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 $\left\{A \subset \mathbb{R} \mid f^{-1}(A) \in \mathcal{F}\right\}$ 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

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

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

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

Rearranging gives the desired result. ∎

Theorem 1.17 (Egorov)

Let E be a measurable set with $\mu(E) < \infty$ and $f_n : E \rightarrow \mathbb{R}$ are measurable functions. If $f_n \rightarrow f$ 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 $f_n \rightarrow f$ uniformly on A_ϵ.

Proof. Consider the case where $f_n \rightarrow f$ everywhere on E since $\{x \in E \mid f_n(x) \nrightarrow f(x)\}$ is of measure zero. For each $n, k \in \mathbb{N}$, let $E_k^n = \{x \in E \mid |f_j(x) - f(x)| < 1/n \text{ for all } j > k\}$. Then fix n and note that $E_k^n \nearrow E$ as $k \rightarrow \infty$. By proposition 1.16, there exists k_n such that $\mu(E - E_{k_n}^n) < 2^{-n}$. Then we have $|f_j(x) - f(x)| < 1/n$ for every $j > k_n$ and $x \in E_{k_n}^n$. Choose N such that $\sum_{n \geq N} 2^{-n} < \epsilon/2$ and let $\hat{A}_\epsilon = \bigcap_{n \geq N} E_{k_n}^n$. Then $\mu(E - \hat{A}_\epsilon) \leq \sum_{n \geq N} \mu(E - E_{k_n}^n) < \epsilon/2$. Also, for any $\delta > 0$, we may pick $n > N$ with $1/n < \delta$ and for $x \in \hat{A}_\epsilon$, $|f_j(x) - f(x)| < \delta$ whenever $j > k_n$. Hence $f_n \rightarrow f$ 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 function $s : X \rightarrow Y$ is called **simple** if it only takes finitely many values.*

Lemma 1.19

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

Proof. Consider $s_n = \sum_{k=0}^{n2^n - 1} k 2^{-n} \chi_{f^{-1}([k2^{-n}, (k+1)2^{-n}))} + n\chi_{f^{-1}([n, \infty])}$. 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 $s_n \rightarrow f$ uniformly. ∎

Theorem 1.20 (Lusin)

Let $E \subset \mathbb{R}$ be a set of finite measure and $f : E \rightarrow \mathbb{R}$ 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 $E_i = E \cap [-i, i]$. 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, $F_j \cap E_i$ are compact and hence f being constant on each $F_j \cap E_i$ is continuous. Note that $F_\epsilon = \bigcup_{i,j=1}^N F_j \cap E_i$ 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 1.19. Now by Egorov's theorem, we may find a closed set $F_\epsilon \subset E$ such that $\mu(E - F_\epsilon) < \epsilon$ and $s_n \rightarrow f$ 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 1.21

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 $B = (E \cap B) \bigcup (B - E)$ and since the later set has outer measure zero, $E \cap B$ is measurable. And since $E - B$ is outer measure zero, $E \cap B = E$ is measurable. ∎

Proposition 1.22

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 1.23

*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 1.24

*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 1.25

*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 1.26

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 1.27 (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 1.28 (Bounded Covergence Theorem)

Suppose $\mu(X) < \infty$. Let $f_n : X \to \mathbb{R}_+$ be a sequence measurable functions such that $f_n \le M$ 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

$$\left| \int f_n d\mu - \int f \, d\mu \right| \leq \int_X |f_n - f| \, d\mu$$

$$= \int_F |f_n - f| \, d\mu + \int_{X-F} |f_n - f| \, d\mu$$

$$\leq \epsilon \mu(F) + 2M\mu(X - F) = \epsilon(\mu(F) + 2M\epsilon).$$

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 1.29 (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 1.30 (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,

$$\int 2g \, d\mu = \int \liminf_n h_n d\mu \leq \liminf_n \int h_n d\mu = \liminf_n \int 2g - |f_n - f| \, d\mu$$

$$= \int 2g \, d\mu - \limsup_n \int |f_n - f| \, d\mu.$$

It follows that
$$0 \le \liminf_{n} \int |f_n - f| \, d\mu \le \limsup_{n} \int |f_n - f| \, d\mu \le 0.$$

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

By the triangle inequality,
$$\left| \int f \, d\mu - \int f_n d\mu \right| \le \int |f - f_n| \, d\mu \to 0.$$

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

∎

Remark

If $\mathrm{supp}(f)$ *has finite measure and* f *is bounded, then*
$$\int f = \inf_{s \ge f} \int s \, d\mu,$$

where s *is simple.*

Definition 1.31

$\mathcal{L}^1 = \{f : X \to \mathbb{R} \mid f \text{ is integrable}\}$ *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 1.32

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

Proof. Let $E_n = \{x \in X \mid f(x) > n\}$. Then by Lebesgue dominated convergence theorem, since $f \chi_{E_n} \le f$,
$$\int_{E_n} f \, d\mu \to 0.$$

For any $\epsilon > 0$, there exists n such that
$$\int_{E_n} f \, d\mu \le \frac{\epsilon}{2}.$$

Pick $\delta \le \epsilon/(2n)$. Then for any measurable E with $\mu(E) \le \delta$,
$$\int_E f \, d\mu = \int_{E \cap E_n} f \, d\mu + \int_{E \cap E_n^c} f \, d\mu \le \int_{E_n} f \, d\mu + n\mu(E) \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

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

Theorem 1.33 (Lebesgue-Vitali)

$f : X \rightarrow \mathbb{R}$ *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

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

where U is open.

We first claim that f is continuous at x if and only if $\text{osc}(f, x) = 0$. 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 $\text{diam}(f(B_\delta(x))) \leq 2\epsilon$. Since ϵ is arbitrary, $\text{osc}(f, x) = 0$. Conversely, if $\text{osc}(f, x) = 0$, then $\forall \epsilon > 0$, $\exists$ open U containing x such that $\text{diam}(f(U)) < \epsilon$. 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 $\text{osc}(f, x) \geq \epsilon > 0$. We claim that D_ϵ is closed. For any convergent sequence $x_k \in D_\epsilon$, let $x_k \rightarrow x$. 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 $\text{diam}(f(U')) \geq \epsilon$. Hence $\text{osc}(f, x) \geq \epsilon$ 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 $\text{U}(f, \mathcal{P}) - \text{L}(f, \mathcal{P}) < \frac{1}{n}$ and $\|\mathcal{P}\| < \frac{1}{n}$. Then

$$\sum_{\substack{Q \in \mathcal{P}, \\ Q \cap D_{\frac{1}{n}} \neq \varnothing}} (\sup_Q f - \inf_Q f)\,|Q| + \sum_{\substack{Q \in \mathcal{P}, \\ Q \cap D_{\frac{1}{n}} = \varnothing}} (\sup_Q f - \inf_Q f)\,|Q|$$

$$= \sum_{Q \in \mathcal{P}} (\sup_Q f - \inf_Q f)\,|Q| = \text{U}(f, \mathcal{P}) - \text{L}(f, \mathcal{P}) < \epsilon.$$

Note that $\sup_Q f - \inf_Q f = \text{diam}(f(Q))$. 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

$$\text{U}(f, \mathcal{P}) - \text{L}(f, \mathcal{P}) = \sum_{Q \in \mathcal{P}} (\sup_Q f - \inf_Q f)\,|Q|$$

$$= \sum_{\substack{Q \in \mathcal{P}, \\ Q \cap D_\epsilon = \varnothing}} (\sup_Q f - \inf_Q f)\,|Q| + \sum_{\substack{Q \in \mathcal{P}, \\ Q \cap D_\epsilon \neq \varnothing}} (\sup_Q f - \inf_Q f)\,|Q|$$

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 1.34

(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 1.19, 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 1.35
*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 1.36
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 1.37

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

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 1.38 (Weak-Type Inequality)

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

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

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 subcollec-

tion $\{B_{i_1}, \ldots, B_{i_k}\}$ with

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

As a result,

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

$$\le \frac{3^d}{\alpha} \int_{\cup_{j=1}^{k} B_{i_j}} |f(y)|\, dy \le \frac{3^d}{\alpha} \int_{\mathbb{R}^d} |f(y)|\, dy.$$

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 1.39 (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

$$\left| \frac{1}{m(B)} \int_B f(y)dy - f(x) \right| = \left| \frac{1}{m(B)} \int_B (f(y) - g(y))dy + \frac{1}{m(B)} \int_B g(y) - g(x)dy + g(x) - f(x) \right|$$

$$\le \frac{1}{m(B)} \int_B |(f(y) - g(y))|\, dy + \frac{1}{m(B)} \int_B |g(y) - g(x)|\, dy + |g(x) - f(x)|$$

$$\le (f - g)^*(x) + \epsilon + |g(x) - f(x)|.$$

Since ϵ can be arbitrary small, we have

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

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,

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

Thus $\mu(E_\alpha) \le \mu(F_\alpha \cup G_\alpha) < \frac{A+1}{\alpha}\epsilon$. 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,

$$\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| = \frac{1}{h}\left| \int_x^{x+h} f(y) - f(x)dy \right| \le \frac{1}{h}\int_x^{x+h} |f(y) - f(x)|\, dy$$

$$\le \frac{1}{h}\int_{x-h}^{x+h} |f(y) - f(x)|\, dy \le 2\frac{1}{2h}\int_{x-h}^{x+h} |f(y) - f(x)|\, dy \to 0$$

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 1.40

*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 1.41

*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 1.42

*Let ν, λ be two measures defined on a measurable space. ν is said to be **absolutely contin-***

uous 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 1.43

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*

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

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

Definition 1.44

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

Definition 1.45

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 1.46 (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 1.47 (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 1.48 (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

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

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

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

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,

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

17

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 $\nu_0(A) \geq \epsilon\lambda(A)$ for each $A \subset P$. Thus

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

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

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

posing a contradiction. Meanwhile,

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

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 $\lambda(X_n) <$

∞ 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 1.49

*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 1.50

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

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

19

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 1.51 (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 1.50,

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

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 $\nu_s(E^c) = \nu(\emptyset) = 0$, $\nu_s \perp \lambda$. 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 1.52
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. $\blacksquare$

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 Bnach 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 $f \in V$,

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

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,

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

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 $\left\|T_i(f) - T_j(f)\right\| \leq \epsilon$ for all $i, j \geq N$. Then for fixed i,

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

for every $f \in V$ and $j \geq N$. Hence $\|T_i - T\| \leq \epsilon$ for all $i \geq N$ and the proof is complete. ∎

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 $\|T f_i\| \to \infty$. Thus

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

Hence T is not continuous.

