Compiled Notes of Course 18.125

Real and Functional Analysis

Instructor: Yaron Ostrover
Notes Author: Dahua Lin

May 01, 2008

Contents

I Measure and Integration Theory **3**

1 Lebesgue Measure Theory **4**
1.1 Why Riemann Theory is Not Enough? 4
1.1.1 Brief Review of Riemann Integration 4
1.1.2 Riemann Theory vs Lebesgue Theory 4
1.1.3 The Basic Idea of Lebesgue Integration 4
1.2 Generic Measure Space 5
1.2.1 A Naive Approach 5
1.2.2 σ-algebra 5
1.2.3 Measurable and Measure Spaces 6
1.2.4 Continuity of Measure 7
1.3 Lebesgue Measure 9
1.3.1 The Measure of Intervals 9
1.3.2 Outer Measure 13
1.3.3 From Outer Measure to Lebesgue Measure 15
1.3.4 Existence of Mon-measurable Sets 20
1.4 Borel Measurable Functions 20
1.4.1 Borel Algebra 20
1.4.2 Measurable Functions 21

2 Integration Theory **24**
2.1 Lebesgue Integration 24
2.1.1 Simple Functions 24
2.1.2 Lebesgue Integral 25
2.1.3 Additivity, Monotone Convergence Theorem, and Induced Measure 27
2.1.4 Lebesgue Integration of Complex Functions 31
2.1.5 Dominated Convergence Theorem 34
2.2 Null Sets and Almost Everywhere 35
2.2.1 Null Sets and Complete Measure 35
2.2.2 Almost Everywhere 36
2.3 Riesz Representation Theorem 38
2.4 Principles of Measure Theory 45

3 Introduction to L^p Spaces **46**
3.1 Important Inequalities 46
3.1.1 Convex Sets and Convex Functions 46
3.1.2 Jensen's Inequality 46
3.1.3 Hölder's Inequality 47
3.1.4 Minkowski's Inequality 48
3.2 L^p Spaces 48
3.2.1 From $\mathcal{L}^p$ space to L^p space 48
3.2.2 L^p Convergence and Completeness 49
3.3 Important Theorems 50
3.3.1 Lusin's Theorem 50

3.3.2 Egoroff's Theorem and Convergence in measure . . . 52

4 Product Measure **54**
4.1 Product Measure Space . . . 54
4.1.1 Product Measurable Space . . . 54
4.1.2 Measure by Extension: Hahn-Kolmogorov theorem . . . 55
4.1.3 Monotone Class . . . 57
4.1.4 Product Measure . . . 58
4.2 Fubini's theorem . . . 59

5 Complex Measures **61**
5.1 The Space of Complex Measures . . . 61
5.2 Lebesgue Decomposition of Measures . . . 62
5.2.1 Absolute continuity and Singularity . . . 62
5.2.2 Lebesgue Decomposition Theorem and Radon-Nikodym Theorem . . . 64

II Fundamentals of Functional Analysis **67**

6 Review of Linear Algebra **68**
6.1 Vector Spaces . . . 68
6.2 Subspaces . . . 69
6.3 Linear Maps . . . 70
6.4 Quotient Spaces . . . 71

7 Normed and Banach Spaces **73**
7.1 Metric Spaces . . . 73
7.1.1 Metric and Metric Spaces . . . 73
7.1.2 Convergence in Metric Spaces . . . 73
7.2 Normed Spaces . . . 75
7.2.1 Norm and Normed Spaces . . . 75
7.2.2 Seminorm and Norm of Quotient Spaces . . . 76
7.3 Banach Spaces . . . 77

8 Hilbert Spaces **81**
8.1 Inner Product and Inner Product Spaces . . . 81
8.2 Orthogonality . . . 83
8.3 Hilbert Spaces and Closed Subspaces . . . 84
8.4 Complete Orthonormal Systems . . . 86

9 Linear Functionals **90**
9.1 The Space of Linear Functionals . . . 90
9.2 Dual Spaces . . . 91
9.2.1 Bounded Linear Functionals . . . 91
9.2.2 Dual Space . . . 92
9.2.3 Riesz Representation Theorem on Hilbert Space . . . 93
9.2.4 Hahn-Banach Theorem . . . 94
9.2.5 Second Dual . . . 95

10 Linear Operators **97**
10.1 The Space of Linear Operators . . . 97
10.2 Inverse Operators . . . 98
10.3 Dual Operators . . . 100
10.4 Convergence of Operators . . . 100

Part I

Measure and Integration Theory

Chapter 1

Lebesgue Measure Theory

1.1 Why Riemann Theory is Not Enough?

1.1.1 Brief Review of Riemann Integration

Let $f : I \to R$ be a real valued function defined on a real interval $I = [a, b]$. A *partition* P of I is defined by $P = a = t_0 < t_1 < \ldots < t_n = b$. A *choice function* ε is a function that chooses a point from each small intervals in the partition: $\xi(P) = (t_1^*, t_2^*, \ldots, t_n^*)$, s.t. $t_i^* \in [t_i, t_{i+1}] \forall i = 1, \ldots, n$. Then, we can define a sum to approximate the integral as follows

$$I(f, P, \xi) = \sum_{i=1}^{n} \sum_{i=1}^{n} f(t_i^*)(t_{i+1} - t_i), \tag{1.1}$$

where n is the number of paritioned intervals in P.

If the sum defined above converges as the maximum length of the partitioned intervals in P approaches zero. The limit of the sum is called the *Riemann integral* of f on I, and f is said to be *Riemann integrable.*

Formally, a real valued function f is said to be *integrable* on I if there exists $v \in \mathbb{R}$ such that

$$\forall \varepsilon > 0, \exists \delta > 0, \text{ s.t.} \forall P, \forall \xi, \text{ s.t. } \max_{1 \le j \le n} |t_j - t_{j-1}| < \delta, \quad |I(f, P, \xi) - v| < \varepsilon,$$

then v is called the *Riemann integral* of f on I, and f is said to be *Riemann integrable.*

1.1.2 Riemann Theory vs Lebesgue Theory

1. *"Riemann integration mainly works for continuous functions".* An example that is not Riemann integrable is the *Dirichlet function* defined by

$$D(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}.$$

   While, *"Lebesgue integration works for every function that you can imagine".*

2. Lebesgue theory readily answers a series of important questions, especially the convergence of integration. Does $f_n \to f$ imply $\int f_n \to \int f$?

3. Lebesgue theory is the foundation of many modern mathematical branches, including modern probability theory and functional analysis.

1.1.3 The Basic Idea of Lebesgue Integration

The theoretical development of the Lebesgue integration in a rigorous manner is very technical, however, the basic idea behind is pretty simple.

"In doing Riemann integration, we divide the domain of a function, while in Lebesgue integration, we try to divide its range."

$$\int f = \sum_{r \in \mathrm{rgn} f} r \cdot sizeof\{x | f(x) = r\}.$$

The key question here is that *how to define the size of a set*, which leads to the study of *measure*.

1.2 Generic Measure Space

1.2.1 A Naive Approach

We intend to assign a nonnegative value to each subset to measure its size. This can be defined as a function from subsets to nonnegative real values:

$$\mu : 2^{\mathbb{R}} \to [0, \infty], \tag{1.2}$$

where $2^{\mathbb{R}}$ refers to the set of all subsets of $\mathbb{R}$.

We hope that the function μ satisfies the following properties:

- $\mu(\emptyset) = 0$.
- $\mu(\mathbb{R}) = \infty$.
- **(countable additivity)** $\mu(\bigsqcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} \mu(A_i)$.
- $\mu([a, b]) = b - a, \ \forall b \geq a \in \mathbb{R}$.
- **(monotonicity)** $A \subseteq B \ \Rightarrow \ \mu(A) \leq \mu(B)$
- **(translation invariance)** $\mu(A + y) = \mu(A), \ \forall y \in \mathbb{R}$.

Then we can construct approximation to the integration by the following series.

$$I_n(f) = \sum_{k=-\infty}^{\infty} \frac{k}{2^n} \mu\{x | f(x) \in [\frac{k}{2^n}, \frac{k+1}{2^n}]\}, \tag{1.3}$$

we can see that this way of construction actually divides the range into intervals of length $1/2^n$. As n approches infinity, we are expecting that I_n approches the integral.

The **Bad news** is that *"there is no such function μ defined on $2^{\mathbb{R}}$ that simultaneously satisfies all conditions above."* More strictly, simultaneous satisfaction of $\mu(\emptyset) = 0$, $\mu([a, b]) = b - a$ and countable additivity is impossible for $\mu : 2^{\mathbb{R}} \to [0, \infty]$.

Clearly we want too much. To address this issue, we try to construct a function defined on only a reasonable subset of $2^{\mathbb{R}}$ that satisfies all the above conditions, rather than making it defined on the entire $2^{\mathbb{R}}$, which is too big.

1.2.2 σ-algebra

To construct such a "reasonable collection of subsets", we introduce an algebraic system of subsets called *σ-algebra.*

Definition 1.1 (σ-algebra)**.** *A **σ-algebra** over a set X is a non-empty class $\mathcal{F}$ of subsets of X that includes the empty set and is closed under complementation and countable unions. Formally, we have*

1. $\emptyset \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. $\forall i \in \mathbb{N}, A_i \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Given any non-empty set X, it is easy to verify that the following are σ-algebras:

1. 2^X

2. $\{\emptyset, X\}$

3. Given any $A \subset X$, $\{\emptyset, A, A^c, X\}$. This σ-algebra is the smallest σ-algebra that contains A, denoted by $\sigma(\{A\})$.

Proposition 1.1 (Properties of σ-algebra). *For any σ-algebra $\mathcal{F}$ over a non-empty set X, we have*

1. ***(inclusion of*** X***)*** $X \in \mathcal{F}$

2. ***(closed under intersection and union)*** $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$, *and* $A \backslash B \in \mathcal{F}$

3. ***(closed under countable intersection)*** $\forall i \in \mathbb{N}, A_i \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{n} A_i \in \mathcal{F}$

4. ***(closed under finite union and intersection)*** $\forall i = 1, \ldots, n, A_i \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{n} A_i \in \mathcal{F}$, *and* $\bigcap_{i=1}^{n} A_i \in \mathcal{F}$.

Proof. 1. $X \in \mathcal{F}$ follows from the fact that $\emptyset \in \mathcal{F}$ and $X = \emptyset^c$.

2. $A, B \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}, B^c \in \mathcal{F}$, hence $A \cap B = (A^c \cup B^c)^c \in \mathcal{F}$ and $A \backslash B = A \cap B^c \in \mathcal{F}$.

3. $\forall i \in \mathbb{N}, A_i \in \mathcal{F} \Rightarrow \forall i \in \mathbb{N}, A_i^c \in \mathcal{F}$. Hence,

$$\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c \right)^c \in \mathcal{F}.$$

4. Given $A_i \in \mathcal{F}, i = 1, \ldots n$, we can augment the finite collection of subsets to an countable collection by letting $A_i = \emptyset, \forall i > n$. Clearly, all sets in the augmented collection are in $\mathcal{F}$. Then,

$$\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

The closeness of finite intersection immediately follows from that of finite union, as

$$\bigcap_{i=1}^{n} A_i = \left(\bigcup_{i=1}^{n} A_i^c \right)^c \in \mathcal{F}.$$

$\square$

1.2.3 Measurable and Measure Spaces

Definition 1.2 (Measurable space). *A non-empty set X together with a σ-algebra $\mathcal{F}$ defined over it is called a **measurable space**, denoted by $(X, \mathcal{F})$.*

*Elements of $\mathcal{F}$ are said to be $\mathcal{F}$**-measurable**, or simply **measurable** if $\mathcal{F}$ is clear from context.*

Definition 1.3 (Measure). *A **measure** of a measurable space $(X, \mathcal{F})$ is a map $\mu : \mathcal{F} \to [0, \infty]$ such that*

1. $\mu(\emptyset) = 0$;

2. ***(countable additivity)*** $\mu(\bigsqcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ *for disjoint sets* $A_1, A_2, \ldots$ *in* $\mathcal{F}$.

Note that the first condition is required to prevent the trivial construction that assigns infinity to every set in $\mathcal{F}$.

Any measurable space admits a measure. A trivial example is to let $\mu(A) = 0$ for each $A \in \mathcal{F}$.

Definition 1.4 (Measure space). *A measurable space $(X, \mathcal{F})$ together with a measure μ is called a **measure space**, denoted by $(X, \mathcal{F}, \mu)$.*

Proposition 1.2 (Properties of measure space). *Given a measure space $(X, \mathcal{F}, \mu)$, the measure μ satisfies the following statements:*

1. ***(finite additivity)*** *Given n disjoint sets $A_1, \ldots, A_n$ in $\mathcal{F}$,* $\mu(\bigsqcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i)$.

2. ***(monotonicity)*** $\forall A, B \in \mathcal{F}$ s.t.$A \subseteq B,\ \mu(A) \leq \mu(B)$.
3. ***(measure of difference set)*** $\forall A, B \in \mathcal{F}$ s.t.$A \subseteq B,\ \mu(A) < \infty,\ \mu(B \backslash A) = \mu(B) - \mu(A)$.
4. ***(inclusion-exclusion principle)*** $\forall A, B \in \mathcal{F}$ s.t.$\mu(A \cap B) < +\infty,\ \mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.

Proof. 1. Augment the finite collection of subsets to a countable collection by letting $A_i = \emptyset$ for all $i > n$. It is obvious that the subsets in the augmented collection are all in $\mathcal{F}$ and mutually disjoint, thus countable additivity can be applied as follows,

$$\mu\left(\bigsqcup_{i=1}^{n} A_i\right) = \mu\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) = \sum_{i=1}^{n} \mu(A_i).$$

The last equality is from the fact that $\mu(A_i) = \mu(\emptyset) = 0, \forall i > n$.

2. We can write B as $B = A \sqcup (B \backslash A)$, where $B \backslash A \in \mathcal{F}$ because $A, B \in \mathcal{F}$. From the finite additivity proved above, we have

$$\mu(B) = \mu(A) + \mu(B \backslash A) \geq \mu(A),$$

by nonnegativity of measure.

3. From the formula above, we have

$$\mu(B \backslash A) = \mu(B) - \mu(A)$$

whenever $\mu(A) < \infty$. Note that $\mu(A) < \infty$ is indispensable as $\infty - \infty$ is undefined in the extended real system.

4. Rewrite $A \cup B$ as

$$A \cup B = A \sqcup (B \backslash A) = A \sqcup (B \backslash (A \cap B)).$$

Clearly, $B \backslash (A \cap B) \in \mathcal{F}$ and it has no overlap with A. By finite additivity and the measure of difference set proved above, we have

$$\mu(A \cup B) = \mu(A) + \mu(B \backslash (A \cap B)) = \mu(A) + \mu(B) - \mu(A \cap B).$$

Here, the condition that $\mu(A \cap B) < \infty$ is necessary due to the reason given above.

□

1.2.4 Continuity of Measure

To study the continuity of a measure, we first define the monotonical sequences of sets.

Definition 1.5 (Monotonical sequence of sets). *A sequence of sets $(E_n)_{n=1}^{\infty}$ is said to **increases to** E if $\forall i \in \mathbb{N}, E_i \subseteq E_{i+1}$, and $\bigcup_{i=1}^{\infty} E_i = E$, denoted by $E_i \uparrow E$. A sequence of sets $(E_n)_{n=1}^{\infty}$ is said to **decreases to** E if $\forall i \in \mathbb{N}, E_i \supseteq E_{i+1}$, and $\bigcap_{i=1}^{\infty} E_i = E$, denoted by $E_i \downarrow E$.*

Lemma 1.1 (Continuity of measure). *Given a measure space $(X, \mathcal{F}, \mu)$, let $(E_i)_{i=1}^{\infty}$ be a sequence of sets in $\mathcal{F}$, then*

1. $E_n \uparrow E \Rightarrow \mu(E) \Rightarrow \lim_{n\to\infty} \mu(E_n)$;
2. $E_n \downarrow E$, *and* $\exists E_i$ s.t. $\mu(E_i) < \infty \Rightarrow \mu(E) = \lim_{n\to\infty} \mu(E_n)$

Proof. 1. This statement is proved by three steps: express the limit set as a disjoint union → write its union as a series by countable additivity → prove the convergence of the series.

   (a) First, we prove that

$$E = \bigcup_{i=1}^{\infty} E_i = E_1 \sqcup \left(\bigsqcup_{i=1}^{\infty} E_{i+1} \backslash E_i\right) \tag{1.4}$$

   For conciseness, we denote the set defined in the right hand by $\tilde{E}$. To establish the equality, we show both $\tilde{E} \subseteq E$ and $E \subseteq \tilde{E}$. The former follows directly from the fact that $E_1 \subseteq E$ and $E_i \backslash E_{i-1} \subseteq E_i \subseteq E$ for all $i = 1, 2, \ldots$.

Now, we prove the other direction. For arbitrary $x \in E$, there exists E_i such that $x \in E_i$. Let k be the minimum integer such that $x \in E_k$. If $k = 1$, then $x \in E_1$, otherwise $k > 1$ and $x \in E_k \backslash E_{k-1}$. In both cases, it is obvious that $x \in E$. As x is arbitrarily choosen from E, we have $E \subseteq \tilde{E}$.

Furthermore, we need to verify that the sets in the right hand side are disjoint. For each positive integer i, we have $E_1 \subseteq E_i$, it follows that

$$E_1 \cap (E_{i+1} \backslash E_i) = (E_1 \cap E_i^c) \cap E_{i+1} = \emptyset \cap E_{i+1} = \emptyset.$$

In addition, for any $i < j \in nsp$, we have $E_i \subseteq E_{i+1} \subseteq E_j$, and thus

$$(E_{i+1} \backslash E_i) \cap (E_{j+1} \backslash E_j) = E_i^c \cap (E_{i+1} \cap E_j^c) \cap E_{j+1} = E_i^c \cap \emptyset \cap E_{j+1} = \emptyset.$$

Therefore, the sets $\{E_1, E_{i+1} \backslash E_i, \forall i \in \mathbb{N}\}$ are mutually disjoint.

(b) Based on (1.4), we have the following by countable additivity

$$\mu(E) = \mu(E_1) + \sum_{i=1}^{\infty} \mu(E_{i+1} \backslash E_i) \tag{1.5}$$

On the other hand, it can be easily shown by induction that

$$E_k = E_1 \sqcup \left(\bigsqcup_{i=1}^{k-1} E_{i+1} \backslash E_i \right). \tag{1.6}$$

by finite additivity, we have

$$\mu(E_k) = \mu(E_1) + \sum_{i=1}^{k-1} \mu(E_{i+1} \backslash E_i). \tag{1.7}$$

Comparing (1.5) and (1.7), we can see that $\mu(E_k)$ is a finite partial sum of the series $\mu(E)$.

(c) When $\mu(E) < +\infty$, the series (1.4) is absolutely convergent (due to nonnegativity of measure), it follows that $\lim_{k\to\infty} \mu(E_k) \to \mu(E)$. Otherwise, $\mu(E) \to \infty$, then given any $v > 0$, there is $k \in \mathbb{N}$ such that the partial sum in (1.7) is larger than v. Hence, $\lim_{k\to\infty} \mu(E_k) = \infty = \mu(E)$.

2. Because $E_i \supseteq E_{i+1}$, $\forall i$, it can be easily shown that for any k,

$$E = \bigcap_{i=1}^{\infty} E_i = \bigcap_{i=k}^{\infty} E_i.$$

Since there exists k_0 such that $\mu(E_{k_0}) < \infty$ and thus $\mu(E_i) < \infty$ for all $i \geq k_0$, we can express the intersection of any sequence of the given condition as the intersection of sequence of finite-measure sets. Hence, we assume $\mu(E_1) < +\infty$ without lossing generality.

(a) First of all, we prove the following identity

$$E = E_1 \cap \left(\bigcap_{i=1}^{\infty} (E_i \backslash E_{i+1})^c \right) = E_1 \backslash \left(\bigcup_{i=1}^{\infty} (E_i \backslash E_{i+1}) \right).$$

We only need to prove the first equality, while the second one immediately follows.
Let $G = \bigcap_{i=1}^{\infty} (E_i \backslash E_{i+1})^c$, then $E \subseteq E_1 \cap G$ follows from the fact that $E \subseteq E_1$ and $E \subseteq E_{i+1} \subseteq (E_i \backslash E_{i+1})^c$. Now, we prove the other direction. For arbitrary $x \in E_1 \cap G$, we have $x \in E_1$ and $x \in E_i \backslash E_{i+1}$ for all $i \in \mathbb{N}$. It can be easily shown by induction that

$$E_k = E_1 \cap \left(\bigcap_{i=1}^{k-1} (E_i \backslash E_{i+1})^c \right). \tag{1.8}$$

Hence, $x \in E_k$ for all $k \in \mathbb{N}$. As a result, $x \in E = \bigcap_{i=1}^{\infty} E_i$, and thus $E_1 \cap G \subseteq E$.

In addition, for all $i < j \in \mathbb{N}$ we have $(E_i \backslash E_{i+1}) \cap (E_j \backslash E_{j+1})$ due to $E_j \subseteq E_{i+1}$. Hence, the sets in form of $E_i \backslash E_{i+1}$ are disjoint. Furthermore, $\bigcup_{i=1}^{\infty}(E_i \backslash E_{i+1}) \subseteq E_1$ due to $E_i \backslash E_{i+1} \subseteq E_i \subseteq E_1$. Therefore, (1.8) can be rewritten as

$$E_1 \backslash E = \bigsqcup_{i=1}^{\infty} (E_i \backslash E_{i+1}). \tag{1.9}$$

And from (1.8), we have

$$E_1 \backslash E_k = \bigsqcup_{i=1}^{k-1} (E_i \backslash E_{i+1}). \tag{1.10}$$

(b) From (1.9) and (1.10), we have $(E_1 \backslash E_k) \uparrow (E_1 \backslash E)$. Applying the conclusion for increasing sequence sets that we have just proved, we get

$$\mu(E_1 \backslash E) = \lim_{n \to \infty} \mu(E \backslash E_n). \tag{1.11}$$

As $E \subseteq E_1$, $E_n \subseteq E_1$ for all $n \in \mathbb{N}$,and $\mu(E_1) < \infty$, the measure of E and $E_n, \forall n$ are finite. Then, the above formula can be further written into

$$\mu(E_1) - \mu(E) = \lim_{n \to \infty} (\mu(E_1) - \mu(E_n)) = \mu(E_1) - \lim_{n \to \infty} \mu(E_n) \tag{1.12}$$

It immediately leads to $\mu(E) = \lim_{n \to \infty} \mu(E_n)$.

□

It is important to note that the existence of a finite-measure set E_i is crucial for the second statement in the above lemma. Here, I give an example. Consider the sequence of sets $(E_i)_{i=1}^{\infty}$ with $E_i = (i, \infty)$. Clearly, $\mu(E_i) = \infty$ for each $i \in \mathbb{N}$, however, this sequences obviously decreases to $E = \emptyset$, and thus $\mu(E) = 0$, which is not the limit of $\mu(E_i)$.

1.3 Lebesgue Measure

In the following, we are trying to construct a measure on $\mathbb{R}$, which is called the *Lebesgue measure*, denoted by m. The goal is that Lebesgue measure should satisfy the following properties

1. $m(\text{interval}) = \text{len}(\text{interval})$;
2. translation invariance: $m(x + A) = m(A)$;
3. σ-additivity.

The plan of constructing Lebesgue measure comprises three stages:

1. define the measure m on some simple sets, i.e. intervals;
2. extend the definition to all reasonable sets by approximating them with finite or countable union of intervals;
3. restrict to a certain σ-algebra to achieve σ-additivity.

1.3.1 The Measure of Intervals

Definition 1.6 (Semi-algebra)**.** *A collection of subsets $S \subseteq 2^X$ is called a* ***semi-algebra*** *over X if it satisfies*

1. *$\emptyset \in S$ and $X \in S$;*
2. *closed under intersection: $A, B \in S \Rightarrow A \cap B \in S$;*

3. *if $A \in S$, then A^c can be expressed as finite union of sets in S, as*

$$X \backslash A = \bigcup_{i=1}^{N} B_i$$

where $B_i \in S, \forall i = 1, \ldots, N$.

Consider the set of all intervals in $\mathbb{R}$, which is defined as

$$\mathrm{Int} = \{I \text{ is in form of } \langle a, b\rangle | a, b \in [-\infty, \infty], a \leq b\}.$$

Here, we say that a set $I \subseteq \mathbb{R}$ is form of $\langle a, b\rangle$ if I is either of the following: $[a, b]$, $(a, b]$, $[a, b)$ or (a, b).

Proposition 1.3. *The set of intervals* Int *is a semi-algebra.*

Proof. 1. $\emptyset = (a, a)$ for arbitrary $a \in \mathbb{R}$, thus $\emptyset \in \mathrm{Int}$, and $\mathbb{R} = (-\infty, +\infty) \in \mathrm{Int}$.

2. Given arbitrarily two intervals respectively in form of $\langle a_1, b_1\rangle$ and $\langle a_2, b_2\rangle$, let $a' = \max(a_1, a_2)$ and $b' = \min(b_1, b_2$. If $a' \leq b'$, their intersection is in form of $\langle a', b'\rangle$, otherwise their intersection is $\emptyset$. In both cases, the intersection is in Int, thus Int is closed under intersection.

3. For $I = [a, b] \in \mathrm{Int}$, we have $I^c = (-\infty, a) \cup (b, \infty) \in \mathrm{Int}$. Similar argument shows that the third condition also applies to other interval forms: $[a, b)$, $(a, b]$ and (a, b).

□

We define the length of an interval by $l(\langle a, b\rangle) = b - a$.

Lemma 1.2. *The length of intervals defined over* Int *satisfies σ-additivity, that is if $I \in \mathrm{Int}$ and $I_1, I_2, \ldots \in$* Int*, and $\{I_i\}_{i=1}^{\infty}$ are disjoint, then*

$$I = \bigsqcup_{k=1}^{\infty} I_k \Rightarrow l(I) = \sum_{k=1}^{\infty} l(I_k). \tag{1.13}$$

Proof. 1. First of all, we proof that the length satisfies finite additivity, that is

$$\forall I, I_k \in S,\ I = \bigsqcup_{k=1}^{N} I_k \Rightarrow l(I) = \sum_{k=1}^{N} l(I_k).$$

Let I be in form of $\langle a, b\rangle$ with $a \leq b$ and I_k be in form of $\langle a_k, b_k\rangle$ with $a_k \leq b_k$. In the case where $I = \sqcup_{k=1}^{N} I_k = \emptyset$, we must have $I = \emptyset$ and $I_k = \emptyset, \forall k = 1, \ldots, N$. Then, $l(I) = 0$ and $\sum_{k=1}^{N} l(I_k) = \sum_{k=1}^{N} 0 = 0$. The statement holds. In the following, we assume $I \neq \emptyset$, and $I_k \neq \emptyset$ for all $k = 1, \ldots, N$. (If there are some empty sets in $\{I_k\}_{k=1}^{N}$ we can simply remove them without affecting the equality as the length of empty sets are zeros.)

We rearrange I_k in the order of a_k such that $a_1 \leq a_2 \leq \cdots \leq a_N$. Now we show that this ordered sequence of intervals satisfies the following properties.

(a) $a_1 = a$. If $a_1 < a$, then for any $x \in (a_1, a)$, $x \in I_1$ but $x \notin I$; if $a_1 > a$, then for any $x \in (a, a_1)$, $x \in I$ but $x \notin I_k, \forall k = 1, \ldots, N$, and thus x is not in their union. Hence, for $I = \bigsqcup_{k=1}^{N} I_k$, it is necessary that $a_1 = a$.

(b) $b_N = b$. The proof of this is similar to above. If $b_N < b$, then for any $x \in (b_N, b)$, $x \in I$ but $x \notin \bigsqcup_{k=1}^{N} I_k$; or if $b_N > b$, then for any $x \in (b, b_N)$, $x \in I_k$ but $x \notin I$.

(c) $b_k = a_{k+1}$ for all $k = 1, \ldots, N-1$. If $b_k < a_{k+1}$, then for any $x \in (b_k, a_{k+1})$, it has $a = a_1 \leq a_k \leq b_k < x < a_{k+1} \leq a_N \leq b_N = b$, thus $x \in I$, but it is clear that x is not in I_k for any k. If $b_k > a_{k+1}$, then the set (a_{k+1}, b_k), which is contained in both I_k and I_{k+1}, is not empty. This violates the condition that I_k and I_{k+1} are disjoint.

Based on the three properties, we have

$$\sum_{i=1}^{N} l(I_k) = \sum_{i=1}^{N} (b_k - a_k) = -a_1 + \sum_{k=1}^{N-1} (b_k - a_{k+1}) + b_N = b_N - a_1 = b - a = l(I). \tag{1.14}$$

Hence, the finite additivity is established.

2. Then, we prove countable sub-additivity.That is to prove

$$\forall I, I_k \in \text{Int}, I = \bigsqcup_{k=1}^{\infty} I_k \Rightarrow l(I) = \sum_{k=1}^{\infty} l(I_k).$$

In doing this, we first prove it for closed and bounded interval $I = [a, b]$. Let I_k be in form of $\langle a_k, b_k \rangle$ with $a_k \leq b_k$. Given arbitrary $\epsilon > 0$, we set $I'_k = (a'_k, b'_k)$ with $a'_k = a_k - 2^{-(k+1)}\epsilon$ and $b'_k = b_k + 2^{-(k+1)}\epsilon$. Clearly, $I_k \subset I'_k$, and thus $I \subseteq \bigcap_{k=1}^{\infty} I'_k$.

The collection $\{I'_k\}_{k=1}^{\infty}$ constitutes a open cover of I. And since $I = [a, b]$ is closed and bounded, there is an finite sub-cover of I due to Heine-Borel theorem. For convenience of notation, we denote the sets in this finite sub-cover by $\{J_k\}_{k=1}^{N}$, which are respectively (c_k, d_k).

