$ implies $x = 0$. Indeed, if $x \neq 0$ satisfies $\lambda x - Ax = 0$, then λ becomes an eigenvalue of A with eigenvector x, which contradicts to our assumption. Hence the solution is unique. The converse is trvial since $N_\lambda^\perp = \mathcal{H}$.

Now suppose that $\lambda = \lambda_j$ for some j, say $j = 1$. Then $N_\lambda = E_{\lambda_1}$. If $z \in E_{\lambda_1}^\perp$, then since $\dim(E_{\lambda_1}) < \infty$ by the spectral theorem, E_{λ_1} is a closed subspace of $\mathcal{H}$ and hence $E_{\lambda_1}^\perp$ by proposition 3.10. Thus for such z, we can write $z = \sum_{j:\lambda_j \neq \lambda_1} \langle z, x_j \rangle x_j + \sum_k \langle z, w_k \rangle w_k$. Set

$$x = \sum_{j:\lambda_j \neq \lambda_1} \frac{1}{\lambda - \lambda_j} \langle z, x_j \rangle x_j + \sum_k \frac{1}{\lambda} \langle z, w_k \rangle w_k.$$

Then

$$\begin{aligned} \lambda x - Ax &= \lambda \sum_{j:\lambda_j \neq \lambda_1} \frac{1}{\lambda - \lambda_j} \langle z, x_j \rangle x_j + \lambda \sum_k \frac{1}{\lambda} \langle z, w_k \rangle w_k - \sum_{j:\lambda_j \neq \lambda_1} \frac{\lambda_j}{\lambda - \lambda_j} \langle z, x_j \rangle x_j \\ &= \sum_{j:\lambda_j \neq \lambda_1} \langle z, x_j \rangle x_j + \sum_k \langle z, w_k \rangle w_k = z. \end{aligned}$$

We verify that such x indeed belongs to $\mathcal{H}$. By exactly the same argument as above,

$$\|x\|^2 \le \sup_{j:\lambda_j \neq \lambda} \frac{1}{|\lambda - \lambda_j|^2} \|z\|^2 + \sup_k \frac{1}{|\lambda|^2} \|z\|^2 < C_\lambda \|z\|^2.$$

Since by the spectral theorem we have $\lambda_j \to 0$ as $j \to \infty$, $C_\lambda < \infty$. We conclude that the equation $\lambda x - Ax = z$ has a solution if $z \in N_\lambda^\perp$. Conversely, if x is a solution, then for every x_{j_i} with $\lambda_{j_i} = \lambda_1 = \lambda$,

$$\begin{aligned} \langle z, x_{j_i} \rangle &= \langle \lambda x - Ax, x_{j_i} \rangle = \langle \lambda x, x_{j_i} \rangle - \langle Ax, x_{j_i} \rangle \\ &= \lambda \langle x, x_{j_i} \rangle - \langle x, Ax_{j_i} \rangle = 0. = \lambda \langle x, x_{j_i} \rangle - \lambda_{j_i} \langle x, x_{j_i} \rangle = 0. \end{aligned}$$

We see that $z \in N_\lambda^\perp$. This completes the proof. ■

Remark

If λ is an eigenvalue of A, we can actually find infinitely many solutions. Since every eigenvector x corresponding to λ would become a homogeneous solution, for any solution x_0 such that $\lambda x_0 - Ax_0 = z$, $x_0 + x$ forms another solution for any $x \in E_\lambda$. In other words, the set of solutions is $x_0 + E_\lambda$.

4. Approximation Theory and Fourier Theory

4.1. Approximation by Polynomials

Proposition 4.1
Let X be a finite-dimensional vector space. Then every norm on X is equivalent.

Proof. This can be seen as a special case of proposition 2.93, as any finite-dimensional vector space is a Banach space. However, we also have a simple proof here.

Let $\{e_1, \dots, e_n\}$ be a basis for X. For any $x \in X$, we can write $x = \sum_{i=1}^n x_i e_i$. For any norm $\|\cdot\|$ on X,

$$\|x\| = \left\| \sum_{i=1}^n x_i e_i \right\| \le \sum_{i=1}^n |x_i| \, \|e_i\| \le \left(\max_{1 \le i \le n} \|e_i\| \right) \sum_{i=1}^n |x_i| = C_1 \|x\|_1 .$$

where $\|\cdot\|_1$ is the ℓ^1 norm. Also, this implies that $\|\cdot\| : X \to \mathbb{R}$ is continuous with respect to the ℓ^1 norm since

$$|\|x\| - \|y\|| \le \|x - y\| \le C_1 \|x - y\|_1 .$$

Now for any $x \neq 0$, the function $f(x) = \left\| \frac{x}{\|x\|_1} \right\|$ is continuous on $S = \{x \in X \mid \|x\|_1 = 1\}$, which is compact. By extreme value theorem, f attains its minimum on S, which leads to

$$\frac{\|x\|}{\|x\|_1} = \left\| \frac{x}{\|x\|_1} \right\| \ge C_2 > 0$$

for some $C_2 > 0$ since $x \neq 0$. Thus $\|x\| \ge C_2 \|x\|_1$ and the norms are equivalent. ■

Remark
Every finite-dimensional normed vector space is complete.

Remark
Every closed ball in a finite-dimensional normed vector space is compact.

Theorem 4.2
Let X be a Banach space and Y be a finite-dimensional subspace. For any $x \in X$, there exists a $y^* \in Y$ such that

$$\|x - y^*\| = \inf_{y \in Y} \|x - y\| .$$

Proof. Since Y is a subspace, $0 \in Y \subset X$. Then $\|x - y^*\| \le \|x\|$. Consider the closed ball $B = \{y \in Y \mid \|x - y\| \le \|x\|\}$. Let $f(y) = \|x - y\|$. Observe that

$$|f(y) - f(z)| = |\|x - y\| - \|x - z\|| \le \|y - z\| ,$$

and f is continuous. Since B is compact, f attains its minimum at some point $y^* \in B \subset Y$. Thus $\|x - y^*\| = \inf_{y \in Y} \|x - y\|$. ■

Proposition 4.3
Let X be a Banach space and $Y \subset X$ be a finite-dimensional subspace. Suppose that for each $x \in X$, there corresponds a unique $y_x \in Y$ such that $\|x - y_x\| = \inf_{y \in Y} \|x - y\|$. Then the map $P : X \to Y$ defined by $P : x \mapsto y_x$ is continuous.

Proof. Let $x_n \to x$ in X. Since

$$\|P(x_n)\| = \|P(x_n) - x_n + x_n\| \leq \|P(x_n) - x_n\| + \|x_n\| \leq 2\|x_n\|,$$

$P(x_n)$ is bounded. By Bolzano-Weierstrass theorem, there is a subsequence x_{n_k} such that $P(x_{n_k}) \to P(\hat{x})$ for some $\hat{x} \in X$. It remains to show that $P(x) = P(\hat{x})$. Indeed,

$$\left\|P(x_{n_k}) - x_{n_k}\right\| \leq \left\|P(x) - x_{n_k}\right\|$$

for all k. Letting $k \to \infty$ gives $\|P(\hat{x}) - x\| \leq \|P(x) - x\|$. Since the minimizer is unique, $P(x) = P(\hat{x})$ and $x = \hat{x}$. Hence P is continuous. ∎

Theorem 4.4
Let X be a Banach space, $Y \subset X$ be a subspace, and $x \in X$. Then the set

$$Y_x = \left\{ y \in Y \;\middle|\; y = \operatorname{argmin}_{y \in Y} \|x - y\| \right\}$$

is a bounded convex set.

Proof. Since $0 \in Y$, for all $y \in Y_x$, $\|y\| \leq \|x - y\| + \|x\| \leq 2\|x\|$. Thus Y_x is bounded. To see the convexity, let $y_1, y_2 \in Y_x$ and $t \in [0,1]$. Then $ty_1 + (1-t)y_2 \in Y$ and

$$\begin{aligned} \|x - (ty_1 + (1-t)y_2)\| &= \|t(x - y_1) + (1-t)(y_2 - x)\| \\ &\leq t\|x - y_1\| + (1-t)\|x - y_2\| \\ &= \|x - y_1\| = \|x - y_2\| \end{aligned}$$

since $y_1, y_2 \in Y_x$. Thus $ty_1 + (1-t)y_2 \in Y_x$ and Y_x is convex. ∎

Definition 4.5
A Banach space $(X, \|\cdot\|)$ has a strictly convex norm if

$$\|x + y\| < \|x\| + \|y\|$$

for all $x, y \in X$ such that $\alpha x \neq \beta y$ for all $\alpha, \beta \in \mathbb{R}$.

Remark
$\mathcal{L}^p$ spaces are strictly convex for $1 < p < \infty$.

Definition 4.6
For any bounded function $f : [0,1] \to \mathbb{R}$, the Bernstein polynomial of degree n for f is defined

as

$$B_n(f)(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}.$$

Theorem 4.7 (Weierstrass)
Let $f \in C([a,b])$. Then for any $\epsilon > 0$, there exists a polynomial p such that $\|f - p\|_\infty < \epsilon$.

Proof. First consider the mapping $\sigma : x \mapsto a + (b-a)x$ for $x \in [a,b]$. Then by replacing f with $f \circ \sigma$, we can assume that $a = 0$ and $b = 1$.

Now consider the Bernstein polynomial $B_n(f)$. Since f is continuous on $[0,1]$, which is compact, f is uniformly continuous. Thus for any $\epsilon > 0$, there exists a $\delta > 0$ such that $|x-y| < \delta$ implies $|f(x) - f(y)| < \epsilon$. Let $F = \left\{k \in \{0,\dots,n\} \,\middle|\, \left|x - \frac{k}{n}\right| < \delta\right\}$. We can compute that

$$\begin{aligned}
|B_n(f)(x) - f(x)| &= \left| f(x) - \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k} \right| \\
&\le \sum_{k\in F}\left|f(x) - f\left(\frac{k}{n}\right)\right|\binom{n}{k}x^k(1-x)^{n-k} + \sum_{k\notin F}\left|f(x) - f\left(\frac{k}{n}\right)\right|\binom{n}{k}x^k(1-x)^{n-k} \\
&\le \sum_{k=0}^{n}\epsilon\binom{n}{k}x^k(1-x)^{n-k} + 2\|f\|_\infty\sum_{k\notin F}\binom{n}{k}x^k(1-x)^{n-k} \\
&\le \epsilon + 2\|f\|_\infty\sum_{k=0}^{n}\mathbb{1}\left\{\left|x - \frac{k}{n}\right| \ge \delta\right\}\binom{n}{k}x^k(1-x)^{n-k} \\
&\le \epsilon + 2\|f\|_\infty\sum_{k=0}^{n}\frac{1}{\delta^2}\left(x - \frac{k}{n}\right)^2\binom{n}{k}x^k(1-x)^{n-k} \\
&= \epsilon + 2\frac{\|f\|_\infty}{\delta^2}\sum_{k=0}^{n}\left(x^2 - \frac{2k}{n}x + \frac{k^2}{n^2}\right)\binom{n}{k}x^k(1-x)^{n-k}.
\end{aligned}$$

Now let

$$S(x,y) = \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k} = (x+y)^n.$$

Then

$$\begin{aligned}
nx(x+y)^{n-1} &= x\frac{\partial S}{\partial x} = \sum_{k=0}^{n} k\binom{n}{k}x^k y^{n-k}, \\
n(n-1)x^2(x+y)^{n-2} &= x^2\frac{\partial^2 S}{\partial x^2} = \sum_{k=0}^{n} k(k-1)\binom{n}{k}x^k y^{n-k}.
\end{aligned}$$

Taking $y = 1 - x$ gives

$$\sum_{k=0}^{n}\left(x^2 - \frac{2k}{n}x + \frac{k^2}{n^2}\right)\binom{n}{k}x^k(1-x)^{n-k} = x^2 - \frac{2}{n}x\cdot nx + \frac{1}{n^2}\left(n(n-1)x^2 + nx\right) = \frac{x(1-x)}{n} \le \frac{1}{4n}.$$

Hence we obtain the estimate

$$|B_n(f)(x) - f(x)| \le \epsilon + \frac{\|f\|_\infty}{2n\delta^2}.$$

Letting $n \to \infty$ and by the arbitrariness of ϵ, we see that $B_n(f) \to f$ uniformly. ∎

Remark

An alternative expression for the Weierstrass theorem is that for such f, there exists a sequence of polynomials p_n such that $p_n \to f$ uniformly.

Remark

As a direct consequence of the Weierstrass theorem, the polynomial space is dense in $C([0,1])$.

Definition 4.8

*A map $T : C([a,b]) \to C([a,b])$ is said to be **positive** if $T(f) \ge 0$ for all $f \ge 0$.*

Proposition 4.9

The map $U : C([0,1]) \to C([0,1])$ defined by $f \mapsto B_n(f)$ is linear, positive, and continuous.

Proof. To show the linearity, let $c \in \mathbb{R}$ and $f, g \in C([0,1])$. Then

$$\begin{aligned} U(cf+g)(x) &= \sum_{k=0}^{n} \left(cf\left(\frac{k}{n}\right) + g\left(\frac{k}{n}\right) \right) \binom{n}{k} x^k (1-x)^{n-k} \\ &= c \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} + \sum_{k=0}^{n} g\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} = cU(f)(x) + U(g)(x). \end{aligned}$$

To show the positivity, let $f \ge 0$. Then $f\left(\frac{k}{n}\right) x^k (1-x)^{n-k} \ge 0$ for all k and $x \in [0,1]$. Then the sum is nonnegative and $U(f) \ge 0$.

To show the continuity, it is enough to show the boundedness of U.

$$\begin{aligned} \|U(f)\|_\infty &= \sup_{x \in [0,1]} \left| \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} \right| \\ &\le \sup_{x \in [0,1]} \sum_{k=0}^{n} \left| f\left(\frac{k}{n}\right) \right| \binom{n}{k} x^k (1-x)^{n-k} \\ &\le \sup_{x \in [0,1]} \sum_{k=0}^{n} \|f\|_\infty \binom{n}{k} x^k (1-x)^{n-k} = \|f\|_\infty . \end{aligned}$$

Hence U is a bounded linear operator and thus continuous. ∎

Theorem 4.10 (Korovkin)

Let $T_n : C([0,1]) \to C([0,1])$ be positive linear maps. Suppose that $T_n(f_i) \to f_i$ uniformly for $i = 0, 1, 2$ with $f_i(x) = x^i$. Then $T_n(f) \to f$ uniformly for all $f \in C([0,1])$.

Proof. Let $\epsilon > 0$. Since f is continuous on a compact set, we can assume that it is Lipschitz with constant L. Now observe that

$$|f(x) - f(a)| \le L|x - a| \le L\epsilon + L\frac{(x-a)^2}{\epsilon}.$$

This can be verify as follows,

$$\begin{cases} L|x-a| \le L\epsilon \le L\epsilon + L\frac{(x-a)^2}{\epsilon} & \text{if } |x-a| \le \epsilon, \\ L|x-a| \le L\frac{(x-a)^2}{\epsilon} \le L\epsilon + L\frac{(x-a)^2}{\epsilon} & \text{if } |x-a| > \epsilon. \end{cases}$$

Next, we apply T_n and note that we have $|T_n(f)| = T_n(|f|)$. Then,

$$\begin{aligned} |T_n(f)(x) - f(a)| &\le |T_n(f)(x) - f(a)T_n(f_0)(x)| + |f(a)T_n(f_0)(x) - f(a)| \\ &= T_n(|f - f(a)|)(x) + |f(a)|\,|T_n(f_0)(x) - f_0| \\ &\le L\left(\epsilon T_n(f_0)(x) + \frac{1}{\epsilon}T_n(f_2 - 2af_1 + a^2 f_0)(x)\right) + \|f\|_\infty \|T_n(f_0) - f_0\|_\infty \\ &\le L\epsilon(T_n(f_0)(x) - f_0(x)) + L\epsilon f_0(x) + \frac{L}{\epsilon}(T_n(f_2)(x) - f_2(x)) + \frac{L}{\epsilon}f_2(x) \\ &\quad - \frac{2aL}{\epsilon}(T_n(f_1)(x) - f_1(x)) - \frac{2aL}{\epsilon}f_1(x) + \frac{a^2L}{\epsilon}(T_n(f_0)(x) - f_0(x)) \\ &\quad + \frac{a^2L}{\epsilon}f_0(x) + \|f\|_\infty \|T_n(f_0) - f_0\|_\infty \\ &\le L\epsilon \|T_n(f_0) - f_0\|_\infty + \frac{L}{\epsilon}\|T_n(f_2) - f_2\|_\infty + \frac{2aL}{\epsilon}\|T_n(f_1) - f_1\|_\infty \\ &\quad + \frac{a^2L}{\epsilon}\|T_n(f_0) - f_0\|_\infty + \|f\|_\infty \|T_n(f_0) - f_0\|_\infty + L\epsilon + \frac{L}{\epsilon}(x-a)^2. \end{aligned}$$

Now taking $a = x$ and then taking supremum over $x \in [0,1]$ gives

$$\begin{aligned} \|T_n(f) - f\|_\infty &\le L\epsilon \|T_n(f_0) - f_0\|_\infty + \frac{L}{\epsilon}\|T_n(f_2) - f_2\|_\infty \\ &\quad + \frac{2L}{\epsilon}\|T_n(f_1) - f_1\|_\infty + \frac{L}{\epsilon}\|T_n(f_0) - f_0\|_\infty + \|f\|_\infty \|T_n(f_0) - f_0\|_\infty + L\epsilon. \end{aligned}$$

By the assumptions, there is N such that $n > N$ implies that

$$\|T_n(f_i) - f_i\|_\infty < \epsilon^2, \quad i = 0, 1, 2.$$

Thus,

$$\|T_n(f) - f\|_\infty < L\epsilon^3 + 5L\epsilon + \|f\|_\infty \epsilon^2.$$

Since ϵ is arbitrary, we obtain that $T_n(f) \to f$ uniformly. ∎

Example

Let $f \in C([0,1])$ and $L_n(f)$ be the polygonal approximation of f with nodes at k/n for $k =$

$0, \ldots, n$, *i.e.*,

$$L_n(f)(x) = f\left(\frac{k}{n}\right) + n\left(f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)\right)\left(x - \frac{k}{n}\right), \quad x \in \left[\frac{k}{n}, \frac{k+1}{n}\right].$$

Now $L_n(1) = 1$, $L_n(x) = x$, *and* $\left\|L_n(x^2) - x^2\right\|_\infty \leq \max_{0 \leq k \leq n-1} \frac{(k+1)^2}{n^2} - \frac{k^2}{n^2} \leq \frac{1}{n} \to 0$. *By the Korovkin theorem, we can conclude that* $L_n(f) \to f$ *uniformly for all* $f \in C([0,1])$.