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

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

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

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

Example

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

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

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

$$|Tg(f)| = \left| \int_0^1 f(t)g(t)dt \right| \le \int_0^1 |f(t)g(t)| \, dt \le \int_0^1 |g(t)| \sup_{[0,1]} |f(t)| \, dt$$

$$= \|f\|_\infty \int_0^1 |g(t)| \, dt \le \int_0^1 |g(t)| \, dt.$$

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

*$\ell^p = \left\{ \{x_i\}_{i \in \mathcal{I}} \mid \|x\|_p < \infty \right\}$, 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 $a, b \geq 0$,

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

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

$$\log(ab) = \log(a) + \log(b) = t \log(a^p) + (1 - t) \log(b^{p'}) \leq \log\left(ta^p + (1 - t)b^{p'}\right)$$

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 $1 \leq p, p' \leq \infty$ with $\frac{1}{p} + \frac{1}{p'} = 1$. Then for all $f \in \ell^p$ and $g \in \ell^{p'}$,

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

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 $p' = \infty$, $|f_i g_i| \leq \|g\|_\infty |f_i|$. 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,

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

Thus

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

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.

Theorem 2.18 (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 $g \in \ell^{p'}$,

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

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,

$$\|\psi g\| = \sup\left\{|\psi g f| \ \big| \ \|f\|_p = 1\right\} = \sup\left\{\left|\sum_i f_i g_i\right| \ \bigg| \ \|f\|_p = 1\right\}$$

$$\leq \sup_{\|f\|_p = 1}\left\{\|g\|_{p'}\right\} \leq \|g\|_{p'}.$$

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. ∎

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

Definition 2.19

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 2.20

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

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

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

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

Theorem 2.21 (Hölder's Inequality in $\mathcal{L}^p$)
Let $1 \le p, p' \le \infty$ with $\frac{1}{p} + \frac{1}{p'} = 1$. Then for all $f \in \mathcal{L}^p$ and $g \in \mathcal{L}^{p'}$,

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

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

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

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,

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

Integrating both sides yields

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

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. $\blacksquare$

Theorem 2.22 (Minkowski's Inequality in $\mathcal{L}^p$)
Let $1 \le p \le \infty$. Then for all $f, g \in \mathcal{L}^p$,

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

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

$$\|f + g\|_p^p = \int |f + g|^p \, d\mu = \int |f + g| \, |f + g|^{p-1} \, d\mu$$

$$\leq \int |f| \, |f + g|^{p-1} \, d\mu + \int |g| \, |f + g|^{p-1} \, d\mu$$

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

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

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 2.23

$1 \leq p \leq \infty$. *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.

Lemma 2.24

$1 \leq p < \infty$. $g_k \in \mathcal{L}^p$ 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 2.25 (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 $\sum_j \|g_j\|_p < \infty$. By the lemma 2.24, 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. ∎

Proposition 2.26
$1 \leq p < \infty$. $1/p + 1/p' = 1$. Let $g \in \mathcal{L}^{p'}(X, \mu)$. Then the mapping $Tg : \mathcal{L}^p(X, \mu) \to \mathbb{R}$ defined by

$$Tg(f) = \int_X f g \, d\mu$$

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 $f \in \mathcal{L}^p$,

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

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

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

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

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

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

Theorem 2.27 (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 2.26, 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 $B_n \cap C_n = \emptyset$,

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

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 1.19, 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

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

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

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

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

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

As a result,

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

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 $Tg(f) = \int f g \, d\mu$. 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 $f_K = \text{sgn}(g)\chi_{A_K}/\mu(A_K)$. Note that $\|f_K\|_1 = 1$. If $g \geq 0$, 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 $\|h_k\| \leq \|h\|$. By the previous result, we can find a unique $g_k \in \mathcal{L}^{p'}(X_k, \mu)$ such that

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

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

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

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 $\delta(f) = \int_{-1}^1 fg \, dx$. Let $f = \chi_A$ where $A = [-1, 1] - [-\epsilon, \epsilon]$ for some small $\epsilon > 0$. Then $f \in \mathcal{L}^\infty$ and by definition,

$$0 = f(0) = \delta(f) = \int_{-1}^1 fg \, dx = \int_A g \, dx.$$

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

Definition 2.28

$M(X)$ 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 2.29

$M(X)$ 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 $\|\nu\| \geq 0$. If $\|\nu\| = 0$, then $\nu^+(X) = \nu^-(X) = 0$, $\nu^+(A)$ and $\nu^-(A)$ are zero for all $A \in \mathcal{A}$,

and hence $v = 0$. Conversely, if $v = 0$, then so are v^+ and v^- and hence $\|v\| = 0$. For $c \in \mathbb{R}$,

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

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

$$\|v + \mu\| = (v + \mu)^+(X) + (v + \mu)^-(X) \leq v^+(X) + \mu^+(X) + v^-(X) + \mu^-(X) = \|v\| + \|\mu\|\,,$$

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

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

$$(v_m - v_n)^+(X) + (v_m - v_n)^-(X) = \|v_m - v_n\| \leq \epsilon$$

for all $m, n \geq N$. Since both $(v_m - v_n)^+$ and $(v_m - v_n)^-$ are positive measures, we have

$$(v_m - v_n)^+(A) \leq (v_m - v_n)^+(X) \leq \epsilon, \quad \text{and} \quad (v_m - v_n)^-(A) \leq (v_m - v_n)^-(X) \leq \epsilon$$

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

$$|v_m(A) - v_n(A)| = \left|(v_m - v_n)^+(A) - (v_m - v_n)^-(A)\right| \leq \epsilon.$$

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

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

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 $v(X) < \infty$, $\sum_j v(A_j) < \infty$ and hence $v(C_n) \to 0$ as $n \to \infty$. Thus

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

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

$$\|v - v_n\| = \lim_{m \to \infty} \|v_m - v_n\| = \lim_{m \to \infty} |v_m(X) - v_n(X)| = |v(X) - v_n(X)| \leq \epsilon$$

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

Definition 2.30

*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 2.31

*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 2.32

$BV([a, b])$ 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

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

$$\leq \sup_{\mathcal{P}} |f(a)| + |g(a)| + \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)| + \sum_{i=0}^{n-1} |g(t_{i+1}) - g(t_i)|$$

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

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 $m, n \geq N$. Given any $x \in [a, b]$, consider the partition $\mathcal{P} = \{a < x < b\}$.

$$|f_m(x) - f_n(x)| = |f_m(x) - f_m(a) + f_m(a) - f_n(a) + f_n(a) - f_n(x)|$$

$$\leq |((f_m(x) - f_n(x))) - (f_m(a) - f_n(a))| + |f_m(a) - f_n(a)|$$

$$\leq V(f_m - f_n) + |f_m(a) - f_n(a)| = \epsilon.$$

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 $\mathcal{P} = \{a = t_0 < \cdots < t_n = b\}$,

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

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

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

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 $m, n \geq N$ and some partition $\mathcal{P}$. Taking $m \to \infty$, we obtain $V_{\mathcal{P}}(f - f_n) < \epsilon$ for all $n \geq N$. Since the partition is arbitrary, we have $V(f - f_n) < \epsilon$ for all $n \geq N$. Thus $f_n \to f$ in $BV([a, b])$ and $BV([a, b])$ is complete. ∎

Theorem 2.33

$M([a, b])$ *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 $\mathcal{P} = \{a = t_0 < \cdots < t_n = b\}$,

$$\sum_{i=0}^{n-1} |\rho(t_{i+1}) - \rho(t_i)| + |\rho(a)| = \sum_{i=0}^{n-1} |\nu([a, t_{i+1}]) - \nu([a, t_i])| + |\nu(\{a\})|$$

$$= \sum_{i=0}^{n-1} |\nu((t_i, t_{i+1}])| + |\nu(\{a\})|$$

$$= \sum_{i=0}^{n-1} |\nu|\,((t_i, t_{i+1}]) + |\nu|\,(\{a\}) = |\nu|\,([a, b]) = \|\nu\|.$$

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 2.34

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 $w, v \in M$,

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

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

$$S(w) - \|S\| \|w + x\| \leq -\lambda \leq S(w) + \|S\| \|w + x\|$$

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

$$|Sw + \lambda| \leq \|S\| \|w + x\|$$

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 2.35 (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.

---

[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.