Then for any $x \in I$, there exists some $1 \leq k \leq N$ such that $x \in J_k$. From the finite sub-cover, we construct a sub collection as follows. In the first step, we choose k_1 such that $a \in J_{k_1}$. The selection procedure continues as follows, after selecting J_{k_i}, if $d_{k_i} > b$ then it is done, otherwise we select k_{i+1} such that $d_{k_i} \in J_{k_{i+1}}$. As the sub-cover is finite, this procedure can be finished with finite steps. It is easy to see that the selected collection has: $c_{k_1} < a$, $c_{k_i} < c_{k_{i+1}} < d_{k_i}, \forall i = 1, \ldots, N' - 1$, and $d_{k_{N'}} > b$, where N' is the number of the selected sets. So, we have

$$\sum_{i=1}^{N'} l(J_{k_i}) = \sum_{i=1}^{N'} (d_{k_i} - c_{k_i}) = -c_{k_1} + \sum_{i=1}^{N'-1} (d_k - c_{k+1}) + d_{k_{N'}} > d_{k_{N'}} - c_{k_1} > b - a = l([b, a]). \quad (1.15)$$

On the other hand,

$$\sum_{i=1}^{N'} l(J_{k_i}) \leq \sum_{k=1}^{N} l(J_k) \leq \sum_{k=1}^{\infty} l(I'_k) = \sum_{k=1}^{\infty} (l(I_k) + 2^{-k}\epsilon) = \left(\sum_{k=1}^{\infty} l(I_k) \right) + \epsilon. \quad (1.16)$$

Combining the above two inequalities, we have

$$l([b, a]) < \left(\sum_{k=1}^{\infty} l(I_k) \right) + \epsilon. \quad (1.17)$$

This holds for arbitrary positive number ϵ, hence

$$l([b, a]) \leq \sum_{k=1}^{\infty} l(I_k). \quad (1.18)$$

Now, the sub-additivity is established for any close interval $[a, b]$. we then continue to show this for other forms of intervals.

For $I = [a, b) = \bigsqcup_{k=1}^{\infty} I_k$, and consider a family of close sets $A_\epsilon = [a, b - \epsilon]$ where $0 < \epsilon < b - a$. It is easy to see that $\bigsqcup_{k=1}^{\infty} I_k$ covers A_ϵ defined above. Since A_ϵ is a close interval, from the sub-additivity for close interval proved above, we have

$$b - a - \epsilon = l(A_\epsilon) \leq \sum_{k=1}^{\infty} l(I_k), \ \forall \ \epsilon \text{ s.t. } 0 < \epsilon < b - a \quad (1.19)$$

thus

$$\sum_{k=1}^{\infty} l(I_k) \geq b - a = l(I). \quad (1.20)$$

With similar argument, we can as well establish sub-additivity for $(a, b]$ and (a, b).

3. We finally prove inequality in opposite direction, that is

$$\forall I, I_k \in \text{Int}, I = \bigsqcup_{k=1}^{\infty} I_k \Rightarrow l(I) \geq \sum_{i=1}^{\infty} l(I_k).$$

First of all, we show that for any $N > 0$, I can be written as

$$I = \left(\bigsqcup_{k=1}^{N} I_k\right) \bigsqcup \left(\bigsqcup_{i=1}^{M} J_i\right) \tag{1.21}$$

for some disjoint intervals $J_1, \ldots, J_M \in \mathrm{Int}$. From De-Morgan's rule,

$$I \backslash \left(\bigcup_{k=1}^{N} I_k\right) = I \cap \left(\bigcup_{k=1}^{N} I_k\right)^c = I \cap \left(\bigcap_{k=1}^{N} I_k^c\right). \tag{1.22}$$

As Int is a semi-algebra, for each I_k we have

$$I_k^c = \bigsqcup_{j=1}^{n_k} B_{k,j},$$

for some disjoint intervals $B_{k,j} \in \mathrm{Int}$, and thus

$$\bigcap_{k=1}^{N} I_k^c = \bigcap_{k=1}^{N} \left(\bigcup_{j=1}^{n_k} B_{k,j}\right) = \bigcup_{j_1=1}^{n_1} \cdots \bigcup_{j_N=1}^{n_N} (B_{1,j_1} \cap B_{2,j_2} \cap \cdots \cap B_{N,j_N}) \tag{1.23}$$

For two distinct terms in this union, $B_{j_1} \cap \cdots \cap B_{j_N}$ and $B_{j_1'} \cap \cdots \cap B_{j_N'}$, there exists l such that $j_l \neq j_l'$, thus $B_{l,j_l} \cap B_{l,j_l'} = \emptyset$, and these two distinct terms are disjoint. Hence, we can write this union by rearranging tht terms

$$\bigcap_{k=1}^{N} I_k^c = \bigsqcup_{i=1}^{M} C_i. \tag{1.24}$$

Here, $M = n_1 n_2 \cdots n_N$. C_i are derived by reindexing the terms in (1.23), and thus they are disjoint. Pluging this back into (1.22), we have

$$I \backslash \left(\bigcup_{k=1}^{N} I_k\right) = I \cap \left(\bigsqcup_{i=1}^{M} C_i\right) = \bigsqcup_{i=1}^{M} (I \cap C_i) = \bigsqcup_{i=1}^{M} J_i, \tag{1.25}$$

where $J_i = I \cap C_i \in \mathrm{Int}$ for $i = 1, \ldots, M$. Note that I_k are mutually disjoint, and from (1.25), $\sqcup_{i=1}^{M} J_i$ are contained in I and disjoint from $\sqcup_{k=1}^{N} I_k$, so

$$I = \left(\bigsqcup_{k=1}^{N} I_k\right) \bigsqcup \left(\bigsqcup_{i=1}^{M} J_i\right), \tag{1.26}$$

where $J_i \in \mathrm{Int}$ for $i = 1, \ldots, M$. The identity of (1.21) is established. Then, from finite additivity proved above, we have

$$l(I) = \sum_{i=1}^{N} l(I_k) + \sum_{j=1}^{M} l(J_k) \geq \sum_{i=1}^{N} l(I_k). \tag{1.27}$$

Note that this holds for any $N > 0$. Taking the limit as $N \to \infty$, we get

$$l(I) \geq \lim_{N \to \infty} \sum_{k=1}^{N} l(I_k) = \sum_{k=1}^{\infty} l(I_k). \tag{1.28}$$

Now, both direction of the inequality has been proved, as a result, the countable additivity of length in Int is thus established.

□

Corollary 1.1. *Let $I \in \mathrm{Int}$ be an interval, and $\{J_i\}_{i=1}^{\infty}$ with $J_i \in \mathrm{Int}$ for all $i \in \mathbb{N}$ be a cover of I, i.e. $I \subseteq \bigcup_{i=1}^{\infty} J_i$, then*

$$l(I) \leq \sum_{i=1}^{\infty} l(J_i).$$

Proof. Let $K_1 = J_1$ and $K_i = J_i \backslash J_{i-1}$ for $i \geq 2$. It is easy to see that the sets in $\{K_i\}_{i=1}^{\infty}$ are disjoint. And, since Int is a semi-algebra, each K_i can be written into a finite disjoint union of intervals as

$$K_i = \bigsqcup_{j=1}^{n_i} L_{i,j}. \tag{1.29}$$

We extend the definition of length to finite disjoint union of intervals by $l(\bigsqcup_{k=1}^{n} I_k) = \sum_{k=1}^{n} l(I_k)$, then

$$l(K_i) = \sum_{j=1}^{n_i} l(L_{i,j}). \tag{1.30}$$

On the other hand, $J_i = (J_i \cap J_{i-1}) \sqcup K_i$, and $J_i \cap J_{i-1} \in \text{Int}$, thus by additivity of length,

$$l(J_i) = l(J_i \cap J_{i-1}) + l(K_i) \geq l(K_i). \tag{1.31}$$

Now, as $I \subseteq \bigsqcup_{i=1}^{\infty} K_i$, we can write

$$I = \bigsqcup_{i=1}^{\infty} (I \cap K_i) = \bigsqcup_{i=1}^{\infty} \bigsqcup_{j=1}^{n_i} (I \cap L_{i,j}). \tag{1.32}$$

Here, $I \cap L_{i,j}$ are intervals, then by σ-additivity of lengths, we have

$$l(I) = \sum_{i=1}^{\infty} \sum_{j=1}^{n_i} l(I \cap L_{i,j}) \leq \sum_{i=1}^{\infty} \sum_{j=1}^{n_i} l(L_{i,j}) = \sum_{i=1}^{\infty} l(K_i) \leq \sum_{i=1}^{\infty} l(J_i). \tag{1.33}$$

□

1.3.2 Outer Measure

Definition 1.7 (Outer measure)**.** *A function $\mu^* : 2^X \to [0, \infty]$ is said to be an **outer measure** over X if it satisfies*

1. $\mu^*(\emptyset) = 0$;
2. $\mu^*(A) \leq \mu^*(B)$ *whenever* $A \subseteq B$;
3. ***(sub-additivity)*** $\mu^*(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$.

Definition 1.8 (Lebesgue's outer measure)**.** ***Lebesgue's outer measure*** *is a function $m^* : 2^X \to [0, \infty]$ given by*

$$m^*(A) = \inf \left\{ \sum_{k=1}^{\infty} l(I_k) \mid A \subseteq \bigcup_{i=1}^{\infty} I_k, I_k \in \text{Int} \right\}.$$

Proposition 1.4. *Lebesgue's outer measure is an outer measure over $\mathbb{R}$.*

Proof. We proof Lebesgue's outer measure satisfies the three conditions of outer measure.

1. First of all, Lebesgue's outer measure is non-negative, which follows from the non-negativity of interval length. For any $a \in \mathbb{R}$, we can write
$$\emptyset = \bigcup_{i=1}^{\infty} A_i,$$
with $A_i = (a, a)$. Then,
$$m^*(\emptyset) \leq \sum_{i=1}^{\infty} l(A_i) = \sum_{i=1}^{\infty} 0 = 0.$$
So, $m^*(\emptyset) = 0$.
2. When $A \subseteq B$, any cover of B is also a cover of A, then based on the definition of m^* as an infimium, $m^*(A) \leq m^*(B)$ directly follows.

3. Given arbitrary $\epsilon > 0$, for any $i \in \mathbb{N}$, there exists $I_i^{(j)} \subseteq X$ for all $j = 1, \ldots, \infty$ such that

$$A_i \subseteq \bigcup_{j=1}^{\infty} I_i^{(j)},$$

and

$$\sum_{j=1}^{\infty} m^*(I_i^{(j)}) < m^*(A_i) + \frac{\epsilon}{2^i}.$$

Hence,

$$\bigcup_{i=1}^{\infty} A_i \subseteq \bigcup_{i=1}^{\infty} \bigcup_{j=1}^{\infty} I_i^{(j)},$$

and

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} m^*(I_i^{(j)}) < \sum_{i=1}^{\infty} \left(m^*(A_i) + \frac{\epsilon}{2^i} \right) = \sum_{i=1}^{\infty} m^*(A_i) + \epsilon.$$

Note that countable union of countable set is countable, then from the definition of m^*, we have

$$m^* \left(\bigcup_{i=1}^{\infty} A_i \right) < \sum_{i=1}^{\infty} m^*(A_i) + \epsilon,$$

for arbitrary $\epsilon > 0$. Hence,

$$m^* \left(\bigcup_{i=1}^{\infty} A_i \right) \leq \sum_{i=1}^{\infty} m^*(A_i).$$

□

The following lemma shows that the definition of outer measure is consistent with the length of intervals.

Lemma 1.3. *Let $A = \sqcup_{k=1}^{\infty} I_k$ where $I_k \in \mathrm{Int}$ for all $k \in \mathbb{N}$, then $m^*(A) = \sum_{i=1}^{\infty} l(I_k)$. In particular, for each interval $I \in \mathrm{Int}$, we have $m^*(I) = l(I)$.*

Proof. First, $m^*(A) \leq \sum_{i=1}^{\infty} l(I_k)$ immediately follows from the definition of outer measure. So, to establish the equality, it suffices to show

$$m^*(A) \geq \sum_{i=1}^{\infty} l(I_k).$$

For arbitrary $\epsilon > 0$, there exists a cover of of A by $\{J_i\}_{i=1}^{\infty}$ with $J_i \in \mathrm{Int}$, $\forall i \in \mathbb{N}$, such that

$$\sum_{i=1}^{\infty} l(J_i) \leq m^* \left(\bigsqcup_{k=1}^{\infty} I_k \right) + \epsilon. \tag{1.34}$$

Since $A = \sqcup_{k=1}^{\infty} I_k \subseteq \bigcup_{i=1}^{\infty} J_i$, we have $I_k \subseteq \bigcup_{i=1}^{\infty} J_i \cap I_k$. Note here that each $J_i \cap I_k$ is also an interval, hence

$$l(I_k) \leq \sum_{i=1}^{\infty} l(J_i \cap I_k), \quad \forall k \in \mathbb{N}. \tag{1.35}$$

It follows that

$$\sum_{k=1}^{\infty} l(I_k) \leq \sum_{k=1}^{\infty} \sum_{i=1}^{\infty} l(J_i \cap I_k) = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} l(J_i \cap I_k). \tag{1.36}$$

(**TODO:** the exchange of summation order need to be justified).

Note that $J_i \supseteq \bigsqcup_{k=1}^{\infty} (J_i \cap I_k)$, we have

$$l(J_i) \geq \sum_{k=1}^{\infty} l(J_i \cap I_k). \tag{1.37}$$

Combining (1.36), (1.37), and (1.34), we get

$$\sum_{k=1}^{\infty} l(I_k) \leq \sum_{i=1}^{\infty} l(J_i) \leq m^* \left(\bigsqcup_{k=1}^{\infty} I_k \right) + \epsilon. \tag{1.38}$$

As $\epsilon > 0$ is arbitrary,

$$\sum_{k=1}^{\infty} l(I_k) \leq m^* \left(\bigsqcup_{k=1}^{\infty} I_k \right) = m^*(A). \tag{1.39}$$

□

Proposition 1.5 (Translation-Invariance of m^*)**.** *The Lebesgue's outer measure m^* is translation invariant, i.e. for any $E \subseteq \mathbb{R}$ and any $x \in \mathbb{R}$, $m^*(E) = m^*(E + x)$.*

Proof. Note that fact that $\{I_k\}_{k=1}^{\infty}$ is a cover of $E \subseteq \mathbb{R}$, if and only if $\{I_k + x\}_{k=1}^{\infty}$ is a cover of $E + x$. In addition, length of intervals is translation invariant, i.e. $l(I + x) = l(I)$ for any $I \in \text{Int}$ and $xin\mathbb{R}$. Then, we have for any $x \in \mathbb{R}$,

$$\begin{aligned} m^*(E) &= \inf \left\{ \sum_{k=1}^{\infty} l(I_k) | E \subseteq \bigcup_{k=1}^{\infty} I_k, I_k \in \text{Int} \right\} \\ &= \inf \left\{ \sum_{k=1}^{\infty} l(I_k) | E + x \subseteq \bigcup_{k=1}^{\infty} (I_k + x), I_k \in \text{Int} \right\} \\ &= \inf \left\{ \sum_{k=1}^{\infty} l(I_k + x) | E + x \subseteq \bigcup_{k=1}^{\infty} (I_k + x), I_k \in \text{Int} \right\} \\ &= m^*(E + x). \end{aligned} \tag{1.40}$$

□

Proposition 1.6. *The Lebesgue's outer measure of countable set in $\mathbb{R}$ is 0.*

Proof. Let $A = \{x_1, x_2, \ldots\} \in \mathbb{R}$ be a countable set, given arbitrary $\epsilon > 0$, we can construct a collection of intervals that cover A as follows. Let $\mathcal{C}_\epsilon = \{I_k\}$ with $I_k = (x_k - 2^{-k}\epsilon, x_k + 2^{-k}\epsilon)$. It is obvious that $A \subseteq \bigcup_{k=1}^{\infty} I_k$. The total length of this collection is

$$\sum_{i=1}^{\infty} l(I_k) = 2 \sum_{i=1}^{\infty} 2^{-k}\epsilon = 2\epsilon. \tag{1.41}$$

Hence, $m^*(A) \leq 2\epsilon$ for any $\epsilon > 0$, as a result, $m^*(A) = 0$. □

As immediate corollary of this proposition, we have $m^*(\mathbb{Q}) = 0$ and the outer measure of countable union of countable sets is zero.

Proposition 1.7. *The Lebesgue's measure of Cantor set is zero.*

Proof. Here gives a brief sketch of the proof. Consider the construction of Cantor set as an infinite refinement process. At the k-step, it ends up with 2^k intervals with total length $\left(\frac{2}{3}\right)^k$. The intervals derived after any finite steps form a cover of the Cantor set. Therefore, by the definition of m^*, we have $m^*(C) \leq \left(\frac{2}{3}\right)^k$ for any $k \in \mathbb{N}$, where C denotes the Cantor set. Taking $k \to \infty$ results in $m^*(C) = 0$. □

1.3.3 From Outer Measure to Lebesgue Measure

We have defined an outer measure and shown its consistency with the definition of length on intervals. In the final step, we are going to restrict the domain of the outer measure to achieve σ-additivity, which eventually leads to the *Lebesgue measure* on $\mathbb{R}$.

The restricted domain is characterized by the following criterion.

Definition 1.9 (Lebesgue measurable set)**.** *A subset $E \subseteq \mathbb{R}$ is said to be* ***Lebesgue measurable*** *if it satisfies the* ***Carathéodory's condition*** *given by*

$$\forall T \subseteq \mathbb{R}, \quad m^*(T) = m^*(T \cap E) + m^*(T \cap E^c).$$

In the following, we use $\mathcal{B}_0$ to denote the collection of all Lebesgue measurable sets for conciseness. And, we denote the restriction of the outer measure to $\mathcal{B}_0$ by m, i.e. $m = m^*|_{\mathcal{B}_0}$. We will show later that $\mathcal{B}_0$ is a σ-algebra, and m is a measure over $\mathcal{B}_0$.

Theorem 1.1 (Properties of Lebesgue measurable sets)**.** *The Lebesgue measurable sets defined above have*

1. *$\emptyset, \mathbb{R} \in \mathcal{B}_0$ with $m(\emptyset) = 0$ and $m(\mathbb{R}) = \infty$.*
2. *$E \in \mathcal{B}_0 \Rightarrow E^c \in \mathcal{B}_0$.*
3. *$E_1, E_2 \in \mathcal{B}_0 \Rightarrow E_1 \cup E_2,\ E_1 \cap E_2 \in \mathcal{B}_0$, which immediately follows that $E_1 \backslash E_2 \in \mathcal{B}_0$. By induction, we further have $\mathcal{B}_0$ is closed under finite intersection and finite union.*
4. *Let $\{E_i\}_{i=1}^N \subset \mathcal{B}_0$ be disjoint, then $m^*\left(T \cap \bigsqcup_{i=1}^N E_i\right) = \sum_{i=1}^N m^*(T \cap E_i)$. This is a generalization of the Carathéodory's condition.*
5. *m is translation invariant, i.e. $E \in \mathcal{B}_0 \Rightarrow m(E + x) = m(E),\ \forall x \in \mathbb{R}$.*
6. *$\mathcal{B}_0$ is closed under countable union and intersection. $\{E_i\}_{i=1}^\infty \subset \mathcal{B}_0 \Rightarrow \bigcup_{i=1}^\infty E_i,\ \bigcap_{i=1}^\infty E_i \in \mathcal{B}_0$.*
7. *m is countable additive. $m\left(\bigsqcup_{i=1}^\infty E_i\right) = \sum_{i=1}^\infty m(E_i)$ for any disjoint collection of E_i in $\mathcal{B}_0$.*

The properties given above immediately leads to the definition of Lebesgue measure.

Definition 1.10 (Lebesgue measure)**.** *The collection $\mathcal{B}_0$ of all Lebesgue measurable sets is a σ-algebra, and m defined above is a measure over $\mathcal{B}_0$, called the* ***Lebesgue measure****. The triple $(\mathbb{R}, \mathcal{B}_0, m)$ is called the* ***Lebesgue measure space*** *on $\mathbb{R}$.*

Now, we prove the theorem.

Proof. Let $\mathcal{B}_0$ be Lebesgue measurable sets, i.e. the sets satisfying the Caratheòodory's condition, and m be the restriction of outer measure to $\mathcal{B}_0$.

1. For any $T \subseteq R$,
$$m^*(T \cap \emptyset) + m^*(T \cap \emptyset^c) = m^*(\emptyset) + m^*(T) = m^*(T), \tag{1.42}$$
and
$$m^*(T \cap \mathbb{R}) + m^*(T \cap \mathbb{R}^c) = m^*(T) + m^*(\emptyset) = m^*(T). \tag{1.43}$$
So, $\emptyset, \mathbb{R} \in \mathcal{B}_0$, and thus $m(\emptyset) = m^*(\emptyset) = 0$ and $m(\mathbb{R}) = m^*(\mathbb{R}) = \infty$.
2. For any $T \subseteq R$, and $E \in \mathcal{B}_0$,
$$m^*(T \cap E^c) + m^*(T \cap (E^c)^c) = m^*(T \cap E^c) + m^*(T \cap E) = m^*(T), \tag{1.44}$$
so $E^c \in \mathcal{B}_0$.
3. For any $T \subseteq R$, and $E_1, E_2, \in \mathcal{B}_0$, We first prove that $E_1 \cup E_2 \in \mathcal{B}_0$. Note that
$$T \cap (E_1 \cup E_2) = T \cap (E_1 \cup (E_2 \cap E_1^c)) = (T \cap E_1) \cup (T \cap E_1^c \cap E_2). \tag{1.45}$$
By sub-additivity of m^*, we have
$$m^*(T \cap (E_1 \cup E_2)) \leq m^*(T \cap E_1) + m^*(T \cap E_1^c \cap E_2). \tag{1.46}$$
On the other hand, that $E_2 \in \mathcal{B}_0$ results in
$$m^*(T \cap E_1^c \cap E_2) + m^*(T \cap E_1^c \cap E_2^c) = m^*(T \cap E_1^c). \tag{1.47}$$

Combining (1.46) and (1.47), and considering $E_1 \in \mathcal{B}_0$, we have

$$\begin{aligned}
&m^*(T \cap (E_1 \cup E_2)) + m^*(T \cap (E_1 \cup E_2)^c) \\
=&m^*(T \cap (E_1 \cup E_2)) + m^*(T \cap E_1^c \cap E_2^c) \\
\leq&m^*(T \cap E_1) + m^*(T \cap E_1^c \cap E_2) + m^*(T \cap E_1^c \cap E_2^c) \\
=&m^*(T \cap E_1) + m^*(T \cap E_1^c) \\
=&m^*(T).
\end{aligned} \tag{1.48}$$

The opposite direction of the inequality, that is

$$m^*(T) \leq m^*(T \cap (E_1 \cup E_2)) + m^*(T \cap (E_1 \cup E_2)^c),$$

directly follows from sub-additivity of m^*. Hence, the equality is established, and $E_1 \cup E_2 \in \mathcal{B}_0$.

While the closeness under intersection can be directly obtained based on that of set complement and union. The argument is briefly given as follows. Since $E_1, E_2 \in \mathcal{B}_0$, $E_1^c, E_2^c \in \mathcal{B}_0$, then $E_1 \cap E_2 = (E_1^c \cup E_2^c)^c \in \mathcal{B}_0$.

4. We prove this property by induction. When $N = 1$, the equality trivially holds. Suppose the property holds for $N = k$, then we show that it holds for $N = k + 1$. For conciseness, we let $G_k = \bigsqcup_{i=1}^k E_i$. As $E_{k+1} \in \mathcal{B}_0$,
$$m^*(T \cap G_{k+1}) = m^*(T \cap G_{k+1} \cap E_{k+1}) + m^*(T \cap G_{k+1} \cap E_{k+1}^c). \tag{1.49}$$
Here,
$$G_{k+1} \cap E_{k+1} = E_{k+1}, \quad \text{and} \quad G_{k+1} \cap E_{k+1}^c = G_k.$$
Then, we have
$$m^*(T \cap G_{k+1}) = m^*(T \cap G_k) + m^*(T \cap E_{k+1}) \tag{1.50}$$
Due to the induction assumption,
$$m^*(T \cap G_k) = \sum_{i=1}^{k} m^*(T \cap E_i), \tag{1.51}$$
Combining (1.50) and (1.51) leads to
$$m^*(T \cap G_{k+1}) = \sum_{i=1}^{k+1} m^*(T \cap E_i). \tag{1.52}$$
To sum up, this property holds for all $N \in \mathbb{N}$.

5. We first prove several identities for set translation,
$$(A + x) \cap (B + x) = (A \cap B) + x. \tag{1.53}$$
$$(A + x)^c = A^c + x. \tag{1.54}$$
$$(A + x) \cap (B + x)^c = (A \cap B^c) + x. \tag{1.55}$$
This can be shown by
$$p \in (A + x) \cap (B + x) \Leftrightarrow p - x \in A \cap B \Leftrightarrow p \in A \cap B + x. \tag{1.56}$$
and
$$p \in (A + x)^c \Leftrightarrow p \notin A + x \Leftrightarrow p - x \notin A \Leftrightarrow p - x \in A^c \Leftrightarrow p \in A^c + x. \tag{1.57}$$
The third identity immediately follows the first two.

Then, we have for any $E, T \in 2^{\mathbb{R}}$,
$$T \cap (E + x) = ((T - x) + x) \cap (E + x) = ((T - x) \cap E) + x. \tag{1.58}$$

and similarly, we have

$$T \cap (E+x)^c = ((T-x) \cap E^c) + x \tag{1.59}$$

Due to the translation invariance of m^*, we have

$$m^*(T \cap (E+x)) = m^*(((T-x) \cap E) + x) = m^*((T-x) \cap E) \tag{1.60}$$

and likewise

$$m^*(T \cap (E+x)^c) = m^*((T-x) \cap E^c). \tag{1.61}$$

Whenever $E \in \mathcal{B}_0$, we have for any T,

$$m^*(T\cap(E+x))+m^*(T\cap(E+x)^c) = m^*((T-x)\cap E)+m^*((T-x)\cap E^c) = m^*(T-x) = m^*(T). \tag{1.62}$$

This shows that $E+x \in \mathcal{B}_0$. Because m is a restriction of m^* and m^* is translation-invariant, it follows directly that

$$m(E+x) = m^*(E+x) = m^*(E) = m(E). \tag{1.63}$$

6. Given $\{E_i\}_{i=1}^{\infty}$, let $B_1 = E_1$ and $B_i = E_i \backslash \left(\bigcup_{j=1}^{i-1} E_i\right)$ for $i \geq 2$, then it is easy to see that $\{B_i\}_{i=1}^{\infty}$ is disjoint, and

$$\bigcup_{i=1}^{\infty} E_i = \bigsqcup_{i=1}^{\infty} B_i.$$

As $\mathcal{B}_0$ is closed under union and set difference, $B_i \in \mathcal{B}_0$ for all $i \in \mathbb{N}$.

In the following, we are going to show that the right hand side is Lebesgue measurable, that is to show for any $T \subseteq \mathbb{R}$,

$$m^*(T) = m^*\left(T \cap \left(\bigsqcup_{i=1}^{\infty} B_i\right)\right) + m^*\left(T \backslash \left(\bigsqcup_{i=1}^{\infty} B_i\right)\right). \tag{1.64}$$

The inequality ($\leq$) directly follows from σ-sub-additivity of m^*. To establish the equality, it suffices to show the other direction ($\geq$). For any $k \in \mathbb{N}$, $\bigsqcup_{i=1}^{k} B_i \in \mathcal{B}_0$, then

$$\begin{aligned} m^*(T) &= m^*\left(T \cap \left(\bigsqcup_{i=1}^{k} B_i\right)\right) + m^*\left(T \backslash \left(\bigsqcup_{i=1}^{k} B_i\right)\right) \\ &= \sum_{i=1}^{k} m^*(T \cap B_i) + m^*\left(T \backslash \left(\bigsqcup_{i=1}^{k} B_i\right)\right) \\ &\geq \sum_{i=1}^{k} m^*(T \cap B_i) + m^*\left(T \backslash \left(\bigsqcup_{i=1}^{\infty} B_i\right)\right). \end{aligned} \tag{1.65}$$

Here, the first equality is due to the 4th property proved above, while the second inequality is due to monotonicity of m^*. Since (1.65) holds for every $k \in \mathbb{N}$, taking $k \to \infty$, we get

$$\begin{aligned} m^*(T) &\geq \sum_{i=1}^{\infty} m^*(T \cap B_i) + m^*\left(T \backslash \left(\bigsqcup_{i=1}^{\infty} B_i\right)\right) \\ &\geq m^*\left(T \cap \left(\bigsqcup_{i=1}^{\infty} B_i\right)\right) + m^*\left(T \backslash \left(\bigsqcup_{i=1}^{\infty} B_i\right)\right). \end{aligned} \tag{1.66}$$

7. First, $m\left(\bigsqcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} m(E_i)$ directly follows the σ-sub-additivity of m^*. For the other direction of inequality, we have for arbitrary $N > 0$,

$$m\left(\bigsqcup_{i=1}^{\infty} E_i\right) \geq m\left(\bigsqcup_{i=1}^{N} E_i\right) = \sum_{i=1}^{N} m(E_i). \tag{1.67}$$

Here, the first inequality is due to the monotonicity of m^*, while the second equality follows from the 4th property proved above by letting $T = \mathbb{R}$.

Taking $N \to \infty$, we get

$$m\left(\bigsqcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} m(E_i). \tag{1.68}$$

$\square$

Proposition 1.8. *Intervals are Lebesgue measurable, i.e.* Int $\subset \mathcal{B}_0$.

Proof. Here, we need to prove that given $I \in \text{Int}$, for any $T \subseteq \mathbb{R}$, $m^*(T) = m^*(T \cap I) + m^*(T \cap I^c)$.