Definition 4.11

Let f *be a bounded function on* $[a, b]$. *The* ***modulus of continuity*** *of* f *is defined as*

$$\omega_f(\delta) = \sup_{\substack{|x-y| \leq \delta \\ x,y \in [a,b]}} |f(x) - f(y)|.$$

Definition 4.12

A function f *is said to be* ***Lipschitz of order*** α *if*

$$|f(x) - f(y)| \leq M |x - y|^\alpha$$

for some $M > 0$ *and all* $x, y \in [a, b]$.

Proposition 4.13

Let f *be a bounded function on* $[a, b]$. *Then*

(a) $\omega_f(\delta_1) \leq \omega_f(\delta_2)$ *for all* $\delta_1 \leq \delta_2$.

(b) *If* f' *exists and is bounded, then* $\omega_f(\delta) \leq M\delta$ *for some* M.

(c) *If* f *is Lipschitz of order* α, *then* $\omega_f(\delta) \leq M\delta^\alpha$ *for some* M *and all* $\delta > 0$.

Proof. For (a), note that we have $|x - y| \leq \delta_1 \leq \delta_2$ for all $x, y \in [a, b]$.

For (b), from the mean value theorem, we have that if $|x - y| \leq \delta$, then

$$|f(x) - f(y)| = |f'(c)(x - y)| \leq M |x - y| \leq M\delta$$

for some $c \in [a, b]$ and some $M > 0$.

For (c), we have that

$$|f(x) - f(y)| \leq M |x - y|^\alpha \leq M\delta^\alpha$$

for $|x - y| \leq \delta$. ∎

Lemma 4.14

Let f *be a bounded function on* $[a, b]$ *and* $\delta > 0$. *Then*

(a) $\omega_f(n\delta) \leq n\omega_f(\delta)$ *for all* $n \in \mathbb{N}$.

(b) $\omega_f(\lambda\delta) \leq (1 + \lambda)\omega_f(\delta)$ *for all* $\lambda > 0$.

Proof. For (a), let $x < y$ be such that $|x - y| \leq n\delta$. We can split $[x, y]$ into n intervals of length at most δ, say $[z_0, z_1], \ldots, [z_{n-1}, z_n]$. Then $|z_i - z_{i-1}| \leq \delta$ for all i and

$$|f(x) - f(y)| \leq \sum_{i=1}^{n} |f(z_i) - f(z_{i-1})| \leq n\omega_f(\delta).$$

For (b), let $n \in \mathbb{N}$ be such that $n - 1 \leq \lambda \leq n$. Then

$$\omega_f(\lambda\delta) \leq \omega_f(n\delta) \leq n\omega_f(\delta) \leq (1 + \lambda)\omega_f(\delta).$$

∎

Theorem 4.15

For any $f \in C([0, 1])$, the Bernstein polynomial $B_n(f)$ satisfies

$$\|B_n(f) - f\|_\infty \leq \frac{3}{2}\omega_f\left(\frac{1}{\sqrt{n}}\right).$$

Proof. By lemma 4.14, setting $\delta = 1/\sqrt{n}$ and $\lambda = \sqrt{n}\left|x - \frac{k}{n}\right|$, then

$$\begin{aligned}
|f(x) - B_n f(x)| &\leq \sum_{k=0}^{n} \left|f(x) - f\left(\frac{k}{n}\right)\right| \binom{n}{k} x^k (1-x)^{n-k} \leq \sum_{k=0}^{n} \omega_f\left(\left|x - \frac{k}{n}\right|\right) \binom{n}{k} x^k (1-x)^{n-k} \\
&\leq \sum_{k=0}^{n} \omega_f\left(\frac{1}{\sqrt{n}}\right)\left(1 + \sqrt{n}\left|x - \frac{k}{n}\right|\right) \binom{n}{k} x^k (1-x)^{n-k} \\
&\leq \omega_f\left(\frac{1}{\sqrt{n}}\right)\left\{1 + \sqrt{n}\sum_{k=0}^{n}\left|x - \frac{k}{n}\right| \binom{n}{k} x^k (1-x)^{n-k}\right\} \\
&\leq \omega_f\left(\frac{1}{\sqrt{n}}\right)\left\{1 + \sqrt{n}\left(\sum_{k=0}^{n}\left|x - \frac{k}{n}\right|^2 \binom{n}{k} x^k (1-x)^{n-k}\right)^{1/2}\left(\sum_{k=0}^{n}\binom{n}{k} x^k (1-x)^{n-k}\right)^{1/2}\right\} \\
&\leq \omega_f\left(\frac{1}{\sqrt{n}}\right)\left\{1 + \sqrt{n}\frac{1}{2\sqrt{n}}\right\} = \frac{3}{2}\omega_f\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

The fourth inequality follows from the Cauchy-Schwarz inequality. ∎

Theorem 4.16

Let X be a metric space and $Y \subset X$ be a compact subset. Then for any $f \in X$, there exists a $p^* \in Y$ such that $d(f, p^*) \leq d(f, q)$ for all $q \in Y$.

Proof. Let d^* be the shortest distance from f to Y, i.e., $d^* = \inf_{q \in Y} d(f, q)$. Then there exists a sequence $q_n \in Y$ such that $d(f, q_n) \to d^*$. From the compactness of Y, there exists a subsequence q_{n_k} such that $q_{n_k} \to p^* \in Y$. We claim that p^* is the desired point. Indeed, for any $\epsilon > 0$, there is an N such that $d(f, q_{n_k}) \leq d^* + \epsilon$ and $d(q_{n_k}, p^*) \leq \epsilon$ for all $k \geq N$. Then

$$d(f, p^*) \leq d(f, q_{n_k}) + d(q_{n_k}, p^*) \leq d^* + 2\epsilon.$$

Since ϵ is arbitrary, we have $d(f, p^*) \le d^*$, which completes the proof. ∎

Theorem 4.17
If X is a strictly convex Banach space, and Y is a convex compact subset of X, then for any $f \in X$, there exists a unique $p^* \in Y$ such that $\|f - p^*\| = \inf_{q \in Y} \|f - q\|$.

Proof. The exsistence of such p^* follows from theorem 4.16. Denote the shortest distance from f to Y by d^*. To show that p^* is unique, suppose that there are two such points p_1^* and p_2^*. Then by the convexity of Y, $\frac{1}{2}p_1^* + \frac{1}{2}p_2^* \in Y$ and from the strict convexity of X,

$$\left\| f - \frac{1}{2}p_1^* - \frac{1}{2}p_2^* \right\| < \frac{1}{2}\left\| f - p_1^* \right\| + \frac{1}{2}\left\| f - p_2^* \right\| = d^*.$$

This contradicts the minimality of d^* and thus p^* is unique. ∎

Definition 4.18
*A function g on $[a, b]$ satisfies the **equioscillation condition** of degree n if there are $n + 2$ points $a \le x_0 < x_1 \ldots < x_{n+1} \le b$ such that $g(x_i) = (-1)^i \|g\|$ for $i = 0, \ldots, n + 1$.*

Theorem 4.19 (Chebyshev Equioscillation theorem)
Let $f \in C[a, b]$ and $p \in P_n$ be the polynomial of degree n. Let $r = f - p$. Then r satisfies the equioscillation condition of degree n if and only if $\|f - p\|_\infty \le \|f - q\|_\infty$ for all $q \in P_n$.

Proof. First assume that r satisfies the equioscillation condition of degree n. If p is not the best approximation to f in P_n, then there is $q \in P_n$ such that $\|f - (p + q)\|_\infty < \|f - p\|_\infty$. This implies that $\|r - q\|_\infty < \|r\|_\infty$. By the equioscillation condition, $|r(x_i) - q(x_i)| < |r(x_i)|$ for all $i = 0, \ldots, n + 1$. This means that q has the same sign with r at each x_i, so q must change sign $n + 1$ times. This contradicts the fact that $q \in P_n$.

Conversely, suppose that $p \in P_n$ is the best approximation to f in P_n in uniform norm. Let $R = \|r\|_\infty$. Since r is uniform continuous on $[a, b]$, we can split $[a, b]$ into subintervals $[t_i, t_{i+1}]$ such that $|r(x) - r(y)| < R/2$ for all $x, y \in [t_i, t_{i+1}]$. Now observe that if $[t_i, t_{i+1}]$ contains a local extremum of r, then r must have same sign in $[t_i, t_{i+1}]$. Denote the intervals by I_k and rearrange them so that r has maximum in $I_1, \ldots, I_{k_1}$ and minimum in $I_{k_1+1}, \ldots, I_{k_1+k_2}$. The rest intervals are denumerated by $I_{k_1+k_2+1}, \ldots, I_{k_1+k_2+k_3}$. By construction we see that the intervals with extremum points are disjoint.

We claim that $k_1 + k_2 \ge n + 2$. Assume that $k_1 + k_2 \le n + 1$. Consider the polynomial

$$q(x) = \pm \prod_{i=1}^{k_1+k_2-1} (x - z_i),$$

where z_i are the points chosen with $\max I_i < z_i < \min I_{i+1}$ for $i = 1, \ldots, k_1 + k_2 - 1$. Notice that $q(x) \ne 0$ for all x lying in I_i for $i = 1, \ldots, k_1 + k_2$. We select the sign of q such that q has the same sign as r in I_i for $i = 1, \ldots, k_1 + k_2$. We show that $p + \lambda q$ gives a better approximation than p for some $\lambda > 0$. Let $S = \cup_{i=1}^{k_1+k_2} I_i$ and $N = \cup_{i=k_1+k_2+1}^{k_1+k_2+k_3} I_i$. Then for $x \in S$,

$$|f(x) - (p(x) + \lambda q(x))| = |r(x) - \lambda q(x)| \leq R - \lambda \min |q(x)| < R.$$

And for $x \in N$,

$$|f(x) - (p(x) + \lambda q(x))| = |r(x) - \lambda q(x)| \leq R + \lambda \max |q(x)| < \frac{R}{2} + \lambda \|q\|_\infty < R$$

by taking $\lambda = \frac{R}{2\|q\|_\infty}$. This contradicts the assumption that p is the best approximation in P_n. Hence $k_1 + k_2 \geq n + 2$. Since p lies in P_n, p has at most n interior extremum plus the two endpoints; we have $n + 2$ extremum points, yielding the equioscillation condition. ∎

Remark
The Chebyshev equioscillation theorem gives us a way to compute the best polynomial approximation to a function in uniform norm. The approach is as follows. Consider the approximation polynomial $p(x) = \sum_{k=0}^{n} a_k x^k$ and the error $h = \|f - p\|_\infty$. Our goal is to find the coefficients a_k and the error h as well. The Chebyshev equioscillation theorem gives the extremum points $x_0, \ldots, x_{n+1}$ such that $f(x_i) = (-1)^i h$. Then we have the system of equations

$$\begin{pmatrix} 1 & x_0 & \cdots & x_0^n & -1 \\ 1 & x_1 & \cdots & x_1^n & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{n+1} & \cdots & x_{n+1}^n & (-1)^n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \\ h \end{pmatrix} = \begin{pmatrix} f(x_0) \\ f(x_1) \\ \vdots \\ f(x_{n+1}) \end{pmatrix}.$$

Since x_i are unknown, we need to guess a set of x_i and solve the system of equations. The iteration continues until $\|f - p\|_\infty = h$.

4.2. Fourier Series

Definition 4.20
*The **Fourier series** of a function f is given by*

$$Sf(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(kx) + b_k \sin(kx),$$

*where the **Fourier coefficients** a_k and b_k are given by*

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) dx, \quad b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(kx) dx.$$

Or, alternatively,

$$Sf(x) = \sum_{k=-\infty}^{\infty} c_k e^{ikx},$$

*where the **Fourier coefficients** c_k are given by*

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-ikx}dx.$$

Definition 4.21
*The **Truncated Fourier series** of a function f is denoted by*

$$S_N f(x) = \frac{a_0}{2} + \sum_{k=1}^{N} a_k\cos(kx) + b_k\sin(kx) = \sum_{k=-N}^{N} c_k e^{ikx}.$$

Proposition 4.22
Let a_k and b_k be the Fourier coefficients of a function f. Then

(a) If $f \in \mathcal{L}^1$, $|a_k|, |b_k| \leq C\|f\|_1$ for some constant $C > 0$.

(b) If $f \in \mathcal{L}^\infty$, $|a_k|, |b_k| \leq C\|f\|_\infty$ for some constant $C > 0$.

Proof. To see (a), compute that

$$|a_k| \leq \frac{1}{\pi}\int_{-\pi}^{\pi} |f(x)|\,|\cos(kx)|\,dx \leq \frac{1}{\pi}\int_{-\pi}^{\pi} |f(x)|\,dx = C\|f\|_1$$

and similarly for b_k. For (b),

$$|a_k| \leq \frac{1}{\pi}\int_{-\pi}^{\pi} |f(x)|\,|\cos(kx)|\,dx \leq \frac{1}{\pi}\int_{-\pi}^{\pi} \|f\|_\infty\,dx = 2\|f\|_\infty = C\|f\|_\infty.$$

The proof for b_k is analogous. ■

Lemma 4.23 (Riemann-Lebesgue I)
Let $f \in \mathcal{L}^1[a,b]$. Then

$$\lim_{n\to\infty}\int_a^b f(x)e^{-inx}dx = 0.$$

Proof. Let $\epsilon > 0$. Since $f \in \mathcal{L}^1[a,b]$, there is a step function g such that $\|f-g\|_1 < \epsilon$. For any interval E,

$$\left|\int_a^b \chi_E(x)e^{-inx}dx\right| \leq \left|\int_E \cos(nx)dx\right| + \left|\int_E \sin(nx)dx\right| \leq \frac{2\pi}{n} \to 0$$

as $n \to \infty$. A step function is a linear combination of characteristic functions of intervals, and thus $\left|\int_a^b g(x)e^{-inx}dx\right| \to 0$ as $n \to \infty$. Therefore,

$$\begin{aligned}\left|\int_a^b f(x)e^{-inx}dx\right| &\leq \left|\int_a^b (f(x)-g(x))e^{-inx}dx\right| + \left|\int_a^b g(x)e^{-inx}dx\right| \\ &\leq \|f-g\|_1 + \left|\int_a^b g(x)e^{-inx}dx\right| \to 0\end{aligned}$$

as $n \to \infty$. ∎

Definition 4.24
The space of piecewise continuous functions on $[a,b]$ is denoted by $PC[a,b]$. The symbol $PC^n[a,b]$ denotes the space of functions having continuous derivatives up to order $n-1$, with the n-th derivative being piecewise continuous.

Proposition 4.25
Let $f \in PC^1[-\pi,\pi]$ and

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(kx) + b_k \sin(kx).$$

Then

$$f'(x) = \sum_{k=1}^{\infty} -ka_k \sin(kx) + kb_k \cos(kx).$$

Proof. Differentiation term by term gives the desired result. ∎

Remark
If $f \in PC^n[-\pi,\pi]$, then

$$|a_k|, |b_k| \le \frac{\|f\|_{PC^n}}{k^n},$$

where $\|f\|_{PC^n} = \sum_{j=0}^{n} \left\|f^{(j)}\right\|_{\infty}$.

Definition 4.26
*Let f be a function on $\mathbb{R}$. The **right-limit** and the **left-limit** of f at x are defined by*

$$f(x^+) = \lim_{h\to 0^+} f(x+h), \quad f(x^-) = \lim_{h\to 0^+} f(x-h).$$

Definition 4.27
*A **kernel** is a function $k : X \times X \to \mathbb{R}$ such that*

(a) *$k(x,y) = k(y,x)$ for all $x, y \in X$,*

(b) *For finitely many points $x_1, \ldots, x_n \in X$ and scalars $a_1, \ldots, a_n \in \mathbb{R}$,*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(x_i, x_j) \ge 0.$$

Definition 4.28
*The **Dirichlet kernel** is defined by*

$$D_N(x) = \frac{1}{2\pi} \sum_{k=-N}^{N} e^{ikx}.$$

Remark
The Dirichlet kernel can be simplified to

$$D_N(x) = \frac{\sin((N+1/2)x)}{2\pi\sin(x/2)}.$$

To see this, note that

$$2\pi D_N(x)(e^{ix}-1) = e^{i(N+1)x} - e^{-iNx} = \frac{e^{i(N+1/2)x} - e^{-i(N+1/2)x}}{e^{ix/2} - e^{-ix/2}}.$$

And thus,

$$D_N(x) = \frac{\sin((N+1/2)x)}{2\pi\sin(x/2)}.$$

Some other properties of the Dirichlet kernel include $D_N(-x) = D_N(x)$ *and* $\int_{-\pi}^{\pi} D_N(x)dx = 1$.