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 2.34. 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. ∎

Theorem 2.36 (Riesz Representation of $C([a, b])$)
$C([a, b])' \cong BV([a, b]) \cong M([a, b])$ *isometrically.*

Proof. In theorem 2.33, 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 $\mathcal{P} = \{a = t_0 < \cdots < t_n = b\}$,

$$V_{\mathcal{P}}(\rho) = \sum_{i=0}^{n-1} |\rho(t_{i+1}) - \rho(t_i)| = \sum_{i=0}^{n-1} \left| L(\chi_{[a,t_{i+1}]}) - L(\chi_{[a,t_i]}) \right|$$

$$= \sum_{i=0}^{n-1} L(\chi_{(t_i,t_{i+1}]}) s_i = L\left(\sum_{i=0}^{n-1} \chi_{(t_i,t_{i+1}]} s_i \right) \leq \|L\| \left\| \sum_{i=0}^{n-1} \chi_{(t_i,t_{i+1}]} s_i \right\|_{\infty} \leq \|L\|$$

by letting $s_i = \mathrm{sgn}(\rho(t_{i+1}) - \rho(t_i))$. Thus $\rho \in BV([a, b])$ and $\|\rho\|_{TV} \leq \|L\| = \|\ell\|$. 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 $\|\ell_\rho\| = \|\rho\|_{TV}$.

2.4. Weak Convergence

Definition 2.37

*Let $(X, \|\cdot\|)$ be a normed space. A sequence $\{x_n\}$ in X is said to **converge weakly** to $x \in X$, denoted by $x_n \rightharpoonup x$, 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 $x_n \rightharpoonup x$. 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 $x_n \rightharpoonup 0$ 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

$$x_n(t) = \begin{cases} nt & \text{if } 0 \leq t \leq 1/n, \\ 2 - nt & \text{if } 1/n \leq t \leq 2/n, \\ 0 & \text{if } 2/n \leq t \leq 1. \end{cases}$$

Then $\|x_n\|_\infty = 1$ and thus $x_n \not\to 0$. Instead, we have $x_n \rightharpoonup 0$. Assume not, then we can find $T \in X'$ and a subsequence $\{x_{n_k}\}$ such that $|T(x_{n_k})| \geq \delta > 0$. For simplicity, we consider the case $T(x_{n_k}) \geq \delta$, but the other case is similar. Since $T \in X'$, $|T(x_{n_k})| \leq \|T\|_{X \to \mathbb{R}} \|x_{n_k}\|_\infty$. Let $y_K = \sum_{k=1}^K x_{n_k}$. Then $T(y_K) = \sum_{k=1}^K T(x_{n_k}) \geq K\delta$ and $T(y_K) \leq \|T\|_{X \to \mathbb{R}} \|y_K\|_\infty$. This implies that y_K cannot be bounded. Now consider x_{n_k} with $n_{k+1} \geq 2n_k$. For $t \in [0, 1/n_K]$, $x_{n_k}(t) = n_k t$.

$$y_K(t) = \sum_{k=1}^K n_k t \leq \sum_{k=1}^K n_k/n_K \leq 1 + \sum_{k=1}^K 2^{K-k} \leq 1 + \sum_k 2^{-k} = 2.$$

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 $x_n \rightharpoonup 0$.

Proposition 2.38

$(X, \|\cdot\|_X)$ 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 $x_n \rightharpoonup x$ 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

$$|T(x_n) - T(x)| \le |T(x_n) - L(x_n)| + |L(x_n) - L(x)| + |L(x) - T(x)|$$

$$\le \|T - L\|_{X' \to \mathbb{R}} (\|x_n\|_X + \|x\|_X) + |L(x_n) - L(x)| \le 2C\epsilon + \epsilon$$

for all $n \le N$. Since ϵ is arbitrary, $x_n \rightharpoonup x$. ∎

Definition 2.39

*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 2.40 (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 2.41 (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

$$\sup_{\alpha \in A} |f_\alpha(x)| \leq C$$

for all $x \in G$.

Proof. By Baire Category Theorem, X is a Baire space. For each n, let

$$F_n = \left\{ x \in X \ \middle| \ \sup_{\alpha \in A} |f_\alpha(x)| \leq n \right\}.$$

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 $\alpha \in A$, $|f_\alpha(x_k)| \leq n$ for all k and by continuity of f_α,

$$|f_\alpha(x)| = \lim_{k \to \infty} |f_\alpha(x_k)| \leq n.$$

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, $\emptyset = X^c = (\cup_n F_n)^c = \cap_n F_n^c \neq \emptyset$ 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 2.42
*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 2.43 (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 2.41, 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 $\sup_{\alpha \in A} |f_\alpha(z + y)| \le C$ for all $y \in B_r(0)$. Take y with $\|y\| \le r/2$. Then

$$-2C \le f_\alpha(y + z) - f_\alpha(z) \le f_\alpha(y) \le f_\alpha(y + z) + f_\alpha(-z) = f_\alpha(y + z) + f_\alpha(z) \le 2C.$$

Hence $|f_\alpha(y)| \le 2C$ for all y with $\|y\| \le r/2$. Take $x \in X$.

$$|f_\alpha(x)| = \left| f_\alpha \left(\frac{x}{\|x\|} \frac{r}{2} \frac{2}{r} \|x\| \right) \right| = \frac{2}{r} \|x\| \, |f_\alpha(y)| \le \frac{4C}{r} \|x\|.$$

Thus

$$\sup_{\alpha \in A} |f_\alpha(x)| \le \frac{4C}{r} \|x\|$$

for all $x \in X$. ∎

Corollary 2.44

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 $\sup_{\alpha \in A} |L_\alpha(x)| \le M(x)$, then there exists a constant $C < \infty$ such that $\sup_{\alpha \in A} \|L_\alpha\| \le C$.

Proof. Apply theorem 2.43 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 2.43. ∎

Corollary 2.45

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 $\sup_\alpha |L(x_\alpha)| \le M(L)$ and $(X', \|\cdot\|_{X \to \mathbb{R}})$ is a Banach space. Then there exists $C < \infty$ such that $\|x_\alpha\|_X \le C$ for all $\alpha \in A$.

Proof. Apply the theorem 2.43 to $f_\alpha(L) = |L(x_\alpha)|$. First, for $L, T \in X'$,

$$f_\alpha(L + T) = |L(x_\alpha) + T(x_\alpha)| \le |L(x_\alpha)| + |T(x_\alpha)| = f_\alpha(L) + f_\alpha(T).$$

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',

$$|f_\alpha(L_n) - f_\alpha(L)| = |L_n(x_\alpha) - L(x_\alpha)| \le \|L_n - L\|_{X' \to \mathbb{R}} \|x_\alpha\|_X \to 0$$

for each $\alpha \in A$. The conclusion follows from theorem 2.43. ∎

Corollary 2.46

X is a normed space and $x_n \in X$ with $x_n \rightharpoonup x$ in X. Then there exists $C < \infty$ such that $\|x_n\|_X \le C$ for all n.

Proof. This is a direct consequence of corollary 2.45 with $A = \mathbb{N}$. ∎

Proposition 2.47

Let $f_n \in \mathcal{L}^p(X, \mu)$ and $1 \le p < \infty$. Then $f_n \rightharpoonup f \in \mathcal{L}^p$ 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 $f_n \rightharpoonup f$. ∎

Proposition 2.48

$f_n \in \mathcal{L}^p(X, \mu)$ and $1 \le p < \infty$. If $f_n \rightharpoonup f$ in $\mathcal{L}^p$, then f_n is bounded and

$$\|f_n\|_p \le \liminf_{n\to\infty} \|f_n\|_p .$$

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 2.46, f_n is bounded uniformly in n. ∎

Proposition 2.49

$1 \le p < \infty$ and $1/p + 1/p' = 1$. Suppose $f_n \rightharpoonup f$ 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,

$$\left| \int f_n g_n \, d\mu - \int f g \, d\mu \right| \le \left| \int f_n (g_n - g) \, d\mu \right| + \left| \int (f_n - f) g \, d\mu \right|$$

$$\le \|f_n\|_p \|g_n - g\|_{p'} + \|f_n - f\|_p \|g\|_{p'} .$$

Note that by proposition 2.48, 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 $f_n \rightharpoonup f$ in $\mathcal{L}^p$ and $g_n \rightharpoonup g$ 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 $f_n \rightharpoonup 0$, 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 $f_n \rightharpoonup 0$ in $\mathcal{L}^2([0,\pi])$. However, f_n does not converge to 0 strongly in $\mathcal{L}^2([0,\pi])$ since $\|f_n\|_2 = 1 \ne 0$ for all n.