First, given arbitrary $\epsilon > 0$, there exist $\{I_k : I_k \in \text{Int}\}_{k=1}^{\infty}$ that covers T sucht that

$$m^*(T) \geq \sum_{k=1}^{\infty} l(I_k) - \epsilon = \sum_{k=1}^{\infty} m^*(I_k) - \epsilon. \tag{1.69}$$

As $T \cap I \subseteq \bigcup_{k=1}^{\infty}(I_k \cap I)$,

$$m^*(T \cap I) \leq \sum_{k=1}^{\infty} m^*(I_k \cap I). \tag{1.70}$$

Likewise,

$$m^*(T \cap I^c) \leq \sum_{k=1}^{\infty} m^*(I_k \cap I^c). \tag{1.71}$$

Combining (1.70) and (1.71),

$$m^*(T \cap I) + m^*(T \cap I^c) \leq \sum_{k=1}^{\infty} \left(m^*(I_k \cap I) + m^*(I_k \cap I^c)\right). \tag{1.72}$$

Note that $I_k \cap I \in \text{Int}$ for any $k \in \mathbb{N}$, and due to that Int is a semi-algebra, we can write $I_k \cap I^c = \bigsqcup_{i=1}^{n_k} J_{k,i}$ with $J_{k,i} \in \text{Int}$ for any k, i. Then,

$$\begin{aligned} m^*(I_k \cap I) + m^*(I_k \cap I^c) &= m^*(I_k \cap I) + \sum_{i=1}^{n_k} m^*(I_k \cap J_{k,i}) \\ &= l(I_k \cap I) + \sum_{i=1}^{n_k} l(I_k \cap J_{k,i}) \\ &= l\left((I_k \cap I) \sqcup \left(\bigsqcup_{i=1}^{n_k} J_{k,i}\right)\right) \\ &= l(I_k) = m^*(I_k). \end{aligned} \tag{1.73}$$

Plugging (1.73) into (1.72) and (1.69), we have

$$m^*(T \cap I) + m^*(T \cap I^c) \leq \sum_{k=1}^{\infty} m^*(I_k) \leq m^*(T) + \epsilon. \tag{1.74}$$

As $\epsilon > 0$ is arbitrary, it is necessary that

$$m^*(T \cap I) + m^*(T \cap I^c) \geq m^*(T). \tag{1.75}$$

The other direction of the inequality follows directly from m^*'s σ-sub-additivity. $\square$

Definition 1.11 (Borel algebra)**.** *The Borel algebra on* $\mathbb{R}$ *is the smallest sigma-algebra that includes all open subsets of* $\mathbb{R}$*, denoted by* $\mathcal{B}(\mathbb{R})$.

Proposition 1.9. *All sets in Borel algebra of* $\mathbb{R}$ *are Lebesgue measurable, i.e.* $\mathcal{B}(\mathbb{R}) \subset \mathcal{B}_0$.

To prove this proposition, we make use of the Lindelöf's Lemma.

Lemma 1.4 (Lindelöf's Lemma)**.** *Every open subset of* $\mathbb{R}$ *is a countable union of open intervals.*

Proof. Since $\mathcal{B}_0$ is a σ-algebra over $\mathbb{R}$, it suffices to prove that every open set is in $\mathbb{R}$. From Lindelöf's Lemma, each open set can be written as countable union of open intervals, and open intervals are in $\mathcal{B}_0$, so each open set is in $\mathcal{B}_0$. $\square$

Corollary 1.2. *All open and closed sets in* $\mathbb{R}$ *is Lebesgue measurable.*

1.3.4 Existence of Mon-measurable Sets

Though $\mathcal{B}_0$ is a very large collection, but it does not contain everything, i.e. $\mathcal{B}_0 \neq 2^{\mathbb{R}}$.

To "construct" a non-measurable set, we need *Axiom of Choice*, which has many equivalent form.

Theorem 1.2 (Axiom of Choice (AoC)). *For any collection $\mathcal{C}$ of non-empty sets, there exists a set that contains exactly one element from each set in $\mathcal{C}$.*

Equivalently, for any collection of $\mathcal{C}$ of non-empty sets, there exists a "choice function" f defined on C, such that for any $A \in \mathcal{C}$, $f(A) \in A$.

Equivalently, for any collection of $\mathcal{C}$ of non-empty sets, the Cartesian product $\prod_{A\in\mathcal{C}} A$ is not empty.

Theorem 1.3 (Existence of non-measurable sets). *There exists a set $E \subset \mathbb{R}$ such that $E \notin \mathcal{B}_0$.*

Proof. First, we define an equivalence relation on $\mathbb{R}$ by

$$x \sim y \;\Leftrightarrow\; x - y \in \mathbb{Q}, \quad \forall x, y \in \mathbb{R}.$$

It is trivial to verify that $\sim$ defined above is really an equivalence relation. Let $[x]$ denote the equivalence class containing $x \in \mathbb{R}$. Then, $[x] \cap (0,1) \neq \emptyset$, due to the obvious fact that there exist rationales in $(x-1, x)$ for any $x \in \mathbb{R}$.

Let $E \subset [0,1]$ be a set with a single representative from every equivalence class defined above. **By axiom of choice**, such a set exists. In the following, we show that it is non-measurable.

Note that

$$\mathbb{R} = \bigsqcup_{r\in\mathbb{Q}} (E + r).$$

As $\mathbb{Q}$ is countable. Assume E is measurable, by countable additivity and translation invariance of m, we have

$$\infty = m(\mathbb{R}) = m\left(\bigsqcup_{r\in\mathbb{Q}} (E+r)\right) = m\left(\bigsqcup_{r\in\mathbb{Q}} E\right), \tag{1.76}$$

which follows that $m(E) > 0$. On the other hand, let

$$F = \bigsqcup_{r\in\mathbb{Q}\cap(0,1)} (E+r),$$

then $F \subseteq [0,2]$, and thus

$$2 \geq m(F) \geq \sum_{r\in\mathbb{Q}\cap(0,1)} m(E) = \infty. \tag{1.77}$$

Hence, E is not measurable. □

1.4 Borel Measurable Functions

1.4.1 Borel Algebra

Lemma 1.5. *Let S a collection of σ-algebra over X, then $\bigcap_{\mathcal{M}\in S} \mathcal{M}$ is also a σ-algebra over X.*

Proof. Let $\mathcal{M}^* = \bigcap_{\mathcal{M}\in S} \mathcal{M}$, then

1. $\emptyset, X \in \mathcal{M}, \forall \mathcal{M} \in S \Rightarrow \emptyset, X \in \mathcal{M}^* = \bigcap_{\mathcal{M}\in S} \mathcal{M}$;
2. $A \in \mathcal{M}^* \Rightarrow A \in \mathcal{M}, \forall \mathcal{M} \in S \Rightarrow A^c \in \mathcal{M}, \forall \mathcal{M} \in S \Rightarrow A^c \in \mathcal{M}^*$;
3. $\{A_i\}_{i=1}^\infty \subseteq \mathcal{M}^* \Rightarrow \{A_i\}_{i=1}^\infty \subseteq \mathcal{M}, \forall \mathcal{M} \in S, \Rightarrow \bigcup_{i=1}^\infty A_i \subseteq \mathcal{M}, \forall \mathcal{M} \in S \Rightarrow \bigcup_{i=1}^\infty A_i \subseteq \mathcal{M}^*$.

□

Lemma 1.6 (σ-algebra generated by $\mathcal{F}$). *Given a set X and any collection of subset $\mathcal{F} \subseteq 2^X$, there exists a smallest σ-algebra that contains $\mathcal{F}$, which is called the **σ-algebra generated by** $\mathcal{F}$, and denoted by $\sigma(\mathcal{F})$.*

Here, when we say $\mathcal{M}^*$ is a "smallest σ-algebra", it means that if $\mathcal{M}$ is also a σ-algebra satisying the same condition, then $\mathcal{M}^* \subseteq \mathcal{M}$. Hence, we can see that if such a "smallest σ-algebra" exists, it must be unique.

Proof. Let S be the class of all σ-algebra that contains $\mathcal{F}$. First of all S is not empty, for any non-emtpy set X and collection $\mathcal{F} \subseteq 2^X$, it is easy to see that 2^X is a σ-algebra over X that contains $\mathcal{F}$.

Define $\mathcal{M}^* = \bigcap_{\mathcal{M}\in S} \mathcal{M}$, we are going to show that $\mathcal{M}^*$ is the smallest σ-algebra that is desired.

1. $\mathcal{F} \subseteq \mathcal{M}^* = \bigcap_{\mathcal{M}\in S} \mathcal{M}$, as $\mathcal{F} \subseteq \mathcal{M},\ \forall \mathcal{M} \in S$.
2. $\forall \mathcal{M} \in S,\ \mathcal{M}^* \subseteq \mathcal{M}$ by its definition.
3. As $\mathcal{M}^*$ is an intersection of σ-algebras over X, by the lemma above, it is itself a σ-algebra.

Together, they imply that $\mathcal{M}^*$ is the smallest σ-algebra containing $\mathcal{F}$. □

Definition 1.12 (Borel σ-algebra)**.** *Let $(X, \mathcal{T})$ be a topological space, then $\sigma(\mathcal{T})$, the σ-algebra over X generated by the topology $\mathcal{T}$, is called the **Borel σ-algebra** of the topological space.*

Conventionally, if the topology of a space is implicit, we use $\mathcal{B}(X)$ to denote the Borel σ-algebra over the default topological space of X.

The following are some examples of Borel σ-algebras for different topological space.

1. For the indiscrete topology $\mathcal{T} = \{\emptyset, X\}$, $\sigma(\mathcal{T}) = \{\emptyset, X\}$;
2. For the discrete topology $\mathcal{T} = 2^X$, $\sigma(\mathcal{T}) = 2^X$.
3. For the topology $\mathcal{T} = \{\emptyset, U, X\}$, $\sigma(\mathcal{T}) = \{\emptyset, U, U^c, X\}$.
4. For the topology $\mathcal{T} = \{\emptyset, U_1, U_2, X\}$, such with $U_1 \subset U_2$, $\sigma(\mathcal{T}) = \{\emptyset, U_1, U_2, U_1^c, U_2^c, U_2\backslash U_1, (U_2\backslash U_1)^c\}$. (You can derive this set by constructing with union, intersection, and complement, and carefully remove the identical sets)
5. $\mathcal{B}(\mathbb{R})$ and $\mathcal{B}(\mathbb{R}^n)$ are respectively borel algebra over real field and real vector spaces, which contains all open and closed sets, and all countable unions of closed sets F_σ, and all countable intersections of open sets G_δ, and so on. Note that the sets in the borel algebra are not restricted to those can be expressed in these ways.

Proposition 1.10. *Let $\mathcal{I} = \{(-\infty, a) : a \in \mathbb{R}\}$, then $\sigma(\mathcal{I}) = \mathcal{B}(\mathbb{R})$.*

Proof. It is obvious that $\mathcal{I} \in \mathcal{B}(\mathbb{R})$, which immediately follows that $\sigma(\mathcal{I}) \subseteq \mathcal{B}(\mathbb{R})$. Then, we are going to prove the inclusion in other direction.

1. For each $\alpha \in \mathbb{R}$, $[\alpha, +\infty) = (-\infty, \alpha)^c \in \sigma(\mathcal{I})$;
2. For each $\alpha < \beta \in \mathbb{R}$, $[\alpha, \beta) = (-\infty, \beta) \cap [\alpha, \infty) \in \sigma(\mathcal{I})$;
3. Thus, $(\alpha, \beta) = \bigcup_{n=1}^{\infty} [\alpha + 1/n, \beta) \in \sigma(\mathcal{I})$.

While, in $\mathbb{R}$ all open sets are countable union of open intervals, so all open sets are in $\sigma(\mathcal{I})$, which leads to that the natural topology of $\mathbb{R}$ is contained in $\sigma(\mathcal{I})$, and hence $\mathcal{B}(\mathbb{R}) \subseteq \sigma(\mathcal{I})$. □

1.4.2 Measurable Functions

Before going into the definition, we first show two useful propositions in the following lemma.

Lemma 1.7. *Let $(X, \mathcal{F}_X)$ and $(Y, \mathcal{F}_Y)$ be two measurable spaces, and $f : X \to Y$ be any function, then*

1. *$\mathcal{G}_X = \{f^{-1}(A) | A \in \mathcal{F}_Y)$ is a σ-algebra over X;*
2. *$\mathcal{G}_Y = \{A \subseteq Y | f^{-1}(A) \in \mathcal{F}_X\}$ is a σ-algebra over Y.*

Proof. The verification of these statements are trivial by noting that $f^{-1}(\emptyset) = \emptyset$, $f^{-1}(X) = X$, and f^{-1} preserves basic set operations including arbitrary union, arbitrary intersection, and set complement. □

Definition 1.13 ((Generic) Measurable function). *Let $(X, \mathcal{F}_X)$ and $(Y, \mathcal{F}_Y)$ be measurable spaces, then a function $f : X \to Y$ is called a* ***measurable function*** *if $A \in \mathcal{F}_Y \Rightarrow f^{-1}(A) \in \mathcal{F}_X$, in other words, the pre-image of any measurable set is measurable.*

Definition 1.14 ((Borel) Measurable function). *Let $(X, \mathcal{F})$ be a measurable space, and $(Y, \mathcal{T})$ be a topological space, then a function $f : X \to Y$ is called a* ***(Borel) measurable function*** *when it is measurable in generic sense for Y endowed with the Borel σ-algebra $\sigma(\mathcal{T})$.*

The borel measurability of a function can be equivalently defined as follows.

Let $(X, \mathcal{F})$ be a measurable space, and $(Y, \mathcal{T})$ be a topological space a function $f : X \to Y$ is called a ***(Borel) measurable function*** *if for any open set $A \subset Y$, $f^{-1}(A)$ is measurable.*

The following lemma shows that the two definitions of "borel measurable function" are indeed equivalent.

Lemma 1.8. *Let $(X, \mathcal{F})$ be a measurable space, and $(Y, \mathcal{T})$ be a topological space endowed with a Borel σ-algebra $\mathcal{B}(\mathcal{T})$, give a function $f : X \to Y$, $f^{-1}(A) \in \mathcal{F}$, $\forall A \in \sigma(\mathcal{T})$ iff $f^{-1}(A) \in \mathcal{F}$, $\forall A \in \mathcal{T}$.*

Proof. It is trivial to see that $f^{-1}(A) \in \mathcal{F}$, $\forall A \in \mathcal{B}(\mathcal{T}) \Rightarrow f^{-1}(A) \in \mathcal{F}$, $\forall A \in \mathcal{T}$ as $\mathcal{T} \subseteq \mathcal{B}(\mathcal{T})$. In the following, we show the converse.

Suppose, f has $f^{-1}(A) \in \mathcal{F}$ for all open set A in Y. Let $\mathcal{G} = \{A \subseteq Y | f^{-1}(A) \in \mathcal{F}\}$, then by the lemma above, $\mathcal{G}$ is a σ-algebra over Y. Clearly, all open sets are in $\mathcal{G}$, in other words, $\mathcal{T} \subseteq \mathcal{G}$, thus $\mathcal{B}(Y) = \sigma(\mathcal{T}) \subseteq \mathcal{G}$, which implies that for each $A \in \sigma(\mathcal{T})$, $f^{-1}(A) \in \mathcal{F}$. □

Proposition 1.11. *Let $(X, \mathcal{F})$ be a measurable space, and $f : X \to \mathbb{R}$ be a function satisfying $f^{-1}((-\infty, a)) \in \mathcal{F}$ for all $a \in \mathbb{R}$, then f is a borel measurable.*

Proof. We have shown in previous section that any open set in $\mathbb{R}$ can be derived from set complement, intersection, and countable union of sets in form of $(-\infty, a)$, which follows that for any open set $A \subseteq \mathbb{R}$, its preimage $f^{-1}(A)$ can be expressed in form of the countable union, intersection, and set complements of the sets in form of $f^{-1}((-\infty, a))$ which are measurable. Hence, the preimage of any open set in $\mathbb{R}$ is measurable in $(X, \mathcal{F})$, thus f is borel measurable. □

Proposition 1.12. *Let X and Y be topological spaces, then any continuous function $f : X \to Y$ is measurable.*

Proof. This immediately follows from the fact that in a topological space (by convention endowed with Borel σ-algebra), any open set is measurable. □

Proposition 1.13. *Let X be a measurable space, Y and Z be topological spaces. Given any function $f : X \to Y$ and continuous function $g : Y \to Z$, we have*

1. *if f is continous then $g \circ f$ is continuous;*
2. *if f is borel measurable then $g \circ f$ is borel measurable.*

Proof. Given any open set $V \subset Z$, $g^{-1}(V)$ is open due to the continuity of g. Note that $(g \circ f)^{-1}(V) = f^{-1}(g^{-1}(V))$, so if f is continuous, then $f^{-1}(g^{-1}(V))$ is open, thus $g \circ f$ is continous; or if f is measurable, then $f^{-1}(g^{-1}(V))$ is measurable, thus $g \circ f$ is measurable. □

Lemma 1.9. *Let X be a measurable space, and Y be a topological space. Let $u : X \to \mathbb{R}$ and $v : X \to \mathbb{R}$ be measurable functions, and $\phi : \mathbb{R}^2 \to Y$ be continuous functions. Then, $h : X^2 \to Y$ defined by $h(x, y) = \phi(u(x), v(y))$ is measurable.*

Proof. We define $f : X^2 \to \mathbb{R}^2$ by $f(x, y) = (u(x), v(y))$, then $h = \phi \circ f$. Since ϕ is continuous, from the proposition above, we can see that it suffices to show that f is measurable.

Given any open box in form of $V = I_1 \times I_2 \subseteq \mathbb{R}^2$ with I_1 and I_2 being open intervals in $\mathbb{R}$, then $f^{-1}(V) = u^{-1}(I_1) \cap v^{-1}(I_2)$, where $u^{-1}(I_1)$ and $v^{-1}(I_2)$ are both measurable due to the measurability of u and v, thus $f^{-1}(V)$ is measurable.

Since open boxes constitute basis of the standard topology of $\mathbb{R}^2$, and $\mathbb{R}^2$ is second countable under this topology, we can see that any open set $U \subset \mathbb{R}^2$ can be written as countable union of open boxes as $U = \bigcup_{i=1}^{\infty} V_i$, then $f^{-1}(U) = \bigcup_{i=1}^{\infty} f^{-1}(V_i)$ is measurable. □

Theorem 1.4. *For measurable functions, we have*

1. *Let u and v be real measurable functions, then $f = u + iv$ is complex measurable.*
2. *Let $f = u + iv$ be complex measurable function, then u, v, $|f|$ are real measurable.*
3. *Let f, g be real (complex) measurable functions, then $f + g$ and fg are real (complex) measurable.*
4. *Let E be a measurable set, then its indicator function χ_E is measurable.*

Proof. These statements are immediate corollaries of the above propositions.

1. We can write $f(x) = \phi(u(x), v(x))$ with $\phi : \mathbb{R}^2 \to \mathbb{C}$ given by $\phi(u, v) = u + i * v$ which is continuous and thus measurable.
2. This follows from the fact that taking real or imaginary part, or taking magnitude, are all continuous and thus measurable functions.
3. We can write $(f + g)(x) = \phi_+(f(x), g(x))$ with $\phi_+(u, v) = u + v$ and $(fg)(x) = \phi_\times(f(x), g(x))$ with $\phi_\times(u, v) = uv$. Both ϕ_+ and $\phi_\times$ are continuous and thus measurable.
4. The pre-image of χ_E^{-1} can only be either one of the following: $\emptyset$, E, E^c, and X, which are all measurable whenever E is measurable.

□

Theorem 1.5. *Let $(X, \mathcal{F})$ be a measurable space, and $f_n : X \to \mathbb{R}$ be a sequence of Borel measurable functions, then $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$ and $\liminf_n f_n$ are all Borel measurable.*

Proof. We first look at $\sup_n f_n$. We claim that for each $a \in \mathbb{R}$,

$$\left(\sup_n f_n\right)^{-1}((a, +\infty)) = \bigcup_{n=1}^{\infty} f_n^{-1}((a, +\infty)). \tag{1.78}$$

This can be shown as follows

$$\begin{aligned} x \in \left(\sup_n f_n\right)^{-1}((a, +\infty)) &\Rightarrow \sup_n f_n(x) > a \Rightarrow \exists m \in \mathbb{N}, f_m(x) > a \\ &\Rightarrow x \in f_m^{-1}((a, +\infty)) \subseteq \bigcup_{n=1}^{\infty} f_n^{-1}((a, +\infty)); \end{aligned} \tag{1.79}$$

and

$$\begin{aligned} x \in \bigcup_{n=1}^{\infty} f_n^{-1}((a, +\infty)) &\Rightarrow \exists m \in \mathbb{N}, x \in f_m^{-1}((a, +\infty)) \\ &\Rightarrow f_m(x) > a \Rightarrow \sup_n f_n(x) > a \\ &\Rightarrow x \in \left(\sup_n f_n\right)^{-1}((a, +\infty)). \end{aligned} \tag{1.80}$$

The claim is shown, which readily leads to the conclusion that $\sup_n f_n$ is borel measurable.

Similar argument shows that $\inf_n f_n$ is also borel measurable. While, $\limsup_n f_n = \inf_n \sup_{kgen} f_n$ and $\liminf_n f_n = \sup_n \inf_{k \geq n} f_n$, so they are also borel measurable. □

Chapter 2

Integration Theory

2.1 Lebesgue Integration

2.1.1 Simple Functions

Definition 2.1 (Simple function)**.** *A function $s : X \to \mathbb{R}$ is called a* ***simple function*** *if $s(X)$ is finite.*

A simple function s can be uniquely written as

$$s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$$

where $\alpha_i \in \mathbb{R}$ and $A_i = s^{-1}(\alpha_i) = \{x \in X | s(x) = \alpha_i\}$.

Proposition 2.1. *A simple function s given by $s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ is measurable iff A_i is measurable for all $i = 1, \ldots, n$.*

Proof. 1. Given any set of $B \in \mathbb{R}$, we have

$$\chi_{A_i}^{-1}(B) = \begin{cases} \emptyset & (0 \in A_i \text{ and } 1 \in A_i), \\ A_i & (0 \notin A_i \text{ and } 1 \in A_i), \\ A_i^c & (0 \in A_i \text{ and } 1 \notin A_i), \\ X & (0 \in A_i \text{ and } 1 \in A_i). \end{cases} \tag{2.1}$$

Hence, if A_i is measurable, then χ_{A_i} is measurable. As a finite sum of measurable function, s is measurable.

2. Since s is measurable, $A_i = s^{-1}(\{\alpha_i\})$ is measurable.

□

Definition 2.2 (Monotonical function sequence)**.** *Let $f_1, f_2, \ldots$ be a sequence of real valued functions defined on X, such that for every $x \in X$, $f_1(x), f_2(x), \ldots$ is an increasing sequence, then $f_1, f_2, \ldots$ is called an* ***increasing sequence of functions****.*

Similarly, we can define ***decreasing sequence of functions****.*

Theorem 2.1. *Each non-negative measurable function can be approximated (pointwisely) by an increasing sequence of simple measurable functions.*

Formally, let $f : X \to [0, +\infty]$ be a measurable function, then there exists a increasing sequence of simple measurable functions $s_1 \le s_2 \le \cdots$, such that

$$\lim_{n\to\infty} s_n(x) = f(x), \quad \forall x \in X.$$

Proof. We construct a function $\Psi_n : [0, +\infty] \to [0, +\infty]$ in as follows. First, we partition $[0, n]$ into $n2^n$ intervals of length $\delta_n = 2^{-n}$. Then for each $t \ge 0$, there exists a unique k_t such that $k_t\delta_n < t \le (k+1)\delta_n$. We then define

$$\Psi_n(t) = \begin{cases} k\delta_n & (0 \le t < n), \\ n & (n \le t \le +\infty). \end{cases}$$

Note that Ψ_n is a simple Borel measurable function as it can only take integer values from between 0 and $n2^n$, and $\Psi_1, \Psi_2, \ldots$ is an increasing sequence of functions.

In addition, $\forall t \in [0, +\infty],\ t - \delta_n < \Psi_n(t) < t$, hence $\Psi_n(t)$ converges to t pointwisely as $n \to +\infty$.

Define $s_n = \Psi_n \circ f$. Note that s_n is a simple measurable function, as both Ψ_n and f are measurable, and Ψ_n is a simple function. Then $(s_n)_{n=1}^{\infty}$ is an increasing sequence of simple measurable functions that converges to f pointwisely. □

In the following, we use the term **positive measure** to refer to a measure $\mu : \mathcal{M} \to [0, +\infty]$ such that there exists a nonempty set $A \in \mathcal{M}$ with $\mu(A) < +\infty$.

2.1.2 Lebesgue Integral

We first define the Lebesgue integral on simple functions, and then extend it to all measurable functions.

Definition 2.3 (Lebesgue Integral)**.** *Given a measure space $(X, \mathcal{M}, \mu)$, and a simple measurable function $s : X \to [0, +\infty]$. Note that s can be uniquely written as $s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ with $A_1, \ldots, A_n$ being mutually disjoint. Then, its **Lebesgue integral** on a measurable set $E \in \mathcal{M}$, denoted by $\int_E s d\mu$ is defined by*

$$\int_E s d\mu = \sum_{i=1}^{n} \alpha_i \mu(E \cap A_i).$$

*Let $f : X \to [0, +\infty]$ be non-negative measurable function, then its **Lebesgue integral** is given by*

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

Proposition 2.2. *Given a measure space $(X, \mathcal{M}, \mu)$, non-negative measurable functions $f, g : X \to [0, +\infty]$, measurable sets $E, A, B \in \mathcal{M}$, and $c \in [0, \infty]$, then we have*

1. $f \le g \Rightarrow \int_E f d\mu \le \int_E g d\mu$;
2. $A \subset B \Rightarrow \int_A f d\mu \le \int_B f d\mu$;
3. $\int_E c f d\mu = c \int_E f d\mu$;
4. $f|_E \equiv 0 \Rightarrow \int_E f d\mu = 0$ *(This holds even when $\mu(E) = +\infty$);*
5. $\mu(E) = 0 \Rightarrow \int_E f d\mu = 0$ *(This holds even when $f|_E = +\infty$);*
6. $\int_E f d\mu \Rightarrow \int_X \chi_E f d\mu$.

Proof. We prove these properties respectively. For convenience, we denote

$$\mathcal{S}_f = \{s \mid s \text{ is simple measurable, and } 0 \le s \le f\}.$$

Then,

$$\int_E f d\mu = \sup_{s \in \mathcal{S}_f} \int_E s d\mu. \tag{2.2}$$

1. As $f \le g$, $s \le f \Rightarrow s \le g$, it means that $\mathcal{S}_f \subset \mathcal{S}_g$, which follows that

$$\int_E f d\mu = \sup_{s \in \mathcal{S}_f} \int_E s d\mu \le \sup_{s \in \mathcal{S}_g} \int_E s d\mu = \int_E g d\mu. \tag{2.3}$$

2. Let $s = \sum_{i=1}^{n} \alpha_i \chi_{C_i}$ be a non-negative simple measurable function. When $A \subset B$,

$$\int_A s d\mu = \sum_{i=1}^{n} \alpha_i \mu(A \cap C_i) \le \sum_{i=1}^{n} \alpha_i \mu(B \cap C_i) = \int_B s d\mu. \tag{2.4}$$

Hence,

$$\int_A f d\mu = \sup_{s \in \mathcal{S}_f} \int_A s d\mu \le \sup_{s \in \mathcal{S}_f} \int_B s d\mu = \int_B f d\mu. \tag{2.5}$$

3. Let $s = \sum_{i=1}^n \alpha_i \chi_{C_i}$ be a non-negative simple measurable function, then when $c \geq 0$, $cs = \sum_{i=1}^n (c\alpha_i)\chi_{C_i}$ is also a non-negative simple measurable function, which has
$$\int_E csd\mu = \sum_{i=1}^n (c\alpha_i)\mu(C_i) = c\sum_{i=1}^n \alpha_i \mu(C_i) = c\int_E sd\mu. \tag{2.6}$$
In addition, we note that $s \leq f \Leftrightarrow cs \leq cf$, which means $\mathcal{S}_{cf} = \{cs|s \in \mathcal{S}_f\}$. Hence,
$$\int_E cfd\mu = \sup_{s\in\mathcal{S}_{cf}} \int_E sd\mu = \sup_{s\in\mathcal{S}_f} \int_E csd\mu = c \cdot \sup_{s\in\mathcal{S}_f} \int_E sd\mu = c\int_E fd\mu. \tag{2.7}$$
4. For each $s \in \mathcal{S}_f$, $0 \leq s(x) \leq f(x), \forall x \in X$, hence, when $f(x) = 0, \forall x \in E \subset X$, $s(x) = 0, \forall x \in E$. For each s satisfying the above condition, we have
$$\int_E sd\mu = 0 \cdot \mu(E \cap s^{-1}(0)) = 0. \tag{2.8}$$
As a result,
$$\int_E fd\mu = \sup_{s\in\mathcal{S}_f} \int_E sd\mu = \sup_{s\in\mathcal{S}_f} 0 = 0. \tag{2.9}$$
5. For each $s \in \mathcal{S}_f$, write it in form of $s = \sum_{i=1}^n \alpha_i \chi_{A_i}$. And, note that $\mu(E) = 0$, and thus $\mu(E \cap A_i) = 0, \forall i = 1, \ldots, n$, then
$$\int_E sd\mu = \sum_{i=1}^n \alpha_i \mu(E \cap A_i) = \sum_{i=1}^n \alpha_i \cdot 0 = 0. \tag{2.10}$$
Hence,
$$\int_E fd\mu = \sup_{s\in sset_f} \int_E sd\mu = \sup_{s\in\mathcal{S}_f} 0 = 0. \tag{2.11}$$
6. Let $s = \sum_{i=1}^n \alpha_i \chi_{A_i}$ be a simple non-negative measurable function. We note that
$$\chi_E s = \sum_{i=1}^n \alpha_i \chi_{E\cap A_i}. \tag{2.12}$$
It follows that
$$\int_X \chi_E sd\mu = \sum_{i=1}^n \alpha_i \mu(E \cap A_i) = \int_E sd\mu. \tag{2.13}$$
In addition, $s \leq f \Rightarrow \chi_E s \leq \chi_E f$, in other words, $s \in \mathcal{S}_f \Rightarrow \chi_E s \in \mathcal{S}_{\chi_E f}$
$$\int_E fd\mu = \sup_{s\in\mathcal{S}_f} \int_E sd\mu = \sup_{s\in\mathcal{S}_f} \int_X \chi_E sd\mu \leq \sup_{s'\in\mathcal{S}_{\chi_E f}} \int_X s'd\mu = \int_X \chi_E fd\mu. \tag{2.14}$$
On the other hand, for every $s \in \mathcal{S}_{\chi_E f}$, we have $s(x) = 0, \forall x \in E^c$. Write s into $s = \sum_{i=1}^n \alpha_i \chi_{C_i}$, then $\alpha_i > 0 \Rightarrow C_i \in E \Leftrightarrow C_i \cap E = C_i$, therefore,
$$\int_X sd\mu = \sum_{i=1}^n \alpha_i \mu(C_i) = \sum_{i=1}^n \alpha_i \mu(C_i \cap E) = \int_E sd\mu. \tag{2.15}$$
Hence,
$$\int_X \chi_E fd\mu = \sup_{s\in\mathcal{S}_{\chi_E f}} \int_X sd\mu = \sup_{s\in\mathcal{S}_{\chi_E f}} \int_E sd\mu = \int_E \chi_E fd\mu. \tag{2.16}$$
Note that $\chi_E f \leq f$, by monotonicity shown above
$$\int_E \chi_E fd\mu \leq \int_E fd\mu. \tag{2.17}$$
As a result,
$$\int_X \chi_E fd\mu \leq \int_E fd\mu. \tag{2.18}$$

□

2.1.3 Additivity, Monotone Convergence Theorem, and Induced Measure

We first prove finite additivity of simple functions, and then generalize it to countable additivity of generic measurable functions.