Definition 4.29
Let $f, g : X \to \mathbb{R}$*. The **convolution** of* f *and* g *is defined by*

$$(f * g)(x) = \int_X f(x-y)g(y)dy.$$

Proposition 4.30
For any 2π*-periodic function* $f \in PC$,

$$S_N f = D_N * f.$$

Proof. Compute that

$$\begin{aligned} S_N f(x) &= \sum_{k=-N}^{N} c_k e^{ikx} = \sum_{k=-N}^{N} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)e^{-iky}dye^{ikx} \\ &= \int_{-\pi}^{\pi} f(y)\frac{1}{2\pi}\sum_{k=-N}^{N} e^{ik(x-y)}dy = \int_{-\pi}^{\pi} f(y)D_N(x-y)dy = (D_N * f)(x). \end{aligned}$$

■

Theorem 4.31 (Dirichlet-Jordan)
Let f *be a* 2π*-periodic function and piecewise Lipschitz. Then*

$$\lim_{N\to\infty} S_N f(x) = \frac{f(x^+) + f(x^-)}{2}.$$

In particular, if f *is continuous at* x*, then*

$$\lim_{N\to\infty} S_N f(x) = f(x).$$

Proof. Since f is 2π-periodic,

$$\begin{aligned} S_N f(x) &= \int_{-\pi}^{\pi} D_N(x-y) f(y) dy = \int_{-\pi}^{\pi} D_N(y) f(x-y) dy \\ &= \int_0^{\pi} D_N(y) f(x-y) dy + \int_{-\pi}^{0} D_N(y) f(x-y) dy \\ &= \int_0^{\pi} D_N(y) f(x-y) dy + \int_0^{\pi} D_N(-y) f(x+y) dy \\ &= \int_0^{\pi} D_N(y)(f(x-y)+f(x+y)) dy. \end{aligned}$$

Notice that

$$\frac{1}{2}\big(f(x^+)+f(x^-)\big) = \int_0^{\pi} D_N(y)\big(f(x^+)+f(x^-)\big) dy.$$

Thus for given x, we have

$$\begin{aligned} \left| S_N f(x) - \frac{f(x^+)+f(x^-)}{2} \right| &\le \left| \int_0^{\pi} D_N(y)\big(f(x+y)-f(x^+)\big) dy \right| \\ &\quad + \left| \int_0^{\pi} D_N(y)(f(x-y)-f(x^-)) dy \right|. \end{aligned}$$

We claim that

$$\int_0^{\pi} D_N(y) \left| f(x+y) - f(x^+) \right| dy \to 0$$

as $N \to \infty$ and the other integral is similar. There is a $\delta > 0$ such that f is continuous on $[x, x+\delta]$. Thus f is uniformly continuous on $[x, x+\delta]$, and $|f(x+y) - f(x^+)| < Cy$ for some constant $C > 0$ and $y \in [0, \delta]$. Then

$$\int_0^{\delta} |D_N(y)| \left| f(x+y) - f(x^+) \right| dy \le C \int_0^{\delta} y\, |D_N(y)|\, dy \le C \int_0^{\delta} dy = C\delta,$$

because $|D_N(t)| \le 1/|t|$. On the other hand,

$$\begin{aligned} \int_{\delta}^{\pi} |D_N(y)| \left| f(x+y) - f(x^+) \right| dy &\le \frac{1}{2\pi} \int_{\delta}^{\pi} \frac{\sin((N+1/2)y)}{\sin(y/2)} \left| f(x+y) - f(x^+) \right| dy \\ &\le \frac{1}{2\pi \sin(\delta/2)} \int_{\delta}^{\pi} \sin((N+1/2)y) g(y) dy \to 0, \end{aligned}$$

as $N \to \infty$ by the Riemann-Lebesgue lemma, where $g(y) = |f(x+y) - f(x^+)|$ is a continuous function on $[\delta, \pi]$. Hence we have

$$\left| \int_0^{\pi} D_N(y)\big(f(x+y) - f(x^+)\big) dy \right| \to 0 \quad \text{as } N \to \infty.$$

We now see that

$$\left| S_N f(x) - \frac{f(x^+)+f(x^-)}{2} \right| \to 0 \quad \text{as } N \to \infty.$$

The pointwise convergence is achieved whenever f is continuous since $f(x^+) = f(x^-) = f(x)$. ■

Definition 4.32

The series $\sigma_N f$ is defined by

$$\sigma_N f(x) = \frac{1}{N+1}\sum_{n=0}^{N} S_n f(x).$$

Remark

The series $\sigma_N f$ is the Cesaro's mean of the Fourier series of f.

Definition 4.33

*The **Fejer kernel** is defined by*

$$F_N(x) = \frac{1}{N+1}\sum_{k=0}^{N} D_k(x).$$

Remark

The Fejer kernel can be simplified to

$$F_N(x) = \frac{\sin^2\left(\frac{N+1}{2}x\right)}{2\pi(N+1)\sin^2(x/2)}.$$

To see this, note that

$$\begin{aligned} F_N(t) &= \frac{1}{N+1}\sum_{k=0}^{N} D_k(t) = \frac{1}{2\pi(N+1)}\sum_{k=0}^{N}\frac{\sin((k+1/2)t)}{\sin(t/2)} \\ &= \frac{1}{2\pi(N+1)\sin^2(t/2)}\sum_{k=0}^{N}\sin((k+1/2)t)\sin(t/2) \\ &= \frac{1}{2\pi(N+1)\sin^2(t/2)}\sum_{k=0}^{N}\cos(kt)-\cos((k+1)t) \\ &= \frac{1}{4\pi(N+1)\sin^2(t/2)}(1-\cos((N+1)t)) = \frac{\sin^2\left(\frac{N+1}{2}t\right)}{2\pi(N+1)\sin^2(t/2)}. \end{aligned}$$

Some other properties of the Fejer kernel include that if $f = 1$, then

$$\sigma_N f = \int_{-\pi}^{\pi} F_N(x)dx = 1,$$

and that $F_N(-x) = F_N(x), F_N \geq 0$. Also,

$$\sigma_N f(x) = \frac{1}{N+1}\sum_{k=0}^{N}\int_{-\pi}^{\pi} D_k(x-y)f(y)dy = \int_{-\pi}^{\pi}\left(\frac{1}{N+1}\sum_{k=0}^{N} D_k(x-y)\right)f(y)dy = F_N * f(x).$$

Definition 4.34
$C_{2\pi}$ denote the space of 2π-periodic continuous functions. $C_{2\pi}^k$ denotes the space of 2π-periodic functions having continuous derivatives up to order k.

Theorem 4.35 (Fejer)
Let $f \in C_{2\pi}$. Then $\sigma_N f \to f$ uniformly.

Proof. Let $\epsilon > 0$. There is a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$. Observe that if $|t| \geq \delta$, then

$$F_N(t) \leq \frac{1}{2\pi(N+1)\sin^2(\delta/2)} \to 0$$

as $N \to \infty$. Hence

$$\begin{aligned}
|\sigma_N f(x) - f(x)| &= \left| \int_{-\pi}^{\pi} F_N(x-y) f(y) dy - f(x) \right| \leq \int_{-\pi}^{\pi} F_N(x-y) \left| f(y) - f(x) \right| dy \\
&\leq \int_{|x-y| \leq \delta} F_N(x-y) \left| f(y) - f(x) \right| dy + \int_{\delta \leq |x-y| \leq \pi} F_N(x-y) \left| f(y) - f(x) \right| dy \\
&\leq \epsilon \int_{-\pi}^{\pi} F_N(x-y) dy + 2 \left\| f \right\|_\infty \int_{\delta \leq |x-y| \leq \pi} F_N(x-y) dy \\
&= \epsilon + 2 \left\| f \right\|_\infty \frac{\pi - \delta}{2\pi(N+1)\sin^2(\delta/2)} \to 0
\end{aligned}$$

as $N \to \infty$. ■

Definition 4.36
The ***trigonometric polynomial*** *of degree N is a function of the form*

$$TP_N(x) = \sum_{k=0}^{N} a_k \cos(kx) + b_k \sin(kx).$$

The ***trigonometric polynomial space*** *is denoted by $TP = \cup_N TP_N$.*

Theorem 4.37
Under $\mathcal{L}^2[-\pi, \pi]$, $PC_{2\pi} \subset \overline{TP}$.

Proof. Since continuous functions are dense in $PC_{2\pi}$, it suffices to show that continuous functions can be approximated by trigonometric polynomials. Let $f \in C_{2\pi}$. By the Fejer theorem, $\sigma_N f \to f$ uniformly. Since $\sigma_N f$ is a trigonometric polynomial, f can be approximated by trigonometric polynomials. ■

Definition 4.38
The ***best approximation error*** *of a function f by a trigonometric polynomial is defined by $\tilde{E}_N(f) = \inf_{p \in TP_N} \|p - f\|_\infty$.*

Definition 4.39
For $f, g > 0$, $f \lesssim g$ if there is some constant $c > 0$ such that $f \leq cg$.

Theorem 4.40
For $f \in C_{2\pi}$,

$$\|S_N f - f\|_\infty \lesssim (1 + \log N)\tilde{E}_N(f).$$

Proof. Recall that $S_N f = D_N * f$. Then

$$|S_N f(x)| = \left|\int_{-\pi}^{\pi} D_N(x-y) f(y) dy\right| \le \int_{-\pi}^{\pi} |D_N(x-y)|\,|f(y)|\, dy \le \|f\|_\infty \int_{-\pi}^{\pi} |D_N(t)|\, dt.$$

Observe that

$$|D_N(t)| = \frac{1}{2\pi}\left|\frac{\sin((N+1/2)t)}{\sin(t/2)}\right| \le \min\left\{\frac{2N+1}{2\pi}, \frac{1}{2\,|t|}\right\}.$$

Thus

$$\begin{aligned}\int_{-\pi}^{\pi} |D_N(t)|\, dt &\le \int_{|t|\le \pi/(2N+1)} \frac{2N+1}{2\pi} dt + \int_{\pi/(2N+1)\le|t|\le\pi} \frac{1}{2\,|t|} dt \\ &\le \frac{2N+1}{2\pi}\frac{2\pi}{2N+1} + 2\cdot\frac{1}{2}\log(2N+1) \lesssim (1+\log N).\end{aligned}$$

Now let q^* be the best approximation to f in TP_N. Notice that $S_N q^* = q^*$ and S_N is a linear operator. Then

$$\|S_N f - q^*\|_\infty = \|S_N f - S_N q^*\|_\infty \le \|S_N\|\,\|f - q^*\|_\infty \lesssim (1+\log N)\tilde{E}_N(f)$$

as desired. ∎

Theorem 4.41
If $f \in C_{2\pi}$ is L-Lipschitz, then

(a) $\|\sigma_N f - f\|_\infty \lesssim \frac{1+\log N}{N} L$,

(b) $\|S_N f - f\|_\infty \lesssim \frac{(1+\log N)^2}{N} L$.

Proof. For (a) we have

$$\sigma_N f(x) = (F_N * f)(x) = \int_{-\pi}^{\pi} F_N(x-y) f(y) dy.$$

And thus

$$\begin{aligned}|\sigma_N f(x) - f(x)| &\le \int_{-\pi}^{\pi} F_N(x-y)\,|f(y) - f(x)|\, dy \\ &= \int_{-\pi}^{\pi} F_N(u)\,|f(x-u) - f(x)|\, du \le L\int_{-\pi}^{\pi} F_N(u)\,|u|\, du.\end{aligned}$$

Observe that

$$|F_N(u)| = \frac{1}{2\pi(N+1)}\left|\frac{\sin^2\left(\frac{N+1}{2}u\right)}{\sin^2(u/2)}\right| \le \min\left\{\frac{N+1}{2\pi}, \frac{\pi}{2(N+1)\,|u|^2}\right\}.$$

Then

$$
\begin{aligned}
L\int_{-\pi}^{\pi} F_N(u)\,|u|\,du &\le L\int_{|u|\le\pi/(N+1)} \frac{N+1}{2\pi}\,|u|\,du + L\int_{\pi/(N+1)\le|u|\le\pi} \frac{\pi}{2(N+1)\,|u|^2}\,|u|\,du \\
&\le L\int_{|u|\le\pi/(N+1)} \frac{N+1}{2\pi}\frac{\pi}{N+1}du + \frac{L\pi}{2(N+1)}\cdot 2\log(N+1) \\
&= \frac{L\pi}{N+1} + L\pi\frac{\log(N+1)}{N+1} \lesssim \frac{1+\log N}{N}L,
\end{aligned}
$$

proving (a).

For (b), from theorem 4.40 we have

$$
\|S_N f - f\|_\infty \lesssim (1+\log N)\tilde{E}_N(f) \lesssim \|\sigma_N f - f\|_\infty \lesssim \frac{(1+\log N)^2}{N}L,
$$

proving (b). ■

Definition 4.42
*The **Chebyshev polynomials** are defined by* $T_n(x) = \cos(n\cos^{-1}(x))$ *for* $n = 0, 1, \ldots$ *on* $[-1, 1]$.

Remark
The Chebyshev polynomials have the following recurrence property,

$$
\begin{aligned}
T_{n+1}(x) &= \cos((n+1)\cos^{-1}(x)) = \cos(n\cos^{-1}(x))\cos(\cos^{-1}(x)) - \sin(n\cos^{-1}(x))\sin(\cos^{-1}(x)) \\
&= xT_n(x) - \frac{1}{2}\cos((n-1)\cos^{-1}(x)) + \frac{1}{2}\cos((n+1)\cos^{-1}(x)) \\
&= xT_n(x) - \frac{1}{2}T_{n-1}(x) + \frac{1}{2}T_{n+1}(x).
\end{aligned}
$$

Then

$$
T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x).
$$

Immediately we see that $T_n(x) \in P_n[-1, 1]$.

Proposition 4.43
$\{T_n\}_{n=0}^{\infty}$ *forms an orthogonal set with respect to the inner product*

$$
\langle f, g\rangle_T = \frac{1}{\pi}\int_{-1}^{1} f(x)g(x)\frac{dx}{\sqrt{1-x^2}}.
$$

Proof. Using the change of variable $x = \cos(\theta)$, a direct computation gives

$$
\begin{aligned}
\langle T_m, T_n\rangle_T &= \frac{1}{\pi}\int_{-1}^{1}\cos(m\cos^{-1}(x))\cos(n\cos^{-1}(x))\frac{dx}{\sqrt{1-x^2}} = \frac{1}{\pi}\int_{0}^{\pi}\cos(m\theta)\cos(n\theta)d\theta \\
&= \frac{1}{2\pi}\int_{0}^{\pi}\cos((m+n)\theta) + \cos((m-n)\theta)d\theta = 0
\end{aligned}
$$

for $m \ne n$. ■

Proposition 4.44
Let $E[-\pi, \pi]$ be the subspace of $C[-1,1]$ consisting of all even continuous function on $[-\pi, \pi]$. Consider the mapping $\Phi : C[-1,1] \to E[-\pi, \pi]$ defined by $\Phi : f \to f \circ \cos$. Then the followings are true.

(a) *Φ is well-defined and is an isomorphism.*

(b) $(\Phi T_n)(\theta) = \cos(n\theta)$.

(c) $\Phi(P_n) = E[-\pi, \pi] \cap TP_n$.

(d) *Φ is isometric.*

(e) *$E_n(f) = \tilde{E}_n(\Phi f)$ for all $f \in C[-1,1]$.*

(f) *$\langle f, g\rangle_T = \langle \Phi f, \Phi g\rangle$ for all $f, g \in C[-1,1]$.*

Proof. For (a), since $(\Phi f)(-x) = f(\cos(-x)) = f(\cos(x)) = (\Phi f)(x)$ and both f and cos are continuous, $f \circ \cos$ is also continuous, we conclude that $\Phi f \in E[-\pi, \pi]$ and Φ is well-defined. Now if $\Phi f = 0$, then $\|f\|_\infty = \|\Phi f\|_\infty = 0$ and thus $f = 0$. Hence Φ is injective. For the sujectivity, let $g \in E[-\pi, \pi]$. $\Phi(g \circ \cos^{-1}) = g$ and $g \circ \cos^{-1}$ is continuous. Thus Φ is surjective.

(b) is immediate from the definition of Φ. $(\Phi T_n)(\theta) = \cos(n \cos^{-1}(\cos(\theta))) = \cos(n\theta)$.