Proposition 2.50

$1 \le p < \infty$. Let $f_n \in \mathcal{L}^p(X, \mu)$ be a bounded sequence of functions. Then $f_n \rightharpoonup f$ 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.

$$f_n \rightharpoonup f \iff \int f_n g\, d\mu \to \int f g\, d\mu \text{ for all } g \in \mathcal{L}^{p'}$$

$$\iff \int_A f_n s\, d\mu \to \int_A f s\, d\mu \text{ for all simple } s \in \mathcal{L}^{p'}$$

$$\iff \int_A f_n d\mu = \int f_n \chi_A d\mu \to \int f \chi_A d\mu = \int_A f d\mu$$

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 2.51

$1 < p < \infty$. Let $f_n \in \mathcal{L}^p(X, \mu)$ be a sequence with $\|f_n\|_p \le M$ and $f_n \to f$ pointwise a.e. Then $f_n \rightharpoonup f$ 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 2.50, 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 1.32, 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,

$$\begin{aligned}
\int_A |f_n - f| \, d\mu &\leq \int_{F_\epsilon} |f_n - f| \, d\mu + \int_{A-F_\epsilon} |f_n - f| \, d\mu \\
&\leq \int_A \epsilon \, d\mu + \int_{A-F_\epsilon - E} |f_n - f| \, d\mu + \int_E |f_n - f| \, d\mu \\
&\leq \epsilon\mu(A) + \mu(A - F_\epsilon) + \int_{A-F_\epsilon} |f_n - f|^p \, d\mu \leq \epsilon\mu(A) + \epsilon + \epsilon.
\end{aligned}$$

Hence $f_n \rightharpoonup f$. $\blacksquare$

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,

$$\int_0^1 f_n(x) \, dx = 1 \neq 0 = \int_0^1 0 \, dx.$$

Thus f_n does not converge weakly to 0 in $\mathcal{L}^1$.

Theorem 2.52 (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 $f_n \rightharpoonup f$ 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 $\|f_n\|_p \leq \|f_n - f\|_p + \|f\|_p$ and thus

$$\left| \|f_n\|_p - \|f\|_p \right| \leq \|f_n - f\|_p \to 0$$

by the strong convergence. Conversely, suppose that $\|f_n\|_p \to \|f\|_p$ and $f_n \rightharpoonup f$ in $\mathcal{L}^p$.

Assume $p \geq 2$. For any $y \in \mathbb{R}$, notice that $|1 + y|^p \geq 1 + py + c\,|y|^p$ 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,

$$\left|\frac{f_n}{f}\right|^p \geq 1 + p\left(\frac{f_n - f}{f}\right) + c\left|\frac{f_n - f}{f}\right|^p$$

Thus

$$|f_n|^p \geq |f|^p + p(f_n - f)\,|f|^{p-1}\,\mathrm{sgn}(f) + c\,|f_n - f|^p\,.$$

Rearranging the inequality and integrating both sides on E^c gives

$$c\int_{E^c} |f_n - f|^p\,d\mu \leq \int_{E^c} |f_n|^p - |f|^p\,d\mu - p\int_{E^c} |f|^{p-1}\,\mathrm{sgn}(f)(f_n - f)d\mu$$

Note that as shown in the proof of proposition 2.48, $|f|^{p-1}\,\mathrm{sgn}(f) \in \mathcal{L}^{p'}$. 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

$$\int_{E_n^c} |f_n - f|^p\,d\mu \leq \frac{1}{c}\int_{E_n^c} |f_n|^p - |f|^p\,d\mu - \frac{p}{c}\int_{E_n^c} |f|^{p-1}\,\mathrm{sgn}(f)(f_n - f)d\mu$$

as $n \to \infty$. On E_n, replacing z by $(f_n - f)/f$,

$$\left|\frac{f_n}{f}\right|^p \geq 1 + p\frac{f_n - f}{f} + c'\left(\frac{f_n - f}{f}\right)^2 \implies |f_n|^p \geq |f|^p + p(f_n - f)\,|f|^{p-1}\,\mathrm{sgn}(f) + c'\,|f_n - f|^2\,|f|^{p-2}$$

for some $c' > 0$. Thus

$$\int_{E_n} |f_n - f|^2\,|f|^{p-2}\,d\mu \leq \frac{1}{c'}\int_{E_n} |f_n|^p - |f|^p\,d\mu - \frac{p}{c'}\int_{E_n} |f|^{p-1}\,\mathrm{sgn}(f)(f_n - f)d\mu.$$

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

$$\int_{E_n} |f_n - f|^p \, d\mu \leq \int_{E_n} |f_n - f| \, |f|^{p-1} \, d\mu \leq \left(\int_{E_n} |f_n - f|^2 \, |f|^{p-2} \, d\mu \right)^{1/2} \left(\int_{E_n} |f|^p \, d\mu \right)^{1/2}$$

$$\leq \left(\int_{E_n} |f_n - f|^2 \, |f|^{p-2} \, d\mu \right)^{1/2} \|f\|_p^{p/2} \to 0.$$

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 $f_n \rightharpoonup f$ 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.

2.5. Open Mapping Theorem and Closed Graph Theorem

Proposition 2.53

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 $G \subseteq F_n$.

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 $U \cap G_n = \varnothing$. This implies $U \subseteq F_n$. ∎

Theorem 2.54 (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 2.53, 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 $\|y - T(\sum_{k=1}^n x_k)\| < c_n$. 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 2.55 (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 $z = \frac{2}{r}\|z\| Tx$,

$$T^{-1}z = \frac{2}{r}\|z\| x \implies \|T^{-1}z\| \le \frac{2}{r}\|z\| \|x\|.$$

Note that $\|x\| \leq 1$. We see that $\left\|T^{-1}\right\|$ is bounded by $2/r$. $\blacksquare$

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 2.56

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 2.57

*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 2.58 (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,

$$\|S(x, Tx)\|_X = \|x\|_X \leq \|x\|_X + \|Tx\|_Y = \|(x, Tx)\|_\Gamma.$$

Thus $\|S\| \leq 1$. 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. $\blacksquare$

Definition 2.59

*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 2.60

*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 2.61

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 $x = y$. $\blacksquare$

Proposition 2.62

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, $\|x\|_2 = \|Ix\|_2 \leq c_1 \|x\|_1$ for some $c_1 > 0$. Applying the same argument exchanging the roles of $\|\cdot\|_1$ and $\|\cdot\|_2$ gives $\|x\|_1 \leq c_2 \|x\|_2$ for some $c_2 > 0$. Hence

$$\frac{1}{c_2} \|x\|_1 \leq \|x\|_2 \leq c_1 \|x\|_1$$

and the norms are equivalent. $\blacksquare$

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

$$0 \leq \left\langle t \frac{|\langle x, y \rangle|}{\langle x, y \rangle} x + y, \ t \frac{|\langle x, y \rangle|}{\langle x, y \rangle} x + y \right\rangle$$

$$= t^2 \|x\|^2 + 2t \Re \left(\frac{|\langle x, y \rangle|}{\langle x, y \rangle} \langle x, y \rangle \right) + \|y\|^2 = t^2 \|x\|^2 + 2t \, |\langle x, y \rangle| + \|y\|^2 \, .$$

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

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \|x\|^2 + 2\mathfrak{R}(\langle x, y \rangle) + \|y\|^2,$$
$$\|x - y\|^2 = \langle x - y, x - y \rangle = \|x\|^2 - 2\mathfrak{R}(\langle x, y \rangle) + \|y\|^2.$$

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,

$$|\langle x, y \rangle| \leq \|x\| \|y\| \implies \left| \left\langle x, \frac{y}{\|y\|} \right\rangle \right| \leq \|x\| \implies \sup_{\|y\|=1} |\langle x, y \rangle| \leq \|x\|.$$

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

$$|\langle x_n, y_n \rangle - \langle x_m, y_m \rangle| \leq |\langle x_n, y_n \rangle - \langle x_n, y_m \rangle| + |\langle x_n, y_m \rangle - \langle x_m, y_m \rangle|$$
$$\leq \|x_n\| \|y_n - y_m\| + \|x_n - x_m\| \|y_m\| \to 0$$

as $n, m \to \infty$ 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

$$|\langle x_n^1, y_n^1 \rangle - \langle x_n^2, y_n^2 \rangle| \leq |\langle x_n^1, y_n^1 \rangle - \langle x_n^1, y_n^2 \rangle| + |\langle x_n^1, y_n^2 \rangle - \langle x_n^2, y_n^2 \rangle|$$
$$\leq \|x_n^1\| \|y_n^1 - y_n^2\| + \|x_n^1 - x_n^2\| \|y_n^2\| \to 0.$$

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

$$\langle cx + y, z \rangle_* = \lim_{n \to \infty} \langle cx_n + y_n, z_n \rangle = \lim_{n \to \infty} c \langle x_n, z_n \rangle + \langle y_n, z_n \rangle$$
$$= c \lim_{n \to \infty} \langle x_n, z_n \rangle + \lim_{n \to \infty} \langle y_n, z_n \rangle = c \langle x, z \rangle_* + \langle y, z \rangle_* \, .$$

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 1.34, 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,

$$2\left(\|x - y_n\|^2 + \|x - y_m\|^2\right) = \|2x - y_n - y_m\|^2 + \|y_n - y_m\|^2$$

Rearranging gives

$$\begin{aligned}
\frac{1}{4}\|y_n - y_m\|^2 &= \frac{1}{2}\|x - y_n\|^2 + \frac{1}{2}\|x - y_m\|^2 - \left\|x - \frac{y_n + y_m}{2}\right\|^2 \\
&\leq \frac{1}{2}(d(x, K) + \epsilon)^2 + \frac{1}{2}(d(x, K) + \epsilon)^2 - d(x, K)^2 \\
&= \epsilon^2 + 2\epsilon d(x, K)
\end{aligned}$$

for all $m, n \geq N$ 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,

$$4d(x, K)^2 = 2\|x - z_1\|^2 + 2\|x - z_2\|^2 = \|z_1 - z_2\|^2 + \|2x - z_1 - z_2\|^2$$
$$= \|z_1 - z_2\|^2 + 4\left\|x - \frac{z_1 + z_2}{2}\right\|^2.$$