Lemma 2.1. *Each non-negative simple measurable function induces a measure.*

Formally, let s be a non-negative simple measurable function defined in a measure space $(X, \mathcal{M}, \mu)$, and define $\Psi : \mathcal{M} \to [0, +\infty]$ by

$$\Psi(E) = \int_E s d\mu, \ \forall E \in \mathcal{M},$$

then Ψ is a measure.

Proof. We show that Ψ satisfies the condition of a measure.

1. $\Psi(\emptyset) = \int_\emptyset s d\mu = 0$ due to $\mu(\emptyset) = 0$.
2. We write $s = \sum_{i=1}^n \alpha_i \chi_{A_i}$ Let $E_1, E_2, \ldots$ be disjoint measurable sets, and let $E = \bigsqcup_{k=1}^\infty E_k$, then

$$\begin{aligned} \Psi(E) &= \sum_{i=1}^n \alpha_i \mu(E \cap A_i) = \sum_{i=1}^n \alpha_i \mu\left(\left(\bigsqcup_{k=1}^\infty E_k\right) \cap A_i\right) = \sum_{i=1}^n \alpha_i \mu\left(\bigsqcup_{k=1}^\infty (E_k \cap A_i)\right) \\ &= \sum_{i=1}^n \alpha_i \left(\sum_{k=1}^\infty \mu(E_k \cap A_i)\right) = \sum_{k=1}^\infty \sum_{i=1}^n \alpha_i \mu(E_k \cap A_i) \\ &= \sum_{k=1}^\infty \int_{E_k} s d\mu = \sum_{k=1}^\infty \Psi(E_k). \end{aligned} \tag{2.19}$$

□

Proposition 2.3. *Lebesgue Integral is additive on simple measurable functions.*

Formally, let s and t be two non-negative simple measurable functions defined in a measure space $(X, \mathcal{M}, \mu)$, then

$$\int_X (s+t) d\mu = \int_X s d\mu + \int_X t d\mu.$$

As an immediate corollary, we have for each measurable set $E \in \mathcal{M}$,

$$\int_E (s+t) d\mu = \int_X (\chi_E s + \chi_E t) d\mu = \int_X \chi_E s d\mu + \int_X \chi_E t d\mu = \int_E s d\mu + \int_E t d\mu.$$

Proof. We define for each $E \in \mathcal{M}$, $\Psi_{s+t}(E) = \int_E (s+t) d\mu$, $\Psi_s(E) = \int_E s d\mu$, $\Psi_t(E) = \int_E t d\mu$, then the functions Ψ_{s+t}, Ψ_s and Ψ_t are all measures. Let $s = \sum_{i=1}^m \alpha_i \chi_{A_i}$ and $t = \sum_{j=1}^n \beta_j \chi_{B_j}$, such that $\bigsqcup_{i=1}^m A_i = \bigsqcup_{j=1}^n B_j = X$. Let $E_{ij} = A_i \cap B_j$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. It is obvious that E_{ij} are measurable and mutually disjoint, and they have

$$X = \bigsqcup_{1 \le i \le m, \ 1 \le n \le n} E_{ij}.$$

We note

$$\int_{E_{ij}} (s+t) d\mu = (\alpha_i + \beta_j)\mu(E_{ij}) = \alpha_i \mu(E_{ij}) + \beta_j \mu(E_{ij}) = \int_{E_{ij}} s d\mu + \int_{E_{ij}} t d\mu. \tag{2.20}$$

In addition,

$$\begin{aligned}
\int_X (s+t)d\mu = \Psi_{s+t}(X) = \Psi_{s+t}\left(\bigsqcup_{i,j} E_{ij}\right) &= \sum_{i=1}^{m}\sum_{j=1}^{n}\Psi_{s+t}(E_{ij}) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\int_{E_{ij}}(s+t)d\mu = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\int_{E_{ij}} sd\mu + \int_{E_{ij}} td\mu\right) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n}\Psi_s(E_{ij}) + \sum_{i=1}^{m}\sum_{j=1}^{n}\Psi_t(E_{ij}) = \Psi_s\left(\bigsqcup_{i,j} E_{ij}\right) + \Psi_t\left(\bigsqcup_{i,j} E_{ij}\right) \\
&= \Psi_s(X) + \Psi_t(X) = \int_X sd\mu + \int_X td\mu. \qquad (2.21)
\end{aligned}$$

□

Before generalizing the additivity to generic measurable functions, we still need to prove the following theorem, namely the Lebesgue Monotone Convergence Theorem, which in itself is a very important theorem in integration theory.

Theorem 2.2 (Montotone Convergence Theorem (MCT))**.** *Let $f_1, f_2, \dots$ be an increasing sequence of non-negative measurable functions defined on a measure space $(X, \mathcal{M}, \mu)$, assume that it converges pointwisely to a function f, i.e*

$$\lim_{n\to\infty} f_n(x) = f(x),\ \forall x \in X,$$

then f is measurable and

$$\int_X fd\mu = \int_X \lim_{n\to\infty} fd\mu. = \lim_{n\to\infty}\int_X f_n d\mu.$$

Proof. By monotonicity of the Lebesgue integral, we have $\int_X f_i d\mu \le \int_X f_{i+1}d\mu$ for all $i \in \mathbb{N}$ due to $f_i \le f_{i+1}$. Hence, the real value sequence $\left(\int_X f_n d\mu\right)_{n=1}^{\infty}$ is an increasing sequence, and thus have a unique limit $\alpha \in [0, \infty]$.

1. As $(f_n)_{n=1}^{\infty}$ is increasing, we have $f_n \le f$, which follows that

$$\int_X f_n d\mu \le \int_X fd\mu, \quad \forall n \in \mathbb{N}. \qquad (2.22)$$

   Take the limit for left hand side, we get

$$\alpha = \lim_{n\to\infty}\int_X f_n d\mu \le \int_X fd\mu. \qquad (2.23)$$

2. Let s be a simple measurable function satisfying $0 \le s \le f$, and c be any real value with $0 < c < 1$. Define $\Psi : \mathcal{M} \to [0, +\infty] : E \mapsto \int_E sd\mu$, then Ψ is a measure.

   On the other hand, let $E_n = \{x \in X : f_n(x) \ge cs(x)\}$, then it is easy to see that E_n is measurable (note $E_n = (f_n - cs)^{-1}[0, \infty]$). In addition, $f_n(x) \ge cs(x) \Rightarrow f_{n+1}(x) \ge cs(x)$, it follows that $E_1 \subseteq E_2 \subseteq \cdots$. Furthermore, we claim that $\bigcup_{n=1}^{\infty} E_n = X$. This claim is briefly shown as follows. First, $E_n \subseteq X \Rightarrow \bigcup_{n=1}^{\infty} E_n \subseteq X$. In showing the other direction, for each $x \in X$, if $f(x) = 0$, then $s(x) = 0$, thus it is obvious that $x \in E_n, \forall n \in \mathbb{N}$. Otherwise, $f(x) > cs(x)$, as $f_n(x) \uparrow f(x)$, there exists n such that $f(x) - f_n(x) < f(x) - cs(x)$, i.e $f_n(x) > cs(x)$, thus $x \in E_n$. The claim is proved.

   By continuity of measure, we have

$$\lim_{n\to\infty}\Psi(E_n) = \Psi(E) \quad \Leftrightarrow \quad \lim_{n\to\infty}\int_{E_n} sd\mu = \int_X sd\mu. \qquad (2.24)$$

   From the basic property of Lebesgue integral, we have

$$\int_X f_n d\mu \ge \int_{E_n} f_n d\mu \ge \int_{E_n} csd\mu = c\int_{E_n} sd\mu. \qquad (2.25)$$

Take the limit as $n \to \infty$ for both sides, then

$$\alpha = \lim_{n\to\infty} \int_X f_n d\mu \geq c \lim_{n\to\infty} \int_{E_n} s d\mu = c \int_X s d\mu. \tag{2.26}$$

This holds for any $0 < c < 1$, which follows that

$$\alpha = \lim_{n\to\infty} \int_X f_n d\mu \geq \int_X s d\mu. \tag{2.27}$$

□

Theorem 2.3 (Countable Additivity of Lebesgue Integral)**.** *Let $f_1, f_2, \dots$ of non-negative measurable functions on a measure space $(X, \mathcal{M}, \mu)$, and define $f : X \to [0, +\infty]$ by*

$$f(x) = \sum_{n=1}^{\infty} f_n(x), \ \forall x \in X,$$

then

$$\int_X f d\mu = \int_X \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int_X f_n d\mu.$$

Proof.

1. We first prove for the sum of two functions. Let f and g be two non-negative measurable functions, then there exist increasing sequences of non-negative simple measurable functions $(s_i)_{i=1}^{\infty} \uparrow f$ and $(t_i)_{i=1}^{\infty} \uparrow g$, then $(s_i + t_i)_{i=1}^{\infty} \uparrow f + g$. Based on the additivity proved on simple functions, we have

$$\int_X (s_i + t_i) d\mu = \int_X s_i d\mu + \int_X t_i d\mu. \tag{2.28}$$

By MCT, we get

$$\int_X (f+g) d\mu = \int_x \lim_{i\to\infty} (s_i+t_i) d\mu = \lim_{i\to\infty} \int_X (s_i+t_i) d\mu = \lim_{i\to\infty} \int_X s_i d\mu + \lim_{i\to\infty} \int_X t_i d\mu = \int_X f d\mu + \int_X g d\mu. \tag{2.29}$$

2. By induction, we can show the finite additivity as

$$\int_X \sum_{i=1}^{n} f_i d\mu = \sum_{i=1}^{n} \int_X f_i d\mu. \tag{2.30}$$

3. Let $g_n = \sum_{i=1}^{n} f_i$, then $g_n \uparrow g_\infty = f$. By finite additivity and MCT,

$$\sum_{i=1}^{\infty} \int_X f_i d\mu = \lim_{n\to\infty} \sum_{i=1}^{n} \int_X f_i d\mu = \lim_{n\to\infty} \int_X \sum_{i=1}^{n} f_i d\mu = \lim_{n\to\infty} \int_X g_n d\mu = \int_X \lim_{n\to\infty} g_n d\mu = \int_X f d\mu. \tag{2.31}$$

□

As an immediate corollary, we can show that the order of infinite sum can be exchanged for non-negative terms.

Corollary 2.1. *Consider a non-negative function $a : \mathbb{N} \times \mathbb{N} \to [0, +\infty]$, we have*

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a(i, j) = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a(i, j).$$

Proof. Suppose we are working with a measure space on $\mathbb{N}$ with counting measure μ. Define a sequence of non-negative function $f_1, f_2, \dots$ by $f_j(i) = a(i, j), \forall i, j \in \mathbb{N}$. Then f_j are measurable functions, hence

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a(i, j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_j(i) = \int_{\mathbb{N}} \sum_{j=1}^{\infty} f_j d\mu = \sum_{j=1}^{\infty} \int_{\mathbb{N}} f_j d\mu = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} f_j(i) = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a(i, j). \tag{2.32}$$

□

The following lemma states the relation between lim inf and Lebesgue integral, which has important utility.

Lemma 2.2 (Fatou's Lemma). *Let $f_1, f_2, \dots$ be a sequence of non-negative measurable functions defined on a measure space $(X, \mathcal{M}, \mu)$, then*

$$\int_X \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_X f_n d\mu.$$

Proof. For each $k \in \mathbb{N}$, define $g_k : X \to [0, +\infty]$ by $g_k(x) = \inf_{i\geq k} f_i(x)$. Then, $g_1, g_2, \dots$ is an increasing sequence of non-negative measurable functions, and it has

$$\lim_{k\to\infty} g_k(x) = \sup_{k\in\mathbb{N}} g_k(x) = \sup_{k\in\mathbb{N}} \inf_{i\geq k} f_i(x) = \liminf_{n\to\infty} f_n(x). \tag{2.33}$$

By MCT, we have

$$\lim_{k\to\infty} \int_X g_k d\mu = \int_X \lim_{k\to\infty} g_k d\mu = \int_X \liminf_{n\to\infty} f_n d\mu. \tag{2.34}$$

In addition, by definition, $g_n \leq f_n$, thus by monotonicity,

$$\int_X g_n d\mu \leq \int_X f_n d\mu. \tag{2.35}$$

Taking lim inf as $n \to \infty$ of both sides, we get

$$\int_X \liminf_{n\to\infty} f_n d\mu = \lim_{k\to\infty} \int_X g_k d\mu = \liminf_{n\to\infty} \int_X g_n d\mu \leq \liminf_{n\to\infty} \int_X f_n d\mu. \tag{2.36}$$

□

Theorem 2.4. *Given a measure space $(X, \mathcal{M}, \mu)$, every non-negative measurable function $f : X \to [0, +\infty]$ defines a measure $\nu : \mathcal{M} \to [0, +\infty]$ by*

$$\nu(E) = \int_E f d\mu, \ \forall E \in \mathcal{M}.$$

And, for each measurable function $g : X \to [0, +\infty]$, we have

$$\int_X g d\nu = \int_X g f d\mu.$$

This equation can be summarized as the rule $d\nu = f d\mu$.

Proof.

1. We first prove that ν is a measure, by showing that it satisfies the conditions of a measure.
   (a) $\nu(\emptyset) = \int_\emptyset f d\mu = 0$.
   (b) Let $\{E_i\}_{i=1}^\infty$ be a sequence of pairwisely disjoint measurable sets, and let $E = \bigcup_{i=1}^\infty E_i$. Note that

   $$\chi_E f = \sum_{i=1}^\infty \chi_{E_i} f. \tag{2.37}$$

   Then by countable additivity of Lebesgue integral, we get

   $$\nu(E) = \int_E f d\mu = \int_X \chi_E f d\mu = \int_X \sum_{i=1}^\infty \chi_{E_i} f d\mu = \sum_{i=1}^\infty \int_X \chi_{E_i} f d\mu = \sum_{i=1}^\infty \int_{E_i} f d\mu = \sum_{i=1}^\infty \nu(E_i). \tag{2.38}$$

2. Then, we prove the second statement, namely $d\nu = f d\mu$.

   We first show that this holds for each characteristic functions of measurable set E.

   $$\int_X \chi_E d\nu = \nu(E) = \int_E f d\mu = \int_X \chi_E f d\mu. \tag{2.39}$$

Then for each non-negative simple measurable function $s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$. Then by additivity,

$$\int_X s d\nu = \int_X \sum_{i=1}^{n} \alpha_i \chi_{A_i} d\nu = \sum_{i=1}^{n} \alpha_i \int_X \chi_{A_i} d\nu = \sum_{i=1}^{n} \alpha_i \int_X \chi_{A_i} f d\mu = \int_X \sum_{i=1}^{n} \alpha_i \chi_{A_i} f d\mu = \int_X s f d\mu. \tag{2.40}$$

Finally, for arbitrary non-negative measurable function g, there exists an increasing sequence of non-negative simple measurable functions $(s_1, s_2, \ldots)$ such that $s_n \uparrow g$ (pointwisely). It follows that $(s_n f) \uparrow gf$. Then, by MCT

$$\int_X g d\nu = \int_X \lim_{n\to\infty} s_n d\nu = \lim_{n\to\infty} \int_X s_n d\nu = \lim_{n\to\infty} \int_X s_n f d\mu = \int_X \lim_{n\to\infty} s_n f d\mu = \int_X g f d\mu. \tag{2.41}$$

□

2.1.4 Lebesgue Integration of Complex Functions

In the following, we generalize the concept of Lebesgue integral from non-negative real functions to generic complex functions.

Definition 2.4 (L^1 functions)**.** *Given a measure space $(X, \mathcal{M}, \mu)$, we denote*

$$L^1(X, \mu) = \left\{ f : X \to \mathbb{C} \,\middle|\, f \text{ is measurable, and } \int_X |f| d\mu < +\infty \right\}.$$

Each function in $L^1(X, \mu)$ is called an L^1 ***function****. In other words, an L^1 function is a measurable function f such that $|f|$ has finite Lebesgue integral.*

$L^1(X, \mu)$ is sometimes written as $L^1(X)$ or $L^1(\mu)$, when the underlying measure or the universe is clear from context.

Definition 2.5 (Lebesgue integral of complex functions)**.** *Let $f = u + iv$ be a complex measurable function defined on a measure space $(X, \mathcal{M}, \mu)$, where u and v are respectively the real and imaginary part of f, such that u, v are measurable, and $f \in L^1(X, \mu)$. Then for each $E \in \mathcal{M}$, we define*

$$\int_E f d\mu = \int_E u d\mu + i \int_E v d\mu = \left(\int_E u^+ d\mu - \int_E u^- d\mu \right) + i \left(\int_E v^+ d\mu - \int_E v^- d\mu \right),$$

where $u^+ = \max(u, 0)$, $u^- = \min(u, 0)$, $v^+ = \max(v, 0)$, and $v^- = \min(v, 0)$.

(Note that as $|u^\pm| \leq |u| \leq |f|$ and $|v^\pm| \leq |v| \leq |f|$, hence $u^\pm$ and $v^\pm$ are all non-negative L^1-functions.)

The following proposition states that Lebesgue integration is a linear functional acting on a L^1 function.

Proposition 2.4 (Linearity of Lebesgue integration)**.** *Given a measure space $(X, \mathcal{M}, \mu)$. Let $f, g \in L^1(X, \mu)$, and $\alpha, \beta \in \mathbb{C}$, then $\alpha f + \beta g \in L^1(X, \mu)$ and*

$$\int_X (\alpha f + \beta g) d\mu = \alpha \int_X f d\mu + \beta \int_X g d\mu.$$

To prove the linearity, we first note the following simple fact

Lemma 2.3. *Given a measure space $(X, \mathcal{M}, \mu)$ and let f, g, f', g' be non-negative L^1 functions such that $f - g = f' - g'$, then*

$$\int_X f d\mu - \int_X g d\mu = \int_X f' d\mu - \int_X g' d\mu.$$

Proof. This can be shown by addivitity of Lebesgue integral of non-negative functions, as

$$\begin{aligned} & f - g = f' - g' \Rightarrow f + g' = f' + g \\ & \Rightarrow \int_X (f + g') d\mu = \int_X (f' + g) d\mu \Rightarrow \int_X f d\mu + \int_X g' d\mu = \int_X f' d\mu + \int_X g d\mu \\ & \Rightarrow \int_X f d\mu - \int_X g d\mu = \int_X f' d\mu - \int_X g' d\mu. \end{aligned} \tag{2.42}$$

□

Now we come back to prove the linearity.

Proof. We first show $\int_X (f+g)d\mu = \int_X f d\mu + \int_X g d\mu$, and then show $\int_X \alpha f d\mu = \alpha \int_X f d\mu$, then linearity immediately follows from these two properties.

1. The statement that $f+g$ is an L^1 function follows from the fact that $|f+g| \leq |f|+|g|$ and monotonicity of Lebesgue integral.

   Decompose f and g into real and imaginary parts as $f = u_f + iv_f$ and $g = u_g + iv_g$, then $f+g = (u_f+u_g)+i(v_f+v_g)$, where u_f+u_g and v_f+v_g are respectively the real and imaginary parts of $f+g$. Then we have

$$\int_X (f+g)d\mu = \left(\int_X (u_f+u_g)^+ d\mu - \int_X (u_f+u_g)^- d\mu\right) + i\left(\int_X (v_f+v_g)^+ d\mu - \int_X (v_f+v_g)^- d\mu\right). \tag{2.43}$$

   Note that

$$(u_f+u_g)^+ - (u_f+u_g)^- = \mathrm{Re}(f) + \mathrm{Re}(g) = (u_f^+ - u_f^-) + (u_g^+ - u_g^-) = (u_f^+ + u_g^+) - (u_f^- + u_g^-); \tag{2.44}$$

$$(v_f+v_g)^+ - (v_f+v_g)^- = \mathrm{Im}(f) + \mathrm{Im}(g) = (v_f^+ - v_f^-) + (v_g^+ - v_g^-) = (v_f^+ + v_g^+) - (v_f^- + v_g^-). \tag{2.45}$$

   Then, by the lemma above,

$$\begin{aligned}\int_X (u_f+u_g)^+ d\mu - \int_X (u_f+u_g)^- d\mu &= \int_X (u_f^+ + u_g^+)d\mu - \int_X (u_f^- + u_g^-)d\mu \\ &= \int_X u_f^+ d\mu + \int_X u_g^+ d\mu - \int_X u_f^- d\mu - \int_X u_f^- d\mu \\ &= \left(\int_X u_f^+ d\mu - \int_X u_f^- d\mu\right) + \left(\int_X u_g^+ d\mu - \int_X u_g^- d\mu\right). \end{aligned} \tag{2.46}$$

   Likewise, we have

$$\int_X (v_f+v_g)^+ d\mu - \int_X (v_f+v_g)^- d\mu = \left(\int_X v_f^+ d\mu - \int_X v_f^- d\mu\right) + \left(\int_X v_g^+ d\mu - \int_X v_g^- d\mu\right). \tag{2.47}$$

   Combining the results above, we get

$$\begin{aligned}\int_X (f+g)d\mu &= \left(\int_X u_f^+ d\mu - \int_X u_f^- d\mu\right) + i\left(\int_X v_f^+ d\mu - \int_X v_f^- d\mu\right) \\ &\quad + \left(\int_X u_g^+ d\mu - \int_X u_g^- d\mu\right) + i\left(\int_X v_g^+ d\mu - \int_X v_g^- d\mu\right) \\ &= \int_X f d\mu + \int_X g d\mu. \end{aligned} \tag{2.48}$$

2. The statement that αf is an L^1 function follows from the fact of $|\alpha f| \leq |\alpha||f|$. Let $f = u + iv$, where $u = \mathrm{Re}(f)$ and $v = \mathrm{Im}(f)$.

   If $\alpha \in \mathbb{R}$ and $\alpha \geq 0$, then $(\alpha u)^\pm = \alpha u^\pm$ and $(\alpha v)^\pm = \alpha v^\pm$, thus

$$\begin{aligned}\int_X \alpha f d\mu = \int_X (\alpha u + i\alpha v)d\mu &= \left(\int_X (\alpha u)^+ d\mu - \int_X (\alpha u)^- d\mu\right) - i\left(\int_X (\alpha v)^+ d\mu - \int_X (\alpha v)^- d\mu\right) \\ &= \alpha\left(\left(\int_X u^+ d\mu - \int_X u^- d\mu\right) - i\left(\int_X v^+ d\mu - \int_X v^- d\mu\right)\right) \\ &= \alpha \int_X f d\mu. \end{aligned} \tag{2.49}$$

In addition, we claim $\int_X (-f)d\mu = -\int_X f d\mu$. Combining the result above and this claim leads to that $\int_X \alpha f d\mu = \alpha \int_X f d\mu$ for all $\alpha \in \mathbb{R}$. This claim is briefly shown below

$$\begin{aligned}\int_X (-f)d\mu &= \left(\int_X (-u)^+ d\mu - \int_X (-u)^- d\mu\right) + i\left(\int_X (-v)^+ d\mu - \int_X (-v)^- d\mu\right) \\ &= \left(\int_X u^- d\mu - \int_X u^+ d\mu\right) + i\left(\int_X v^- d\mu - \int_X v^+ d\mu\right) \\ &= -\left(\left(\int_X u^+ d\mu - \int_X u^- d\mu\right) + i\left(\int_X v^+ d\mu - \int_X v^- d\mu\right)\right) = -\int_X f d\mu. \end{aligned} \tag{2.50}$$

Now we generalize the conclusion from the case with $\alpha \in \mathbb{R}$ to $\alpha \in \mathbb{C}$. Write $\alpha = a + ib$, then $\alpha f = (au - bv) + i(bu + av)$. Hence,

$$\begin{aligned}\int_X \alpha f d\mu &= \int_X (au - bv)d\mu + i\int_X (bu + av)d\mu = \int_X aud\mu + \int_X (-b)vd\mu + i\int_X bud\mu + i\int_X avd\mu \\ &= a\int_X ud\mu - b\int_X vd\mu + ib\int_X ud\mu + ia\int_X vd\mu \\ &= (a + ib)(\int_X ud\mu + \int_X vd\mu) = \alpha\int_X f d\mu. \end{aligned} \tag{2.51}$$

□

By linearity, we can extend the monotonicity of Lebesgue integral from non-negative functions to all real-valued L^1 functions.

Proposition 2.5. *Given a measure space $(X, \mathcal{M}, \mu)$ and $f, g : X \to \mathbb{R} \in L^1(X, \mu)$, then*

$$f(x) \le g(x),\ \forall x \in X \Rightarrow \int_X f d\mu \le \int_X g d\mu.$$

Proof.

$$f \le g \Rightarrow g - f \ge 0 \Rightarrow \int_X g d\mu - \int_X f d\mu = \int_X (g - f)d\mu \ge 0 \Rightarrow \int_X f d\mu \le \int_X g d\mu. \tag{2.52}$$

□

Proposition 2.6. *Given a measure space $(X, \mathcal{M}, \mu)$ and $f \in L^1(X, \mu)$, then*

$$\left|\int_X f d\mu\right| \le \int_X |f| d\mu.$$

Proof. Let $z = \int_X f d\mu$. It is obvious that $\int_X |f|d\mu \ge 0$, hence if $z = 0$, the statement trivially holds. Now we consider the case in which $z \ne 0$. Let $\alpha = \bar{z}/|z|$, then $|\alpha| = 1$ and $\alpha z = |z|$. Thus,

$$\left|\int_X f d\mu\right| = \alpha\int_X f d\mu = \int_X \alpha f d\mu. \tag{2.53}$$

Write $\alpha f = u_a + iv_a$, then $|u_a(x)| \le |\alpha f(x)| = |f(x)|, \forall x \in X$. By definition,

$$|z| = \int_X \alpha f d\mu = \int_X u_a d\mu + i\int_X v_a d\mu. \tag{2.54}$$

since $|z| \in \mathbb{R}$, $\int_X v_a d\mu = 0$. Then by monotonicity,

$$\left|\int_X f d\mu\right| = |z| = \int_X u_a d\mu \le \int_X |f| d\mu. \tag{2.55}$$

□

2.1.5 Dominated Convergence Theorem

Dominated Convergence Theorem (DCT) introduced below is one of the most important theorem in measure and integration theory, which establishes a widely applicable condition under which the order of limit and integration can be exchanged.

Theorem 2.5 (Dominated Convergence Theorem (DCT)). *Let $(f_n)_{n=1}^{\infty}$ be a sequence of complex measurable function defined in a measure space $(X, \mathcal{M}, \mu)$ and f is a complex function satisfying*

$$f(x) = \lim_{n\to\infty} f_n(x), \forall x \in X.$$

If there exists a positive function $g \in L^1(\mu)$, such that $|f_n(x)| \leq g(x), \forall x \in X$, then $f \in L^1(\mu)$ and

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

thus

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

Proof.

1. We first show that f is integrable, i.e $f \in L^1(\mu)$. Since $|f_n(x)| \leq g(x)$, $\forall x \in X$, $\forall n \in \mathbb{N}$, $|f(x)| = \lim_{n\to\infty} |f_n(x)| \leq g(x)$, hence

$$\int_X |f| d\mu \leq \int_X g d\mu < \infty, \tag{2.56}$$

which implies that $f \in L^1(\mu)$.

2. Note that $|f_n(x) - f(x)| \leq 2g(x)$, we define a real valued function $h_n : X \to \mathbb{R}$ for each $n \in \mathbb{N}$ by $h_n(x) = 2g(x) - |f_n(x) - f(x)|$. It is clear that h_n is positive function for each n. In addition, as $\lim_{n\to\infty} f_n = f$, $\lim_{n\to\infty} h_n = 2g$. Then, we have

$$\begin{aligned}
\int_X 2g d\mu &= \int_X \lim_{n\to\infty} (2g - |f_n - f|) d\mu = \int_X \liminf_{n\to\infty} (2g - |f_n - f|) d\mu \\
&\leq \liminf_{n\to\infty} (2g - |f_n - f|) d\mu \quad \text{(by Fatou's Lemma)} \\
&= \int_X 2g d\mu + \liminf_{n\to\infty} \left(- \int_X |f_n - f| d\mu \right) \\
&= \int_X 2g d\mu - \limsup_{n\to\infty} \int_X |f_n - f| d\mu.
\end{aligned} \tag{2.57}$$

Since $g \in L^1(\mu)$, $\int_X 2g d\mu < +\infty$, hence it is necessary that

$$\limsup_{n\to\infty} \int_X |f_n - f| d\mu = \lim_{n\to\infty} \int_X |f_n - f| d\mu = 0. \tag{2.58}$$

3. Finally, we have for each n,

$$\lim_{n\to\infty} \left| \int_X f d\mu - \int_X f_n d\mu \right| \leq \lim_{n\to\infty} \int_X |f - f_n| d\mu = 0 \tag{2.59}$$

It follows that

$$\int_X f d\mu = \lim_{n\to\infty} \int_X f_n d\mu. \tag{2.60}$$

□

2.2 Null Sets and Almost Everywhere

2.2.1 Null Sets and Complete Measure

Definition 2.6 (Null set)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space, a set $A \subset X$ is called a* ***null set****, or* ***zero measure set****, if it is contained in a set with measure zero, i.e there exists $B \in \mathcal{M}$ such that $\mu(B) = 0$ and $A \subseteq B$.*

Definition 2.7 (Complete measure)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space, μ is said to be a* ***complete measure*** *if for every $E \in \mathcal{M}$ with $\mu(E) = 0$, one has $A \subseteq E \Rightarrow A \in \mathcal{M}$.*

In other words, a measure is complete if every null set is measurable.