We now prove (c). From (a) we have Φ is an isomorphism. Also, from (b) we have that $\Phi(P_n) \subset TP_n$. For any even trigonometric polynomial $p \in TP_n$, $p(x) = \sum_{k=0}^n a_k \cos(kx)$. Then consider $g = \sum_{k=0}^n a_k T_k$. Then

$$(\Phi g)(x) = \sum_{k=0}^{n} a_k (T_k \circ \cos)(x) = \sum_{k=0}^{n} a_k \cos(k \cos^{-1}(\cos(x))) = \sum_{k=0}^{n} a_k \cos(kx) = p(x).$$

Hence $TP_n \subset \Phi(P_n)$ and (c) is proven.

For (d),

$$\|\Phi f\|_\infty = \sup_{x \in [-\pi, \pi]} |f(\cos(x))| = \sup_{x \in [-1,1]} |f(x)| = \|f\|_\infty .$$

(e) is an immediate consequence of (d).

Finally, for (f), by changing the variable $x = \cos\theta$ and the fact that $f(\cos(\theta))g(\cos(\theta))$ is even, we have

$$\langle f, g\rangle_T = \frac{1}{\pi} \int_{-1}^{1} f(x) g(x) \frac{dx}{\sqrt{1-x^2}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\cos(\theta)) g(\cos(\theta)) d\theta = \langle \Phi f, \Phi g\rangle ,$$

showing (f). ∎

Theorem 4.45
If $f \in C[-1,1]$ admits a Chebyshev series

$$f(x) = \sum_{n=0}^{\infty} a_n T_n(x),$$

then

$$\tau_N f = \sum_{k=0}^{N}(1 - \frac{k}{N+1})a_k T_k \to f$$

uniformly on $[-1,1]$. *If* f *is Lipschitz and* $P_N^C f$ *is the truncated Chebyshev series to* f, *then*

$$\left\|P_N^C f - f\right\|_\infty \lesssim (1 + \log N)E_N(f).$$

Proof. Consider the transformation $\Phi : C[-1,1] \to E[-\pi,\pi]$ defined by $\Phi : f \to f \circ \cos$. Then from proposition 4.44 we have

$$(\Phi\tau_N f)(\theta) = \sum_{k=0}^{N}\left(1 - \frac{k}{N+1}\right)a_k \cos(k\theta) = \frac{1}{N+1}\sum_{j=0}^{N}\sum_{k=0}^{j} a_j \cos(j\theta) = \sigma_N(\Phi f)(\theta).$$

By the Fejer theorem,

$$\|\tau_N f - f\|_\infty = \|\Phi\tau_N f - \Phi f\|_\infty = \|\sigma_N(\Phi f) - \Phi f\|_\infty \to 0$$

as $N \to \infty$.

To see the second part, suppose that f is L-Lipschitz. Then

$$|\Phi f(\alpha) - \Phi f(\beta)| \le L\,|\cos(\alpha) - \cos(\beta)| \le L\,|\alpha - \beta|$$

since the derivative of cos is bounded by 1. Thus Φf is L-Lipschitz and

$$\left\|P_N^C f - f\right\|_\infty = \left\|\Phi P_N^C f - \Phi f\right\|_\infty = \|S_N(\Phi f) - \Phi f\|_\infty \lesssim (1 + \log N)\tilde{E}_N(\Phi f) = (1 + \log N)E_N(f)$$

by theorem 4.40 and part (e) of proposition 4.44. ■

Theorem 4.46 (Jackson)
For $f \in C[-1,1]$,

$$E_N(f) \lesssim \omega(f, \frac{1}{N}).$$

Proof. Let $\phi(\theta) = \sum_{k=0}^{N} c_k \cos(k\theta) \in TP_N$, where $c_k \in \mathbb{R}$ such that $\phi \ge 0$. For any 2π-periodic f, define Ψ by

$$\Psi f(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta - t)\phi(t)dt = \frac{1}{2\pi}(f * \phi)(\theta).$$

Next we make the following observations. First, $\Psi\mathbf{1} = \mathbf{1}$ where $\mathbf{1}$ is the constant function $\mathbf{1}(\theta) = 1$. Second, Ψ is linear and positive. To see this, note that

$$\Psi(cf + g) = \frac{1}{2\pi}\int_{-\pi}^{\pi}(cf + g)(\theta - t)\phi(t)dt = c\Psi f + \Psi g$$

for any $c \in \mathbb{R}$ and $f, g \in C[-1,1]$. Also, if $f \geq 0$, then

$$\Psi f(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta - t)\phi(t)dt \geq 0.$$

Third, $\Psi f \in TP_N$ for any $f \in C[-1,1]$. Indeed,

$$\begin{aligned}
\Psi f(\theta) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\phi(\theta - t)dt = \frac{1}{2\pi} \sum_{k=0}^{N} \int_{-\pi}^{\pi} f(t) c_k \cos(k(\theta - t))dt \\
&= \frac{1}{2\pi} \sum_{k=0}^{N} B_k \cos(k\theta) + D_k \sin(k\theta) \in TP_N
\end{aligned}$$

with

$$B_k = \frac{c_k}{2\pi} \int_{-\pi}^{\pi} f(t)\cos(kt)dt, \quad \text{and} \quad D_k = \frac{c_k}{2\pi} \int_{-\pi}^{\pi} f(t)\sin(kt)dt.$$

Now we have the last claim that

$$\|\Psi f - f\|_\infty \leq \omega(f, \frac{1}{N})\left(1 + \frac{N\pi}{2}\sqrt{2 - C}\right)$$

for some constant $C > 0$, which will be determined later. Since f is uniformly continuous,

$$|f(\theta - t) - f(\theta)| \leq \omega(f, |t|) \leq (1 + N|t|)\omega\left(f, \frac{1}{N}\right).$$

Using the first observation $\Psi\mathbf{1} = \mathbf{1}$, we have

$$|\Psi f(\theta) - f(\theta)| = \left|\frac{1}{2\pi} \int_{-\pi}^{\pi} (f(\theta - t) - f(\theta))\phi(t)dt\right| \leq \frac{1}{2\pi}\omega\left(f, \frac{1}{N}\right) \int_{-\pi}^{\pi} (1 + N|t|)\phi(t)dt.$$

Also,

$$\begin{aligned}
\frac{1}{2\pi} \int_{-\pi}^{\pi} (1 + N|t|)\phi(t)dt &= 1 + \frac{N}{2\pi} \int_{-\pi}^{\pi} |t|\, \phi(t)dt \\
&\leq 1 + N\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |t|^2\, \phi(t)dt\right)^{1/2} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(t)dt\right)^{1/2} \\
&= 1 + N\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |t|^2\, \phi(t)dt\right)^{1/2}
\end{aligned}$$

by the Cauchy-Schwarz inequality. Notice that

$$1 - \cos t = 2\sin^2 \frac{t}{2} \geq 2\frac{4}{\pi^2}\frac{t^2}{4} = \frac{2t^2}{\pi^2} \Rightarrow |t|^2 \leq \frac{\pi^2}{2}(1 - \cos t).$$

Then

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}(1+N|t|)\phi(t)dt \le 1+N\left(\frac{1}{2\pi}\frac{\pi^2}{2}\int_{-\pi}^{\pi}(1-\cos t)\phi(t)dt\right)^{1/2} = 1+\frac{N\pi}{2}\sqrt{2-C}.$$

Thus

$$\|\Psi f - f\|_\infty \le \omega(f, \frac{1}{N})\left(1+\frac{N\pi}{2}\sqrt{2-C}\right).$$

Finally, we want to pin down our constant C so that $\sqrt{2-C}$ is minimized and $\Psi\mathbf{1} = \mathbf{1}$. We conjecture that $\phi(\theta) = C_1|p(\theta)|^2$, where $p(\theta) = \sum_{k=0}^N a_k e^{ik\theta}$, $a_k = \sin\left(\frac{k+1}{N+2}\pi\right)$. Compute that

$$\begin{aligned}
C_1|p(\theta)|^2 &= C_1 p(\theta)\overline{p(\theta)} = C_1\sum_{k=0}^N a_k e^{ik\theta}\sum_{j=0}^N a_j e^{-ij\theta}\\
&= C_1\sum_{k=0}^N\sum_{j=0}^N a_k a_j e^{i(k-j)\theta} = C_1\sum_{k=0}^N a_k^2 + C_1\sum_{s=1}^N\sum_{k=0}^{N-s} a_k a_{k+s}\left(e^{is\theta}+e^{-is\theta}\right)\\
&= C_1\sum_{k=0}^N a_k^2 + 2C_1\sum_{s=1}^N\sum_{k=0}^{N-s} a_k a_{k+s}\cos(s\theta).
\end{aligned}$$

Take $C = \left(\sum_{k=0}^N a_k^2\right)^{-1}$, then

$$\phi(\theta) = 1+\sum_{s=1}^N 2C_1 b_s\cos(s\theta), \quad \text{where } b_s = \sum_{k=0}^{N-s} a_k a_{k+s}.$$

Now

$$\begin{aligned}
2b_1 &= \sum_{k=0}^{N-1} 2\sin\left(\frac{k+1}{N+2}\pi\right)\sin\left(\frac{k+2}{N+2}\pi\right) = \sum_{k=1}^{N} 2\sin\left(\frac{k}{N+2}\pi\right)\sin\left(\frac{k+1}{N+2}\pi\right)\\
&= \sum_{k=0}^{N-1} 2\sin\left(\frac{k}{N+2}\pi\right)\sin\left(\frac{k+1}{N+2}\pi\right).
\end{aligned}$$

Combining the first expression and the third one allows us to write

$$\begin{aligned}
2b_1 &= \sum_{k=0}^{N-1}\sin\left(\frac{k+1}{N+2}\pi\right)\left(\sin\left(\frac{k}{N+2}\pi\right)+\sin\left(\frac{k+2}{N+2}\pi\right)\right)\\
&= \sum_{k=0}^{N-1}\sin^2\left(\frac{k+1}{N+2}\pi\right)\cos\frac{2\pi}{N+2} = \cos\left(\frac{2\pi}{N+2}\right)\sum_{k=0}^{N-1} a_k^2 = C_1^{-1}\cos\left(\frac{2\pi}{N+2}\right).
\end{aligned}$$

Now

$$C = 2C_1 b_1 = 2\cos(\frac{2\pi}{N+2}) \Rightarrow 2-C \lesssim \frac{1}{N^2}$$

by the Taylor expansion of cos. It now follows from the last claim that

$$\|\Psi f - f\|_\infty \leq \omega(f, \frac{1}{N})\left(1 + \frac{N\pi}{2}\sqrt{2 - C}\right) \lesssim \omega(f, \frac{1}{N}).$$

The proof is complete. ■

4.3. Fourier Transform

Definition 4.47
*For $f \in \mathcal{L}^1(\mathbb{R})$, its **Fourier transform** is defined as*

$$\hat{f}(t) = \mathcal{F} f = \int_{\mathbb{R}} f(x) e^{-2\pi i t x} dx. \tag{1}$$

Remark
The Fourier series coefficients can be viewed as discrete Fourier transform $f \mapsto \{a_n\}_{n\in\mathbb{Z}}$, with

$$a_n = \int_{-1}^{1} f(x) e^{-2\pi i n x} dx. \tag{2}$$

The inverse discrete Fourier transform is then given by

$$f(x) = \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n x}. \tag{3}$$

Example

$$\hat{\chi}_{[a,b]}(t) = \int_a^b e^{-2\pi i t x} dx = \begin{cases} b - a & \text{if } t = 0, \\ \frac{-1}{2\pi i t}\left(e^{-2\pi i t b} - e^{-2\pi i t a}\right) & \text{if } t \neq 0. \end{cases}$$

Lemma 4.48 (Riemann-Lebesgue II)
Let $f \in \mathcal{L}^1(\mathbb{R})$. Then $\hat{f}$ is uniformly continuous on $\mathbb{R}$, satisfying $\left\|\hat{f}\right\|_\infty \leq \|f\|_1$, and

$$\lim_{|t|\to\infty} \hat{f}(t) = 0.$$

Proof. We first prove the uniform continuity of $\hat{f}$. Let $t_n \to t$. Then since $\left|e^{-2\pi i t_n x} f(x)\right| \leq |f(x)|$, we may apply the Lebesgue dominated convergence theorem to obtain

$$\lim_{n\to\infty} \hat{f}(t_n) = \lim_{n\to\infty} \int_{\mathbb{R}} f(x) e^{-2\pi i t_n x} dx = \int_{\mathbb{R}} f(x) e^{-2\pi i t x} dx = \hat{f}(t).$$

Hence $\hat{f}$ is uniformly continuous.

To see the second property, we have

$$\left|\hat{f}(t)\right| = \left|\int_{\mathbb{R}} f(x) e^{-2\pi i t x} dx\right| \leq \int_{\mathbb{R}} |f(x)| \left|e^{-2\pi i t x}\right| dx = \int_{\mathbb{R}} |f(x)|\, dx = \|f\|_1$$

for any $t \in \mathbb{R}$ and thus $\|\hat{f}\|_\infty \le \|f\|_1$.

Finally, if $f = \chi_E$ where $E = [a, b]$ is an interval, then

$$\hat{f}(t) = \begin{cases} b - a & \text{if } t = 0, \\ \frac{-1}{2\pi i t}\left(e^{-2\pi i t b} - e^{-2\pi i t a}\right) & \text{if } t \neq 0. \end{cases}$$

Clearly $\hat{f}(t) \to 0$ as $|t| \to \infty$. Since step functions are finite linear combinations of such characteristic functions, the result holds for stpe functions. For any integrable function, we can find a sequence of step functions f_n such that $\|f_n - f\|_1 \to 0$. Then

$$\left|\hat{f}(t) - \hat{f}_n(t)\right| = \left|\int_{\mathbb{R}} (f(x) - f_n(x))e^{-2\pi i t x} dx\right| \le \int_{\mathbb{R}} |f(x) - f_n(x)|\, dx = \|f - f_n\|_1 \to 0$$

as $n \to \infty$. Since $\hat{f}_n(t)$ is uniformly continuous and $\hat{f}_n(t) \to 0$ as $|t| \to \infty$, we have $\hat{f}(t) \to 0$ as well. ■

Proposition 4.49

Let $\hat{f}$ be the Fourier transform of f.

(a) If $f \in \mathcal{L}^1(\mathbb{R})$ and $g(x) = xf(x) \in \mathcal{L}^1(\mathbb{R})$ as well, then $\hat{f} \in C^1(\mathbb{R})$ and $\hat{f}'(t) = -2\pi i \hat{g}(t)$.

(b) If $f \in \mathcal{L}^1(\mathbb{R}) \cap C^1(\mathbb{R})$ and $f' \in \mathcal{L}^1(\mathbb{R})$, then

$$\widehat{(f')}(t) = 2\pi i t \hat{f}(t).$$

Proof. For (a),

$$\frac{1}{s}\left(\hat{f}(t+s) - \hat{f}(t)\right) = \frac{1}{s}\int_{\mathbb{R}} f(x)e^{-2\pi i(t+s)x - e^{-2\pi i t x}} dx = \int_{\mathbb{R}} f(x)e^{-2\pi i t x}\frac{e^{-2\pi i s x} - 1}{s} dx.$$

Observe that

$$\left|f(x)e^{-2\pi t x}\frac{1}{s}(e^{-2\pi i s x} - 1)\right| \lesssim |xf(x)| = |g(x)| \in \mathcal{L}^1(\mathbb{R}).$$

By the Lebesgue dominated convergence theorem,

$$\frac{1}{s}\left(\hat{f}(t+s) - \hat{f}(t)\right) = \int_{\mathbb{R}} f(x)e^{-2\pi i t x}\frac{e^{-2\pi i s x} - 1}{s} dx \to -2\pi i \int_{\mathbb{R}} g(x)e^{-2\pi i t x} dx = -2\pi i \hat{g}(t).$$

For (b), using integration by parts,

$$\widehat{(f')}(t) = \int_{\mathbb{R}} f'(x)e^{-2\pi i t x} dx = f(x)e^{-2\pi i t x}\Big|_{-\infty}^{\infty} + 2\pi i t \int_{\mathbb{R}} f(x)e^{-2\pi i t x} dx = 2\pi i t \hat{f}(t).$$

■

Proposition 4.50

Let $f \in \mathcal{L}^1(\mathbb{R})$ and $b, t \in \mathbb{R}$. Then *(a) If* $g(x) = f(x-b)$, $\hat{g}(t) = e^{-2\pi ibt}\hat{f}(t)$.

(b) If $g(x) = e^{2\pi ibx}f(x)$, $\hat{g}(t) = \hat{f}(t-b)$.

(c) If $g(x) = f(bx)$, $\hat{g}(t) = \frac{1}{|b|}\hat{f}\left(\frac{t}{b}\right)$.