Hence
$$\|z_1 - z_2\|^2 = 4d(x, K)^2 - 4\left\|x - \frac{z_1 + z_2}{2}\right\|^2 \leq 4d(x, K)^2 - 4d(x, K)^2 = 0$$

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 $(Y^\perp)^\perp = Y$. $\blacksquare$

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*

$$\text{span}\{x_\alpha\} = \left\{ \sum_{\alpha \in A} c_\alpha x_\alpha \;\middle|\; c_\alpha \in \mathbb{F} = \mathbb{R} \text{ or } \mathbb{C} \right\},$$

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 $Y = \overline{\text{span}(\{x_\alpha\})} \subset \mathcal{H}$ 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

$$\langle x, z \rangle = \lim_{M \to \infty} \left\langle x, \sum_{j=1}^{M} c_{\alpha_j} x_{\alpha_j} \right\rangle = \lim_{M \to \infty} \sum_{j=1}^{M} \overline{c_{\alpha_j}} \langle x, x_{\alpha_j} \rangle = 0.$$

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 $\langle x_\alpha, x_\beta \rangle = \delta_{\alpha\beta}$ 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 $\overline{\text{span}(\{x_\alpha\})} = \mathcal{H}$.*

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

$$Y = \left\{ \sum_j c_j x_{\alpha_j} \;\middle|\; \sum_j |c_j|^2 < \infty, \alpha_j \in A \right\}.$$

Proof. Let $S = \left\{ \sum_j c_j x_{\alpha_j} \;\middle|\; \sum_j |c_j|^2 < \infty, \alpha_j \in A \right\}$. For $x \in S$, $x = \sum_j c_j x_{\alpha_j}$ with $\sum_j |c_j|^2 < \infty$. 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 $\sum_j |c_j^n|^2 < \infty$ 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 $\sum_j c_j x_{\alpha_j} \mapsto \{c_j\}$. 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

$$\left\| \sum_j c_j x_{\alpha_j} \right\|^2 = \sum_j |c_j|^2 = \|\{c_j\}\|^2.$$

For $z_n \in S$, $z_n \to z$. Since z_n is Cauchy and T is isometric, $\{c_j^n\}$ is Cauchy in ℓ^2 and thus converges to some $\{c_j\} \in \ell^2$. 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 $\hat{z} = \sum_j \langle z, x_{\alpha_j} \rangle x_{\alpha_j}$. For each x_{α_k},

$$\langle z - \hat{z}, x_{\alpha_k} \rangle = \lim_{m \to \infty} \left\langle z - \sum_{j=1}^m \langle z, x_{\alpha_j} \rangle x_{\alpha_j}, x_{\alpha_k} \right\rangle$$

$$= \lim_{m \to \infty} \langle z, x_{\alpha_k} \rangle - \sum_{j=1}^m \langle z, x_{\alpha_j} \rangle \delta_{jk}$$

$$= \langle z, x_{\alpha_k} \rangle - \langle z, x_{\alpha_k} \rangle = 0.$$

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 $\{x_\alpha\} \cup \{y\}$ 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 O. Define a partial order as the inclusion of sets. Let $C \subset O$ be a chain. We claim that $B = \bigcup_{\{x_\alpha\} \in C} \{x_\alpha\}$ is an upper bound of C. By construction we have $\{x_\alpha\} \subset B$ for all $\{x_\alpha\} \in C$. We need to show that $B \in O$. For distinct $x_\alpha, x_\beta \in B$, they belong to a common set $C \in C \subset O$. 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 $C \in C$ and thus $\|x_\alpha\| = 1$. It follows that B is also orthonormal. By Zorn's lemma, there exists a maximal element in O, say B, such that if $C \in O$ 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 $\{x_\alpha\} \cup \{y\}$ 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_α $x_\beta \in S$, $\|x_\alpha - x_\beta\| = \sqrt{2}$. Consider the open balls $B_{1/2}(x_\alpha)$. They are clearly disjoint since if y lies in two such balls, then $\sqrt{2} = \|x_\alpha - x_\beta\| \leq \|x_\alpha - y\| + \|y - x_\beta\| < 1$, 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 $\{x_\alpha \mid \alpha \in A\}$, $\{y_\beta \mid \beta \in B\}$ be two orthonormal bases in $\mathcal{H}$. Then $\operatorname{card}(A) = \operatorname{card}(B)$.

Proof. Fixed an $\alpha \in A$, $B_\alpha = \{\beta \in B \mid \langle y_\beta, x_\alpha \rangle \neq 0\}$ 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 $y_\beta = \sum_k \langle y_\beta, x_{\alpha_k} \rangle x_{\alpha_k}$ with at least one $\langle y_\beta, x_{\alpha_k} \rangle \neq 0$. 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

$$\|x\|^2 = \sum_j \left| \langle x, x_{\alpha_j} \rangle \right|^2 .$$

Proof. Let $x \in \mathcal{H}$. Write $x = \sum_j c_j x_{\alpha_j}$ with $\sum_j |c_j|^2 < \infty$. Then

$$\langle x, x_{\alpha_k} \rangle = \lim_{M \to \infty} \left\langle \sum_{j=1}^M c_j x_{\alpha_j}, x_{\alpha_k} \right\rangle = \lim_{M \to \infty} \sum_{j=1}^M c_j \langle x_{\alpha_j}, x_{\alpha_k} \rangle = c_k.$$

It follows that

$$\|x\|^2 = \lim_{M \to \infty} \left\langle \sum_{j=1}^M c_j x_{\alpha_j}, \sum_{j=1}^M c_j x_{\alpha_j} \right\rangle = \lim_{M \to \infty} \sum_{j=1}^M |c_j|^2 = \sum_j |c_j|^2 = \sum_j \left| \langle x, x_{\alpha_j} \rangle \right|^2 .$$

∎

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) $\mathcal{H} = \text{span}(\{w\}) \oplus \ker(T)$ *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 $v \in \text{span}(\{w\}) \cap \ker(T)$, $v = cw$ and $Tv = 0$. Then $cTw = Tv = 0$; $c = 0$ and thus $v = 0$. Therefore, $\text{span}(\{w\}) \cap \ker(T) = \{0\}$ and $\mathcal{H} = \text{span}(\{w\}) \oplus \ker(T)$.

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 $\mathcal{H} = \text{span}(\{w\}) \oplus \ker(S) = \text{span}(\{w\}) \oplus \ker(T)$. 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 $\mathcal{H} = \text{span}(\{w\}) \oplus \ker(T)$ with $\ker(T)$ closed. Also, $\mathcal{H} = \ker(T) \oplus \ker(T)^\perp$ by proposition 3.10. We claim that $\ker(T)^\perp = \text{span}(\{w\})$. 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 $Tx = cSx = c\langle x, w \rangle = \langle x, \overline{c}w \rangle$. Set $x^* = \overline{c}w$ 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. ∎

Definition 3.26

*Let X, Y be vector spaces. $T : X \to Y$ is called **skew-linear** if $T(cx + y) = \bar{c}Tx + Ty$ for all $x, y \in X$ and $c \in \mathbb{F}$.*

Definition 3.27

*$\mathbb{F} = \mathbb{R}$ 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 3.28

*A bilinear form $B : \mathcal{H} \times \mathcal{H} \to \mathbb{F}$ is called **bounded** if there exists $C < \infty$ such that $|B(x, y)| \le C \|x\| \|y\|$ for all $x, y \in \mathcal{H}$.*

Definition 3.29

*A bilinear form $B : \mathcal{H} \times \mathcal{H} \to \mathbb{F}$ is called **coercive** if there exists $\delta > 0$ such that $B(x, x) \ge \delta \|x\|^2$ for all $x \in \mathcal{H}$.*

Theorem 3.30 (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

$$\langle y, T(cx + z) \rangle = B(y, cx + z) = \bar{c}B(y, x) + B(y, z) = \bar{c} \langle y, Tx \rangle + \langle y, Tz \rangle = \langle y, cTx + Tz \rangle.$$

Hence $T(cx + z) = cTx + Tz$. For boundedness, by proposition 3.5,

$$\|Tx\| = \sup_{\|y\|=1} |\langle y, Tx \rangle| = \sup_{\|y\|=1} |B(y, x)| \le \sup_{\|y\|=1} C \|x\| \|y\| = C \|x\|.$$

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 $A^{\perp} \neq \{0\}$ 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 $0 = B(z, z) \geq \delta \|z\|^2$ 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 3.31
*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 3.32 (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 $x \in \mathcal{H}$,

$$-M < \frac{1}{2}\delta \|x\|^2 - \|L\| \, \|x\| \leq \frac{1}{2}B(x, x) - |Lx| \leq F(x) \leq \left| \frac{1}{2}B(x, x) - Lx \right| \leq \frac{1}{2}C \|x\|^2 + \|L\| \, \|x\| < \infty$$

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

$$
\begin{aligned}
F(u_m) + F(u_n) &= \frac{1}{2}B(u_m, u_m) - Lu_m + \frac{1}{2}B(u_n, u_n) - Lu_n \\
&= 2\left(B\left(\frac{u_m}{2}, \frac{u_m}{2}\right) + B\left(\frac{u_n}{2}, \frac{u_n}{2}\right)\right) - 2L\left(\frac{u_m + u_n}{2}\right) \\
&= 2\left[\frac{1}{2}\left(B\left(\frac{u_m + u_n}{2}, \frac{u_m + u_n}{2}\right) + B\left(\frac{u_m - u_n}{2}, \frac{u_m - u_n}{2}\right)\right) - L\left(\frac{u_m + u_n}{2}\right)\right] \\
&= 2F\left(\frac{u_m + u_n}{2}\right) + B\left(\frac{u_m - u_n}{2}, \frac{u_m - u_n}{2}\right) \geq 2\alpha + \frac{\delta}{4}\|u_m - u_n\|^2 .
\end{aligned}
$$