Theorem 2.6 (Completion of a measure)**.** *Given a measure space $(X, \mathcal{M}, \mu)$, define $\widetilde{M} = \{E \subset X | \exists A, B \in \mathcal{M}, A \subseteq E \subseteq B, \text{ and } \mu(B \backslash A) = 0\}$, and $\tilde{\mu} : \widetilde{M} \to [0, +\infty]$, such that for each $E \in \widetilde{\mathcal{M}}$ as defined above, $\tilde{\mu}(E) = \tilde{\mu}(A)$. Then, $\widetilde{M}$ is a σ-algebra over X, and $\tilde{\mu}$ is well-defined and it is a complete measure.*

Proof. 1. First of all, we show that $\widetilde{\mathcal{M}}$ is a σ-algebra.

Note that $\mathcal{M} \subseteq \widetilde{\mathcal{M}}$, since for each $E \in \mathcal{M}$, we can choose $A = E$ and $B = E$, which have $A \subseteq E \subseteq B$ and $B \backslash A = \emptyset$ (thus $\mu(B \backslash A) = 0$). Hence, $E \in \widetilde{\mathcal{M}}$. Then, we verify that $\widetilde{\mathcal{M}}$ satisfies the three conditions of a σ-algebra.

(a) $\emptyset \in \mathcal{M} \subseteq \widetilde{\mathcal{M}}$.

(b) Suppose $E \in \widetilde{\mathcal{M}}$, there exists $A, B \in \mathcal{M}$ such that $A \subseteq E \subseteq B$ and $\mu(B \backslash A) = 0$. Then for E^c, we have $B^c \subseteq E^c \subseteq A^c$, where both A^c and B^c are in $\mathcal{M}$. In addition, $A^c \backslash B^c = A^c \cap B = B \backslash A$, thus $\mu(A^c \backslash B^c) = 0$. Hence, $E^c \in \widetilde{\mathcal{M}}$.

(c) Suppose $\{E_n\}_{n=1}^{\infty} \subset \mathcal{M}$, there exists, $A_1, A_2, \ldots \in \mathcal{M}$ and $B_1, B_2, \ldots \in \mathcal{M}$ such that $A_n \subseteq E_n \subseteq B_n$ and $\mu(B_n \backslash A_n) = 0$ for each $n \in \mathbb{N}$. Then for $\bigcup_{n=1}^{\infty} E_n$, we have

$$\bigcup_{n=1}^{\infty} A_n \subseteq \bigcup_{n=1}^{\infty} E_n \subseteq \bigcup_{n=1}^{\infty} B_n \tag{2.61}$$

where $\bigcup_{n=1}^{\infty} A_n \in \mathcal{M}$ and $\bigcup_{n=1}^{\infty} B_n \in \mathcal{M}$. In addition,

$$\left(\bigcup_{n=1}^{\infty} B_n\right) \backslash \left(\bigcup_{n=1}^{\infty} A_n\right) \subseteq \bigcup_{n=1}^{\infty} (B_n \backslash A_n). \tag{2.62}$$

It implies that

$$\mu\left(\left(\bigcup_{n=1}^{\infty} B_n\right) \backslash \left(\bigcup_{n=1}^{\infty} A_n\right)\right) \leq \sum_{n=1}^{\infty} \mu(B_n \backslash A_n) = 0. \tag{2.63}$$

Therefore, $\bigcup_{n=1}^{\infty} E_n \in \widetilde{\mathcal{M}}$.

2. We then show that $\tilde{\mu}$ is well defined.

Formally, this is equivalent to the following statement: *Given $E \in \widetilde{\mathcal{M}}$, and there exists $A_1, B_1, A_2, B_2 \in \mathcal{M}$ such that $A_1 \subseteq E \subseteq B_1$, $A_2 \subseteq E \subseteq B_2$, and $\mu(B_1 \backslash A_1) = \mu(B_2 \backslash A_2) = 0$, then $\mu(A_1) = \mu(A_2)$.*

Note that $A_2 \subseteq B_1$, then we have $A_2 \backslash A_1 \subseteq B_1 \backslash A_1$, hence $\mu(A_2 \backslash A_1) \leq \mu(B_1 \backslash A_1) = 0$. Consequently, $\mu(A_2) \leq \mu(A_1) + \mu(A_2 \backslash A_1) = \mu(A_1)$, likewise, we have $\mu(A_1) \leq \mu(A_2)$. It follows that $\mu(A_1) = \mu(A_2)$.

3. In the following, we continue to show that $\tilde{\mu}$ is a measure of the measurable space $(X, \widetilde{\mathcal{M}})$.

(a) For $\emptyset$, we can choose $A = B = \emptyset$, thus $\tilde{\mu}(\emptyset) = \mu(\emptyset) = 0$.

(b) Let $\{E_n\}_{n=1}^{\infty} \subset \widetilde{\mathcal{M}}$ be a countable collection of disjoint sets which are measurable w.r.t $\widetilde{\mathcal{M}}$, and let $E = \bigcup_{n=1}^{\infty} E_n$. Then we can find $A_1, A_2, \ldots \in \mathcal{M}$ and $B_1, B_2, \ldots \in \mathcal{M}$ such that $A_n \subseteq E_n \subseteq B_n$ and $\mu(B_n \backslash A_n) = 0$ for each $n \in \mathbb{N}$. Let $A = \bigcup_{n=1}^{\infty} A_n$ and $B = \bigcup_{n=1}^{\infty} B_n$. We have shown above that $A \subseteq E \subseteq B$ and $\mu(B \backslash A) = 0$. Hence, $\tilde{\mu}(E) = \mu(A)$, and $\tilde{\mu}(E_n) = \mu(A_n)$.

$\{E_n\}_{n=1}^{\infty}$ are disjoint and $A_n \subseteq E_n$, it follows that $\{A_n\}_{n=1}^{\infty}$ are disjoint, hence $\mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$. As a result, $\tilde{\mu}(E) = \sum_{n=1}^{\infty} \tilde{\mu}(E_n)$.

4. Finally, we show that μ is complete.

   Let $E \subset X$ such that there exists $B \in \mathcal{M}$ such that $E \subseteq B$ and $\mu(B) = 0$, then $E \in \widetilde{\mathcal{M}}$, which immediately follows from the definition of $\widetilde{\mathcal{M}}$ by choosing $A = \emptyset$.

□

2.2.2 Almost Everywhere

Definition 2.8 (Almost everywhere)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space, and P be a property of X, then we say that P holds* ***almost everywhere*** *with respect to μ if there exists $N \in \mathcal{M}$ with $\mu(N) = 0$ such that P holds for each $x \in X \backslash N$. This is notated by P holds a.e.$[\mu]$. Here, $[\mu]$ can be omitted if the measure is clear from context.*

In other words, P holds almost everywhere, if it holds over the entire X except for a null set.

Definition 2.9 (Almost everywhere equality)**.** *Let f and g be two measurable functions defined on a measure space $(X, \mathcal{M}, \mu)$, we say that f equals g almost everywhere if $\mu\{x|f(x) \neq g(x)\} = 0$, notated by $f = g$, a.e.$[\mu]$.*

Proposition 2.7. *Almost everywhere equality is an equivalence relation between measurable functions.*

It is trivial to check this.

Theorem 2.7. *Let $(X, \mathcal{M}, \mu)$ be a measure space, f and g be two measurable functions, then*

$$f = g, \ a.e.[\mu] \quad \Rightarrow \quad \int_X f d\mu = \int_X g d\mu.$$

Proof. Let $N = \{x|f(x) \neq g(x)\}$. If $N = \emptyset$, then f is the same as g, then $\int_X f d\mu = \int_X g d\mu$ trivially holds. Otherwise, we have $\int_X f d\mu = \int_{X\backslash N} f d\mu + \int_N f d\mu$. Since $\mu(N) = 0$, $\int_N f d\mu = 0$, thus $\int_X f d\mu = \int_{X\backslash N} f d\mu$. Likewise, $\int_X g d\mu = \int_{X\backslash N} g d\mu$. And f equals g on $X\backslash N$, we thus have $\int_{X\backslash N} f d\mu = \int_{X\backslash N} g d\mu$. Hence, the equality is established. □

In the following, we extend the concept of measurable function to those defined almost everywhere (but not necessarily the entire space).

Definition 2.10 (measurable function (defined almost everywhere))**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space, $E \in \mathcal{M}$ and $\mu(E^c) = 0$, Y be a topological space, a function $f : E \to Y$ is said to be* ***measurable on*** *X if $f^{-1}(V) \cap E$ is measurable for each open set $V \subseteq Y$.*

In addition, we define its Lebesgue integral over X by

$$\int_X f d\mu := \int_E f d\mu$$

and we say $f \in L^{-1}(X, \mu)$ if $\int_E |f| d\mu < +\infty$.

Lemma 2.4. *Integrable function is finite almost everywhere. Formally, let $(X, \mathcal{M}, \mu)$ be a measure space, and $f \in L^1(\mu)$, then*

$$|f(x)| < +\infty, \ a.e.[\mu].$$

Proof. Let $S_n = \{x \in X||f(x)| > n\}$, and $M = \int_X |f| d\mu$, since $f \in L^1(\mu)$, $M < +\infty$. Then, we have

$$M = \int_X |f| d\mu \geq \int_{S_n} |f| d\mu \geq \int_{S_n} n d\mu = n\mu(S_n). \tag{2.64}$$

Hence, $\mu(S_n) \leq M/n$, thus $\lim_{n\to\infty} \mu(S_n) = 0$. Let $R = \{x \in X||f(x)| = +\infty\}$, then $R \subseteq S_n$, $\forall n \in \mathbb{N}$, it follows that $\mu(R) \leq \mu(S_n)$, $\forall n \in \mathbb{N}$, therefore, $\mu(R) = 0$. □

The following theorem states some important facts about infinite series of integrable functions.

Theorem 2.8. *Let $(f_n)_{n=1}^{\infty}$ be a sequence of measurable functions defined almost everywhere on a measure space $(X, \mathcal{M}, \mu)$, such that $\sum_{n=1}^{\infty} \int_X |f_n| d\mu < +\infty$ then. we have*

1. *$f(x) = \sum_{n=1}^{\infty} f_n(x)$ converges almost everywhere w.r.t μ;*

2. *f is integrable, i.e $f \in L^1(X, \mu)$;*

3. *$\int_X f d\mu = \sum_{n=1}^{\infty} \int f_n d\mu$.*

Note that even when each of f_n is defined on the entire X, their sum converges almost everywhere (not necessarily the entire X).

Proof. Let S_n is the domain of f_n, in which f_n is defined, then $\mu(S_n^c) = 0$. Let $S = \cap_{n=1}^{\infty} S_n$, it is easy to see that $\mu(S^c) = 0$. We define $\phi : S \to [0, +\infty]$ by $\phi(x) = \sum_{n=1}^{\infty} |f_n(x)|$, then by MCT,

$$\int_S \phi d\mu = \int_S \sum_{n=1}^{\infty} |f_n| d\mu = \sum_{n=1}^{\infty} \int_S |f_n| d\mu < +\infty \tag{2.65}$$

Hence, $\phi \in L^1(\mu)$. Let $E = \{x \in S | \phi(x) < +\infty\}$, by the lemma above, $\mu(E^c) = 0$. Note that f is absolutely convergent on E and $|f(x)| le \phi(x)$, $\forall x \in E$, therefore, $f \in L^1(\mu)$.

Let $g_n = \sum_{i=1}^{N} f_i(x)$, then $|g_n(x)| \leq \phi(x)$, $\forall x \in E$. By definition of infinite series, $\lim_{n\to\infty} g_n(x) = f(x)$, $\forall x \in E$. Hence, by DCT (g_n is dominated by $\phi \in L^1(\mu)$),

$$\int_E f d\mu = \lim_{n\to\infty} \int_E g_n d\mu = \lim_{n\to\infty} \sum_{i=1}^{n} \int_E f_i d\mu = \sum_{n=1}^{\infty} \int_E f_n d\mu. \tag{2.66}$$

□

Proposition 2.8. *Let $(X, \mathcal{M}, \mu)$ be a measure space, then*

1. *If $f : X \to [0, +\infty]$ is measurable such that $\int_E f d\mu = 0$ for $E \in \mathcal{M}$, then $f = 0$, a.e.$[\mu]$ on E.*

2. *Let $f \in L^1(\mu)$, such that $\int_E f d\mu = 0, \forall E \in \mathcal{M}$, then $f = 0$, a.e.$[\mu]$ on X.*

3. *Let $f \in L^1(\mu)$, such that $|\int_X f d\mu| = \int_X |f| d\mu$, then there exists $\alpha \in \mathbb{C}$ such that $\alpha f = |f|$, a.e. on X.*

Proof.

1. Let $A_n = \{x \in E | f(x) > 1/n\}$, then

$$0 = \int_E f d\mu \geq \int_{A_n} f d\mu \geq \int_{A_n} \frac{1}{n} d\mu = \frac{1}{n}\mu(A_n). \tag{2.67}$$

Hence, $\mu(A_n) = 0$. Let $S = \{x \in E | f(x) > 0\}$, note that $S = \bigcup_{n=1}^{\infty} A_n$. As a result,

$$\mu(S) \leq \sum_{n=1}^{\infty} \mu(A_n) = 0. \tag{2.68}$$

It means that $f(x) = 0$, a.e.$[\mu]$ on E.

2. Decompose f into $f = (u^+ - u^-) + i(v^+ - v^-)$, since f is measurable, u^+, u^-, v^+, v^- are measurable. Let $A = (u^+)^{-1}((0, +\infty])$. And, we have

$$0 = \text{Re} \int_A f d\mu = \int_A u^+ d\mu \tag{2.69}$$

By the statement shown above, $u^+ = 0$ almost every on A, thus $\mu(A) = 0$. Likewise, the measure of $(u^-)^{-1}((0, +\infty])$, $(v^+)^{-1}((0, +\infty])$, and $(v^-)^{-1}((0, +\infty])$ are all zeros. Note that

$$f^{-1}(\mathbb{R}\backslash\{0\}) \subset (u^+)^{-1}((0, +\infty]) \cup (u^-)^{-1}((0, +\infty]) \cup (v^+)^{-1}((0, +\infty]) \cup (u^-)^{-1}((0, +\infty]). \tag{2.70}$$

Hence, $\mu(f^{-1}(\mathbb{R}\backslash\{0\}) = 0$, which means that $f = 0$, a.e.$[\mu]$ on X.

3. Let $z = \int_X f d\mu$, and choose $\alpha = \bar{z}/|z|$, then $\alpha z = |z| \in [0, +\infty]$. Then

$$|z| = \int_X \alpha f d\mu = \int_X |f| d\mu. \tag{2.71}$$

Note that $|\alpha| = 1$, $|alpha f| = |f|$, thus $\text{Re}(\alpha f) \leq |f|$ on X, i.e $\text{Re}(|f| - \alpha f) \geq 0$ on X, let $h = |f| - \alpha f$, we can write $h = u^+ + iv$, since $u^- = 0$ on X. From (2.71), $\int_X h d\mu = 0$, it follows that $\int_X u^+ d\mu = 0$, thus $u^+ = 0$, a.e. on X, i.e $|f| = Re(\alpha f)$, a.e. on X, thus $|f| = \alpha f$, a.e. on X.

□

Theorem 2.9. *Let $(X, \mathcal{M}, \mu)$ be a measure space with locally finite measure, $f \in L^1(\mu)$, and S be a closed subset of $\mathbb{C}$, if for each $E \in \mathcal{M}$ with $0 < \mu(E) < +\infty$, one has*

$$A_E(f) = \frac{1}{\mu(E)} \int_E f d\mu \in S,$$

then $f(x) \in S$, a.e.$[\mu]$ on X.

Proof. Since S is closed, S^c is open. If, $S^c = \emptyset$, then we are done. Otherwise, for each $c \in S^c$, there exists $r > 0$ such that $B_r(c) \subset S^c$, where $B_r(c)$ is the open ball of radius r centered at c.

Note that we need prove $f^{-1}(S^c)$ has zero measure, and $f^{-1}(S^c)$ is countable union of the sets in form of $f^{-1}(B_r(c)$ with $B_r(c) \subseteq S^c$. Hence, it is enough to show that $\mu(f^{-1}(B_r(c))) = 0$ for each $B_r(c) \subseteq S^c$.

By contradition, we assume that there exists $B_r(c) \subseteq S^c$ such that $\mu(E) > 0$ where $E = f^{-1}(B_r(c))$. (If $\mu(E) = +\infty$, we can find any finite-measure subset in it. This can be done since μ is locally finite.) Then, we have

$$|A_E(f) - c| = \left| \frac{1}{\mu(E)} \int_E f d\mu - \frac{1}{\mu(E)} \int_E c d\mu \right| \leq \frac{1}{\mu(E)} \int_E |f - c| d\mu \leq \frac{1}{\mu(E)} \int_E r d\mu = r. \tag{2.72}$$

Since, $A_E(f) \in S$, and $B_r(c) \subseteq S^c$, hence $\forall x \in S$, $|x - c| > r$, thus $A_E(f) > r$, leading to contradition. □

(Note that in the lecture, the condition is given as $\mu(X) < +\infty$, actually, as we have seen that this can be relaxed to a mild condition, namely local finiteness, such that it applies to more cases.)

Theorem 2.10. *Let $\{E_n\}_{n=1}^{\infty}$ be a collection of measurable sets in a measure space $(X, \mathcal{M}, \mu)$ such that $\sum_{n=1}^{\infty} \mu(E_n) < +\infty$, then almost every $x \in X$ is covered by at most finitely many E_n.*

Proof. Let $A = \{x \in X | x \text{ is in infinitely many } E_n\}$, we need show $\mu(A) = 0$. Define, $g = \sum_{n=1}^{\infty} \chi_{E_n}$, then it is equivalent to proving $g < +\infty$, *a.e.* on X as $x \in A \Leftrightarrow g(x) = +\infty$. This readily follows from

$$\int_X g d\mu = \int_X \sum_{n=1}^{\infty} \chi_{E_n} d\mu = \sum_{n=1}^{\infty} \int_X \chi_{E_n} d\mu = \sum_{n=1}^{\infty} \mu(E_n) < +\infty. \tag{2.73}$$

□

Note that in this proof, A can be formally written as $A = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_n$.

Remark: it is claimed in the lecture that $\mu(A) = 0 \Leftrightarrow \sum_{n=1}^{\infty} \mu(E_n) < +\infty$ when $\mu(X) < +\infty$. I notice that *this claim is not necessarily true.*

2.3 Riesz Representation Theorem

Definition 2.11 (positive linear functional)**.** *A* ***positive linear functional*** *defined on a function space is a linear functional that yields non-negative value for each non-negative function.*

Definition 2.12 (Regular measure)**.** *Let X be a topological space, and $\mathcal{M}$ is a σ-algebra over X, then a measure $\mu : \mathcal{M} \to [0, +\infty]$ is called a* ***regular measure****, if it satisfies*

1. *(outer regularity)*
$$\mu(E) = \inf\{\mu(V) | E \subseteq V,\ V \textit{ is open}\}, \quad \forall E \in \mathcal{M};$$
2. *(inner regularity)*
$$\mu(E) = \sup\{\mu(K) | K \subseteq E,\ K \textit{ is compact}\}, \quad \forall E \in \mathcal{M}.$$

In particular, we say that μ satisifies *outer regularity* at E if $\mu(E) = \inf\{\mu(V) | E \subseteq V,\ V \text{ is open}\}$, and that is satisfies *inner regularity* at K if $\mu(E) = \sup\{\mu(K) | K \subseteq E,\ K \text{ is compact}\}$.

Theorem 2.11 (Riesz representation theorem in a σ-compact space)**.** *Let X be a Hausdorff, locally compact, and σ-compact topological space, Λ is a positive linear functional on $C_C(X)$, then there exists a unique regular Borel measure μ such that $\Lambda f = \int_X f d\mu$, $\forall f \in C_c(X)$.*

A more general version of Resiz representation theorem (without the requirement of σ-compactness) is

Theorem 2.12 (Riesz representation theorem). *Let X be a Hausdorff, locally compact topological space, Λ be a positive linear functional on $C_C(X)$, then there exists a σ-algebra $\mathcal{M}$ that contains the Borel σ-algebra, and there exists a unique positive measure μ on $\mathcal{M}$ such that it satisfies the following five conditions:*

C.1: $\Lambda f = \int_X f d\mu, \ \forall f \in C_c(X)$;

C.2: $\mu(K) < +\infty$ *for each compact set* K;

C.3: (outer regularity) $\mu(E) = \inf\{\mu(V) | E \subseteq V, \ V \text{ is open}\}$ *for all* $E \in \mathcal{M}$;

C.4: ((semi) inner regularity) $\mu(E) = \sup\{\mu(K) | K \subseteq E, \ K \text{ is compact}\}$ *for all* $E \in \mathcal{M}$ *with* $\mu(E) < +\infty$ *or* E *is open;*

C.5: (completeness) μ *is a complete measure, i.e.* $\mu(E) = 0 \ \Rightarrow \ A \in \mathcal{M}, \ \forall A \subseteq E$.

We divide the proof into two parts: the first part is to show the existence of such μ; the second part is to show that the measure μ that satisfies all the properties is unique.

In addition, for conciseness, we introduce the following notations to represent the relation between a function $f : X \to [0,1] \in C_c(X)$ and a set A. $f \prec A$ means supp$f \ \subseteq A$; while $f \succ V$ means supp$(1-f) \ \subset A$, i.e. $f(x) = 1, \ \forall x \in A$.

Proof of Existence. We first construct the σ-algebra $\mathcal{M}$ and the measure μ explicitly, and then prove that the constructed objects satisfy the required conditions.

First of all, we define a non-negative real valued function $\mu : 2^X \to [0, +\infty]$ as follows. For each open set V,

$$\mu_0(V) = \sup\{\Lambda f | f \prec V\},$$

and then it can be extended to each subset E of X as

$$\mu(E) = \inf\{\mu_0(V) | E \subseteq V, \ V \text{ is open}\}.$$

Actually, μ as defined above is an outer measure that on which we are going to construct the σ-algebra and derive the measure by restriction.

Claim (1) μ is an outer measure on X with $\mu(E) = \mu_0(E)$ when E is open, i.e. μ is an extension of μ_0.

Proof of Claim (1). The proof takes several steps,

1. First of all, we show that μ_0 satisfies the following three conditions.

   (a) ($\mu_0(\emptyset) = 0$) It is easy to see that the only function f that has $f \prec \emptyset$ is the zero function, thus $\Lambda f = 0$, and $\mu_0(\emptyset) = 0$.

   (b) (monotonicity) Given two open sets $V_1 \subset V_2$, $f \prec V_1 \Rightarrow f \prec V_2$, and thus $\mu_0(V_1) \le \mu_0(V_2)$ by definition, as V_2 may admit a larger set of functions whose supports contained in it.

   (c) (sub-additivity) We first show that $\mu_0(V_1 \cup V_2) \le \mu_0(V_1) + \mu_0(V_2)$ for any two open sets V_1, V_2, and then extend it to finite sub-additivity.
   Consider any function $g \prec V_1 \cup V_2$, by partition of unity, there exist h_1 and h_2 such that $h_1 \prec V_1$, $h_2 \prec V_2$ and $h_1(x) + h_2(x) = 1, \ \forall x \in \text{supp } g$. As a result, $h_1 g \prec V_1$, $h_2 g \prec V_2$, and $h_1 g + h_2 g = g$. By linearity of Λ,

   $$\Lambda g = \Lambda(h_1 g + h_2 g) = \Lambda(h_1 g) + \Lambda(h_2 g) \le \mu_0(V_1) + \mu_0(V_2), \tag{2.74}$$

   This holds for any $g \prec V_1 \cup V_2$, consequently,

   $$\mu_0(V_1 \cup V_2) = \sup_{g \prec V_1 \cup V_2} \Lambda g \le \mu_0(V_1) + \mu_0(V_2). \tag{2.75}$$

   By induction, this results can be generalized to finite sub-additivity as for any open sets $V_1, \dots, V_n$,

   $$\mu_0\left(\bigcup_{i=1}^n V_i\right) \le \sum_{i=1}^n \mu_0(V_i). \tag{2.76}$$

2. Then, we show that μ is an outer measure based on the results above.
   (a) ($\mu(\emptyset) = 0$) The smallest open set that contains an empty set is obviously an empty set. Thus, by definition, we have $\mu(\emptyset) = 0$.
   (b) (monotonicity) Given any subsets E_1 and E_2 of X with $E_1 \subset E_2$, it is clearly that $E_2 \subseteq V \Rightarrow E_1 \subseteq V$. It follows that $\mu(E_1) \leq \mu(E_2)$ by definition.
   (c) (countable sub-additivity) Consider a countable collection of sets $\{E_i\}_{i=1}^{\infty}$, and let $E = \bigcup_{i=1}^{\infty} E_i$. We assume that $\mu(E_i) < +\infty$ for all i, otherwise, the sub-additivity trivially holds as both sides are infinity. By definition of $\mu(E)$, given any $\epsilon > 0$, for each i, there is an open set V_i with $E_i \subseteq V_i$ such that $\mu_0(V_i) \leq \mu(E_i) + 2^{-i}\epsilon$. Let $f : X \to [0, 1] \in C_c(X)$ have $f \prec E$. Then, $\{V_i\}_{i=1}^{\infty}$ is an open cover of supp f, which is a compact set, hence there is a finite number N such that $f \prec \bigcup_{i=1}^{N} V_i$. It follows that
   $$\Lambda f \leq \mu_0 \left(\bigcup_{i=1}^{N} V_i \right) \leq \sum_{i=1}^{N} \mu_0(V_i) \leq \sum_{i=1}^{\infty} \mu(E_i) + \epsilon. \tag{2.77}$$
   As this holds for all open sets $V \supseteq \bigcup_{i=1}^{\infty} E_i$ and all $f \prec V$, we have
   $$\mu(E) = \sup\{\Lambda f | f \prec V \supseteq E\} \leq \sum_{i=1}^{\infty} \mu(E_i) + \epsilon. \tag{2.78}$$
   Since it holds for any $\epsilon > 0$, the sub-additivity is established.
3. Finally, we show that μ and μ_0 agrees on open sets. Note that when E is open, the smallest open set that contains E is itself. By monotonicity of μ_0, we have
   $$\mu(E) = \inf\{\mu_0(V) | E \subseteq V,\ V \text{ is open}\} = \mu_0(E). \tag{2.79}$$

The claim (1) is finally proved. □

For the convenience of discussing inner and outer regularity, we define

$$\mu_*(E) = \sup\{\mu(K) | K \subseteq E,\ K \text{ is compact}\}, \quad \text{and} \quad \mu^*(E) = \inf\{\mu(V) | E \subseteq V,\ V \text{ is open}\}.$$

Hence, $\mu(E)$ satisfies *outer regularity* at E when $\mu(E) = \mu^*(E)$, and satisfies *inner regularity* when $\mu(E) = \mu_*(E)$. We can see that μ is inherently outer regular by its definition. ($C.3$ automatically holds)

Then, we define

$$\mathcal{M}_F = \{E | \mu(E) < +\infty \text{ and } \mu(E) = \mu_*(E)\},$$

and

$$\mathcal{M} = \{E | K \cap E \in \mathcal{M}_F,\ \forall K \text{ compact}\}.$$

We are going to show that $\mathcal{M}$ is an σ-algebra with μ being a measure over it. To this end, we first establish several important results about $\mathcal{M}_F$.

Claim (2) if K is compact, then $K \in \mathcal{M}_F$ and $\mu(K) = \inf\{\Lambda f | K \prec f\}$.
Claim (3) if V is open, then $\mu(V) = \mu_*(V)$, i.e. $\mu(V) < +\infty \Rightarrow V \in \mathcal{M}_F$.

We note that C.4 is established when these two claims are proved.

Proof of Claim (2). First, it immediately follows from the monotonicity of μ that for each compact set K,

$$\mu(K) = \sup\{\mu(K') | K' \subseteq K,\ K' \text{ is compact}\}. \tag{2.80}$$

Therefore, $K \in \mathcal{M}_F$. In the following, we will show the identity that $\mu(K) = \inf\{\Lambda f | f \succ K\}$.

For a given compact set K, let $f : X \to [0, 1] \in C_c(X)$ satisfy $f \succ K$ and $0 < \alpha < 1$, we define $V_\alpha = f^{-1}((\alpha, +\infty))$, which is open due to the continuity of f. We note that $K \subseteq V_\alpha$ for any $\alpha \in (0, 1)$ and $\alpha g \leq f,\ \forall g \prec V_\alpha$. By claim (1), we have

$$\mu(K) \leq \mu(V_\alpha) = \sup_{g \prec V_\alpha} \Lambda g \leq \alpha^{-1} \Lambda f, \tag{2.81}$$

this holds any $f \succ K$ and any $\alpha \in (0,1)$, thus $\mu(K) \leq \inf\{\Lambda f | K \prec f\}$, since $\Lambda f < +\infty$ for all $f \in C_c(f)$ (why?), $\mu(K) < +\infty$. On the other hand, by definition of μ and claim (1), given $\epsilon > 0$, there is an open set V with $K \subset V$ such that $\mu(V) < \mu(K) + \epsilon$. By Urysohn lemma, there is a $f \in C_C(X)$ with $K \prec f \prec V$, hence,

$$\Lambda f \leq \mu(V) \leq \mu(K) + \epsilon. \tag{2.82}$$

As this holds for any $\epsilon > 0$, we have $\mu(K) \geq \inf_{f \succ K} \Lambda f$, and thus $\mu(K) = \inf\{\Lambda f | K \prec f\}$. □

Proof of Claim (3). Given any open set V, first of all $\mu(V) \geq \mu_*(V)$, which directly follows from the monotonicity of μ and the definition of μ^*. For the other direction, we need to show that for any open set V,

$$\sup\{\Lambda f | f \prec V\} \leq \sup\{\mu(K) | K \subseteq V\}. \tag{2.83}$$

It suffices to show that *for each $f \prec V$, there exists a compact set $K \subset V$ such that $\Lambda f \leq \mu(K)$*. To show this, we let $K = \text{supp } f$, clearly, $K \subseteq V$ due to $f \prec V$. And, for each $f' \succ K$, we have $\Lambda f \leq \Lambda f'$, due to positiveness of Λ, thus $\Lambda f \leq \mu(K) = \inf\{\Lambda f' | f' \succ K\}$ (the last equality is due to claim (2)). □

An important property with $\mathcal{M}_F$ is that μ satisfies countable additivity on $\mathcal{M}_F$. This property will be used in proving other claims.