(d) If $f, g \in \mathcal{L}^1(\mathbb{R})$, *then*

$$\int \hat{f}(t)g(t)dt = \int f(t)\hat{g}(t)dt.$$

Proof. For (a), using a translation,

$$\hat{g}(t) = \int f(x-b)e^{-2\pi xt}dx = \int f(x)e^{-2\pi(x+b)t}dt = e^{-2\pi ibt}\int f(t)e^{-2\pi ixt}dx = e^{-2\pi ibt}\hat{f}(t).$$

For (b), using a translation,

$$\hat{g}(t) = \int f(x)e^{2\pi bx}e^{-2\pi ixt}dx = \int f(x)e^{-2\pi ix(t-b)}dx = \hat{f}(t-b).$$

For (c), using a dilation,

$$\hat{g}(t) = \int f(bx)e^{-2\pi ixt}dx = \frac{1}{|b|}\int f(x)e^{-2\pi ixt/b}dx = \frac{1}{|b|}\hat{f}\left(\frac{t}{b}\right).$$

For (d), using Fubini theorem,

$$\int \hat{f}(t)g(t)dt = \int\int f(x)g(t)e^{-2\pi ixt}dxdt = \int\int f(x)g(t)e^{-2\pi ixt}dtdx = \int f(x)\hat{g}(x)dx.$$

The use of Fubini theorem is justified as follows:

$$\int\int \left|f(x)g(t)e^{-2\pi ixt}\right|dxdt = \int |g(t)|\,dt\int f(x)dx = \|f\|_1\,\|g\|_1 < \infty,$$

since $f, g \in \mathcal{L}^1(\mathbb{R})$. ■

Theorem 4.51 (Convolution Theorem)

(a) For $p \in [1,\infty]$, *if* $f \in \mathcal{L}^1(\mathbb{R})$ *and* $g \in \mathcal{L}^p(\mathbb{R})$, *then* $\|f * g\|_p \le \|f\|_1\,\|g\|_p$.

(b) If $f, g \in \mathcal{L}^1(\mathbb{R})$, *then* $\widehat{f * g} = \hat{f}\cdot\hat{g}$.

Proof. We first prove (a). For the case $p = \infty$,

$$|(f * g)(x)| = \int f(y)g(x-y)dy \le \|f\|_1\,\|g\|_\infty\,.$$

For the case $p = 1$, by Tonelli theorem,

$$\begin{aligned}\|f * g\|_1 &\le \int\int |f(y)g(x-y)|\,dydx = \int\int |f(y)|\,|g(x-y)|\,dxdy\\ &= \|g\|_1\int |f(y)|\,dy = \|f\|_1\,\|g\|_1\,.\end{aligned}$$

For the general case where $p \in (1, \infty)$, with $1/p + 1/p' = 1$,

$$\begin{aligned}
\|f * g\|_p^p &= \int \left| \int f(x-y)g(y)dy \right|^p dx \le \int \left(\int |f(x-y)g(y)|\, dy \right)^p dx \\
&\le \int \left(\int |f(x-y)|\, dy \right)^{p/p'} \int |f(x-y)|\, |g(y)|^p \, dydx \\
&= \|f\|_1^{p/p'} \|f * (g^p)\|_1 \le \|f\|_1^{p/p'} \|g^p\|_1 \|f\|_1 = \|f\|_1^{p/p'} \|f\|_1 \|g\|_p^p .
\end{aligned}$$

The second line uses the Hölder inequality and the inequality in the third line uses the result for $p = 1$. Now we obtain that

$$\|f * g\|_p = \|f\|_1 \|g\|_p$$

For (b), using Fubini theorem,

$$\begin{aligned}
\widehat{f * g}(t) &= \int \int f(x-y)g(y)dy e^{-2\pi i x t} dx = \int \int f(x-y)g(y)e^{-2\pi i x t} dxdy \\
&= \int \int f(x-y)g(y)e^{-2\pi i (x-y)t} d(x-y) e^{-2\pi i y t} dy \\
&= \int g(y)\hat{f}(t)e^{-2\pi y t} dy = \hat{f}(t)\hat{g}(t).
\end{aligned}$$

We verify that $(x, y) \mapsto \left| f(x-y)g(y)e^{-2\pi i x t} \right|$ is integrable. Indeed,

$$\begin{aligned}
\int \int \left| f(x-y)g(y)e^{-2\pi i x t} \right| dydx &= \int \int |f(x-y)|\, |g(y)|\, dxdy \\
&= \int |g(y)|\, dy \int |f(x-y)|\, dx = \|f\|_1 \|g\|_1 < \infty
\end{aligned}$$

by Tonelli theorem. The proof is complete. ■

Definition 4.52

*Given $\epsilon > 0$, the **Poisson kernel** is defined as*

$$P_\epsilon(x) = \frac{1}{\pi} \frac{\epsilon}{x^2 + \epsilon^2}.$$

Proposition 4.53

Let P_ϵ be the Poisson kernel. Then

(a) $P_\epsilon(x) \ge 0$ *for all* $x \in \mathbb{R}$ *and* $\epsilon > 0$.

(b) For any $\epsilon > 0$,

$$\int P_\epsilon(x)dx = 1.$$

(c) $\sup_\epsilon \|P_\epsilon\|_1 \le M < \infty$ *for some* $M > 0$.

(d) For any given $\eta > 0$,

$$\lim_{\epsilon \to 0} \int_{|x| > \eta} P_\epsilon(x)dx = 0.$$

Proof. (a) is trivial. For (b),

$$\int P_\epsilon(x)dx = \frac{1}{\pi}\epsilon^{-2}\epsilon^2\pi = 1.$$

(c) follows immediately from (b). For (d), let $\eta > 0$ be given. Then

$$\int_{|x|>\eta} P_\epsilon(x)dx = \frac{1}{\pi}\int_{|x|>\eta}\frac{\epsilon}{x^2+\epsilon^2}dx = \frac{2\epsilon}{\pi}\int_\eta^\infty \frac{1}{x^2+\epsilon^2}dx = \frac{2\epsilon}{\pi}\frac{1}{\epsilon}\left(\frac{\pi}{2} - \tan^{-1}\left(\frac{\eta}{\epsilon}\right)\right) \to 0$$

as $\epsilon \to 0$. ■

Remark

The properties (b)-(d) are sometimes referred to as the ***good kernel*** *property. (d) is used to approximate the dirac δ function.*

Lemma 4.54

Let P_ϵ be the Poisson kernel. Then

(a) If f is uniformly continuous and bounded on $\mathbb{R}$, then $\|P_\epsilon * f - f\|_\infty \to 0$ as $\epsilon \to 0$.

(b) If $f \in \mathcal{L}^p(\mathbb{R})$ where $1 \leq p < \infty$, then

$$\|P_\epsilon * f - f\|_p \to 0 \quad \text{as } \epsilon \to 0.$$

Proof. For (a), we shall proceed with a similat approach in Fejer kernel and Dirichlet kernel. Write

$$\begin{aligned}|P_\epsilon * f(x) - f(x)| &= \left|\int P_\epsilon(x-y)f(y)dy - f(x)\right| = \left|\int P_\epsilon(x-y)(f(y)-f(x))dy\right| \\ &\leq \int P_\epsilon(x-y)\,|f(y)-f(x)|\,dy.\end{aligned}$$

By the uniform continuity of f, for any $\delta > 0$, there exists $\eta > 0$ such that on $[x-\eta, x+\eta]$, $|f(y) - f(x)| < \delta$. Also, by (d) of proposition 4.53, we can choose ϵ small enough such that

$$\int_{|x-y|>\eta} P_\epsilon(x-y)dy < \delta.$$

Then we have

$$\begin{aligned}|P_\epsilon * f(x) - f(x)| &\leq \int_{|x-y|\leq\eta} P_\epsilon(x-y)\,|f(y)-f(x)|\,dy + \int_{|x-y|>\eta} P_\epsilon(x-y)\,|f(y)-f(x)|\,dy \\ &\leq \delta\int_{|x-y|\leq\eta} P_\epsilon(x-y)dy + 2\,\|f\|_\infty \int_{|x-y|>\eta} P_\epsilon(x-y)dy \\ &\leq \delta + 2\,\|f\|_\infty\,\delta = \delta(1 + 2\,\|f\|_\infty)\end{aligned}$$

by the boundedness of f. Since δ is arbitrary, we obtain that $\|P_\epsilon * f - f\|_\infty \to 0$ as $\epsilon \to 0$.

For (b),

$$\begin{aligned}\|P_\epsilon * f - f\|_p^p &= \int \left|\int (f(x)-f(x-y))P_\epsilon(y)dy\right|^p dx\\ &\le \int \left(\int |(f(x)-f(x-y))|\, P_\epsilon(y)dy\right)^p dx\\ &\le \int\int |f(x)-f(x-y)|^p\, P_\epsilon(y)dydx\end{aligned}$$

by Jensen inequality with $d\mu = P_\epsilon(y)dy$ and proposition 4.53 (b). Next, by Fubini theorem, letting $g(y) = \int |f(x)-f(x-y)|^p\, dx$,

$$\begin{aligned}\int\int |f(x)-f(x-y)|^p\, P_\epsilon(y)dydx &= \int\int |f(x)-f(x-y)|^p\, P_\epsilon(y)dxdy\\ &= \int P_\epsilon(y)\int |f(x)-f(x-y)|^p\, dxdy\\ &= \int P_\epsilon(y)g(y)dy = (P_\epsilon * g)(0) \longrightarrow 0\end{aligned}$$

as $\epsilon \longrightarrow 0$ by (a). Thus we conclude that

$$\|P_\epsilon * f - f\|_p \longrightarrow 0$$

as $\epsilon \longrightarrow 0$ for any $1 \le p < \infty$. ■

Theorem 4.55 (Fourier Inversion Theorem)
Suppose $f, \hat{f} \in \mathcal{L}^1(\mathbb{R})$*. Then*

$$f(x) = \int \hat{f}(t)e^{2\pi itx}dt$$

for almost every $x \in \mathbb{R}$.

Proof. Consider

$$I_\epsilon(x) = \int \hat{f}(t)e^{-2\pi\epsilon|t|}e^{2\pi itx}dt.$$

Letting $g_\epsilon(t;x) = e^{-2\pi\epsilon|t|}e^{2\pi itx}$, we have

$$I_\epsilon(x) = \int g_\epsilon(t;x)\hat{f}(t)dt = \int f(t)\hat{g}_\epsilon(t;x)dt$$

since g_ϵ is clearly integrable and this follows from proposition 4.50 (d). Compute that

$$\begin{aligned}\hat{g}_\epsilon(\xi;x) &= \int e^{-2\pi\epsilon|t|}e^{2\pi itx}e^{-2\pi i\xi t}dt = \int_0^\infty e^{2\pi t(i(x-\xi)-\epsilon)}dt + \int_{-\infty}^0 e^{2\pi t(i(x-\xi)+\epsilon)}dt\\ &= \frac{-1}{2\pi(i(x-\xi)-\epsilon)} + \frac{1}{2\pi(i(x-\xi)+\epsilon)} = \frac{1}{\pi}\frac{\epsilon}{(x-\xi)^2+\epsilon^2} = P_\epsilon(x-\xi).\end{aligned}$$

Thus

$$\int f(t)\hat{g}_\epsilon(t;x)dt = \int f(t)P_\epsilon(x-t)dt = (f * P_\epsilon)(x).$$

It follows that $\|P_\epsilon * f - f\|_1 \longrightarrow 0$ as $\epsilon \longrightarrow 0$ by lemma 4.54 (b). It follows that by theorem 2.25 there is a subsequence $I_{\epsilon_k}(x) \longrightarrow f(x)$ almost everywhere. On the other hand,

$$I_\epsilon(x) = \int \hat{f}(t)e^{-2\pi\epsilon|t|}e^{2\pi itx}dt \longrightarrow \int \hat{f}(t)e^{2\pi itx}dt$$

as $\epsilon \longrightarrow 0$ by Lebesgue dominated convergence theorem since

$$\left|\hat{f}(t)e^{-2\pi\epsilon|t|}e^{2\pi itx}\right| \leq \left|\hat{f}(t)\right| \in \mathcal{L}^1(\mathbb{R}).$$

Thus

$$f(x) = \int \hat{f}(t)e^{2\pi itx}dt.$$

This completes the proof. ■

Remark

We may also write

$$\hat{\hat{f}}(x) = f(-x).$$

Definition 4.56

If $f \in \mathcal{L}^2(\mathbb{R})$, we define its fourier transform as

$$\hat{f}(t) = \lim_{N\to\infty} \int_{-N}^{N} f(x)e^{-2\pi itx}dx.$$

Theorem 4.57 (Plancherel)

For $f \in \mathcal{L}^2(\mathbb{R}) \cap \mathcal{L}^1(\mathbb{R})$, $\left\|\hat{f}\right\|_2 = \|f\|_2$.

Proof. Directly write

$$\begin{aligned}\|f\|_2^2 &= \int |f(x)|^2\,dx = \int f(x)\overline{f(x)}dx = \int f(-x)\overline{f(-x)}dx \\ &= \int \hat{\hat{f}}(x)\overline{f(-x)}dx = \int \hat{f}(t)\overline{\hat{f}(t)}dt = \int \left|\hat{f}(t)\right|^2 dt = \left\|\hat{f}\right\|_2^2.\end{aligned}$$

The second equality in the second line follows from proposition 4.50 (d) and the following fact:

$$\widehat{\overline{f(-x)}}(t) = \int_{-\infty}^{\infty} \overline{f(-x)}e^{-2\pi itx}dx = -\int_{\infty}^{-\infty} \overline{f(u)}e^{2\pi itu}du = \int_{-\infty}^{\infty} \overline{f(u)} \cdot \overline{e^{-2\pi itu}}du = \overline{\hat{f}(t)},$$

where we have used the change of variable $u = -x$. ■

Definition 4.58

*We denote the **fourier transform operator** as*

$$\mathcal{F}f(t) = \int_{\mathbb{R}} f(x)e^{-2\pi itx}dx$$

for $f \in \mathcal{L}^1(\mathbb{R})$. If $f \in \mathcal{L}^2(\mathbb{R})$, we define

$$\mathcal{F}f(t) = \lim_{N\to\infty} \int_{-N}^{N} f(x)e^{-2\pi itx}dx$$

instead.

Remark
From Plancherel theorem, it is immediate that $\mathcal{F}$ is a bounded linear operator.

Definition 4.59
*The **Schwartz space** $\mathcal{S}(\mathbb{R})$ is the space of all functions $f \in C^\infty(\mathbb{R})$ such that*

$$\sup_{x\in\mathbb{R}} \left|x^k D^m f(x)\right| < \infty$$

for all $k, m \in \mathbb{N}$, where D^m is the m-th differentiation operator.

Proposition 4.60
Let $\mathcal{S}(\mathbb{R})$ be the Schwartz space.

(a) $\mathcal{S}(\mathbb{R})$ is a vector space over $\mathbb{R}$.

(b) If $f \in \mathcal{S}(\mathbb{R})$, then $x^k f^{(m)}(x) \in \mathcal{S}(\mathbb{R})$ for all $k, m \in \mathbb{N} \cup \{0\}$.

(c) If $f \in \mathcal{S}(\mathbb{R})$, then $f \in \mathcal{L}^p(\mathbb{R})$ for all $p \geq 1$.

Proof. For (a), we check that $\mathcal{S}(\mathbb{R})$ is closed under addition and scalar multiplication. Let $f, g \in \mathcal{S}(\mathbb{R})$ and $c \in \mathbb{R}$. Then $cf + g$ is also smooth and for $k, l \in \mathbb{N} \cup \{0\}$,

$$\sup_{x\in\mathbb{R}} \left|x^k (cf+g)^{(l)}(x)\right| \leq |c| \sup_{x\in\mathbb{R}} |x|^k \left|f^{(l)}(x)\right| + \sup_{x\in\mathbb{R}} \left|x^k g^{(l)}(x)\right| < \infty$$

by the definition of Schwartz space. Then $cf + g \in \mathcal{S}(\mathbb{R})$, so $\mathcal{S}(\mathbb{R})$ is a vector space over $\mathbb{R}$.

To prove (b), we only need to show the following two facts: first, for any $f \in \mathcal{S}(\mathbb{R})$, $xf(x) \in \mathcal{S}(\mathbb{R})$; second, for any $f \in \mathcal{S}(\mathbb{R})$, $f'(x) \in \mathcal{S}(\mathbb{R})$. Suppose that $f \in \mathcal{S}(\mathbb{R})$. Then for any $k, l \in \mathbb{N} \cup \{0\}$,

$$\sup_{x\in\mathbb{R}} \left|x^k (xf(x))^{(l)}\right| = \sup_{x\in\mathbb{R}} \left|x^k \left(\sum_{i=0}^{l} \binom{l}{i} x^{(i)} f^{(l-i)}(x)\right)\right| \leq \sup_{x\in\mathbb{R}} \left|x^{k+1} f^{(l)}(x)\right| + n \sup_{x\in\mathbb{R}} \left|x^k f^{(l-1)}(x)\right| < \infty$$

by the Leibniz formula. Also,

$$\sup_{x\in\mathbb{R}} \left|x^k (f'(x))^{(l)}\right| = \sup_{x\in\mathbb{R}} \left|x^k f^{(l+1)}(x)\right| < \infty.$$

Thus $xf(x), f'(x) \in \mathcal{S}(\mathbb{R})$. In general, the function of the form $x^k f^{(l)}(x) \in \mathcal{S}(\mathbb{R})$ can be proved by using the above two facts finitely many times.