For arbitrary $\epsilon > 0$, there is $N \in \mathbb{N}$ such that for all $m, n \geq N$, $F(u_m), F(u_n) \leq \alpha + \epsilon$. Then

$$
2\alpha + \frac{\delta}{4}\|u_m - u_n\|^2 \leq F(u_m) + F(u_n) \leq 2\alpha + 2\epsilon \implies \|u_m - u_n\|^2 \leq \frac{8\epsilon}{\delta}.
$$

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,

$$
|B(u_n, v_n) - B(u, v)| \leq |B(u_n - u, v_n)| + |B(u, v_n - v)| \leq C\|u_n - u\|\|v_n\| + C\|u\|\|v_n - v\| \to 0
$$

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. $\blacksquare$

Theorem 3.33
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

$$\phi(t) = F(x_0 + ty) = \frac{1}{2}B(x_0 + ty, x_0 + ty) - L(x_0 + ty)$$
$$= \frac{1}{2}B(x_0, x_0) + tB(x_0, y) + \frac{1}{2}t^2 B(y, y) - L(x_0) - tL(y).$$

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 3.34
$\mathcal{D}(A) \subset \mathcal{H}$ *is dense in* $\mathcal{H}$. *A linear operator* $A : \mathcal{D}(A) \to \mathcal{H}$ *is said to be **symmetric** if* $\langle Ax, y \rangle = \langle x, Ay \rangle$ *for all* $x, y \in \mathcal{D}(A)$.

Remark
Note that the domain $\mathcal{D}(A)$ *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 3.35
$\lambda \in \mathbb{F}$ *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 3.36
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 3.37

*A linear operator $A : \mathcal{H} \to \mathcal{H}$ is called **bounded** if*

$$\|A\| = \sup_{\|x\|=1} \|Ax\| < \infty.$$

Proposition 3.38

$A : \mathcal{H} \to \mathcal{H}$ is a symmetric bounded linear operator. Then

$$\|A\| = \sup_{\|x\|=1} |\langle Ax, x \rangle|.$$

Proof. Assume $\|x\| = 1$. By Cauchy-Schwarz inequality, $|\langle Ax, x \rangle| \leq \|Ax\| \|x\| = \|Ax\|$. Taking supremum,

$$\sup_{\|x\|=1} |\langle Ax, x \rangle| \leq \sup_{\|x\|=1} \|Ax\| = \|A\|.$$

To see the reverse inequality, note that $\|Ax\|^2 = \langle Ax, Ax \rangle = \langle A^2 x, x \rangle$. 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,

$$
\begin{aligned}
\langle A^2 x, x \rangle &= \left\langle A\left(\frac{\lambda}{2}(x^+ - x^-)\right), \frac{1}{2\lambda}(x^+ + x^-) \right\rangle \\
&= \frac{1}{4} \left\langle Ax^+ - Ax^-, x^+ + x^- \right\rangle \\
&= \frac{1}{4} \left(\langle Ax^+, x^+ \rangle + \langle Ax^+, x^- \rangle - \langle Ax^-, x^+ \rangle - \langle Ax^-, x^- \rangle \right)
\end{aligned}
$$

Notice that $\langle A^2 x, x \rangle$, $\langle Ax^+, x^+ \rangle$ and $\langle Ax^-, x^- \rangle$ are real numbers by proposition 3.36; hence $\mathfrak{I}(\langle Ax^+, x^- \rangle - \langle Ax^-, x^+ \rangle) = 0$. Also, $\langle Ax^-, x^+ \rangle = \overline{\langle x^+, Ax^- \rangle} = \overline{\langle Ax^+, x^- \rangle}$. We have $\mathfrak{I}(\langle Ax^+, x^- \rangle) =$

0 and $\langle Ax^+, x^-\rangle - \langle Ax^-, x^+\rangle = 0$. Thus, letting $C = \sup_{\|x\|=1} |\langle Ax, x\rangle|$,

$$\langle A^2 x, x\rangle = \frac{1}{4}(\langle Ax^+, x^+\rangle + \langle Ax^-, x^-\rangle)$$

$$\leq \frac{1}{4}C\left(\|x^+\|^2 + \|x^-\|^2\right)$$

$$= \frac{1}{4}C(\langle x^+, x^+\rangle + \langle x^-, x^-\rangle)$$

$$= \frac{1}{4}C\left(2\lambda^2\|x^2\| + \frac{2}{\lambda^2}\|Ax\|^2\right) = \frac{1}{2}C\left(\lambda^2\|x\|^2 + \frac{1}{\lambda^2}\|Ax\|^2\right).$$

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

$$\|Ax\|^2 = \langle A^2 x, x\rangle \leq \frac{1}{2}C\inf_{\lambda \neq 0}\lambda^2\|x\|^2 + \frac{1}{\lambda^2}\|Ax\|^2 \leq C\|Ax\|\|x\|.$$

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 3.39
*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 $x_n \to x \in [0, 1]$.

$$|Au(x_n) - Au(x)| = \left|\int_0^1 K(x_n, y)u(y)dy - \int_0^1 K(x, y)u(y)dy\right|$$

$$\leq \int_0^1 |K(x_n, y) - K(x, y)|\,|u(y)|\,dy \leq \|K(x_n, \cdot) - K(x, \cdot)\|_\infty \|u\|_\infty.$$

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 $\|u_n\|_\infty \le M$ for all n. Also, since K is continuous on a compact set, we have $K(x, y) \le C$ for all $x, y \in [0, 1]$. Then

$$
\begin{aligned}
\|Au_n\|_\infty &= \sup_{x \in [0,1]} \left| \int_0^1 K(x, y) u_n(y) dy \right| \\
&\le \sup_{x \in [0,1]} \int_0^1 |K(x, y)| \, |u_n(y)| \, dy \\
&\le \sup_{x \in [0,1]} \int_0^1 C \, \|u_n\|_\infty \, dy = C \, \|u_n\|_\infty \le CM.
\end{aligned}
$$

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 $n \in \mathbb{N}$,

$$
\begin{aligned}
|Au_n(x) - Au_n(z)| &= \left| \int_0^1 K(x, y) u_n(y) dy - \int_0^1 K(z, y) u_n(y) dy \right| \\
&\le \int_0^1 |K(x, y) - K(z, y)| \, |u_n(y)| \, dy \le \frac{\epsilon}{M} \|u_n\|_\infty \le \epsilon.
\end{aligned}
$$

Thus $\{Au_n\}$ is uniformly equicontinuous. By Arzelà-Ascoli theorem, $\{Au_n\}$ has a convergent subsequence, i.e., A is compact.

Definition 3.40

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 3.41 (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 $\lambda_i \ne \lambda_j$, then $\langle x_i, x_j \rangle = 0$.*

(c) *For any $\lambda \ne 0$, $\dim(E_\lambda) < \infty$.*

(d) *If $\dim(\mathcal{H}) = \infty$, then either $\lambda_j \to 0$ or there is only finitely many $\lambda_j \ne 0$.*

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 3.38, $\|A\| = \sup_{\|x\|=1} |\langle Ax, x \rangle|$. 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 $\lambda_1 = \mathrm{sgn}(\langle Az_n, z_n \rangle) \|A\|$ 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 $W_1 = \mathrm{span}(\{x_1\})$ 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 $\|x_2\| = 1$, $|\lambda_2| = \|A_1\| \leq \|A\| = |\lambda_1|$ 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 $|\lambda_j| \leq |\lambda_{j-1}|$ and $\|x_j\| = 1$. Furthermore, observe that $\langle x_i, x_j \rangle = 0$ for all $i \neq j$ 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 $\lambda \neq 0$. 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.

$$2\left|\lambda\right|^2 = \left|\lambda\right|^2 \left\|x_{n_i}\right\|^2 + \left|\lambda\right|^2 \left\|x_{n_j}\right\|^2 = \left\|\lambda x_{n_i} - \lambda x_{n_j}\right\|^2 = \left\|Ax_{n_i} - Ax_{n_j}\right\|^2 \to 0.$$

Hence $|\lambda| = 0$, a contradiction. Thus $\dim(E_\lambda) < \infty$.

Lastly, we show that $\overline{\mathrm{span}(\{x_j\})} = \mathcal{H}$. 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

$$\left\langle x - z_n, x_j \right\rangle = \left\langle x - \sum_{i=1}^n \langle x, x_i \rangle x_i, x_j \right\rangle = \left\langle x, x_j \right\rangle - \left\langle x, x_j \right\rangle = 0.$$

Hence $x - z_n \in W_n^\perp$. Thus since $\|x - z_n\|$ is bounded,

$$\|Ax - Az_n\| = \|A(x - z_n)\| = \|A_n(x - z_n)\| \leq \|A_n\| \|x - z_n\| = |\lambda_{n+1}| \|x - z_n\| \to 0$$

Hence $Az_n \to Ax$ and thus we can write $Ax = \sum_j \lambda_j \left\langle x, x_j \right\rangle x_j$. If $y = \sum_j \left\langle x, x_j \right\rangle x_j$, then $Ay = \sum_j \lambda_j \left\langle x, x_j \right\rangle x_j = Ax$. Because A has zero kernel, $x = y = \sum_j \left\langle x, x_j \right\rangle x_j$. Thus we have $\overline{\mathrm{span}(\{x_j\})} = \mathcal{H}$.