Claim (4) Let $\{E_i\}_{i=1}^{\infty}$ be a countable collection of disjoint sets in $\mathcal{M}_F$, let $E = \bigsqcup_{i=1}^{\infty} E_i$, then $\mu(E) = \sum_{i=1}^{\infty} \mu(E_i)$. In addition, if $\mu(E) < +\infty$, $E \in \mathcal{M}_F$.

Proof. We first show that for any two compact sets K_1 and K_2, we have $\mu(K_1 \sqcup K_2) = \mu(K_1) + \mu(K_2)$. From claim (1), we immediately have $\mu(K_1 \sqcup K_2) \leq \mu(K_1) + \mu(K_2)$ due to sub-additivity. Hence, it suffices to show the inequality in the other direction, i.e. $\mu(K_1 \sqcup K_2) \geq \mu(K_1) + \mu(K_2)$.

First, by Urysohn lemma, there exist $f : X \to [0,1] \in C_c(X)$ such that $f(x) = 1, \forall x \in K_1$ and $f(x) = 0, \forall x \in K_2$. On the other hand, from claim (2), given any $\epsilon > 0$, there is a $g : X \to [0,1] \in C_c(X)$ such that $K_1 \sqcup K_2 \prec g$ and $\Lambda g < \mu(K_1 \sqcup K_2) + \epsilon$. As a result,

$$\mu(K_1) + \mu(K_2) \leq \Lambda(fg) + \Lambda((1-f)g) = \Lambda g \leq \mu(K_1 \sqcup K_2) + \epsilon. \tag{2.84}$$

As this holds for any $\epsilon > 0$, $\mu(K_1) + \mu(K_2) \leq \mu(K_1 \sqcup K_2)$, thus $\mu(K_1 \sqcup K_2) = \mu(K_1) + \mu(K_2)$. By induction, this generalizes to any finite union of compact sets.

Then, we consider a countable collection $\{E_i\}_{i=1}^{\infty} \subset \mathcal{M}_F$ where E_i are disjoint. Let $E = \bigsqcup_{i=1}^{\infty} E_i$, Likewise, it is enough to prove $\mu(E) \geq \sum_{i=1}^{\infty} \mu(E_i)$. Given $\epsilon > 0$, for each i, since $E_i \in \mathcal{M}_F$, there exists a compact set $H_i \subseteq E_i$ such that $\mu(H_i) \geq \mu(E_i) - 2^{-i}\epsilon$. Let $G_n = \bigsqcup_{i=1}^{n} H_i$, then G_n is a compact set for each n, and $\mu(G_n) = \sum_{i=1}^{n} \mu(H_i)$ due to the conclusion shown right above.

$$\mu(E) \geq \mu(G_n) = \sum_{i=1}^{n} \mu(H_i) \geq \mu_{i=1}^{n} \mu(E_i) - \epsilon. \tag{2.85}$$

As this holds for any $n \in \mathbb{N}$ and $\epsilon > 0$, take $n \to \infty$, we have $\mu(E) \geq \sum_{i=1}^{n} \mu(E_i)$. □

We have shown that compact sets and open sets with finite μ are all in $\mathcal{M}_F$. Actually, we will show that any set in $\mathcal{M}_F$ can be approximated by compact sets from below, and by open sets from above.

Claim (5) If $E \in \mathcal{M}_F$, for any $\epsilon > 0$, there exist a compact set K and an open set V such that $K \subseteq E \subseteq V$ and $\mu(V \backslash K) < \epsilon$.

Proof of Claim (5). By definition of $\mathcal{M}_F$, μ_* and μ^*, for each $E \in \mathcal{M}_F$, there exist K and V, such that

$$\mu(V) - \frac{\epsilon}{2} < \mu(E) < \mu(V) + \frac{\epsilon}{2}. \tag{2.86}$$

Note that $V \backslash K$ is open, and $\mu(V) < \mu(K) + \epsilon < +\infty$, by claim (3), $V \backslash K \in \mathcal{M}_F$. On the other hand, by claim (4), $\mu(V) = \mu(K) + \mu(V \backslash K)$, thus $\mu(V \backslash K) < \epsilon$. □

Now, we have sufficiently characterize what are in $\mathcal{M}_F$.
Claim (6) Let $A, B \in \mathcal{M}_F$, then $A \cup B, A \cap B, A \backslash B \in \mathcal{M}_F$.

Proof of Claim (6). From the definition of $\mathcal{M}_F$ and monotonicity of μ, we can see that to show $E \in \mathcal{M}_F$, it suffices to show there exists a compact set $K \subseteq E$ such that $\mu(E) \leq \mu(K) + \epsilon$ for any given $\epsilon > 0$.

Let $A, B \in \mathcal{M}_F$, then there exist K_1, K_2, V_1 and V_2 with $K_1 \subseteq A \subseteq V_1$ and $K_2 \subseteq B \subseteq V_2$ such that $\mu(V_i \backslash K_i) < \epsilon$ for any given $\epsilon > 0$, according to claim (5).

1. ($A \cup B \in \mathcal{M}_F$). Clearly, $K_1 \cup K_2 \subseteq A \cup B \subseteq V_1 \cup V_2$, $K_1 \cup K_2$ is compact, and $V_1 \cup V_2$ is open. In addition, we have
$$(V_1 \cup V_2) \backslash (K_1 \cup K_2) \subseteq (V_1 \backslash K_1) \cup (V_2 \backslash K_2). \tag{2.87}$$
As μ is an outer measure,
$$\mu((V_1 \cup V_2) \backslash (K_1 \cup K_2)) \leq \mu(V_1 \backslash K_1) + \mu(V_2 \backslash K_2) \leq 2\epsilon. \tag{2.88}$$

2. ($A \cap B \in \mathcal{M}_F$). Clearly, $K_1 \cap K_2 \subseteq A \cap B \subseteq V_1 \cap V_2$, $K_1 \cap K_2$ is compact, and $V_1 \cap V_2$ is open. And,
$$(V_1 \cap V_2) \backslash (K_1 \cap K_2) = V_1 \cap V_2 \cap (K_1^c \cup K_2^c) = (V_1 \cap K_1^c \cap V_2) \cup (V_1 \cap V_2 \cap K_2^c) \subseteq (V_1 \backslash K_1) \cup (V_2 \backslash K_2), \tag{2.89}$$
$$\mu((V_1 \cap V_2) \backslash (K_1 \cap K_2)) \leq \mu(V_1 \backslash K_1) + \mu(V_2 \backslash K_2) \leq 2\epsilon. \tag{2.90}$$

3. ($A \backslash B \in \mathcal{M}_F$). Note that
$$A \backslash B \subseteq V_1 \backslash V_2 \subseteq (V_1 \backslash K_1) \cup (K_1 \backslash V_2) \cup (V_2 \backslash K_2), \tag{2.91}$$
thus
$$\mu(A \backslash B) \leq \mu(K_1 \backslash V_2) + 2\epsilon, \tag{2.92}$$
where $K_1 \backslash V_2$ is compact set contained in $A \backslash B$.

To sum up, $\mathcal{M}_F$ is closed under finite union, finite intersection, and set difference. □

By extending $\mathcal{M}_F$ to $\mathcal{M} = \{E | K \cap E \in \mathcal{M}_F, \forall K \text{ compact}\}$, we obtain an σ-algebra.

Claim (7) $\mathcal{M}$ is a σ-algebra that contains the Borel σ-algebra.

Proof of Claim (7). We first show that $\mathcal{M}$ is a σ-algebra.

1. $\emptyset \in \mathcal{M}$, because $\emptyset \cap K = \emptyset \in \mathcal{M}_F$ for every compact set K.

2. Let $A \in \mathcal{M}$, i.e. for each compact set K, $A \cap K \in \mathcal{M}_F$, thus $A^c \cap K = K \backslash (A \cap K) \in \mathcal{M}_F$, as $K \in \mathcal{M}_F$ and $\mathcal{M}_F$ is closed under set difference. Therefore, $A^c \in \mathcal{M}$.

3. Let $\{A_i\}_{i=1}^{\infty} \subset \mathcal{M}$, then for for any compact set K, $A_i \cap K \in \mathcal{M}_F$ for each i. Let $B_1 = A_1 \cap K$ and $B_{n+1} = (A_{n+1} \cap K) \backslash \bigcup_{i=1}^{n} B_i$ for each $n \geq 1$. It is easy to see that $B_n \in \mathcal{M}_F$ for each n due to the fact that $\mathcal{M}_F$ is closed under union and set difference. We note that $A \cap K = \bigsqcup_{n=1}^{\infty} B_n$, and $\mu(A \cap K) \leq \mu(K) < +\infty$, by claim (4), $A \cap K \in \mathcal{M}_F$. As this holds for any compact set K, $A \in \mathcal{M}$.

Now, we can conclude that $\mathcal{M}$ is a σ-algebra. In the following, we show that it contains the Borel σ-algebra. It suffices to show that it contains every closed set. For each closed set C, and any compact set K, $C \cap K$ is compact, and thus $C \cap K \in \mathcal{M}_F$, which follows that $C \in \mathcal{M}$. □

Claim (8) $\mathcal{M}_F = \{A \in \mathcal{M} | \mu(A) < +\infty\}$, i.e. $A \in \mathcal{M}_F$ if and only if $A \in \mathcal{M}$ and $\mu(A) < +\infty$.

Proof of Claim (8). The proof is conducted in two directions respectively.

1. $A \in \mathcal{M}_F \Rightarrow A \in \mathcal{M}$ and $\mu(A) < +\infty$. Let $A \in \mathcal{M}_F$, then for any compact set K, $K \in \mathcal{M}_F$ by claim (2), and $A \cap K \in \mathcal{M}_F$ by claim (6), thus $A \in \mathcal{M}$. The statement $\mu(A) < +\infty$ is directly from the definition of $\mathcal{M}_F$.

2. $A \in \mathcal{M}$ and $\mu(A) < +\infty \Rightarrow A \in \mathcal{M}_F$. It suffices to show that given any $\epsilon > 0$, there exists a compact set $K \subseteq A$ such that $\mu(A) \leq \mu(K) + \epsilon$.

   As $\mu(A) = \inf\{\mu(V)|A \subseteq V \text{ and } V \text{ is open}\}$, fix $\epsilon > 0$, there exists an open set $V \supseteq A$ such that $\mu(V) < +\infty$, by claim (5), there exists an open set $K \subseteq V$ such that $\mu(V\backslash K) < \epsilon/2$. As $A \in \mathcal{M}$, $A \cap K \in \mathcal{M}_F$, then there exists another compact set K' such that $K' \subseteq A \cap K$ and $\mu((A\cap K)\backslash K') < \epsilon/2$. Due to the sub-additivity of μ, we have

$$\mu(A) \leq \mu(K') + \mu(A\backslash K'), \tag{2.93}$$

   where

$$A\backslash K' \subseteq ((A \cap K)\backslash K') \cup (A\backslash K) \subseteq ((A \cap K)\backslash K') \cup (V\backslash K) \leq \epsilon/2 + \epsilon/2 = \epsilon. \tag{2.94}$$

   Hence, we can conclude that $A \in \mathcal{M}_F$.

□

Claim (9) The restriction of μ to $\mathcal{M}$ is a complete measure.

Proof. We have shown in claim (1) that $\mu(\emptyset) = 0$, thus we only need to show μ satisfies σ-additivity on $\mathcal{M}$.

Let $\{E_i\}_{i=1}^{\infty}$ be a collection of disjoint sets in $\mathcal{M}$, and let $E = \bigsqcup_{i=1}^{\infty} E_i$. If there exist i with $\mu(E_i) = +\infty$, then by monotonicity $\mu(E) = +\infty$, and the countability of μ trivially holds in this case. Assume $\mu(E_i) < +\infty$ for each i, then every E_i is in $\mathcal{M}$ by claim (8). The countable additivity in this case has been established by claim (4).

Finally, we show that μ is complete. Let $A \in \mathcal{M}$ with $\mu(A) = 0$, and $B \subset A$. By claim (8), $A \in \mathcal{M}_F$. To show that B is also in $\mathcal{M}_F$, it is enough to show that for every compact set K with $K \subseteq B$, $\mu(K) = 0$, this directly follows from the monotonicity of μ. □

Claim (10) $\Lambda f = \int_X f d\mu,\ \forall f \in C_c(X)$. (This corresponds to C.1)

Proof. First of all we prove that for any real valued function $f : X \to \mathbb{R} \in C_c(X)$, we have $\Lambda f = \int_X f d\mu$.

Let $K = \text{supp } f$. Since f is continuous, $f(K)$ is compact, and thus there exists a closed interval $[a, b]$ with $f(X) \subseteq [a, b]$. Given any $\epsilon > 0$, choose a finite sequence of values $y_0, y_1, \dots, y_n$ such that $y_0 < a \leq y_1 \leq \dots \leq y_n = b$ such that $y_{i+1} - y_i < \epsilon$ for $i = 0, \dots, n-1$. Define $E_i = f^{-1}((y_{i-1}, y_i]) \cap K$, then $E_1, \dots, E_n$ are a collection of disjoint Borel sets, with $\bigsqcup_{i=1}^{n} E_i = K$. Therefore, for each i, there exists an open set V_i with $E_i \subseteq V_i$ such that $\mu(V_i) < \mu(E_i) + \epsilon/n$, and $f(x) < y_i + \epsilon,\ forall x \in V_i$. Then $\{V_i\}_{i=1}^{n}$ form a open cover of K, by partition of unity, For V_i, there is $h_i \prec V_i$, such that $\sum_{i=1}^{n} h_i(x) = 1, \forall x \in K$. Hence, $f = \sum_{i=1}^{n} h_i f$. Note that

$$\mu(K) \leq \Lambda\left(\sum_{i=1}^{n} h_i\right) = \sum_{i=1}^{n} \Lambda h_i, \tag{2.95}$$

and $h_i f \leq (y_i + \epsilon)h_i$, thus

$$\begin{aligned}\Lambda f &\leq \sum_{i=1}^{n} \Lambda(h_i f) \leq \sum_{i=1}^{n} \Lambda((y_i + \epsilon)h_i) = \sum_{i=1}^{n}(y_i + \epsilon)\Lambda h_i \\ &= \sum_{i=1}^{n}(|a| + y_i + \epsilon)\Lambda h_i - |a| \sum_{i=1}^{n} \Lambda h_i\end{aligned} \tag{2.96}$$

Here, for each i, $\Lambda h_i \leq \mu(V_i) < \mu(E_i) + \epsilon/n$, as $h_i \prec V_i$, and $\sum_{i=1}^{n} \Lambda h_i \geq mu(K)$, we then have

$$\Lambda f \leq \sum_{i=1}^{n}(|a| + y_i + \epsilon)(\mu(E_i) + \epsilon/n) - |a|\mu(K). \tag{2.97}$$

Note $\mu(K) = \sum_{i=1}^{n} \mu(E_i)$, we further have

$$\Lambda f \leq \sum_{i=1}^{n}(y_i - \epsilon)\mu(E_i) + 2\epsilon\mu(K) + \frac{\epsilon}{n} \sum_{i=1}^{n}(|a| + y_i + \epsilon). \tag{2.98}$$

Here, $\frac{1}{n}\sum_{i=1}^{n} y_i \le |b|$, and $\sum_{i=1}^{n}(y_i - \epsilon)\mu(E_i) \le \int_X f d\mu$ (by monotonicity of Lebesgue integral), thus

$$\Lambda f \le \int_X f d\mu + \epsilon(2\mu(K) + |a| + |b| + \epsilon). \tag{2.99}$$

As this holds for any $\epsilon > 0$, $\Lambda f \le \int_X f d\mu$. From this, we also have $\Lambda(-f) \le \int_X (-f) d\mu$, leading to the inequality in opposite direction, and resulting in the equality $\Lambda f = \int_X f d\mu$ for any real valued function f. It is not straightforward to extend this equality to general complex functions with compact supports, by respectively considering the real and imaginary parts. □

Up to here, the proof of existence is completed, in the following, we will prove the uniqueness. □

Proof of Uniqueness. Suppose $\mathcal{M}$ is a σ-algebra, and μ_1, μ_2 are two measures on $\mathcal{M}$ that satisfy the conditions $C.1$ to $C.5$. Since they both satisfy $C.4$, if they agree on compact sets, then they are identical, thus it suffices to prove $\mu_1 = \mu_2$ for each compact set.

For each compact set K, given $\epsilon > 0$, by $C.3$, there is an open set V with $\mu_i(V) \le \mu_i(K) + \epsilon$ for $i = 1, 2$. By Urysohn lemma, there is $f : X \to [0, 1] \in C_c(X)$ with $K \prec f \prec V$, then

$$\mu_l(K) = \int_X \chi_K d\mu_1 \le \int_X f d\mu_1 = \Lambda f = \int_X f d\mu_2 \le \int_X \chi_V d\mu_2 = \mu_2(V) \le \mu_2(K) + \epsilon \tag{2.100}$$

As this holds for arbitrary $\epsilon > 0$, $\mu_1(K) \le \mu_2(K)$, likewise, $\mu_2(K) \le \mu_1(K)$, thus $\mu_1(K) = \mu_2(K)$. □

2.4 Principles of Measure Theory

There are four important results in measure theory, which we call the *principles of measure theory*, whose proofs will be given in later lectures.

Let $(X, \mathcal{M}, \mu)$ be a measure space, then

1. ***Every measurable set is nearly Borel***. For each measurable set A, there exist F_σ, G_δ and N such that $A = F_\sigma \cup N = G_\delta \cup N$, where F_σ is a countable union of closed sets, G_δ is a countable intersection of open sets, and N is a null set.

2. ***Every measurable set is nearly open***. For each measurable set A, given $\epsilon > 0$, there is an open set U with $A \subseteq U$ such that $\mu(U) < \mu(A) + \epsilon$.

3. ***Every measurable function is nearly continuous***. This is Lusin's theorem.

4. ***Every pointwise convergent sequence of measurable function is nearly uniform convergent***. This is Egorov's theorem.

Chapter 3

Introduction to L^p Spaces

3.1 Important Inequalities

3.1.1 Convex Sets and Convex Functions

Definition 3.1 (Convex set). *Let C be a subset of a complex vector space X, then C is called a **convex set** if $x, y \in C \Rightarrow (1-\lambda)x + \lambda y \in C,\ \forall \lambda \in [0,1]$.*

Definition 3.2 (Convex function). *Let $f : X \to \mathbb{R}$ be a real valued function defined on a complex vector space X, then f is called a **convex function** if $f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y),\ \forall x, y \in X,\ \lambda \in [0,1]$. In particular, if the equality holds iff $x = y$, then f is said to be **strictly convex**.*

Some important convex functions defined on $\mathbb{R}$ include $(-\log)$, $\exp$, and $x \mapsto |x|^\alpha$ with $\alpha \geq 1$.

Proposition 3.1. *Each seminorm is convex. In particular, each norm is convex.*

Proof. Let X be a vector space with a seminorm p defined on it. Given $x, y \in X$, then when $\lambda \in [0,1]$, $p((1-\lambda)x + \lambda y) \leq (1-\lambda)p(x) + \lambda p(y)$, which directly follows from the definition of seminorm. □

3.1.2 Jensen's Inequality

In the following, we introduce several important inequalities.

Theorem 3.1 (Jensen's Inequality). *Let $(X, \mathcal{M}, \mu)$ be a probability space (a measure space with $\mu(X) = 1$), $f : X \to \mathbb{R} \in L^1(X, \mu)$, and $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function, then*

$$\varphi\left(\int_X f d\mu\right) \leq \int_X (\varphi \circ f) d\mu.$$

Proof. Since φ is convex, at each $x_0 \in \mathbb{R}$, there exist $a, b \in \mathbb{R}$ such that $\varphi(x_0) = ax_0 + b$ and $\varphi(x) \geq ax + b$, $\forall x \in \mathbb{R}$, (here, $y = ax + b$ defines a supporting plane of the epigraph of φ at x_0). Let $x_0 = \int_X f d\mu$, then we have

$$\varphi\left(\int_X f d\mu\right) = \varphi(x_0) = ax_0 + b = a\int_X f d\mu + b = \int (af + b) d\mu \leq \int (\phi \circ f) d\mu. \tag{3.1}$$

The proof is completed. □

By considering measure on finite sets, we immediately derive the following finite form of the inequality.

Corollary 3.1 (Jensen's Inequality in finite form). *Let $g : X \to \mathbb{R}$ be a convex function on vector space, $x_1, \ldots, x_n \in X$, and $\alpha_1, \ldots, \alpha_n \in [0,1]$ with $\sum_{i=1}^n \alpha_i = 1$, then*

$$g\left(\sum_{i=1}^n \alpha_i x_i\right) \leq \sum_{i=1}^n \alpha_i g(x_i).$$

3.1.3 Hölder's Inequality

Definition 3.3 (Hölder's conjugate pair)**.** *Let $p, q \in [1, +\infty]$, they are said to be **conjugate** if $1/p + 1/q = 1$. In particular, $p = 2$ is self-conjugate, and $(1, +\infty)$ is a conjugate-pair.*

Definition 3.4 (Essential supremum)**.** *Let $f : X \to \mathbb{R}$ be a real valued measurable function defined on a measure space $(X, \mathcal{M}, \mu)$, then the **essential supremum** of f on a measurable subset S of X is*

$$\operatorname*{ess\,sup}_S f = \inf\{a \in \mathbb{R} : \mu(\{x : f(x) > a\}) = 0\}.$$

In particular, when $S = X$, we denote $\operatorname{ess\,sup}_X f$ by $\operatorname{ess\,sup} f$, which is called the essential supremum of f.

Definition 3.5 ($||\cdot||_p$ function)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space, let $p \in [0, +\infty]$, we define $||\cdot||_p : X \to \mathbb{R}$ as follows: when $p < +\infty$,*

$$||f||_p = \left(\int_X |f|^p d\mu\right)^{1/p}$$

and when $p = +\infty$,

$$||f||_\infty = \operatorname*{ess\,sup}_X |f|.$$

Theorem 3.2 (Hölder's Inequality)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space, $p, q \in [1, +\infty]$ be Hölder conjugate (i.e. $1/p + 1/q = 1$), and $f, g : X \to \mathbb{C}$ be measurable functions,*

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

In particular, if $p, q < +\infty$, it can be written as

$$\int_X |fg| d\mu \leq \left(\int_X |f|^p d\mu\right)^{1/p} \left(\int_X |g|^q d\mu\right)^{1/q}.$$

Proof. We first consider the case where $p, q < +\infty$. Let $a = ||f||_p$ and $b = ||g||_q$, if either a or b is infinite, then we are done. Hence, it suffices to assume that $a, b < +\infty$. In addition, without losing generality, we assume that $fg \neq 0,\ a.e.[\mu]$. (otherwise, we can just consider the subset within which this is satisfied). Let $f' = f/a$ and $g' = g/a$, then

$$\int_X |f'|^p d\mu = 1 \text{ and } \int_X |g'|^p d\mu = 1 \tag{3.2}$$

Note that $0 < |f'(x)| < +\infty,\ a.e.[\mu]$ and so is g. For each x satisfying this condition, let $s(x) = p \log |f'(x)|$ and $t(x) = q \log |g'(x)|$, which are defined almost everywhere. Then

$$|f'(x)| = \exp\left(\frac{s(x)}{p}\right)\ a.e.[\mu] \quad \text{and} \quad |g'(x)| = \exp\left(\frac{t(x)}{q}\right),\ a.e.[\mu] \tag{3.3}$$

Note that exp is convex, and $1/p + 1/q = 1$, we have

$$|f'(x)g'(x)| = \exp\left(\frac{s(x)}{p} + \frac{t(x)}{q}\right) \leq \frac{e^{s(x)}}{p} + \frac{e^{t(x)}}{q} = \frac{1}{p}|f'(x)|^p + \frac{1}{q}|g'(x)|^q. \tag{3.4}$$

Taking integration of both sides, we have

$$\int_X |f'g'| d\mu \leq \frac{1}{p}\int_X |f'| d\mu + \frac{1}{q}\int_X |g'| d\mu = \frac{1}{p} + 1q = 1. \tag{3.5}$$

As a result,

$$||fg||_1 = \int_X |fg| d\mu \leq ab = ||f||_p ||g||_q. \tag{3.6}$$

When $p = 1$ and $q = \infty$. Let $a = ||f||_1$ and $b = ||g||_q$, then $|g(x)| \leq q, a.e.[\mu]$. Assume that $b < +\infty$, otherwise we are done. As a result,

$$||fg||_1 = \int_X |fg| d\mu \leq \int_X b|f| d\mu = b\int_X |f| d\mu = ||g||_\infty ||f||_1. \tag{3.7}$$

The proof is completed. □

3.1.4 Minkowski's Inequality

Theorem 3.3 (Minkowski's Inequality). *Let $(X, \mathcal{M}, \mu)$ be a measure space, $f, g : X \to \mathbb{C}$ be measurable functions, and $p \in [1, +\infty]$, then*

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

In particular, when $p < +\infty$, it can be written as

$$\left(\int_X |f + g|^p d\mu\right)^{1/p} = \left(\int_X |f|^p d\mu\right)^{1/p} + \left(\int_X |g|^p d\mu\right)^{1/p}.$$

Proof. It suffices to assume that $||f||_p < +\infty$ and $||g||_p < +\infty$ and $||f + g||_p > 0$, otherwise we are done. When $p < +\infty$, we have

$$||f + g||_p^p = \int_X |f + g|^p d\mu \leq \int_X (|f| + |g|)|f + g|^{p-1} d\mu = \int_X |f||f + g|^{p-1} d\mu + \int_X |g||f + g|^{p-1} d\mu. \tag{3.8}$$

With Hölder's inequality, let $q = p/(p-1)$, we have

$$\begin{aligned} \int_X |f||f + g|^{p-1} d\mu &\leq \left(\int_X |f|^p\right)^{1/p} \left(\int_X |f + g|^{(p-1)/q}\right)^{1/q} = \left(\int_X |f|^p\right)^{1/p} \left(\int_X |f + g|^p\right)^{1-1/p} \\ &= ||f||_p ||f + g||_p^{p-1}. \end{aligned} \tag{3.9}$$

Likewise, we have

$$\int_X |g||f + g|^{p-1} d\mu \leq ||g||_p ||f + g||_p^{p-1}. \tag{3.10}$$

Hence,

$$||f + g||_p^p \leq (||f||_p + ||g||_p)||f + g||_p^{p-1}, \tag{3.11}$$

thus

$$||f + g||_p \leq ||f||_p + ||g||_p. \tag{3.12}$$

When $p = +\infty$, let $a = ||f||_\infty = \operatorname{ess\,sup} |f|$, and $b = ||g||_\infty = \operatorname{ess\,sup} |g|$, then given $\epsilon > 0$, $\mu\{x||f(x)| > a+\epsilon/2\} = 0$, and $\mu\{x||g(x)| > b+\epsilon/2\} = 0$. Since $|f(x)+g(x)| > a+b+\epsilon \Rightarrow |f(x)| > a+\epsilon/2$, or $|g(x)| > b+\epsilon/2$,

$$\mu\{x||f(x) + g(x)| > a + b + \epsilon\} \leq \mu\{x||f(x)| > a + \epsilon/2\}\mu\{x||g(x)| > b + \epsilon/2\} = 0. \tag{3.13}$$

Hence, $a + b + \epsilon$ is an essential upper bound of $|f + g|$. This holds for all $\epsilon > 0$, thus

$$||f + g||_\infty = \operatorname{ess\,sup} |f + g| \leq a + b = ||f||_\infty + ||g||_\infty. \tag{3.14}$$

□

3.2 L^p Spaces

3.2.1 From $\mathcal{L}^p$ space to L^p space

Definition 3.6 ($\mathcal{L}^p$ space). *Let $(X, \mathcal{M}, \mu)$ be a measure space, and*

$$\mathcal{L}^p(X, \mathcal{M}, \mu) = \{f : X \to \mathbb{C} \mid f \text{ is measurable, and } ||f||_p < +\infty\},$$

*by linearity of Lebesgue integration and Minkowski's inequality, we can see that when $p \in [1, +\infty]$, $\mathcal{L}^p(X, \mathcal{M}, \mu)$ is a vector space with seminorm $||\cdot||_p$, called the $\mathcal{L}^p$ **space**. When the measure space is clear from the context, the following simplified notations are often used: $\mathcal{L}^p(X, \mu)$, $\mathcal{L}^p(X)$, $\mathcal{L}^p(\mu)$, or $\mathcal{L}^p$.*

By Hölder's inequality, we immediately have

Proposition 3.2. *Let $p, q \in [1, +\infty]$ be Hölder conjugate (i.e. $1/p + 1/q = 1$), if $f \in \mathcal{L}^p$ and $g \in \mathcal{L}^q$, then $fg \in \mathcal{L}^1$.*

Proposition 3.3. *Let $f, g : X \to \mathbb{C}$ are both measurable functions defined on a measure space $(X, \mathcal{M}, \mu)$, then $||f - g||_p = 0$ if and only if $f = g$ a.e.$[\mu]$.*

Proof. 1. $(f = g \ a.e.[\mu] \Rightarrow ||f - g||_p = 0)$.