For (c), let $E = [-1, 1]$. By the smoothness of f, we know that there is some M such that $\sup_{x\in E} |f(x)| \leq M$. Also, from the definition of Schwartz space, $\sup_{x\in\mathbb{R}} \left|x^2 f(x)\right| \leq C$ for some constant C. Then

$$\begin{aligned}
\int |f(x)|^p \, dx &= \int_E |f(x)|^p \, dx + \int_{E^c} |f(x)|^p \, dx \\
&= 2M^p + \int_{E^c} \left|\frac{x^2 f(x)}{x^2}\right|^p \, dx \\
&\leq 2M^p + C^p \int_{E^c} \frac{1}{x^2} dx = 2M^p + 2C^p < \infty.
\end{aligned}$$

Thus $\|f\|_p < \infty$ for all $p \geq 1$ and $p \neq \infty$. We check that f is bounded on $\mathbb{R}$. By the continuity of f, we have that f is always bounded on a compact set. Now if f does not vanish at infinity, then there is some $\delta > 0$ and a sequence x_n such that $|x_n| \to \infty$ and $|f(x_n)| > \delta$. Then $\sup_{x\in\mathbb{R}} |xf(x)| \geq \delta \sup_{x\in\mathbb{R}} |x| = \infty$, posing a contradiction. Thus f vanishes at infinity. We can find some compact interval E such that $\sup_{x\in E} |f(x)| \geq \sup_{x\in E^c} |f(x)|$. Then by the extreme value theorem, f is bounded on E and hence on $\mathbb{R}$. We conclude that $f \in \mathcal{L}^p(\mathbb{R})$ for all $p \geq 1$. ■

Proposition 4.61

Let $S(\mathbb{R})$ be the Schwartz space. If $f \in \mathcal{S}(\mathbb{R})$, then $\hat{f} \in \mathcal{S}(\mathbb{R})$.

Proof. To see this, let $f \in \mathcal{S}(\mathbb{R})$ be given. From proposition 4.60 (b), we know that f and $g(x) = xf(x) \in \mathcal{S}(\mathbb{R}) \subset \mathcal{L}^1(\mathbb{R})$. Thus $\hat{f} \in C^1(\mathbb{R})$ and $\hat{f}'(t) = -2\pi i\hat{g}(t)$. Since $g \in \mathcal{S}(\mathbb{R}) \subset \mathcal{L}^1(\mathbb{R})$, we can repeat the argument to obtain that $\hat{f} \in C^2$ and $\hat{f}''(t) = (-2\pi i)^2 \hat{G}(t)$, where $G(x) = x^2 f(x)$. Apply the same argument repeatedly, we have that $\hat{f} \in C^\infty(\mathbb{R})$ and $\hat{f}^{(l)}(t) = (-2\pi i)^l \hat{h}(t)$, where $h(x) = x^l f(x)$ for all $l \in \mathbb{N} \cup \{0\}$. Also,

$$\sup_{x\in\mathbb{R}} \left|x^k \hat{f}^{(l)}(x)\right| = \sup_{x\in\mathbb{R}} \left|x^k (-2\pi i)^l \hat{h}(x)\right| \leq (2\pi)^l \sup_{x\in\mathbb{R}} \left|x^k \hat{h}(x)\right| < \infty.$$

The last inequality follows from the fact that $h \in \mathcal{S}(\mathbb{R}) \subset \mathcal{L}^1(\mathbb{R})$ and the Riemann-Lebesgue lemma guarantees that $\hat{h}$ vanishes at infinity. We conclude that $f \in \mathcal{S}(\mathbb{R})$ implies $\hat{f} \in \mathcal{S}(\mathbb{R})$. ■

Proposition 4.62

$\mathcal{S}(\mathbb{R})$ is dense in $\mathcal{L}^p(\mathbb{R})$ for $1 \leq p < \infty$.

Proof. Since continuous functions with compact support are dense in $\mathcal{L}^p(\mathbb{R})$, it suffices to show that $\mathcal{S}(\mathbb{R})$ is dense in the space of continuous functions with compact support. Without loss of generality, we can assume that f is supported on $[-a, a]$ for some $a > 0$. By the Weierstrass theorem, we can find a polynomial q such that $\|f - q\|_\infty < \epsilon/2$. Consider the function

$$\phi_n(t) = \begin{cases} e^{-\frac{1}{n(t^2-a^2)}} & \text{if } |t| < a \\ 0 & \text{if } |t| \geq a. \end{cases}$$

Note that $\phi_n \to \chi_{(-a,a)}$ pointwisely as $n \to \infty$ and bounded by 1. We verify that $\phi_n \in \mathcal{S}(\mathbb{R})$ for all $n \in \mathbb{N}$. Indeed, for any $k, l \in \mathbb{N} \cup \{0\}$, since $D^l \phi_n$ will result in

$$t^k D^l \phi_n(t) = r(t; n, k, l) e^{-\frac{1}{n(t^2-a^2)}}$$

on $[-a, a]$ for some rational function $r(t; n, k, l)$ having singularities only at $t = \pm a$, we have that

$$\sup_{t \in \mathbb{R}} \left| t^k D^l \phi_n(t) \right| < \infty.$$

Hence, $\phi_n \in \mathcal{S}(\mathbb{R})$ for all $n \in \mathbb{N}$.

Now it follows from proposition 4.60 (b) that $q\phi_n \in \mathcal{S}(\mathbb{R})$ by extending the polynomial q on $\mathbb{R}$. Then

$$\begin{aligned} \int |f - q\phi_n|^p \, d\mu = \int_{-a}^{a} |f - q\phi_n|^p \, d\mu &\leq 2^{p-1}\left(\int_{-a}^{a} |f - q|^p \, d\mu + \int_{-a}^{a} |q - q\phi_n|^p \, d\mu\right) \\ &\leq 2^{p-1}\left(2a\epsilon^p + \int_{-a}^{a} |q - q\phi_n|^p \, d\mu\right) \to 0 \end{aligned}$$

as $n \to \infty$ by the Lebesgue dominated convergence theorem using $|q - q\phi_n|^p \to 0$ pointwisely a.e. and $|q - q\phi_n|^p \leq 2^p |q|^p$ is integrable. The last inequaltiy comes from the convexity $(x/2 + y/2)^p \leq x^p + y^p$ for $x, y \geq 0$ and $p \geq 1$. We conclude that $\mathcal{S}(\mathbb{R})$ is dense in $\mathcal{L}^p(\mathbb{R})$ for $1 \leq p < \infty$. ■

Definition 4.63
A linear operator $T : \mathcal{H}_1 \to \mathcal{H}_2$ is said to be ***unitary*** *if*

- *(a) T is invertible.*
- *(b) $\|Tf\|_2 = \|f\|_1$ for all $f \in \mathcal{H}_1$.*

Proposition 4.64
Let $\mathcal{F}$ be the Fourier transform operator on $\mathcal{L}^2(\mathbb{R})$.

- *(a) $\mathcal{F}$ is unitary on $\mathcal{L}^2(\mathbb{R})$.*
- *(b) $\mathcal{F}^4 = I$.*

Proof. (a) is directly from the Plancherel theorem. For (b), using proposition 4.61,

$$\mathcal{F}^4 f(t) = \mathcal{F}^2 f(-t) = f(t),$$

by the Fourier inversion theorem for Schwartz functions. Since $\mathcal{F}$ is unitary, it is also a bounded linear operator, and hence a continuous operator. It now follows from proposition 4.62 that for any $f \in \mathcal{L}^2(\mathbb{R})$, there is a sequence $f_n \in \mathcal{S}(\mathbb{R})$ such that

$$\|f_n - f\|_2 \to 0.$$

Then

$$\mathcal{F}^4 f(t) = \mathcal{F}^4 \lim_{n\to\infty} f_n(t) = \lim_{n\to\infty} \mathcal{F}^4 f_n(t) = \lim_{n\to\infty} f_n(t) = f(t)$$

for any $f \in \mathcal{L}^2(\mathbb{R})$ by the continuity of $\mathcal{F}$. ■

Example

We can use the Fourier transform to solve some PDEs. Consider the Laplace equation

$$\begin{cases} \nabla^2 \cdot u = 0 & \text{for } u : \mathbb{R}^2 \to \mathbb{R}, x \in \mathbb{R}, y > 0, \\ u(x,0) = f(x) & \text{for } x \in \mathbb{R}. \end{cases}$$

Apply the Fourier transform in x direction, the original PDE becomes

$$\begin{cases} 4\pi^2 t^2 \hat{u}(t,y) + \hat{u}_{yy}(t,y) = 0 & \text{for } t \in \mathbb{R}, y > 0, \\ \hat{u}(t,0) = \hat{f}(t) & \text{for } t \in \mathbb{R}. \end{cases}$$

Fix t, conjecture that $\hat{u}(t,y) = A(t)e^{-2\pi|t|y} + B(t)e^{2\pi|t|y}$. Then we have

$$\begin{cases} \hat{u}(t,y) = A(t)e^{-2\pi|t|y}, \\ A(t) = \hat{f}(t). \end{cases}$$

Since

$$\hat{u}(t,y) = \hat{f}(t)e^{-2\pi|t|y} = \hat{f}(t)\hat{P}_y(t) = \widehat{f * P_y}(t),$$

we obtain

$$\begin{cases} u(x,y) = f * P_y(x), \\ \lim_{y\to 0} u(x,y) = f(x). \end{cases}$$

5. Further Topics

5.1. Linear Operators

Definition 5.1

Let $\{M_n\} \subset B(X,Y)$ be a sequence of bounded linear operators. M_n ***converges strongly*** *if for any $x \in X$, $\|M_n x - y\|_Y \to 0$ for some $y \in Y$.*

Proposition 5.2

If $\{M_n\} \subset B(X,Y)$ converges strongly, then there is an $M \in B(X,Y)$ such that $\|M_n x - Mx\|_Y \to 0$ for all $x \in X$.

Proof. Set $Mx = \lim_{n\to\infty} M_n x$ for all $x \in X$. We check that $M \in B(X,Y)$. Linearity is trivial; we check the boundedness. Let $f_n(x) = \|M_n x\|_Y$. Then f_n is sub-additive and $f_n(\alpha x) = |\alpha| f_n(x)$ for all $\alpha \in \mathbb{R}$. If $x_k \to x$,

$$f_n(x_k) = \|M_n x_k\|_Y \to \|Mx_k\|_Y = f_n(x)$$

for any fixed n as $k \to \infty$; f_n is continuous. Now for any fixed $x \in X$, $\sup_n f_n(x) = \sup_n \|M_n x\| \leq C(x)$ by the strong convergence. It follows from the uniform boundedness principle that there is $C_0 < \infty$ such that $|f_n(x)| \leq C_0 \|x\|_X$ for all n. Thus

$$\|Mx\|_Y = \lim_{n\to\infty} \|M_n x\|_Y \leq C_0 \|x\|_X .$$

Hence $\|M\| \leq C_0$ and $M \in B(X,Y)$. ∎

Definition 5.3

A sequence $\{M_n\} \subset B(X,Y)$ ***converges weakly*** *if for all $x \in X$, $M_n x \xrightarrow{w} y \in Y$ for some y.*

Proposition 5.4

If $\{M_n\} \subset B(X,Y)$ converges weakly, then there is an $M \in B(X,Y)$ such that $M_n x \xrightarrow{w} Mx$ for all $x \in X$.

Proof. Set $Mx = \lim_{n\to\infty} M_n x$ for all $x \in X$. We check that $M \in B(X,Y)$. Linearity is trivial; we check the boundedness. Without loss of genrality, we can assume that $\|x\|_X = 1$. Observe that $\{M_n x\} \subset Y$ is a weakly convergence sequence and hence weakly sequentially compact; by proposition 2.75, it is bounded. Thus there exists $C < \infty$ such that $\|M_n x\|_Y \leq C = C\|x\|$ for all n. Taking the limit, we have $\|Mx\|_Y = \lim_{n\to\infty} \|M_n x\|_Y \leq C \|x\|_X$. Hence $M \in B(X,Y)$.

Lastly, we check the weak convergence. Indeed, for any $\ell \in Y'$, $\ell(M_n x) = \ell(Mx)$ for all $x \in X$ since $M_n x \to Mx$ in Y. Thus $M_n x \xrightarrow{w} Mx$ in Y. ∎

Lemma 5.5

Let X be a reflexive and $\{T_n\} \subset B(X,Y)$. Then $T_n \xrightarrow{w} T$ implies $T_n' \xrightarrow{w} T'$.

Proof. For all $x \in X$ and $\ell \in Y'$, $\ell T_n x \to \ell T x$. We need to show that $T_n'\ell \to T'\ell$ for all $\ell \in Y'$. Note that $T_n'\ell \in X'$ and X is reflexive, so we only need to check $T_n'\ell(x) \to T'\ell(x)$ for all $x \in X$. But this is essentially

$$T_n'\ell(x) = \ell T_n x \to \ell T x = T'\ell(x).$$

■

Remark

The statement of the lemma fails if we replace weak convergence with strong convergence, i.e., $T_n \to T$ does not imply $T_n' \to T'$ in general. Consider $X = \ell^2(\mathbb{N})$ and $T_n : \ell^2 \to \ell^2$ be the operator

$$T_n(x_1, \ldots) = (x_n, 0, \ldots)$$

for $x = (x_1, x_2, \ldots) \in \ell^2$. Since X is a Hilbert space, it is reflexive. Also, $T_n \to 0$ strongly, since

$$\|T_n x - 0\|_2^2 = |x_n|^2 \to 0$$

as $n \to \infty$ for all $x \in X$. However, $T_n' : (\ell^2)' \to (\ell^2)'$ is the operator defined by $\ell \mapsto \ell T_n$. For any $\ell \in (\ell^2)'$, $\ell(x) = \langle x, y_\ell \rangle$ for a unique $y_\ell \in \ell^2$. Then $T_n'\ell(x) = \ell T_n x = \langle y_\ell, T_n x \rangle = (y_\ell)_1 \cdot x_n$. Thus

$$T_n'\ell = (y_\ell)_1 \cdot e_n.$$

Pick any $y_\ell \in \ell^2$ such that $(y_\ell)_1 \neq 0$ will give

$$\left\|T_n'\ell - 0\right\| = |(y_\ell)_1| \not\to 0.$$

Theorem 5.6

Let $T_n \in B(X, Y)$ be a sequence of bounded linear operators such that

(a) $\|T_n\| \leq C < \infty$ *for all n;*

(b) $T_n x \to T x$ *for all $x \in D \subset X$ where D is a dense subset of X.*

Then $T_n \to T$ strongly.

Proof. We claim that for any $z \in X$, the sequence $\{T_n z\}$ is Cauchy in Y. Let $\epsilon > 0$ be given. Since D is dense in X, there exists a $x \in D$ such that $\|z - x\|_X < \frac{\epsilon}{3C}$. Then

$$\begin{aligned}\|T_n z - T_m z\|_Y &\leq \|T_n(z - x)\|_Y + \|T_m(z - x)\|_Y + \|T_n x - T_m x\|_Y \\ &\leq (\|T_n\| + \|T_m\|) \cdot \|z - x\|_X + \|T_n x - T_m x\|_Y \leq 2C \cdot \frac{\epsilon}{3C} + \|T_n x - T_m x\|_Y .\end{aligned}$$

Since $T_n x$ converges, it is Cauchy. Thus there exists N such that for all $n, m \geq N$, $\|T_n x - T_m x\|_Y < \frac{\epsilon}{3}$. Hence

$$\|T_n z - T_m z\|_Y < \frac{2\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.$$

Thus $\{T_n z\}$ is Cauchy in Y. Since Y is complete, there exists $y_z \in Y$ such that $T_n z \to y_z$. Define

$Tz = y_z$. We check that $T \in B(X,Y)$. The linearity is trivial; for any $z \in X$,

$$\|Tz\|_Y = \lim_{n\to\infty} \|T_n z\|_Y \le C\,\|z\|_X\,.$$

Thus $\|T\| \le C$ and $T \in B(X,Y)$. Lastly, we check the strong convergence. For any $z \in X$,

$$\|T_n z - Tz\|_Y = \|T_n z - y_z\|_Y \to 0$$

as $n \to \infty$. Thus $T_n \to T$ strongly. ■

Theorem 5.7 (Uniform Boundedness Principle III)
A family of operators $\{T_\alpha\}_{\alpha\in I} \subset B(X,Y)$ satisfies that for all $x \in X$ and $\ell \in Y'$,

$$|\ell(T_\alpha x)| \le C(x,\ell) < \infty$$

for all $\alpha \in I$. Then there exists $C_0 < \infty$ such that

$$\|T_\alpha\| \le C_0 < \infty$$

for all $\alpha \in I$.

Proof. Set

$$f_\alpha(x) = \|T_\alpha x\|_Y = \sup_{\|\ell\|=1} |\ell(T_\alpha x)|\,.$$

We verify the conditions of the uniform boundedness principle. Let $x_k \to x$ in X. Then $T_\alpha x_k \to T_\alpha x$ in Y. Thus $f_\alpha(x_k) \to f_\alpha(x)$ given any $\alpha \in I$. Thus f_α is continuous.

$$f_\alpha(x+y) = \|T_\alpha(x+y)\|_Y \le \|T_\alpha x\|_Y + \|T_\alpha y\|_Y = f_\alpha(x) + f_\alpha(y).$$

f_α is sub-additve. Also, $f_\alpha(cx) = \|T_\alpha(cx)\|_Y = |c|\,\|T_\alpha x\|_Y = |c|\,f_\alpha(x)$ for all $c \in \mathbb{R}$. Lastly, given x, $g_\alpha(\ell) = |\ell(T_\alpha x)|$ is clearly continuous, sub-additive and homogeneous. Also, $|g_\alpha(\ell)| \le C(x,\ell)$. Using the boundedness assumption and applying the uniform boundedness principle, we have that

$$\sup_{\alpha\in I} |g_\alpha(\ell)| \le C_1(x)\,\|\ell\|\,.$$

Now

$$\sup_{\alpha\in I} |f_\alpha(x)| = \sup_{\alpha\in I}\sup_{\|\ell\|=1} |\ell(T_\alpha x)| = \sup_{\|\ell\|=1}\sup_{\alpha\in I} |\ell(T_\alpha x)| \le \sup_{\|\ell\|}\sup_{\alpha\in I} |g_\alpha(\ell)| \le C_1(x).$$

Applying the uniform boundedness principle again on f_α, we have that

$$\sup_{\alpha\in I} f_\alpha(x) \le C_0\,\|x\|_X\,.$$

We conclude that $\|T_\alpha\| \le C_0$ for all $\alpha \in I$. ■

Proposition 5.8
Let $T \in B(X,Y)$, $U \in B(Y,Z)$. Then $UT \in B(X,Z)$ and $(UT)' = T'U'$.