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 3.42
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

$$x = \sum_j \langle x, x_j \rangle x_j + \sum_k \langle x, w_k \rangle w_k.$$

*Such decomposition is called the **spectral decomposition** of A.*

Theorem 3.43 (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 $\{x_j\} \cup \{w_k\}$ 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

$$x = \sum_j \frac{1}{\lambda - \lambda_j} \langle z, x_j \rangle x_j + \sum_k \frac{1}{\lambda} \langle z, w_k \rangle w_k,$$

we see that

$$\lambda x - Ax = \lambda \sum_j \frac{1}{\lambda - \lambda_j} \langle z, x_j \rangle x_j + \lambda \sum_k \frac{1}{\lambda} \langle z, w_k \rangle w_k - \sum_j \frac{\lambda_j}{\lambda - \lambda_j} \langle z, x_j \rangle x_j$$

$$= \sum_j \langle z, x_j \rangle x_j + \sum_k \langle z, w_k \rangle w_k = z.$$

We verify that such x indeed belongs to $\mathcal{H}$.

$$\|x\|^2 = \sum_j \left| \frac{1}{\lambda - \lambda_j} \langle z, x_j \rangle \right|^2 + \sum_k \left| \frac{1}{\lambda} \langle z, w_k \rangle \right|^2 \leq \sup_j \frac{1}{|\lambda - \lambda_j|^2} \|z\|^2 + \sup_k \frac{1}{|\lambda|^2} \|z\|^2 < C_\lambda \|z\|^2$$

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

$$\lambda x - Ax = \lambda \sum_{j:\lambda_j \neq \lambda_1} \frac{1}{\lambda - \lambda_j} \langle z, x_j \rangle x_j + \lambda \sum_k \frac{1}{\lambda} \langle z, w_k \rangle w_k - \sum_{j:\lambda_j \neq \lambda_1} \frac{\lambda_j}{\lambda - \lambda_j} \langle z, x_j \rangle x_j$$
$$= \sum_{j:\lambda_j \neq \lambda_1} \langle z, x_j \rangle x_j + \sum_k \langle z, w_k \rangle w_k = z.$$

We verify that such x indeed belongs to $\mathcal{H}$. By exactly the same argument as above,

$$\|x\|^2 \leq \sup_{j:\lambda_j \neq \lambda} \frac{1}{|\lambda - \lambda_j|^2} \|z\|^2 + \sup_k \frac{1}{|\lambda|^2} \|z\|^2 < C_\lambda \|z\|^2 \,.$$

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 $\lambda_{j_i} = \lambda_1 = \lambda$,

$$\langle z, x_{j_i} \rangle = \langle \lambda x - Ax, x_{j_i} \rangle = \langle \lambda x, x_{j_i} \rangle - \langle Ax, x_{j_i} \rangle$$
$$= \lambda \langle x, x_{j_i} \rangle - \langle x, Ax_{j_i} \rangle = 0. = \lambda \langle x, x_{j_i} \rangle - \lambda_{j_i} \langle x, x_{j_i} \rangle = 0.$$

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

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 2.62, as any finite-dimensional vector space is a Banach space. However, we also have a simple proof here.

Let $\{e_1, \ldots, e_n\}$ 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

$$\left| \|x\| - \|y\| \right| \le \|x - y\| \le C_1 \|x - y\|_1 .$$

Now for any $x \ne 0$, 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 $x \ne 0$. 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

$$|f(y) - f(z)| = \left| \|x - y\| - \|x - z\| \right| \le \|y - z\| ,$$

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

$$\|P(x_n)\| = \|P(x_n) - x_n + x_n\| \le \|P(x_n) - x_n\| + \|x_n\| \le 2 \|x_n\| ,$$

$P(x_n)$ 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,

$$\left\|P(x_{n_k}) - x_{n_k}\right\| \le \left\|P(x) - x_{n_k}\right\|$$

for all k. Letting $k \to \infty$ gives $\|P(\hat{x}) - x\| \le \|P(x) - x\|$. 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

$$Y_x = \left\{ y \in Y \mid y = \mathrm{argmin}_{y \in Y} \|x - y\| \right\}$$

is a bounded convex set.

Proof. Since $0 \in Y$, for all $y \in Y_x$, $\|y\| \le \|x - y\| + \|x\| \le 2\|x\|$. 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

$$
\begin{aligned}
\|x - (ty_1 + (1 - t)y_2)\| &= \|t(x - y_1) + (1 - t)(y_2 - x)\| \\
&\le t\|x - y_1\| + (1 - t)\|x - y_2\| \\
&= \|x - y_1\| = \|x - y_2\|
\end{aligned}
$$

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 $\alpha x \ne \beta y$ 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 $F = \left\{ k \in \{0, \ldots, n\} \mid \left| x - \frac{k}{n} \right| < \delta \right\}$. We can compute that

$$
\begin{aligned}
|B_n(f)(x) - f(x)| &= \left| f(x) - \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k} \right| \\
&\leq \sum_{k \in F} \left| f(x) - f\left(\frac{k}{n}\right) \right| \binom{n}{k}x^k(1-x)^{n-k} + \sum_{k \notin F} \left| f(x) - f\left(\frac{k}{n}\right) \right| \binom{n}{k}x^k(1-x)^{n-k} \\
&\leq \sum_{k=0}^{n} \epsilon \binom{n}{k}x^k(1-x)^{n-k} + 2\left\|f\right\|_\infty \sum_{k \notin F} \binom{n}{k}x^k(1-x)^{n-k} \\
&\leq \epsilon + 2\left\|f\right\|_\infty \sum_{k=0}^{n} \mathbb{1}\left\{ \left| x - \frac{k}{n} \right| \geq \delta \right\} \binom{n}{k}x^k(1-x)^{n-k} \\
&\leq \epsilon + 2\left\|f\right\|_\infty \sum_{k=0}^{n} \frac{1}{\delta^2}\left(x - \frac{k}{n} \right)^2 \binom{n}{k}x^k(1-x)^{n-k} \\
&= \epsilon + 2\frac{\left\|f\right\|_\infty}{\delta^2} \sum_{k=0}^{n} \left(x^2 - \frac{2k}{n}x + \frac{k^2}{n^2} \right)\binom{n}{k}x^k(1-x)^{n-k}.
\end{aligned}
$$

Now let

$$
S(x, y) = \sum_{k=0}^{n} \binom{n}{k}x^k y^{n-k} = (x + y)^n.
$$

Then

$$
nx(x + y)^{n-1} = x\frac{\partial S}{\partial x} = \sum_{k=0}^{n} k\binom{n}{k}x^k y^{n-k},
$$

$$
n(n-1)x^2(x+y)^{n-2} = x^2\frac{\partial^2 S}{\partial x^2} = \sum_{k=0}^{n} k(k-1)\binom{n}{k}x^k y^{n-k}.
$$

Taking $y = 1 - x$ gives

$$
\sum_{k=0}^{n} \left(x^2 - \frac{2k}{n}x + \frac{k^2}{n^2} \right)\binom{n}{k}x^k(1-x)^{n-k} = x^2 - \frac{2}{n}x \cdot nx + \frac{1}{n^2}\left(n(n-1)x^2 + nx \right) = \frac{x(1-x)}{n} \leq \frac{1}{4n}.
$$

Hence we obtain the estimate

$$
|B_n(f)(x) - f(x)| \leq \epsilon + \frac{\left\|f\right\|_\infty}{2n\delta^2}.
$$

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 $T(f) \geq 0$ for all $f \geq 0$.*

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

$$U(cf+g)(x) = \sum_{k=0}^{n} \left(cf\left(\frac{k}{n}\right) + g\left(\frac{k}{n}\right) \right) \binom{n}{k} x^k (1-x)^{n-k}$$

$$= c \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} + \sum_{k=0}^{n} g\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} = cU(f)(x) + U(g)(x).$$

To show the positivity, let $f \geq 0$. Then $f\left(\frac{k}{n}\right) x^k (1-x)^{n-k} \geq 0$ for all k and $x \in [0,1]$. Then the sum is nonnegative and $U(f) \geq 0$.