This direction is proved as follows: when $p < +\infty$,

$$f = g \ a.e.[\mu] \Rightarrow |f - g|^p = 0 \ a.e.[\mu] \Rightarrow ||f - g||_p^p = \int_X |f - g|^p d\mu = 0. \tag{3.15}$$

When $p = +\infty$, $f = g \ a.e.[\mu] \Rightarrow \operatorname{ess\,sup} |f - g| = 0$.

2. $(||f - g||_p = 0 \Rightarrow f = g \ a.e.[\mu])$.

When $p < +\infty$, let $S_\alpha = \{x||f(x) - g(x)|^p > \alpha\}$, it is easy to see that $||f - g||_p^p \geq \alpha\mu(S_\alpha)$, thus $\mu(S_\alpha) = 0$. Let $S = \{x|f(x) \neq g(x)\}$. We note that $S = \bigcup_{n=1}^\infty S_{1/n}$, thus $\mu(S) \leq \sum_{n=1}^\infty \mu(S_{1/n}) = 0$, which follows that $f = g \ a.e.[\mu]$.

When $p = +\infty$, $\operatorname{ess\,sup} |f - g| = 0 \Rightarrow |f - g| = 0 \ a.e.[\mu] \Rightarrow f = g \ a.e.[\mu]$.

□

According to quotient space theorem, we can derive a normed space from $(\mathcal{L}^p, || \cdot ||_p)$ by merging the equivalent functions.

Definition 3.7 (L^p space)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space, $p \in [1, +\infty]$, and $\sim$ be a relation in $\mathcal{L}^p(X, \mathcal{M}, \mu)$ with $f \sim g \Leftrightarrow ||f - g||_p = 0 \Leftrightarrow f = g$ a.e.$[\mu]$, then $L^p(X, \mathcal{M}, \mu) = \{[f] \mid f \in \mathcal{L}^p(X, \mathcal{M}, \mu)\}$ together with the norm $|| \cdot ||_p$ given by $||[f]||_p = ||f||_p$ constitute a normed space, called the L^p **space**, denoted by $L^p(X, \mathcal{M}, \mu)$. The simplified notation $L^p(X, \mu)$, $L^p(X)$, $L^p(\mu)$, and L^p are often used when the measure is clear in context.*

For L^p spaces, we have the following remarks

1. The elements in L^p are equivalence classes, but not functions.
2. It is meaningless to evaluate point-values of elements in L^p, given any $x \in X$, as different functions in the same equivalence class may yield different values at that point.
3. We can take Lebesgue integration of the elements of L^p, as different functions in the same equivalence class yield the same integral value.

3.2.2 L^p Convergence and Completeness

Definition 3.8 (L^p convergence)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space, and $(f_n)_{n=1}^\infty$ be a sequence of functions in $L^p(X, \mathcal{M}, \mu)$, then we say (f_n) **converges to f in L^p-norm** when $\lim_{n\to\infty} ||f_n - f|| = 0$, denoted by $f_n \xrightarrow{L_p} f$.*

Proposition 3.4. *L^∞ convergence is equivalent to uniform convergence on X except for a null set.*

We note that *L_p convergence is neither a sufficient condition nor a necessary condition of almost everywhere convergence.* We give two examples to show this (under Lebesgue measure space of $\mathbb{R}$)c.

1. Let $f_n : \mathbb{R} \to \mathbb{R}$ be defined by $f_n = \chi_{[n,n+1]}$, then for every $p \in [1, +\infty]$, $f_n \in L^p(\mathbb{R}, \mathcal{B}, m)$, and $f_n \to 0$ everywhere, but $||f_n||_p = 1$ for each n.
2. Consider the following sequence of functions: $f_{1,0}, f_{1,1}, f_{2,0}, \ldots, f_{2,3}, \ldots, f_{k,0}, f_{k,2^k-1}, \ldots$ with $f_{k,n} = \chi_{[n/2^k,(n+1)/2^k]}$, then their Lebesgue integration converges to zero, but they converge nowhere.

Definition 3.9 (σ-finite measure)**.** *A measure μ defined on a measurable space $(X, \mathcal{M})$ is called a **σ-finite measure**, if X is a countable union of measurable sets with finite measure. In this case, $(X, \mathcal{M}, \mu)$ is called a **σ-finite measure space**.*

Lemma 3.1. *Every Cauchy sequence $(f_n)_{n=1}^\infty$ in $\mathcal{L}^p$ has a subsequence $(f_{n_k})_{k=1}^\infty$ that converges to a function $f \in \mathcal{L}^p$ almost everywhere.*

Proof. First, we consider the case with $p < +\infty$. Since $(f_n)_{n=1}^{\infty}$ is $\mathcal{L}^p$-Cauchy, we can choose $(n_k)_{k=1}^{\infty}$ such that $||f_{n_k} - f_{n_{k+1}}||_p < 2^{-k}$, we claim this sequence converges to a function $f \in \mathcal{L}^p$ almost everywhere. Let $g_m = |f_{n_1}| + \sum_{k=2}^{m} |f_{n_k} - f_{n_{k-1}}|$, then for each m, $g_m \in \mathcal{L}^p$ as it is a finite sum of $\mathcal{L}^p$ functions. By Fatou's lemma

$$|| \lim_{m\to\infty} g_m||_p^p = \int_X \left(\lim_{m\to\infty} g_m\right)^p d\mu = \int_X \lim_{m\to\infty} g_m^p d\mu \leq \lim_{m\to\infty} \int_X g_m^p d\mu = \lim_{m\to\infty} ||g_m||_p^p. \tag{3.16}$$

Hence,

$$|| \lim_{m\to\infty} g_m||_p \leq \lim_{m\to\infty} ||g_m||_p \leq ||f_{n_1}||_p + \sum_{k=2}^{m} ||f_{n_k} - f_{n_{k-1}}||_p \leq ||f_{n_1}||_p + 1. \tag{3.17}$$

It implies that (g_m) is a non-decreasing sequence of non-negative functons that is bounded above almost everywhere. Let $g = \lim_{m\to\infty} g_m$, then g is defined almost everywhere, and $g \in \mathcal{L}^p$. Define $h_m(x) = f_{n_1}(x) + \sum_{k=2}^{m}(f_{n_k}(x) - f_{n_{k-1}}(x))$, then $(h_m(x))_{m=1}^{\infty}$ is Cauchy almost everywhere, by completeness of $\mathbb{C}$, it converges almost everywhere, let $f(x) = \lim_{m\to\infty} h_m(x)$, then f is defined almost everywhere. Note that $h_m = f_{n_m}$, thus $\lim_{k\to\infty} f_{n_k} = f,\ a.e.[\mu]$.

In addition,

$$||f||_p^p = \int_X |f|^p d\mu = \int_X \lim_{k\to\infty} |f_{n_k}|^p d\mu \leq \liminf_{k\to\infty} \int_X |f_{n_k}|^p d\mu = \lim_{k\to\infty} ||f_{n_k}||_p^p \leq g \in \mathcal{L}^p. \tag{3.18}$$

Thus $f \in \mathcal{L}^p$.

When $p = +\infty$, L^∞-Cauchy implies uniform Cauchy on X except for a null set. Thus, $f_n(x)$ is Cauchy sequence almost everywhere, and thus f_n converges almost everywhere. In addition, uniform Cauchy implies uniform boundedness, thus $f(x) = \lim_{n\to\infty} f_n(x)$ has $||f||_\infty < +\infty$. ☐

Theorem 3.4 (Riesz-Fisher Theorem). *For $p \in [1, +\infty]$, L^p spaces are complete.*

Proof. For each Cauchy sequence $(f_n)_{n=1}^{\infty}$ in $\mathcal{L}^p$, we choose a subsequence $(f_{n_k})_{k=1}^{\infty}$ with $||f_{n_k} - f_{n_{k-1}}|| < 2^{-k}$. From the proof of the above lemma, we know that f_{n_k} converges to some $f \in \mathcal{L}^p$ almost everywhere. It remains to show that $f_n \xrightarrow{L^p} f$. To this end, it suffices to show that $f_{n_k} \xrightarrow{L^p} f$ (Cauchy sequence with convergent subsequence converges),

$$||f - f_{n_m}||_p^p = \int_X |f - f_{n_m}|^p d\mu = \int_X \lim_{k\to\infty} |f_{n_k} - f_{n_m}|^p d\mu \leq \liminf_{k\to\infty} \int_X |f_{n_k} - f_{n_m}|^p d\mu = \lim_{k\to\infty} ||f_{n_k} - f_{n_m}||_p^p \tag{3.19}$$

which converges to 0 as $m \to \infty$. ☐

3.3 Important Theorems

3.3.1 Lusin's Theorem

Lusin's theorem states that *every measurable function is nearly continuous*, or more exactly, *every measurable function with finite measure support is continuous on nearly all its domain.*

Theorem 3.5 (Lusin's theorem). *Let μ be a regular Borel measure on a locally compact Hausdorff space X such that $\mu(K) < +\infty$ for each compact subset K. Suppose $f : X \to \mathbb{C}$ is a complex measurable function with finite measure support (i.e. $\mu(\mathrm{supp} f) < +\infty$), then for each $\epsilon > 0$, there exists a function with compact support $g \in C_C(X)$ such that $\mu(\{x|f(x) \neq g(x)\}) < \epsilon$.*

Proof. We start from a simple case where f is a non-negative function bounded above by 1 and with compact support, and then generalize the conclusion to a wider family of functions.

The basic idea is to decompose f into a series of attenuating differences, approximate each of term by a continuous function using Urysohn lemma, and construct a uniformly convergent series with them.

Assume that $f : X \to \mathbb{R}$ be a measurable function with compact support $A = \mathrm{supp}\ f$ (note $\mu(A) < +\infty$), and $f(X) \subseteq [0, 1]$. Then, we can construct an increasing sequence of non-negative simple functions $(s_n)_{n=1}^{\infty}$, such that $s_n \uparrow f$, and $s_n - s_{n-1} = 2^{-n}\chi_{T_n}$ where T_n is measurable for each n.

Let $t_1 = s_1$, $t_n = s_n - s_{n-1}$ for $n > 1$, then we have $f = \sum_{n=1}^{\infty} t_n$. Since X is locally compact Hausdorff, and the support A is compact, there exists an open set $U \supset A$ such that $\bar{U}$ is compact. By regularity of μ, for each n there is a compact set K_n, and an open set V_n such that $K_n \subset A \subset V_n \subset U$ and $\mu(V_n \backslash K_n) < 2^{-n}\epsilon$. By Urysohn lemma, we can choose $h_n \in C_c(X)$ such that $K_n \prec h_n \prec V_n$.

Then, we define $g_m = \sum_{n=1}^{m} 2^{-n} h_n$. For g_m we have the following two claims:

1. g_m converges uniformly on X. Given $\xi > 0$, choose an integer N with $N > \log_2 \epsilon+1$, then $\sum_{n=N+1}^{\infty} 2^{-n} h_n \leq \sum_{n=N+1}^{\infty} 2^{-n} < \xi$. Hence, g_m converges uniformly. Let $g = \lim_{m\to\infty} g_m = \sum_{n=1}^{\infty} 2^{-n} h_n$. By uniform convergence theorem, g is a continuous function.

2. g has a compact support, since supp $g \subset \bar{V}$, which is compact. Hence, $g \in C_c(X)$.

3. Note that $2^{-n} h_n = t_n$ on K_n. Hence, $f(x) = g(x)$ except on $S = \bigcup_{n=1}^{\infty}(V_n \backslash K_n)$, which has
$$\mu(S) \leq \sum_{n=1}^{\infty} \mu(V_n \backslash K_n) < \sum_{n=1}^{\infty} 2^{-n}\epsilon = \epsilon. \tag{3.20}$$

From these claims we can conclude that the theorem holds for any non-negative functions bounded above by 1 and has compact support.

We then generalize this conclusion through several steps.

1. Let f be a bounded non-negative function with compact support with an upper bound M. Then f/M is bounded above by 1, which can be approximated by $g \in C_c(X)$, then f can be approximated by Mg.

2. Let f be a bounded complex function with compact support, we can write $f = (u^+ - u^-) + i(v^+ - v^-)$ where u^+, u^-, v^+, v^- are all bounded non-negative functions with compact support. Thus, we can approximate them respectively by $g_u^+, g_u^-, g_v^+, g_v^-$, and hence f can be approximated by $g = (g_u^+ - g_u^-) + i(g_v^+ - g_v^-)$.

3. Let f be a bounded complex function with a finite measure support A which is not necessarily compact. Then, by regularity of measure, we can find a compact set $K \subseteq A$ such that $\mu(A\backslash K) < \epsilon/2$. In addition, we can approximate the restrction $f|_K$ which is a bounded function with compact support by $g \in C_c(X)$ such that g differs from $f|_K$ in a set S with $\mu(S) < \epsilon/2$. Hence, g differs from f in the set $S \cup (A\backslash K)$ whose measure is less than ϵ.

4. If f has a finite measure support, but is not necessarily bounded. Let $B_n = \{x||f(x)| > n\}$, Then (B_n) is a non-increasing sequence of sets, with $\bigcap_{n=1}^{\infty} B_n = \emptyset$. Since $\mu(B_1) < \mu(A) < +\infty$, thus by continuity of μ, $\lim_{n\to\infty} \mu(B_n) = 0$. We can choose N such that $\mu(B_N) < \epsilon/2$, then $f' = (1 - \chi_{B_N})f$ is bounded above by N and has a finite measure support, thus we can find g to approximate f' such that g differs from f' in a set S with $\mu(S) < \epsilon/2$. As a result, g differs from f in the set $S \cup B_N$ whose measure is less than ϵ.

The proof of the theorem is completed. □

Based on Lusin's theorem, we can derive the following important result.

Theorem 3.6. *Let X be a locally compact Hausdorff space, μ be a regular Borel measure on X with $\mu(K) < +\infty$ for every compact set K, then $C_c(X)$ is dense in $L^p(X, \mu)$, when $1 \leq p < +\infty$.c*

In other words, given any $f \in L^p(X, \mu)$, there exists a sequence of functions $(g_n)_{n=1}^{\infty}$ in $C_c(X)$ such that g_n converges to f in L^p-norm, i.e. $\lim_{n\to\infty} ||f - g_n||_p = 0$.

Proof. The proof is conducted in two stages. First, we approach f by a sequence of functions with finite measure support, and then approximate them with compactly supported functions by Lusin's theorem.

Let S be the set of all measurable functions with finite measure support, i.e.
$$S = \{s : X \to \mathbb{C} | s \text{ is measurable, and } \mu(\{x|s(x) \neq 0\}) < +\infty\}.$$

We first claim that S is dense in $L^p(X, \mu)$. This is shown as follows. For each $f \in L^p$ that is non-negative, we can choose an increasing sequence of simple measurable functions $(s_n)_{n=1}^{\infty}$ such that $s_n \uparrow f$, *a.e.*. Obviously, $s_n \in L^p$ for each n. An integrable simple function must have finite measure support, thus $s_n \in S$ for each n. In addition, $|f - s_n|^p < |f|^p$ and $|f|^p \in L^1$, by dominated convergence theorem, we have
$$\lim_{n\to\infty} ||f - s_n||_p^p = \lim_{n\to\infty} \int_X |f - s_n|^p d\mu = \int_X \lim_{n\to\infty} |f - s_n| d\mu = 0. \tag{3.21}$$

Since $x \mapsto x^p$ is a continuous function when $p \in [1, \infty)$, thus $\lim_{n\to\infty} ||f - s_n||_p = 0$.

While when f complex measurable function in L^p, we can approximate each component (positive/negative parts or the real/imaginary parts) respectively. To sum up, S is dense in L^p, and thus the claim is proved.

We then claim that $C_c(X)$ is dense in S. This is shown as follows. For every $s \in S$, by Lusin's lemma, we can find $g \in C_c(X)$ such that $|g| \le ||s||_\infty$, and g differs from s on a set with measure less than ϵ for any given $\epsilon > 0$. As a result,

$$||g - s||_p = \left(\int_X |g - s|^p d\mu \right)^{1/p} \le (2||s||_\infty^p \epsilon)^{1/p} = \epsilon^{1/p} (2||s||_\infty^p)^{1/p}. \tag{3.22}$$

It implies that for each s we can find $g \in C_c(X)$ such that $||g - s||_p$ can be arbitrarily small, which follows that $C_c(X)$ is dense in S.

Combining the two claims above, we can conclude that $C_c(X)$ is dense in S. (Let $A \supset B \supset C$ be in a metric space, such that B is dense in A, and C is dense in B, then C is dense in A). □

Remarks:

1. When $p = +\infty$, $C_c(X)$ is not dense in $L^\infty(X)$. Because convergence in L^∞-norm implies uniform convergence, we can immediately see that any non-continuous function cannot be approached by continuous functions in L^∞-norm (due to uniform convergence theorem).

2. The completion of $(C_c(X), ||\cdot||_\infty)$ is the space of all continuous functions on X such that for every $\epsilon > 0$, there exists a compact set $K \subset X$ with $|f(x)| < \epsilon$ on $X \backslash K$.

3.3.2 Egoroff's Theorem and Convergence in measure

The Egoroff's theorem states that *pointwise convergent sequence of measurable functions in a finite measure space is nearly uniformly convergent.*

Theorem 3.7 (Egoroff's theorem). *Let $(X, \mathcal{M}, \mu)$ be a measure space with $\mu(X) < +\infty$, and $(f_n)_{n=1}^\infty$ be a sequence of measurable functions that converges almost everywhere to f. Then, for every $\epsilon > 0$, there exists a subset $E \subset X$ with $\mu(X \backslash E) < \epsilon$ such that (f_n) converges uniformly on E.*

Proof. Let $Y \subseteq X$ be the set of all points at which $f_n(x)$ converges to $f(x)$, then $\mu(Y^c) = 0$, and let

$$S(n,k) = \{x | |f_i(x) - f_j(x)| < 1/k, \ \forall i,j > n\}.$$

It is easy to see that we have $S(n,k) \subseteq S(n+1,k)$ for each n, and thus $(S(n,k))$ is an increasing sequence for each given k. In addition, for each $x \in Y$, $\lim_{n\to\infty} f_n(x) = f(x)$, thus there exists an integer N such that $|f_i(x) - f_j(x)| < 1/k$ when $i,j > N$, therefore, $x \in S(N,k)$. This follows that $Y \subseteq \bigcup_{n=1}^\infty S(n,k)$ for each k. By continuity of the measure μ, we have

$$\mu(X) = \mu(Y) \le \lim_{n\to\infty} \mu(S(n,k)) \le \mu(X). \tag{3.23}$$

Then, for each k, we can choose n_k such that $|\mu(X) - \mu(S(n,k))| < 2^{-k}\epsilon$. Let $E = \bigcap_{k=1}^\infty S(n_k,k)$. Then, we have the following two claims

1. $\mu(X \backslash E) < \epsilon$, this can be seen by

$$\mu(X \backslash E) \le \sum_{k=1}^\infty \mu(X \backslash S(n_k,k)) < \epsilon. \tag{3.24}$$

2. (f_n) converges to f uniformly on E. For each $\delta > 0$, we can find $k > 1/\delta$, then for every $x \in S(n_k,k)$, we have $|f_i(x) - f_j(x)| < 1/k < \delta$ when $i,j > n_k$ by definition. And, note that $E \subseteq S(n_k,k)$. This claim is proved.

Combining the two claims above, we can see that E is the set that we desire. The proof is completed. □

It is important to note that $\mu(X) < +\infty$ is necessary. If μ is not a finite measure, then $f_n = \chi_{[n,n+1]}$ is a pointwise convergent sequence that is clearly not uniformly convergent.

Definition 3.10 (Convergence in measure)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space, $f, f_1, f_2, \dots$ be measurable functions on X, then we say f_n* ***converges to f in measure*** *if*

$$\forall \epsilon > 0,\ \lim_{n\to\infty} \mu(\{x||f_n(x) - f(x)| > \epsilon\}) = 0,$$

or equivalently,

$$\forall \epsilon > 0,\ \exists N \in \mathbb{N}^+ \text{ s.t. } \mu(\{x||f_n(x) - f(x)| > \epsilon\}) < \epsilon,\ \forall n > N.$$

As an important corollary of Egoroff's theorem, *in a finite measure space, almost everywhere convergence implies convergence in measure.*

Proposition 3.5. *Let $(X, \mathcal{M}, \mu)$ be a measure space with $\mu(X) < +\infty$, and $(f_n)_{n=1}^{\infty}$ be a sequence of measurable functions that converges almost everywhere to f, then f_n converges to f in measure.*

Proof. By Egoroff's theorem, we can find a subset E with $\mu(X\backslash E) < \epsilon$ for any given $\epsilon > 0$ such that f_n converges to f uniformly on E. Hence, we can find N, such that $|f_n(x) - f(x)| <= \epsilon$ on E when $n > N$. It follows that

$$\mu(\{x||f_n(x) - f(x)| > \epsilon\}) \le \mu(X\backslash E) < \epsilon. \tag{3.25}$$

The proposition is proved. ☐

Note that the converse is in general NOT true. However, we have the following "weak" converse.

Proposition 3.6. *Let $(X, \mathcal{M}, \mu)$ be a measure space with $\mu(X) < +\infty$, and $(f_n)_{n=1}^{\infty}$ be a sequence of measurable functions that converges to a measurable function f in measure. Then, there exists a subsequence $(f_{n_k})_{k=1}^{\infty}$ such that f_{n_k} converges to f almost everywhere.*

Proof. We choose n_k such that

$$\mu(\{x||f(x) - f_{n_k}(x)| > 2^{-k}\}) < 2^{-k}. \tag{3.26}$$

Let $E_k = \{x||f(x) - f_{n_k}(x)| > 2^{-k}\}$. Then if $x \notin \bigcup_{i=k}^{\infty} E_i$, then $|f_{n_i}(x) - f(x)| < 2^{-i}$ for every $i \ge k$. Thus $f_{n_i}(x) \to f(x)$ when $i \to \infty$. Let $A = \bigcap_{k=1}^{\infty}\bigcup_{i=k}^{\infty} E_i$, if $x \notin A$, then $x \notin \bigcup_{i=k}^{\infty} E_i$ for some k, then $f_{n_i}(x) \to f(x)$ when $i \to \infty$. It means that f_{n_i} converges to f on $X\backslash A$. And, by continuity of μ, we have

$$\mu(A) \le \lim_{k\to\infty} \mu\left(\bigcup_{i=k}^{\infty} E_i\right) = \lim_{k\to\infty}\sum_{i=k}^{\infty} 2^{-i} = \lim_{k\to\infty} 2^{-k+1} = 0. \tag{3.27}$$

Hence, (f_{n_k}) converges to f almost everywhere. ☐

Dominated convergence theorem holds when the condition of almost everywhere convergence is replaced by convergence in measure.

Theorem 3.8 (Dominated convergence theorem with convergence in measure condition)**.** *Let $(f_n)_{n=1}^{\infty}$ be a sequence of measurable functions defined on a σ-finite measure space $(X, \mathcal{M}, \mu)$ with $|f_n| < g$ for some $g \in L^1(\mu)$. If f_n converges to f in measure, then $f \in L^1(\mu)$ and*

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

Proof. Let $A \subset X$ be a measurable subset with $\mu(A) < +\infty$, then we claim that $\int_A f d\mu = \lim_{n\to\infty} \int_A f_n d\mu$. Since f_n converges to f in measure, it is easy to see that $f_n|_A$ converges to $f|_A$ in measure, therefore, any subsequence of $f_n|_A$ converges to $f|_A$ in measure. By the proposition above, we can choose a subsequenc for each subsequence of $f_n|_A$ such that the chosen subsequence converges to $f|_A$ almost everywhere, by the standard form of DCT, we can conclude that the integral of the chosen subsequence converges to $\int_A f d\mu$. From the above argument, we can see that every subsequence of the sequence $(\int_A f_n d\mu)$ in itself contains a subsequence that converges to $\int_A f d\mu$, thus $(\int_A f_n d\mu) \to \int_A f d\mu$ when $n \to \infty$. The claim is proved.

If $\mu(X) < +\infty$, then we are done. If not, since X is σ-finite, we can find a countable collection of disjoint finite measure sets $\{A_k\}_{k=1}^{\infty}$ such that $X = \bigsqcup_{k=1}^{\infty} A_k$. Then,

$$\int_X f d\mu = \sum_{k=1}^{\infty}\int_{A_k} f d\mu = \sum_{k=1}^{\infty}\lim_{n\to\infty}\int_{A_k} f_n d\mu = \lim_{n\to\infty}\sum_{k=1}^{\infty}\int_{A_k} f_n d\mu = \lim_{n\to\infty}\int_X f_n d\mu. \tag{3.28}$$

Since both f and each f_n are dominated by some $g \in L^1$, thus all series in above formula is absolutely convergent, and thus the operations are valid. ☐

Chapter 4

Product Measure

4.1 Product Measure Space

4.1.1 Product Measurable Space

Definition 4.1 (Product measurable space). *Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be two measurable spaces, and denote by $\mathcal{M} \times \mathcal{N}$ the σ-algebra generated by the subsets of form $A \times B$ with $A \in \mathcal{M}$ and $B \in \mathcal{N}$ (i.e. the smallest σ-algebra that contains the sets of form $A \times B$). Then we call $(X \times Y, \mathcal{M} \times \mathcal{N})$ be the* ***product measurable space****.*

Definition 4.2 (Sections of a set). *Let $E \subset X \times Y$, then $E_x = \{y|(x, y) \in E\} \subset Y$ is called a x-**section** of E; $E^y = \{x|(x, y) \in E\} \subset X$ is called a y-**section** of E.*

Proposition 4.1. *Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be two measurable spaces, and E be measurable in their product measurable space, (i.e. $E \in \mathcal{M} \times \mathcal{N}$), then $E_x \in \mathcal{N}\ \forall x \in X$ and $E^y \in \mathcal{N}\ \forall y \in Y$.*

Important strategy of proof: Before the proof, we first note an important strategy that will be repeatedly used in proving a series of propositions. Let $\mathcal{M}$ be a σ-algebra generated by a $\mathcal{S}$, (i.e. $\mathcal{M}$ is the smallest σ-algebra containing $\mathcal{S}$), then to prove some statement P holds for every measurable set in $\mathcal{M}$, it suffices to show that all sets for which P holds constitute a σ-algebra that contains $\mathcal{S}$. Typically, it comprises the following steps:

1. show that P holds for every set in $\mathcal{S}$;
2. define $\mathcal{C}$ to be the collection of all sets for which P holds
3. show that $\mathcal{S} \subset \mathcal{C}$, which is equivalent to showing that P holds for every set in $\mathcal{S}$;
4. show that $\mathcal{C}$ is a σ-algebra by verifying the three conditions;
5. finally, we can conclude that $\mathcal{M} \subset \mathcal{C}$, which means that P holds for every set in $\mathcal{M}$.

Proof. Let $\mathcal{C}$ be the collection of all sets that satisfy $E_x \in \mathcal{N}\ \forall x \in X$ and $E^y \in \mathcal{M}\ \forall y \in Y$. It suffices to prove that $\mathcal{C}$ is a σ-algebra that contains the sets of the form $A \times B$ with $A \in \mathcal{M}$ and $B \in \mathcal{N}$.

First, it is easy to show that $E = A \times B \in \mathcal{C}$ when $A \in \mathcal{M}$ and $B \in \mathcal{N}$. In this case, E_x is either $\emptyset$ or B, and E^y is either $\emptyset$ or A, which are all measurable sets.

Then, we show that $\mathcal{C}$ defined above is a σ-algebra.

1. $\emptyset = \emptyset \times \emptyset \in \mathcal{C}$.
2. Let $E \in \mathcal{C}$. For E^c, we have $(E^c)_x = (E_x)^c \in \mathcal{N}$ and $(E^c)^y = (E^y)^c \in \mathcal{M}$, thus $E^c \in \mathcal{C}$.
3. Let $E = \bigcup_{n=1}^{\infty} E_n$ with $E_n \in \mathcal{C}$, for each n. Then, $E_x = \bigcup_{n=1}^{\infty} (E_n)_x \in \mathcal{N}$, and $E^y = \bigcup_{n=1}^{\infty} (E_n)^y \in \mathcal{M}$, thus $E \in \mathcal{C}$.

Hence, we can conclude that $\mathcal{C}$ is a σ-algebra that contains $\mathcal{M}$. $\square$

Definition 4.3 (Sections of a function). *Given a function $f : X \times Y \to Z$, then $f_x : Y \to Z$ defined by $f_x(y) = f(x, y)$ is called a x-**section** of f; $f^y : X \to Z$ defined by $f^y(x) = f(x, y)$ is called a y-**section** of f.*

Proposition 4.2. *Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be two measurable spaces, and Z be a topological space, $f : (X \times Y, \mathcal{M} \times \mathcal{N}) \to Z$ be a measurable function, then $f_x : Y \to Z$ is measurable on $(Y, \mathcal{N})$, and $f^y : X \to Z$ is measurable on $(X, \mathcal{M})$.*

Proof. Let A be a measurable set in Z, and $E = f^{-1}(A)$ then for any $x \in X$, $(f_x)^{-1}(Z) = \{y|f(x, y) \in A\} = E_x$. since f is measurable, thus $E \in \mathcal{M} \times \mathcal{N}$, and thus $E_x \in \mathcal{N}$. Therefore, f_x is measurable. Likewise, we can show the measurability of f^y. □

4.1.2 Measure by Extension: Hahn-Kolmogorov theorem

Before deriving the concept of product measure, we first introduce the following important theorem about extending a function to a measure.

Theorem 4.1 (Hahn-Kolmogorov theorem). *Let X be a non-empty set, $\mathcal{A}$ be an algebra of subsets of X (i.e. closed under set difference, finite union and finite intersection), then any countable additive function $\mu_0 : \mathcal{A} \to [0, +\infty]$ extends to a measure μ on $\mathcal{M} = \sigma(\mathcal{A})$ (the σ-algebra generated by aset). If μ_0 is σ-finite, then the extension is unique.*

Proof. The proof proceeds by a series of claims.
Claim 1: Define $\mu^* : 2^X \to [0, +\infty]$ by

$$\mu^*(E) = \inf \left\{ \sum_{n=1}^{\infty} \mu_0(A_n) \middle| E \subseteq \bigcup_{n=1}^{\infty} A_n,\ A_n \in \mathcal{A}\ \forall n \in \mathbb{N}^+ \right\}.$$

then μ^* is an outer measure on X.