Proof. We first show that $UT \in B(X,Z)$. For any $c \in \mathbb{R}$ and $x, y \in X$,

$$UT(cx + y) = U(T(cx + y)) = U(cTx + Ty) = cUTx + UTy.$$

Thus UT is linear. Now we check the boundedness. For any $x \in X$,

$$\|UTx\|_Z \leq \|U\| \, \|Tx\|_Y \leq \|U\| \, \|T\| \, \|x\|_X \,.$$

Since $\|U\|, \|T\|$ are finite, the boundedness follows.

Now we check the adjoint. For any $\ell \in Z'$, $(UT)'(\ell) = \ell UT = (U'\ell)T = T'U'\ell$. ■

Definition 5.9
*$T \in B(X,Y)$ is **compact** if for any bounded sequence $\{x_n\} \subset X$, $\{Tx_n\}$ has a convergent subsequence in Y.*

Definition 5.10
*The **compact operator space** is denoted by $B_0(X,Y)$.*

Proposition 5.11
Let $T \in B(X,Y)$ be a compact operator, $S_1 \in B(Y,Z)$ and $S_2 \in B(W,X)$. Then $S_1T \in B_0(X,Z)$ and $TS_2 \in B_0(W,Y)$.

Proof. Let $\{x_n\} \subset X$ be a bounded sequence. Then $\{Tx_n\} \subset Y$ has a convergent subsequence $\left\{Tx_{n_k}\right\}$. Since S_1 is bounded, it is continuous; thus $\left\{S_1Tx_{n_k}\right\}$ is convergent in Z. Hence S_1T is compact.

Now let $\{w_n\} \subset W$ be a bounded sequence. Then $\|S_2w_n\|_X \leq \|S_2\| \, \|w_n\|_W$ is also bounded in X. Thus by the compactness of T, $\{TS_2w_n\}$ has a convergent subsequence. We conclude that TS_2 is compact. ■

Lemma 5.12
Let X be a metric space. If $A_n \subset X$ is a sequence of separable subsets of X and $A_n \nearrow A$, then A is separable.

Proof. Since A_n is separable, there exists a countable dense subset $D_n \subset A_n$. Let $D = \bigcup_{n=1}^{\infty} D_n$. We claim that D is dense in A. Let $x \in A$ be given. Since $A_n \nearrow A$, there exists n_0 such that $x \in A_{n_0}$. Then for any $\epsilon > 0$, there exists $y \in D_{n_0} \subset D$ such that $d(x,y) < \epsilon$. Thus D is dense in A. ■

Theorem 5.13
Let $T \in B_0(X,Y)$. Then $T(X)$ is separable.

Proof. Consider the closed unit ball $B = \{x \in X \mid \|x\|_X \leq 1\}$ in X. Since T is compact, $T(B)$ is sequentially compact. Then $T(B)$ is compact in Y. Because every compact metric space is separable, $T(B)$ is separable. Write $X = \bigcup_n nB$. Then $T(X) = \bigcup_n T(nB) = \bigcup_n nT(B)$. By lemma 5.12, $T(X)$ is separable. ■

Theorem 5.14
Let $T \in B_0(X,Y)$ be a compact operator. Then $T' \in B_0(Y',X')$.

Proof. Suppose first that Y is separable. Let $g_n \in Y'$ be a bounded sequence. $T(X) \subset Y$ is also separable. There exists a countable dense subset $\{y_k\} \subset T(X)$. For y_1, $\{g_n(y_1)\}$ is a bounded sequence in $\mathbb{R}$. By the Bolzano-Weierstrass theorem, there exists a subsequence $\left\{g_n^{(1)}\right\}$ such that $g_n^{(1)}(y_1)$ converges. For y_2, extract from $\left\{g_n^{(1)}\right\}$ to obtain a subsequence $\left\{g_n^{(2)}\right\}$ such that $g_n^{(2)}(y_2)$ converges. Continuing this process, we obtain a sequence $\left\{g_n^{(k)}\right\}$ such that $g_n^{(k)}(y_j)$ converges for all $j \leq k$. Pick $f_n = g_n^{(n)}$. Then for any k, $f_n(y_k)$ converges. Now given any $y \in Y$, we may without loss of generality assume that $y_k \to y$. Then

$$\begin{aligned} |f_n(y) - f_m(y)| &\leq |f_n(y) - f_n(y_k)| + |f_m(y) - f_m(y_k)| + |f_n(y_k) - f_m(y_k)| \\ &\leq (\|f_n\| + \|f_m\|)\, \|y - y_k\|_Y + |f_n(y_k) - f_m(y_k)|\,. \end{aligned}$$

Since $f_n(y_k)$ converges for all k, it is Cauchy; $\{f_n\} \subset \{g_n\}$ is bounded. Thus taking $m,n \to \infty$ and then $k \to \infty$, we see that $|f_n(y) - f_m(y)| \to 0$. Hence $\{f_n(y)\}$ is Cauchy.

Next, we show that f_n is in fact Cauchy in Y'.

$$\|f_n - f_m\| = \sup_{\|y\|_Y = 1} |f_n(y) - f_m(y)|$$

For each m,n, there exists $y \in Y$ such that $\|y\|_Y = 1$ and

$$|f_n(y) - f_m(y)| \geq \frac{1}{2}\, \|f_n - f_m\|\,.$$

But $f_n(y)$ is Cauchy and thus $\|f_n - f_m\| \to 0$ as $n,m \to \infty$. Thus $\{f_n\} \subset Y'$ is Cauchy. Y' is complete, so there exists $f \in Y'$ such that $f_n \to f$ in Y'. Now $f_n(Tx) \to f(Tx)$ for all $x \in X$. Thus

$$\|T'f_n - T'f\| \leq \|T'\|\, \|f_n - f\| \to 0.$$

Hence $T'f_n \to T'f$ in X'. $\{T'f_n\}$ is a convergent subsequence of $\{T'g_n\}$. T' is compact.

In general if Y is not separable, $T(X)$ is a separable subspace of Y (theorem 5.13). The same argument applies and we obtain a sequence $\{f_n\} \subset Y'$ such that

$$\sup_{\|Tx\|_Y = 1} |f_n(Tx) - f_m(Tx)| \to 0.$$

Thus there exists f on $T(X)$ such that

$$\sup_{\|Tx\|_Y=1} |f_n(Tx) - f(Tx)| \to 0$$

by the completeness of $T(X)$. Then

$$\|T'f_n - T'f\| = \sup_{\|Tx\|_Y=1} |f_n(Tx) - f(Tx)| \to 0.$$

Thus $T'f_n \to T'f$ in X'. Hence T' is compact. ■

Definition 5.15

*Let X, Y be Banach spaces. A linear operator T is said to be **densely defined** if $D(T) = \{x \in X \mid Tx \in Y\}$ is dense in X. We denote it as $T : D(T) \overset{d}{\subset} X \to Y$.*

Definition 5.16

*A linear operator $T : D(T) \overset{d}{\subset} X \to Y$ is said to be **bounded** if there is $c < \infty$ such that*

$$\|Tx\|_Y \leq c\,\|x\|_X$$

*for all $x \in D(T)$; T is **unbounded** if for all $c > 0$, there exists $x \in D(T)$ such that*

$$\|Tx\|_Y > c\,\|x\|_X\,.$$

Remark

T is bounded if and only if T is continuous on every point in $D(T)$; T is unbounded if and only if T is not continuous on every point in $D(T)$.

Definition 5.17

*$T : D(T) \subset X \to Y$ is **closed** if its graph*

$$G(T) = \{(x, y) \in X \times Y \mid x \in D(T), Tx = y\}$$

is closed in the norm $\|(x, y)\|_{X \times Y} = \|x\|_X + \|y\|_Y$.

Remark

T is closed if $x_n \to x$ in X, where $x_n \in D(T)$ and $Tx_n \to y$ in Y implies that $x \in D(T)$ and $Tx = y$.

Definition 5.18

*$T_1 : D(T_1) \subset X \to Y$ and $T_2 : D(T_2) \subset X \to Y$ are linear unbounded operators. We say that T_2 is an **extension** of T_1 if $D(T_1) \subset D(T_2)$ and $T_2x = T_1x$ for all $x \in D(T_1)$. Denote it as $T_1 \subset T_2$.*

Definition 5.19

*A linear operator $T : D(T) \subset X \to Y$ is said to be **closable** if there is a closed extension of T.*

Remark
T is closable if and only if for $x_n \in D(T)$, $x_n \to 0$ and $Tx_n \to y$ in Y implies that $y = 0$.

Example
$f : D(T) \in \ell^2 \to \mathbb{R}$, where $D(T) = \operatorname{span}(\{e_n \mid n \in \mathbb{N}\})$ is defined by $Te_n = n$ and extended by linearity. Then T is unbounded. Since T is unbounded, we may take $x_n \to 0$ in $D(T)$ and $|f(x_n)| \geq \epsilon$ for some $\epsilon > 0$.

$$z_n = \frac{x_n}{Tx_n} \to 0 \quad \textit{and} \quad Tz_n = 1.$$

Thus T is not closable.

Example
$T : f \mapsto f'$ on $X = C[0,1]$ and $D(T) = C^1[0,1]$. Then T is closed. If $D(T) = C^\infty[0,1]$, then T is not closed while closable.

5.2. Second Order Ordinary Differential Equations

The goal of this section is to present some solution techniques for solving the second order ODEs that will be intensively used in the next section. The techniques are presnted without proofs.

We first introduce the variation of parameters method. Consider the second order ODE of the form

$$y'' + p(x)y' + q(x)y = f(x),$$

where $a \leq x \leq b$. The first step is to find the solutions of the homogeneous version

$$y'' + p(x)y' + q(x)y = 0.$$

Assume that we can find two linearly independent solutions $y_1(x)$ and $y_2(x)$. Then a particular solution of the non-homogeneous equation is

$$y_p(x) = -y_1(x)\int_a^x \frac{y_2(t)f(t)}{W(y_1,y_2)(t)}dt + y_2(x)\int_a^x \frac{y_1(t)f(t)}{W(y_1,y_2)(t)}dt,$$

where the **Wronskian** W is defined as

$$W(y_1,y_2)(t) = \begin{vmatrix} y_1(t) & y_2(t) \\ y_1'(t) & y_2'(t) \end{vmatrix}.$$

The general solution of the non-homogeneous equation is

$$y(x) = c_1 y_1(x) + c_2 y_2(x) + y_p(x),$$

for some constants c_1 and c_2 that should be determined by the boundary conditions.

The difficulty of the variation of parameters method is that it is not always easy to find two linearly independent solutions of the homogeneous equation. If the coefficients are actually constant, we can consider the corresponding characteristic polynomial

$$\lambda^2 + p\lambda + q = 0.$$

If the roots λ_1 and λ_2 are distinct, then the two linearly independent can be found as

$$y_1(x) = e^{\lambda_1 x}, \quad y_2(x) = e^{\lambda_2 x}.$$

If $\lambda = \lambda_1 = \lambda_2$, the two linearly independent solutions can be found as

$$y_1(x) = e^{\lambda x}, \quad y_2(x) = xe^{\lambda x}.$$

The next method is using the Green's function. Consider the second order ODE

$$y'' + p(x)y' + q(x)y = f(x).$$

The differential operator L is defined as

$$L = D^2 + p(x)D + q(x)I,$$

with boundary conditions

$$Ry = 0,$$

where D is the differential operator and R is a linear operator that represents the boundary conditions. Suppose that the solution has the form

$$y(x) = \int_a^b G(x,t)f(t)dt,$$

where $G(x,t)$ is the **Green's function** characterized by the following differential equation

$$\begin{cases} LG(x,t) = \delta(x-t), & x \in [a,b], \\ RG(x,t) = 0, \\ G(t^+,t) = G(t^-,t), \\ G_x(t^+,t) - G_x(t^-,t) = \frac{1}{r(t)}, \end{cases}$$

where r is the function with

$$L = D(r(x)D) + s(x)$$

being the result of factorization. This form is called the **Sturm-Liouville form** operator.

The form of r and s can be found by the following

$$[D(rD) + s]\, y = ry'' + r'y' + sy = 0 \quad \Leftrightarrow \quad y'' + \frac{r'}{r}y' + \frac{s}{r}y = 0.$$

This means that

$$p = \frac{r'}{r}, \quad q = \frac{s}{r}.$$

So

$$r(x) = e^{\int_a^x p(t)dt}.$$

We can rewrite the characterization of the Green's function as

$$\begin{cases} LG(x,t) = 0, & x \in [a,b], \\ RG(x,t) = 0, & \\ G(t^+,t) = G(t^-,t), & \\ G_x(t^+,t) - G_x(t^-,t) = \exp\Big(-\int_a^t p(s)ds\Big). & \end{cases}$$

5.3. Spectra and Resolvent

Definition 5.20
*Let $T : D(T) \subset X \to X$ be a closed linear operator. The **resolvent set** of T is defined as*

$$\rho(T) = \{\xi \in \mathbb{C} \mid (T - \xi I) \text{ has bounded inverse on } X\}.$$

*The **spectrum** of T is defined as $\sigma(T) = \mathbb{C} \setminus \rho(T)$. $R_T(\xi) = (T - \xi I)^{-1}$ is called the **resolvent operator** of T.*

Remark
$\xi \in \rho(T)$ if and only if $T - \xi I$ has the bounded inverse on X.

Remark
$\xi \in \sigma(T) = \mathbb{C} \setminus \rho(T)$ if either $T - \xi I$ is not invertible or $T - \xi I$ is invertible but has range smaller than X. If $\dim X < \infty$, $\sigma(T) = \{\lambda \in \mathbb{C} \mid Tx = \lambda x$ for some $x \in X \setminus \{0\}\}$.

Example
$X = C[a,b]$. $Tu = u'$ and $D(T) = C^1[a,b]$. T is not invertible since $T(u) = 0$ for every constant function u. Consider the following domains

- $D_1 = \{u \in D(T) \mid u(a) = 0\}$,
- $D_2 = \{u \in D(T) \mid u(b) = 0\}$,
- $D_3 = \{u \in D(T) \mid u(a) = ku(b)\}$,
- $D_0 = \{u \in D(T) \mid u(a) = u(b) = 0\}$.

$T_i = T|_{D_i}$ *are invertible on* D_i *for* $i = 0, 1, 2, 3$, *but the inverses are different. For example,*

$$(T_1^{-1}v)(x) = \int_a^x v(t)dt.$$

Theorem 5.21 (Neumann Series)
Let $T : X \to X$ *be a bounded lienar operator. If* $\|T\| < 1$, *then* $I - T$ *is invertible and*

$$(I - T)^{-1} = \sum_{n=0}^{\infty} T^n.$$

Proof. Denote $S_n = \sum_{k=0}^{n} T^k$. Compute that

$$(I - T)S_n = S_n - S_nT = \sum_{k=0}^{n} T^k - T^{k+1} = I - T^{n+1}.$$

Take the limit as $n \to \infty$:

$$(I - T)S = I - \lim_{n\to\infty} T^{n+1} = I$$

since $\|T\| < 1$ implies that $\lim_{n\to\infty} T^{n+1} = 0$. Thus $(I - T)S = I$. By a similar argument, $S(I - T) = I$. Hence $I - T$ is invertible and $(I - T)^{-1} = S = \sum_{n=0}^{\infty} T^n$. ■

Proposition 5.22 (First Resolvent Identity)
Let $T : D(T) \to X$ *be a closed linear operator. The followings are true.*

(a) For all $\xi_1, \xi_2 \in \rho(T)$,

$$R_T(\xi_1) - R_T(\xi_2) = (\xi_1 - \xi_2)R_T(\xi_1)R_T(\xi_2).$$

(b) For all $\xi \to \xi_0 \in \rho(T)$,

$$\lim_{\xi\to\xi_0} \frac{R_T(\xi) - R_T(\xi_0)}{\xi - \xi_0} = R_T(\xi_0)^2.$$

(c) If $|\xi - \xi_0| < \|R_T(\xi)\|^{-1}$, *then*

$$R_T(\xi) = [I - (\xi - \xi_0)R_T(\xi_0)]^{-1} R_T(\xi_0) = \sum_{n=0}^{\infty} (\xi - \xi_0)^n R_T(\xi_0)^{n+1}.$$

Proof. For (a), write

$$\begin{aligned}
[R_T(\xi_1) - R_T(\xi_2)]\,(T - \xi_2 I) &= (T - \xi_1 I)^{-1}(T - \xi_2 I) - I \\
&= (T - \xi_1 I)^{-1}(T - \xi_1 I) + (T - \xi_1 I)^{-1}(\xi_1 - \xi_2) - I \\
&= (T - \xi_1 I)^{-1}(\xi_1 - \xi_2).
\end{aligned}$$

Rearranging the equation gives

$$R_T(\xi_1) - R_T(\xi_2) = (\xi_1 - \xi_2) R_T(\xi_1) R_T(\xi_2).$$

For (b), using (a),

$$\lim_{\xi \to \xi_0} \frac{R_T(\xi) - R_T(\xi_0)}{\xi - \xi_0} = \lim_{\xi \to \xi_0} R_T(\xi_0) R_T(\xi) = R_T(\xi_0)^2.$$

For (c), (a) implies

$$R_T(\xi) = [I - (\xi - \xi_0) R_T(\xi_0)]^{-1} R_T(\xi_0) = \sum_{n=0}^{\infty} (\xi - \xi_0)^n R_T(\xi_0)^{n+1}$$

since $|\xi - \xi_0| < \|R_T(\xi)\|^{-1}$ by the von Neumann series. ■

Example
$Tu = u'$ on $X = C[a,b]$ with $D(T) = C^1[a,b]$.

$$(T - \xi I)u = 0 \Leftrightarrow u' = \xi u \Leftrightarrow u(x) = Ce^{\xi x}$$

for all $C \in \mathbb{R}$. Thus $(T - \xi I)^{-1}$ does not exists for all $\xi \in \mathbb{C}$. Hence $\rho(T) = \varnothing$ and $\sigma(T) = \mathbb{C}$.