To show the continuity, it is enough to show the boundedness of U.

$$\|U(f)\|_\infty = \sup_{x \in [0,1]} \left| \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} \right|$$

$$\leq \sup_{x \in [0,1]} \sum_{k=0}^{n} \left| f\left(\frac{k}{n}\right) \right| \binom{n}{k} x^k (1-x)^{n-k}$$

$$\leq \sup_{x \in [0,1]} \sum_{k=0}^{n} \|f\|_\infty \binom{n}{k} x^k (1-x)^{n-k} = \|f\|_\infty .$$

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

$$|f(x) - f(a)| \leq L\,|x-a| \leq L\epsilon + L\frac{(x-a)^2}{\epsilon}.$$

This can be verify as follows,

$$\begin{cases} L\,|x-a| \leq L\epsilon \leq L\epsilon + L\frac{(x-a)^2}{\epsilon} & \text{if } |x-a| \leq \epsilon, \\ L\,|x-a| \leq L\frac{(x-a)^2}{\epsilon} \leq L\epsilon + L\frac{(x-a)^2}{\epsilon} & \text{if } |x-a| > \epsilon. \end{cases}$$

Next, we apply T_n and note that we have $|T_n(f)| = T_n(|f|)$. Then,

$$
\begin{aligned}
|T_n(f)(x) - f(a)| &\leq |T_n(f)(x) - f(a)T_n(f_0)(x)| + |f(a)T_n(f_0)(x) - f(a)| \\
&= T_n(|f - f(a)|)(x) + |f(a)|\,|T_n(f_0)(x) - f_0| \\
&\leq L\left(\epsilon T_n(f_0)(x) + \frac{1}{\epsilon} T_n(f_2 - 2af_1 + a^2 f_0)(x)\right) + \|f\|_\infty \|T_n(f_0) - f_0\|_\infty \\
&\leq L\epsilon(T_n(f_0)(x) - f_0(x)) + L\epsilon f_0(x) + \frac{L}{\epsilon}(T_n(f_2)(x) - f_2(x)) + \frac{L}{\epsilon} f_2(x) \\
&\quad - \frac{2aL}{\epsilon}(T_n(f_1)(x) - f_1(x)) - \frac{2aL}{\epsilon} f_1(x) + \frac{a^2 L}{\epsilon}(T_n(f_0)(x) - f_0(x)) \\
&\quad + \frac{a^2 L}{\epsilon} f_0(x) + \|f\|_\infty \|T_n(f_0) - f_0\|_\infty \\
&\leq L\epsilon \|T_n(f_0) - f_0\|_\infty + \frac{L}{\epsilon} \|T_n(f_2) - f_2\|_\infty + \frac{2aL}{\epsilon} \|T_n(f_1) - f_1\|_\infty \\
&\quad + \frac{a^2 L}{\epsilon} \|T_n(f_0) - f_0\|_\infty + \|f\|_\infty \|T_n(f_0) - f_0\|_\infty + L\epsilon + \frac{L}{\epsilon}(x - a)^2.
\end{aligned}
$$

Now taking $a = x$ and then taking supremum over $x \in [0, 1]$ gives

$$
\begin{aligned}
\|T_n(f) - f\|_\infty &\leq L\epsilon \|T_n(f_0) - f_0\|_\infty + \frac{L}{\epsilon} \|T_n(f_2) - f_2\|_\infty \\
&\quad + \frac{2L}{\epsilon} \|T_n(f_1) - f_1\|_\infty + \frac{L}{\epsilon} \|T_n(f_0) - f_0\|_\infty + \|f\|_\infty \|T_n(f_0) - f_0\|_\infty + L\epsilon.
\end{aligned}
$$

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

$$\omega_f(\lambda\delta) \leq \omega_f(n\delta) \leq n\omega_f(\delta) \leq (1 + \lambda)\omega_f(\delta).$$

79

Theorem 4.15

For any $f \in C([0,1])$, the Bernstein polynomial $B_n(f)$ satisfies

$$\|B_n(f) - f\|_\infty \le \frac{3}{2}\omega_f\left(\frac{1}{\sqrt{n}}\right).$$

Proof. By lemma 4.14, setting $\delta = 1/\sqrt{n}$ and $\lambda = \sqrt{n}\left|x - \frac{k}{n}\right|$, then

$$
\begin{aligned}
|f(x) - B_n f(x)| &\le \sum_{k=0}^{n}\left|f(x) - f\left(\frac{k}{n}\right)\right|\binom{n}{k}x^k(1-x)^{n-k} \le \sum_{k=0}^{n}\omega_f\left(\left|x-\frac{k}{n}\right|\right)\binom{n}{k}x^k(1-x)^{n-k} \\
&\le \sum_{k=0}^{n}\omega_f\left(\frac{1}{\sqrt{n}}\right)\left(1+\sqrt{n}\left|x-\frac{k}{n}\right|\right)\binom{n}{k}x^k(1-x)^{n-k} \\
&\le \omega_f\left(\frac{1}{\sqrt{n}}\right)\left\{1+\sqrt{n}\sum_{k=0}^{n}\left|x-\frac{k}{n}\right|\binom{n}{k}x^k(1-x)^{n-k}\right\} \\
&\le \omega_f\left(\frac{1}{\sqrt{n}}\right)\left\{1+\sqrt{n}\left(\sum_{k=0}^{n}\left|x-\frac{k}{n}\right|^2\binom{n}{k}x^k(1-x)^{n-k}\right)^{1/2}\left(\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}\right)^{1/2}\right\} \\
&\le \omega_f\left(\frac{1}{\sqrt{n}}\right)\left\{1+\sqrt{n}\frac{1}{2\sqrt{n}}\right\} = \frac{3}{2}\omega_f\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

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 $d(f, p^*) \le d(f, q)$ 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 $d(f, q_{n_k}) \le d^* + \epsilon$ and $d(q_{n_k}, p^*) \le \epsilon$ for all $k \ge N$. Then

$$d(f, p^*) \le d(f, q_{n_k}) + d(q_{n_k}, p^*) \le d^* + 2\epsilon.$$

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,

$$\left\| f - \frac{1}{2}p_1^* - \frac{1}{2}p_2^* \right\| < \frac{1}{2}\|f - p_1^*\| + \frac{1}{2}\|f - p_2^*\| = d^*.$$

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 $x \in S$,

$$|f(x) - (p(x) + \lambda q(x))| = |r(x) - \lambda q(x)| \le R - \lambda \min |q(x)| < R.$$

And for $x \in N$,

$$|f(x) - (p(x) + \lambda q(x))| = |r(x) - \lambda q(x)| \le R + \lambda \max |q(x)| < \frac{R}{2} + \lambda \|q\|_\infty < R$$

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,

$$\left| \int_a^b f(x) e^{-inx} \, dx \right| \leq \left| \int_a^b (f(x) - g(x)) e^{-inx} \, dx \right| + \left| \int_a^b g(x) e^{-inx} \, dx \right|$$

$$\leq \|f - g\|_1 + \left| \int_a^b g(x) e^{-inx} \, dx \right| \to 0$$

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

$$|a_k|, |b_k| \leq \frac{\|f\|_{PC^n}}{k^n},$$

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 $a_1, \ldots, a_n \in \mathbb{R}$,*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0.$$

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

$$S_N f(x) = \sum_{k=-N}^{N} c_k e^{ikx} = \sum_{k=-N}^{N} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)e^{-iky}dy\, e^{ikx}$$

$$= \int_{-\pi}^{\pi} f(y)\frac{1}{2\pi} \sum_{k=-N}^{N} e^{ik(x-y)}dy = \int_{-\pi}^{\pi} f(y)D_N(x - y)dy = (D_N * f)(x).$$

∎

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

$$
\frac{1}{2}(f(x^+) + f(x^-)) = \int_0^{\pi} D_N(y) (f(x^+) + f(x^-)) dy.
$$

Thus for given x, we have

$$
\left| S_N f(x) - \frac{f(x^+) + f(x^-)}{2} \right| \leq \left| \int_0^{\pi} D_N(y) (f(x + y) - f(x^+)) dy \right|
$$
$$
+ \left| \int_0^{\pi} D_N(y) (f(x - y) - f(x^-)) dy \right|.
$$

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

$$
\int_0^{\delta} |D_N(y)| \left| f(x + y) - f(x^+) \right| dy \leq C \int_0^{\delta} y |D_N(y)| dy \leq C \int_0^{\delta} dy = C\delta,
$$

because $|D_N(t)| \leq 1/|t|$. On the other hand,

$$
\begin{aligned}
\int_{\delta}^{\pi} |D_N(y)| \left| f(x + y) - f(x^+) \right| dy &\leq \frac{1}{2\pi} \int_{\delta}^{\pi} \frac{\sin((N + 1/2)y)}{\sin(y/2)} \left| f(x + y) - f(x^+) \right| dy \\
&\leq \frac{1}{2\pi \sin(\delta/2)} \int_{\delta}^{\pi} \sin((N + 1/2)y) g(y) dy \to 0,
\end{aligned}
$$

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 see that

$$
\left| \int_0^{\pi} D_N(y) f(x + y) - f(x^+) dy \right| \to 0 \quad \text{as } N \to \infty.
$$

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,

$$\sigma_N f(x) = \frac{1}{N+1} \sum_{k=0}^{N} \int_{-\pi}^{\pi} D_k(x-y)f(y)dy = \int_{-\pi}^{\pi} \left(\frac{1}{N+1} \sum_{k=0}^{N} D_k(x-y) \right) f(y)dy = F_N * f(x).$$

87

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

$$
\begin{aligned}
|\sigma_N f(x) - f(x)| &= \left| \int_{-\pi}^{\pi} F_N(x-y)f(y)dy - f(x) \right| \leq \int_{-\pi}^{\pi} F_N(x-y)\,|f(y) - f(x)|\,dy \\
&\leq \int_{|x-y|\leq\delta} F_N(x-y)\,|f(y) - f(x)|\,dy + \int_{\delta\leq|x-y|\leq\pi} F_N(x-y)\,|f(y) - f(x)|\,dy \\
&\leq \epsilon \int_{-\pi}^{\pi} F_N(x-y)dy + 2\,\|f\|_\infty \int_{\delta\leq|x-y|\leq\pi} F_N(x-y)dy \\
&= \epsilon + 2\,\|f\|_\infty \frac{\pi - \delta}{2\pi(N+1)\sin^2(\delta/2)} \to 0
\end{aligned}
$$

as $N \to \infty$. ∎

Theorem 4.36

If $f \in C_{2\pi}^1$, then $S_N f \to f$ uniformly.