Example
Consider $Tu = u'$ on $X = C[0,1]$ and $D(T) = \{u \in C^1[0,1] \mid u(0) = u(1) = 0\}$. Then

$$\begin{cases} (T - \xi I)u = v, & v \in C[0,1], \\ u(0) = u(1) = 0. \end{cases} \Rightarrow \begin{cases} u(x) = e^{-\xi x} \int_0^x e^{-\xi t} v(t)dt \\ u(1) = 0. \end{cases}$$

Clearly, this is impossible for all $v \in C[0,1]$. Hence

$$\rho(T) = \varnothing \quad \text{and} \quad \sigma(T) = \mathbb{C}.$$

Example
Consider $Tu = u'$ on $X = C[0,1]$ and $D(T) = \{u \in C^1[0,1] \mid u(0) = ku(1)\}$. Solving

$$\begin{cases} u' - \xi u = v, & v \in C[0,1], \\ u(0) = ku(1). \end{cases} \Rightarrow \begin{cases} (e^{-\xi x} u)' = e^{-\xi x} v, \\ u(0) = ku(1). \end{cases}$$

So

$$u(x) = c_1 e^{\xi x} \int_0^x e^{-\xi t} v(t)dt + c_2 e^{\xi x} \int_x^1 e^{-\xi t} v(t)dt,$$

for some $c_1, c_2 \in \mathbb{R}$ that should be determined. The boundary condition gives $c_2 = ke^{\xi} c_1$ and

$$u(x) = c_1 e^{\xi x} \left[\int_0^x e^{-\xi t} v(t) dt + k e^{\xi} \int_x^1 e^{-\xi t} v(t) dt \right].$$

Then

$$(e^{-\xi x} u(x))' = c_1 \left[e^{-\xi x} v(x) - k e^{\xi} e^{-\xi x} v(x) \right] = e^{-\xi x} v(x) \quad \Rightarrow \quad c_1 = \frac{1}{1 - k e^{\xi}}.$$

Thus

$$R_T(\xi) v(x) = \frac{e^{\xi x}}{1 - k e^{\xi}} \left[\int_0^x e^{-\xi t} v(t) dt + k e^{\xi} \int_x^1 e^{-\xi t} v(t) dt \right].$$

We see that

$$\sigma(T) = \left\{ \xi \in \mathbb{C} \,\middle|\, 1 - k e^{\xi} = 0 \right\} = \{ \xi \in \mathbb{C} \mid \xi = -\log k + 2\pi i n, n \in \mathbb{Z} \} \quad \textit{and} \quad \rho(T) = \mathbb{C} \setminus \sigma(T).$$

Definition 5.23

*An operator is said to be with **compact resolvent** if there exists $\xi \in \rho(T)$ such that $R_T(\xi)$ is compact.*

Remark

If T has compact resolvent, then for any $\xi \in \rho(T)$, $R_T(\xi)$ is compact. This is because of the first resolvent identity. If $R_T(\xi)$ is compact, then

$$R_T(\xi) = \left[I + (\xi - \xi_0) R_T(\xi) \right] R_T(\xi_0)$$

is also compact.

Theorem 5.24

$T \in B(X)$. Then $\sigma(T)$ is compact and

$$\sup_{\xi \in \sigma(T)} |\xi| \leq \|T\| < \infty.$$

Proof. $\sigma(T)$ is closed if and only if $\rho(T)$ is open. Take $\xi_0 \in \rho(T)$. Consider the ball

$$B = \left\{ \lambda \in \mathbb{C} \,\middle|\, |\lambda - \xi_0| < \|R_T(\xi_0)\|^{-1} \right\}.$$

For $\lambda \in B$,

$$T - \lambda I = (T - \xi_0 I) + (\xi_0 - \lambda) I = (T - \xi_0 I) \left[I + (\xi_0 - \lambda) R_T(\xi_0) \right].$$

Using the Neumann series, $I + (\xi_0 - \lambda) R_T(\xi_0)$ is invertible since $|\xi_0 - \lambda| \, \|R_T(\xi_0)\| < 1$. Hence $(T - \lambda I)^{-1}$ exists and bounded by the bounded inverse theorem. Then $\lambda \in \rho(T)$. $\rho(T)$ is open and hence $\sigma(T)$ is closed. For arbitrary $\lambda > \|T\|$, the Neumann series shows that $T - \lambda I$ is boundedly invertible. Hence $\lambda \notin \sigma(T)$. Thus

$$\sup_{\xi \in \sigma(T)} |\xi| \leq \|T\| < \infty.$$

Using Heine-Borel theorem, $\sigma(T)$ is compact. ■

Theorem 5.25
Let $T : D(T) \overset{d}{\subset} X \to Y$ be a closed linear operator.

(a) *If T^{-1} exists and is bounded, then $(T')^{-1}$ exists and is bounded, and $(T')^{-1} = (T^{-1})'$.*

(b) *If $(T')^{-1}$ exists and is bounded, then T^{-1} exists and is bounded, and $T^{-1} = (T')^{-1}$.*

Proof. (a) Assume first that T^{-1} exists and is bounded. We first check the identity $(T')^{-1} = (T^{-1})'$. For $g \in D(T')$,

$$(T^{-1})'T'g = (T'g)T^{-1} = gTT^{-1} = g \quad \Rightarrow \quad (T^{-1})'T' = I.$$

For the other side, let $f \in X'$.

$$T'(T^{-1})'f = ((T^{-1})'f)T = f(T^{-1}T) = fI = f. \quad \Rightarrow \quad T'T^{-1} = I.$$

Hence $(T')^{-1} = (T^{-1})'$. Now we show that $(T')^{-1}$ is bounded.

$$\begin{aligned}\left\|(T')^{-1}\right\| &= \sup_{\|f\|=1} \left\|(T')^{-1}f\right\| = \sup_{\|f\|=1}\sup_{\|y\|=1} \left|(T')^{-1}f(y)\right| = \sup_{\|f\|=1}\sup_{\|y\|=1} \left|f(T^{-1}y)\right| \\ &\le \sup_{\|f\|=1}\sup_{\|y\|=1} \|f\| \left\|T^{-1}\right\| \|y\| = \left\|T^{-1}\right\|.\end{aligned}$$

(b) can be shown in a similar way. ■

Theorem 5.26
Let $T : D(T) \overset{d}{\subset} X \to X$ be closed linear operator. Then

(a) *$R_{T'}(\overline{\xi}) = R_T(\xi)'$ for all $\xi \in \rho(T)$.*

(b) *$\rho(T') = \left\{\overline{\lambda} \,\middle|\, \lambda \in \rho(T)\right\}$ and $\sigma(T') = \left\{\overline{\lambda} \,\middle|\, \lambda \in \sigma(T)\right\}$.*

Proof. We first prove (a). Let $\lambda \in \rho(T)$. For all $f \in X'$,

$$\langle \lambda f, x\rangle = \left\langle f, \overline{\lambda}Ix\right\rangle \quad \forall x \in X. \quad \Rightarrow \quad f(\lambda I) = \overline{\lambda}f = \overline{\lambda}If.$$

Thus

$$(T - \lambda I)'f = f(T - \lambda I) = fT - f(\lambda I) = fT - \overline{\lambda}If = (T' - \overline{\lambda}I)f.$$

Hence $(T-\lambda I)' = (T'-\overline{\lambda}I)$. Let $x_n \to x$ in X and $(T-\lambda I)x_n \to y$ in X. x lies in $D(T-\lambda I) = D(T)$. By the closedness of T,

$$(T - \lambda I)x_n = Tx_n - \lambda x_n \to Tx - \lambda x = (T - \lambda I)x.$$

On the other hand, $(T-\lambda I)x_n \to y$ so $(T-\lambda I)x = y$ and $T-\lambda I$ is closed. Since $T-\lambda I$ is closed, densely defined and invertible,

$$R_{T'}(\overline{\lambda}) = (T' - \overline{\lambda}I)^{-1} = ((T - \lambda I)')^{-1} = ((T - \lambda I)^{-1})' = R_T(\lambda)'.$$

For (b), let $\lambda \in \rho(T)$. Then $R_T(\lambda)$ exists and is bounded. Then $R_T(\lambda)' : X' \to X'$ defined by $R_T(\lambda)' f = f R_T(\lambda)$ also exists and

$$\|R_T(\lambda)' f\| = \sup_{\|x\|=1} |f R_T(\lambda) x| \leq \sup_{\|x\|=1} \|f\| \, \|R_T(\lambda)\| \, \|x\| = \|f\| \, \|R_T(\lambda)\| \, ,$$

so $R_T(\lambda)'$ is bounded. It now follows from (b) that $R_{T'}(\overline{\lambda}) = R_T(\lambda)'$ exists and is bounded. Thus $\overline{\lambda} \in \rho(T')$.

Now let $\lambda \in \sigma(T)$. If λ is an eigenvalue, then $T - \lambda I$ is not invertible. Thus from theorem 5.25, $T' - \overline{\lambda} I = (T - \lambda I)'$ is not invertible. Hence $\overline{\lambda} \in \sigma(T')$. If λ is such that $R_T(\lambda)$ exists but is not bounded, then $R_T(\lambda)'$ exists but is not bounded either, since

$$\infty = \|R_T(\lambda) x\| = \sup_{\|f\|=1} |f R_T(\lambda) x| \leq \sup_{\|f\|=1} |(R_T(\lambda)' f) x| \leq \sup_{\|f\|=1} \|R_T(\lambda)' f\| \, \|x\| = \|R_T(\lambda)'\| \, \|x\| \, .$$

From the proof of (b), we have seen that $(T - \lambda I)' = (T' - \overline{\lambda} I)$. Thus by exercise 5.1 (b),

$$R_{T'}(\overline{\lambda}) = (T' - \overline{\lambda} I)^{-1} = ((T - \lambda I)')^{-1} = ((T - \lambda I)^{-1})' = R_T(\lambda)'$$

is not bounded either. Hence $\overline{\lambda} \in \sigma(T')$. It follows that $\rho(T')$ contains the mirror image of $\rho(T)$ and also $\sigma(T')$ contains the mirror image of $\sigma(T)$. Since $\rho(T) \cap \sigma(T) = \varnothing$ and $\rho(T) \cup \sigma(T) = \mathbb{C}$, we conclude that $\rho(T')$ and $\sigma(T')$ are exactly the mirror images of $\rho(T)$ and $\sigma(T)$ with respect to the real axis. ∎

Remark

If $X = \mathcal{H}$, then $T' = T^*$, and if $\lambda \in \sigma(T)$, then $\overline{\lambda} \in \sigma(T^*) = \sigma(T')$.

Lemma 5.27 (Riesz)

Let X be a normed vector space with $\dim X = \infty$. Let Y be a proper closed subspace of X. Then for all $\alpha \in (0,1)$, there exists $x \in X$ with $\|x\| = 1$ such that $\|x - y\| \geq \alpha$ for all $y \in Y$.

Proof. Fix $v \in X \setminus Y$. Let $\beta = \inf_{y \in Y} \|v - y\|$. Since y is closed, $\beta > 0$. For all $\alpha \in (0,1)$, there is a $y_0 \in Y$ such that $\beta \leq \|v - y_0\| \leq \beta/\alpha$. Let $z = \frac{v - y_0}{\|v - y_0\|}$ so $\|z\| = 1$. We claim that $\|z - y\| \geq \alpha$ for all $y \in Y$. Indeed,

$$\|z - y\| = \frac{1}{\|v - y_0\|} \|v - y_0 - \|v - y_0\| \, y\| = \frac{1}{\|v - y_0\|} \|v - (y_0 + \|v - y_0\| \, y)\| \geq \frac{1}{\|v - y_0\|} \beta$$

by the definition of β. Hence,

$$\|z - y\| \geq \frac{\beta}{\|v - y_0\|} \geq \frac{\beta}{\beta/\alpha} = \alpha.$$

Since y is arbitrary, z is the desired vector. ∎

Proposition 5.28

Let $T \in B(X)$ be a compact operator. Then $(T - I)(X)$ is closed.

Proof. Let $x_n \in X$ be a sequence such that $(T-I)x_n \to y$. We first show that $d(x_n, \ker(T-I))$ is bounded. Suppose not. We can find a divergent subsequence, say x_n, and define $z_n = x_n / \|x_n + \ker(T-I)\|_{X/\ker(T-I)}$. Now

$$\|x_n + \ker(T-I)\|_{X/\ker(T-I)} = d(x_n, \ker(T-I))$$

is unbounded. Then

$$(T-I)z_n = \frac{(T-I)x_n}{\|x_n + \ker(T-I)\|_{X/\ker(T-I)}} \to 0.$$

Notice that $z_n = Tz_n - (T-I)z_n$. By the compactness of T, we may choose a subsequence z_{n_k} such that $Tz_{n_k} \to z \in X$ and thus $z_{n_k} \to z$. It follows that $(T-I)z = 0$ and $z \in \ker(T-I)$, so $z + \ker(T-I)$ is a zero vector in $X/\ker(T-I)$. On the other hand, z_n is a sequence of unit vectors in $X/\ker(T-I)$, a contradiction. Hence $d(x_n, \ker(T-I))$ is bounded.

Now for x_n, since $d(x_n, \ker(T-I))$ is bounded, we can find a sequence $y_n \in \ker(T-I)$ such that $x_n - y_n$ is bounded. Since T is compact, we can find a subsequence $x_{n_k} - y_{n_k}$ such that

$$x_{n_k} - y_{n_k} = T(x_{n_k} - y_{n_k}) - (T-I)(x_{n_k} - y_{n_k}) = T(x_{n_k} - y_{n_k}) - (T-I)x_{n_k}$$

is convergent, say to x. Then

$$(T-I)x = \lim_{k\to\infty} (T-I)x_{n_k} = y$$

lies in $(T-I)(X)$. Hence $(T-I)(X)$ is closed. ∎

Theorem 5.29 (Spectral Theorem for Compact Operators)
Let $T \in B(X)$ be a compact operator. Then

(a) Every non-zero $\lambda \in \sigma(T)$ is an eigenvalue of T.

(b) For each non-zero $\lambda \in \sigma(T)$, $\dim(E_\lambda) < \infty$.

(c) $\sigma(T)$ is at most countable.

(d) $\sigma(T)$ has no limit point except possibly 0.

(e) If $\dim(X) = \infty$, then $0 \in \sigma(T)$.

Proof. For (a), let $\lambda \in \sigma(T)$ be non-zero. $T - \lambda I = \lambda(\lambda^{-1}T - I)$. T is compact if and only if $\lambda^{-1}T$ is compact and hence the case reduced to the case where $\lambda = 1$.

Now suppose that $\lambda = 1$. If 1 is not an eigenvalue, then $T - I$ is injective and has no bounded inverse. It follows from the bounded inverse theorem and proposition 5.28 that $(T-I)(X)$ is a proper closed subspace of X.

Put $Y_1 = (T-I)(X)$ and $Y_2 = (T-I)^2(X)$. Since $T - I$ is injective, Y_2 is a proper closed subspace of Y_1. Define $Y_n = (T-I)^n(X)$ for $n \geq 1$. We obtain a sequence of proper closed subspaces

$$Y_1 \supset Y_2 \supset Y_3 \supset \cdots .$$

By the Riesz lemma, we can choose a sequence of unit vectors $y_n \in Y_n$ such that $d(y_n, Y_{n+1}) \geq 1/2$. For $m > n$,

$$\|Ty_m - Ty_n\| = \|(T - I)y_m + y_m - (T - I)y_n - y_n\| \geq d(y_n, Y_{n+1}) \geq \frac{1}{2}.$$

On the other hand, y_n is a bounded sequence and hence Ty_n has a Cauchy subsequence, which is absurd. Hence 1 is an eigenvalue of T. Thus every non-zero $\lambda \in \sigma(T)$ is an eigenvalue of T.

For (b), let $\lambda \in \sigma(T)$ be a non-zero eigenvalue. ■