Proof of Claim 1.
1. We consider $\emptyset$ as being covered by an empty collection, thus $\mu^*(\emptyset) = 0$.
2. The monotonicity of μ^* follows from that μ^* is defined as an infimum.
3. For sub-additivity, let $E = \bigcup_{n=1}^{\infty} E_n$, then given $\epsilon > 0$, for each E_n, we can choose a cover $\{A_{ni}\}_{i=1}^{\infty}$ such that $\sum_{i=1}^{\infty} \mu_0(A_{ni}) < \mu^*(E) + 2^{-n}\epsilon$. Then all A_{ni} form a countable cover of E, with $\sum_{n,i} \mu_0(A_{ni}) < \sum_{n=1}^{\infty} \mu^*(E_n) + \epsilon$. Hence, $\mu^*(E) \leq \sum_{n=1}^{\infty} \mu^*(E_n)$.

Then, we can conclude that μ^* is an outer measure. □

Claim 2: μ^* extends μ_0, i.e. $\mu^*(A) = \mu_0(A)$ for every $A \in \mathcal{A}$.

Proof of Claim 2. Let $A \in \mathcal{A}$. Clearly, $A \subseteq A \cup \emptyset \cup \emptyset \cup \cdots$, which follows that $\mu^*(A) \leq \mu_0(A)$. For the other direction ($\mu_0(A) \leq \mu^*(A)$), it suffices to assume that $\mu^*(A) < +\infty$, otherwise we are done. given $\epsilon > 0$, choose $\{A_n\}_{n=1}^{\infty}$ that covers A and $\sum_{n=1}^{\infty} \mu_0(A_n) < \mu^*(A) + \epsilon$. $B_1 = A \cap A_1$ and $B_n = A \cap (A_n \backslash \bigcup_{i<n} A_i)$, then $\{B_n\}_{n=1}^{\infty}$ are in $\mathcal{A}$ and they form a partition of A. By countable addivitity and monotonicity of μ_0 in $\mathcal{A}$, we have

$$\mu_0(A) = \mu_0\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu_0(B_n) \leq \sum_{n=1}^{\infty} \mu_0(A_n) < \mu^*(A) + \epsilon. \tag{4.1}$$

As this holds for every $\epsilon > 0$, $\mu_0(A) \leq \mu^*(A)$. □

Claim 3: Let $\mathcal{M}$ be the collection of all sets satisfying the Carathéodory condition, i.e. $E \in \mathcal{M}$ iff $\mu^*(S) = \mu^*(S \cap E) + \mu^*(S \cap E^c)$, $\forall S \subset X$. Then, $\mathcal{M}$ is a σ-algebra.

Proof of Claim 3.
1. When $E = \emptyset$, for each S, $\mu^*(S \cap E) = \mu^*(\emptyset) = 0$, and $\mu^*(S \backslash E) = \mu^*(S)$. Hence, the condition trivially holds.
2. Let $E \in \mathcal{M}$, then $E^c \in \mathcal{M}$. This can be easily seen by noting the symmetry in Carathéodory condition.

3. Let $E_1, E_2 \in \mathcal{M}$, and $E = E_1 \cup E_2$. Then,
$$\begin{aligned}\mu^*(S\cap(E_1\cup E_2)) + \mu^*(S\cap(E_1^c\cap E_2^c)) &\leq \mu^*(S\cap E_1) + \mu^*(S\cap E_1^c\cap E_2) + \mu^*(S\cap E_1^c\cap E_2^c)\\ &= \mu^*(S\cap E_1) + \mu^*(S\cap E_1^c) = \mu^*(S). \end{aligned}\tag{4.2}$$
Combining this with sub-additivity of μ^*, we derive the Carathéodory condition. Thus $E_1 \cup E_2 \in \mathcal{M}$.
4. Note that $E_1 \cap E_2 = (E_1^c \cup E_2^c)^c$ and $E_1 \backslash E_2 = E_1 \cap E_2^c$, we immediately obtain that $\mathcal{M}$ is closed under finite union and intersection as well as set differences. In other words, $\mathcal{M}$ is an algebra of subsets of X.
5. It remains to show that $\mathcal{M}$ is closed under countable union. Let $E = \bigcup_{n=1}^\infty E_n$ with $E_n \in \mathcal{M}$ for each $n \in \mathbb{N}^+$. Due to sub-additivity of μ^*, it suffices to show that given any $\epsilon > 0$, $\mu^*(S\cap E) + \mu^*(S\backslash E) < \mu^*(S) + \epsilon$. We can assume here that $\mu^*(S) < +\infty$, otherwise we are done.

   Let $B_n = E_n \backslash \bigcup_{i<n} E_i$, then $B_n \in \mathcal{M}$ for each n, since $\mathcal{M}$ is an algebra. And, $E = \bigsqcup_{n=1}^\infty B_n$. Let $G_n = \bigcup_{i=1}^n B_n$, then $\{G_n\}_{n=1}^\infty \in \mathcal{M}$, and hence Given $\epsilon > 0$ and $S \subset X$, by sub-additivity of μ^*, we have
$$\mu^*(S\cap E) \leq \sum_{n=1}^\infty \mu^*(S\cap B_n). \tag{4.3}$$
   Hence, we can find N such that
$$\mu^*(S\cap E) < \sum_{n=1}^N \mu^*(S\cap B_n) + \epsilon. \tag{4.4}$$
   By induction from Carathéodory condition, it is easy to show that
$$\mu^*(S\cap G_N) = \sum_{n=1}^N \mu^*(S\cap B_n). \tag{4.5}$$
   Note that G_N satisfies Carathéodory condition, hence,
$$\mu^*(S\cap E) + \mu^*(S\cap E^c) < \mu^*(S\cap G_N) + \mu^*(S\cap G_N^c) + \epsilon = \mu^*(S) + \epsilon. \tag{4.6}$$

Now, we can conclude that $\mathcal{M}$ is a σ-algebra. □

Claim 4: $\mathcal{A} \subset \mathcal{M}$, i.e. Carathéodory condition holds for every $A \in \mathcal{A}$.

Proof of Claim 4. It suffices to show that for each $A \in \mathcal{A}$, $S \subset X$, and $\epsilon > 0$, $\mu^*(E\cap A) + \mu^*(E\cap A^c) < \mu^*(E) + \epsilon$. We can assume that $\mu^*(E) < +\infty$, otherwise we are done. By definition of μ^*, we can choose $\{A_n\}_{n=1}^\infty \subset \mathcal{A}$ covering A such that $\sum_{n=1}^\infty \mu_0(A_n) < \mu^*(E) + \epsilon$. Then
$$\mu^*(E\cap A) + \mu^*(E\cap A^c) \leq \sum_{n=1}^\infty (\mu^*(A_n\cap A) + \mu^*(A_n\cap A^c)) = \sum_{n=1}^\infty \mu^*(A_n) < \mu^*(E) + \epsilon. \tag{4.7}$$
Hence, $A \in \mathcal{M}$. □

Claim 5: Define $\mu : \mathcal{M} \to [0, +\infty]$ by $\mu = \mu^*|_{\mathcal{M}}$, then μ is a measure on $\mathcal{M}$.

Proof of Claim 5. From claim 1, we directly know that $\mu(\emptyset) = 0$. It remains to show the σ-additivity. Let $\{E_n\}_{n=1}^\infty \subset \mathcal{M}$ be disjoint and $E = \bigsqcup_{n=1}^\infty E_n$, and $G_N = \bigsqcup_{n=1}^N E_n$. It suffices to show that $\mu^*(E) \geq \sum_{n=1}^\infty \mu^*(E_n)$. By induction from Carathéodory induction, it is easy to show that
$$\mu^*(E) \geq \mu^*(G_N) = \sum_{n=1}^N \mu^*(E_n) \tag{4.8}$$
Take the limit as $N \to \infty$, we get
$$\mu^*(E) \geq \sum_{n=1}^N \mu^*(E_n). \tag{4.9}$$
□

We can conclude the existence of μ from the claims above. In the following, we continue to show the uniqueness when μ_0 is σ-finite, which is summarized by the following claim.

Claim 6: Given any measure ν defined on $\mathcal{M}$ that agrees with μ_0 on $\mathcal{A}$, $\nu = \mu^*$.

Proof of Claim 6. First, we can see that $\nu \leq \mu^*$ which immediately follows from monotonicity of ν. Let $E \in \mathcal{M}$, it suffices to show that given $\epsilon > 0$, $\nu(E) > \mu^*(E) - \epsilon$. Assume that $\mu^*(E) < +\infty$, we will generalize it later. $\sum_{n=1}^{\infty} \mu_0(A_n) < \mu^*(E) - \epsilon/2$, and let $G_n = \bigcup_{i=1}^{n} A_n$, and $G = \bigcup_{i=1}^{\infty} A_n$. It has $G \supset E$. Since $G \in \mathcal{M}$, we have

$$\mu^*(G \backslash E) = \mu^*(G) - \mu^*(E) < \frac{\epsilon}{2}. \tag{4.10}$$

On the other hand, $\lim_{n\to\infty} \mu^*(G_n) = \mu(G) \geq \mu(E)$, hence, there exists N such that $\mu^*(G_N) > \mu^*(E) - \epsilon/2$. Since G_N can be expressed as finite union of disjoint sets in $\mathcal{A}$, we can derive that $\nu(G_N) = \mu^*(G_N)$, since they are both measures agreeing on $\mathcal{A}$. Consequently,

$$\nu(E) = \nu(G) - \nu(G\backslash E) \geq \nu(G_N) - \nu(G\backslash E) = \mu^*(G_N) - \nu(G\backslash E) > \mu^*(E) - \epsilon/2 - \epsilon/2 = \mu^*(E) - \epsilon. \tag{4.11}$$

We then consider the case where $\mu^*(E) = +\infty$, due to σ-finiteness, we can find a $\{A_n\}_{n=1}^{\infty}$ in $\mathcal{A}$ that cover E and $\mu^*(A_n) = \mu_0(A_n) < +\infty$. Let $G_n = \bigcup_{i=1}^{n} A_n$, then $G_n \in \mathcal{A}$, and $\mu^*(G_n) < +\infty$, and thus

$$\nu(E \cap G_n) = \mu^*(E \cap G_n). \tag{4.12}$$

Since $\mu^*(E) = +\infty$, given any $M > 0$, we can find an N, such that $\mu^*(E) > M$ and thus $\nu(E) > M$. It implies that $\nu(E)$ also equals ∞. $\square$

The proof of the entire theorem is completed. $\square$

4.1.3 Monotone Class

Definition 4.4 (Monotone class)**.** *A collection of sets $\mathcal{M}$ is called a* ***monotone class*** *if it is closed under monotonical limit,*

1. *let $E_1, E_2, \ldots \in \mathcal{M}$, if $E_1 \subset E_2 \subset \cdots$, then $\bigcup_{i=1}^{\infty} E_i \in \mathcal{M}$;*
2. *let $E_1, E_2, \ldots \in \mathcal{M}$, if $E_1 \supset E_2 \supset \cdots$, then $\bigcap_{i=1}^{\infty} E_i \in \mathcal{M}$.*

It is trivial to see that

Proposition 4.3. *Every σ-algebra is a monotone class.*

Proposition 4.4. *Arbitrary intersection of monotone classes is a monotone class.*

Then "the smallest monotone class" containing some collection $\mathcal{C}$ is defined to be the intersection of all monotone classes that contain $\mathcal{C}$. The following lemma is important in establishing the product measure.

Lemma 4.1. *Let $\mathcal{A}$ be an algebra of subsets of X, then the smallest monotone class containing $\mathcal{A}$ is precisely the σ-algebra generated by $\mathcal{A}$.*

Proof. Let $\mathcal{M}$ denote the smallest monotone class containing $\mathcal{A}$. Then, we are to show $\mathcal{M} = \sigma(\mathcal{A})$. Since $\sigma(\mathcal{A})$ is a monotone class, we have $\mathcal{M} \subseteq \sigma(\mathcal{A})$. For the other direction, it is enough to show that $\mathcal{M}$ is a σ-algebra.

1. $\emptyset \in \mathcal{A} \subset \mathcal{M}$,
2. Let $\mathcal{N} = \{E \in \mathcal{M} | E^c \in \mathcal{M}\}$. By definition $\mathcal{N} \subseteq \mathcal{M}$, and it is easy to verify that $\mathcal{N}$ is a monotone class containing $\mathcal{A}$, since $\mathcal{M}$ is the smallest $\mathcal{M} \subseteq \mathcal{N}$. Thus $\mathcal{M} = \mathcal{N}$, which follows that $\mathcal{M}$ is closed under set complement.
3. Let $F \in \mathcal{M}$, and $\mathcal{D}_F = \{F | F \cap E \in \mathcal{M}\}$. It is easy to verify that $\mathcal{D}_F$ is a monotone class, thus $\mathcal{M} \subset \mathcal{D}_F$ for each $F \in \mathcal{M}$. It implies that $\mathcal{M}$ is closed under finite intersection. Note that $E_1 \cup E_2 = (E_1^c \cap E_2^c)^c$, therefore, $\mathcal{M}$ is also closed under finite union. As a result, we can conclude that $\mathcal{M}$ is a algebra.

It is easy to see from definition that, being both a monotone class and an algebra of subsets, $\mathcal{M}$ is a σ-algebra. $\square$

4.1.4 Product Measure

Proposition 4.5. *Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be two measurable space, and define*

$$\mathcal{A} = \left\{ \bigsqcup_{i=1}^{n} A_i \times B_i \middle| A_i \in \mathcal{M},\ B_i \in \mathcal{N},\ \{A_i \times B_i\}_{i=1}^{\infty} \text{ are disjount} \right\}$$

Then, $\mathcal{A}$ is the smallest algebra of subsets that contains all sets of the form $A \times B$ with $A \in \mathcal{M}$ and $B \in \mathcal{N}$.

Proof. Denote by $\mathcal{A}^*$ the smallest algebra of subsets that contains $\{A \times B | A \in \mathcal{M},\ B \in \mathcal{N}\}$. We are to show that $\mathcal{A} = \mathcal{A}^*$. One direction $\mathcal{A} \subset \mathcal{A}^*$ is trivial, which is directly from the fact that $\mathcal{A}^*$ is closed under finite union. For the other direction, it suffices to show that $\mathcal{A}$ is in itself an algebra, i.e. it is closed under set complement, finite union and intersection. Verification of this is lengthy but not difficult, and the details are omitted here. □

Proposition 4.6. *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be two measure spaces, and $\mathcal{A}$ be the smallest algebra containing $\{A \times B | A \in \mathcal{M},\ B \in \mathcal{N}\}$, then for each $E \in \mathcal{A}$, we can write $E = \bigsqcup_{i=1}^{n} A_i \times B_i$ with $A_i \in \mathcal{M}$ and $B_i \in \mathcal{N}$ for each i. Define $\rho_0 : \mathcal{A} \to [0, +\infty]$ by $\rho_0(E) = \sum_{i=1}^{n} \mu(A_i)\nu(B_i)$, then ρ_0 is well defined and is countable additive on $\mathcal{A}$.*

Proof. Given $E = \bigsqcup_{i=1}^{n} A_i \times B_i$, then for each $x \in X$, $E_x = \bigsqcup_{i:x \in A_i} B_i$, which is clearly measurable in Y. Define $\phi : X \to [0, +\infty]$ by $\phi(x) = \nu(E_x)$, then $\phi(x) = \sum_{i=1}^{n} \nu(B_i)\chi_{A_i}$, which is clearly a measurable function. We then define $\lambda : \mathcal{A} \to [0, +\infty]$ by

$$\lambda(E) = \int_X \phi d\mu.$$

Then, we have

$$\lambda(E) = \int_X \phi d\mu = \sum_{i=1}^{n} \int_X \nu(B_i)\chi_{A_i} d\mu = \sum_{i=1}^{n} \mu(A_i)\nu(B_i) = \rho_0(E). \tag{4.13}$$

This shows that no matter how you partition E into disjoint union of $A_i \times B_i$, $\rho_0(E)$ will be evaluated to $\lambda(E)$, thus it is well defined.

By monotone convergence theorem, we can immediately see that λ satisfies countable additivity, so does ρ_0 since it is identical to λ. □

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be two σ-finite measure spaces, and $\mathcal{A}$ be the smallest algebra containing $\{A \times B | A \in \mathcal{M},\ B \in \mathcal{N}\}$. Let $\rho_0 : \mathcal{A} \to [0, +\infty]$ be defined by $\rho_0 \left(\bigsqcup_{i=1}^{n} A_i \times B_i \right) = \sum_{i=1}^{n} \mu(A_i)\nu(B_i)$, then ρ_0 is countably additive on $\mathcal{A}$. By Hahn-Kolmogorov theorem, ρ_0 uniquely extends to a measure on $\mathcal{M} \times \mathcal{N}$.

The following theorem gives an constructive description of the measure.

Theorem 4.2. *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be two σ-finite measure spaces, and given a measurable set $E \in \mathcal{M} \times \mathcal{N}$, let $\phi_E : X \to [0, +\infty]$ be defined by $\phi_E(x) = \nu(E_x)$ and $\psi_E : Y \to [0, +\infty]$ be defined by $\psi_E(y) = \mu(E^y)$, then ϕ_E and ψ_E are measurable, and*

$$\int_X \phi_E d\mu = \int_Y \psi_E d\nu.$$

In particular, when $E = \bigsqcup_{i=1}^{n} A_i \times B_i$ with $A_i \in \mathcal{M}$ and $B_i \in \mathcal{N}$, then the integral equals $\sum_{i=1}^{n} \mu(A_i)\nu(B_i)$.

Proof. Let $\mathcal{D}$ be all the subsets of $X \times Y$ for which the stated condition holds,

$$\mathcal{D} = \left\{ E \middle| \phi_E, \psi_E \text{ are measurable, and } \int_X \phi_E d\mu = \int_Y \psi_E d\nu \right\}.$$

It suffices to show that $\mathcal{D}$ contains an σ-algebra that contains all sets in form of $A \times B$ with $A \in \mathcal{M}$ and $B \in \mathcal{N}$.

Let $\mathcal{A}$ be the smallest algebra of subsets that contains every $A \times B$ with $A \in \mathcal{M}$ and $B \in \mathcal{N}$. Given $E \in \mathcal{A}$, we can write $E = \bigsqcup_{i=1}^{n} A_i \times B_i$ with $A_i \in \mathcal{M}$ and $B_i \in \mathcal{N}$ for each $i = 1, \ldots, n$. Then $\phi_E = \sum_{i=1}^{n} \nu(B_i)\chi_{A_i}$ and $\psi_E = \sum_{i=1}^{n} \mu(A_i)\chi_{B_i}$. It is easy to verify that $E \in \mathcal{D}$, i.e. the stated conditions holds for E.

Due to lemma 4.1, to complete the proof, it suffices to show that $\mathcal{D}$ is a monotone class. (If this is true, then $\mathcal{D}$ contains the smallest monotone class which is precisely $\mathcal{M} \times \mathcal{N}$.)

Let $\{E_n\}_{n=1}^{\infty} \subset \mathcal{D}$ with $E_1 \subset E_2 \subset \cdots$, and $E = \bigcup_{n=1}^{\infty} E_n$. Then, for each $x \in X$, $E_{n,x} \uparrow E_x$ and for each $y \in Y$, $E_n^y \uparrow E_y$. Due to monotonicity and continuity of measure, we have $\nu(E_{n,x}) \uparrow \nu(E_x)$, and $\mu(E_n^y) \uparrow \mu(E^y)$. Hence, ϕ_E and ψ_E are measurable. By monotone convergence theorem,

$$\int_X \phi_E d\mu = \lim_{n\to\infty} \int_X \phi_{E_n} d\mu = \lim_{n\to\infty} \int_Y \psi_{E_n} d\nu = \int_Y \psi_E d\nu. \tag{4.14}$$

Hence, the stated condition holds for E, and thus $E \in \mathcal{D}$. We can show that $E \in \mathcal{D}$ when $E_n \downarrow E$ in a similar way and using σ-finiteness. The above argument shows that $\mathcal{D}$ is a monotone class, thus it contains $\mathcal{M} \times \mathcal{N}$ by lemma 4.1. It follows that the stated condition holds for every measurable set in $\mathcal{M} \times \mathcal{N}$.

Note that the given integral is precisely the λ defined in the proof of proposition 4.6, hence it equals $\sum_{i=1}^{n} \mu(A_i)\nu(B_i)$ when $E = \bigsqcup_{i=1}^{n} A_i \times B_i$ with $A_i \in \mathcal{M}$ and $B_i \in \mathcal{N}$. □

Definition 4.5 (Product measure)**.** *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be two σ-finite measure spaces, and given a measurable set $E \in \mathcal{M} \times \mathcal{N}$, let $\phi_E : X \to [0, +\infty]$ be defined by $\phi_E(x) = \nu(E_x)$ and $\psi_E : Y \to [0, +\infty]$ be defined by $\psi_E(y) = \mu(E^y)$, then the* ***product measure*** *$\mu \times \nu : \mathcal{M} \to [0, +\infty]$ is defined by*

$$(\mu \times \nu)(E) = \int_X \nu(E_x) d\mu = \int_Y \mu(E^y) d\nu.$$

From theorem 4.2, we know that $\mu \times \nu$ is well defined on $\mathcal{M} \times \mathcal{N}$. While the σ-additivity follows from the properties of Lebesgue integration and monotone convergence theorem.

Definition 4.6 (Product measure space)**.** *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be two σ-finite measure spaces, then $(X \times Y, \mathcal{M} \times \mathcal{N}, \mu \times \nu)$ is called the* ***product measure space****.*

4.2 Fubini's theorem

Fubini's theorem is an important theorem related to Lebesgue integration on measure product spaces.

Theorem 4.3 (Fubini's theorem)**.** *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be two measure spaces, $f : X \times Y \to \mathbb{C}$ be $\mathcal{M} \times \mathcal{N}$-measurable function, then*

1. *If $f(x)$ is a non-negative function, i.e. $f(X) \subseteq [0, +\infty]$, and let*

$$\phi(x) = \int_Y f_x(y) d\nu \quad \text{and} \quad \psi(y) = \int_X f^y(x) d\mu$$

then $\phi : X \to [0, +\infty]$ and $\psi : Y \to [0, +\infty]$ are measurable, and

$$\int_{X\times Y} f d(\mu \times \nu) = \int_X \phi d\mu = \int_Y \psi d\nu. \tag{4.15}$$

This can be written as follows

$$\int_{X\times Y} f(x, y) d\mu(x) d\nu(y) = \int_X \left(\int_Y f(x, y) d\nu(y) \right) d\mu(x) = \int_Y \left(\int_X f(x, y) d\mu(x) \right) d\nu(y). \tag{4.16}$$

2. *If f is a complex function and*

$$\int_X \left(\int_Y |f(x, y)| d\nu(y) \right) d\mu(x) < +\infty \quad \text{or} \quad \int_Y \left(\int_X |f(x, y)| d\mu(x) \right) d\nu(y) < +\infty,$$

then $f \in L^1(\mu \times \nu)$.

3. *If $f \in L^1(\mu \times \nu)$, then $f_x \in L^1(Y, \nu)$ for almost every $x \in X$, and $f^y \in L^1(X, \mu)$ for almost every $y \in Y$, and $\phi \in L^1(\mu)$ and $\psi \in L^1(\nu)$, and Eq.(4.16) holds.*

Proof. Here just gives a very brief sketch. The product measure theorem has essentially established a restricted Fubini's theorem on characteristic functions of measurable sets. This can be readily extended to all non-negative simple functions. By monotone convegence theorem, we can further generalize the results to general non-negative measurable functions, and then to complex integrable functions. □

We note that σ-finiteness is required here, without which the theorem is not true in general. Here gives an example. Let $X = Y = [0, 1]$ and $(X, \mathcal{M}, \mu)$ is Lebesgue measure space, and $(Y, \mathcal{N}, \nu)$ be counting measure space. Define $f(x, y) = \delta_{x,y}$, which is the characteristic function of the measurable set $\{(x, y) \in X \times Y | x = y\}$. We can see that on one hand

$$\int_X \left(\int_Y f(x, y) d\nu(y) \right) d\mu(x) = \int_X 1 d\mu(x) = 1; \tag{4.17}$$

on the other hand,

$$\int_Y \left(\int_X f(x, y) d\mu(x) \right) d\nu(y) = \int_X 0 d\nu(y) = 0. \tag{4.18}$$

Exchanging the order of integration gives rise to different results.

Chapter 5

Complex Measures

5.1 The Space of Complex Measures

Definition 5.1 (Complex measure)**.** *Let $(X, \mathcal{M})$ be a measurable space, then $\mu : \mathcal{M} \to \mathbb{C}$ is called a **complex measure**, if it satisfies the following conditions*

1. $\mu(\emptyset) = 0$*;*
2. *if $E = \bigsqcup_{i=1}^{\infty} E_i$ with $E_i \in \mathcal{M}$ for each i and $E_i \cap E_j = \emptyset$ when $i \neq j$, then $\mu(E) = \sum_{i=1}^{\infty} \mu(E_i)$.*

To make the distinction clear, the measure defined previously whose range is in $[0, +\infty]$ is called **positive measure**. A complex measure whose range lies in $\mathbb{R}$ is called a **real measure** or a **signed measure**.

Definition 5.2 (Total variation)**.** *Let μ be a complex measure on a measurable space $(X, \mathcal{M})$, then its **total variation** $|\mu| : \mathcal{M} \to [0, +\infty]$ is defined by*

$$|\mu|(E) = \sup_{\mathcal{E}} \sum_{E \in \mathcal{E}} |\mu(E)|$$

Here, $\mathcal{E}$ is a countable partition of E comprised of measurable sets.

Proposition 5.1. *Let μ be a complex measure on a measurable space $(X, \mathcal{M})$, then its total variation $|\mu|$ is a positive measure on $(X, \mathcal{M})$.*

Proof. It is trivial to see that $|\mu|(\emptyset) = 0$, just by considering a collection of all emptysets as a partition of an emptyset. The important part is to show that $|\mu|$ satisfies σ-additivity on $\mathcal{M}$.

Let $\{A_n\}_{n=1}^{\infty} \subset \mathcal{M}$ be a collection of disjoint measurable sets and $A = \bigsqcup_{n=1}^{\infty} A_n$. By definition, given $\epsilon > 0$, for each A_n, we can find a countable partition $\{E_{ni}\}_{i=1}^{\infty}$ where $A_n = \bigsqcup_{i=1}^{\infty} E_{ni}$ and $E_{ni} \in \mathcal{M}$, $\forall n, i$, such that $\sum_{i=1}^{\infty} |\mu(E_{ni})| > |\mu|(A_n) - 2^{-n}\epsilon$, Gathering all E_{ni} together, we form a collection $\{E_{ni}, n \in \mathbb{N}^+, i \in \mathbb{N}^+\}$, which is obviously a countable partition of A. Hence, by definition of $|\mu|$, we have

$$|\mu|(A) \geq \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} |\mu(E_{ni})| = \sum_{n=1}^{\infty} (|\mu|(A_n) - 2^{-n}\epsilon) = \left(\sum_{n=1}^{\infty} |\mu|(A_n) \right) - \epsilon \tag{5.1}$$

As this holds for every $\epsilon > 0$,

$$|\mu|(A) \geq \sum_{n=1}^{\infty} |\mu|(A_n). \tag{5.2}$$

For the other direction, let $\{E_k\}_{k=1}^{\infty}$ be a countable partition of A, where each E_k is measurable. Hence, the collection $\{E_k \cap A_n\}_{k=1}^{\infty}$ forms a countable partition of A_n. By defnition, we have

$$\sum_{k=1}^{\infty} |\mu(E_k \cap A_n)| \leq |\mu|(A_n). \tag{5.3}$$

On the other hand, $\{E_k \cap A_n\}_{n=1}^{\infty}$ is a partition of E_k. Consequently, by additivity of complex measure,

$$\sum_{k=1}^{\infty} |\mu(E_k)| = \sum_{k=1}^{\infty} \left| \sum_{n=1}^{\infty} \mu(E_k \cap A_n) \right| \leq \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} |\mu(E_k \cap A_n)| = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} |\mu(E_k \cap A_n)| \leq \sum_{n=1}^{\infty} |\mu|(A_n). \tag{5.4}$$

Here, the exchange of the sum order is justified by the fact that $|\mu(E_k \cap A_n)|$ are non-negative. Since this holds for every partition $\{E_k\}$ of A, it holds for their supremium, as a result

$$|\mu|(A) \leq \sum_{n=1}^{\infty} |\mu|(A_n). \tag{5.5}$$

Hence, the σ-additivity of $|\mu|$ is established, and we can conclude that $|\mu|$ is a positive measure. □

Proposition 5.2. *Let $(X, \mathcal{M})$ be a measurable space, then*

1. *all complex measures on $(X, \mathcal{M})$ forms a vector space;*
2. *all real measures on $(X, \mathcal{M})$ forms a vector space;*
3. *if μ_1, μ_2 are positive measures, then $\mu_1 + \mu_2$ and $\alpha\mu_1$ with $\alpha \geq 0$ are positive measures.*

It is trivial to verify this proposition.

Theorem 5.1 (Hahn-Jordan decomposition)**.** *Let μ be a real measure on a measurable space $(X, \mathcal{M})$, then we can write $\mu = \mu^+ - \mu^-$ with*

$$\mu^+ = \frac{1}{2}(|\mu| + \mu)$$
$$\mu^- = \frac{1}{2}(|\mu| - \mu)$$

*Then both μ^+ and μ^- are positive measures on $(X, \mathcal{M})$. The decomposition given above is called the **Hahn-Jordan decomposition** of μ.*

Proof. Since $|\mu|$ and μ are real measures, thus μ^+ and μ^- are also real measures. The non-negativeness of μ^+ and μ^- readily follows from the definition of $|\mu|$. □

5.2 Lebesgue Decomposition of Measures

5.2.1 Absolute continuity and Singularity

Definition 5.3 (Absolute continuity)**.** *Let μ and λ respectively be positive and complex measures on a measurable space $(X, \mathcal{M})$. λ is said to be **absolutely continuous** w.r.t μ, denoted by $\lambda \ll \mu$ if $\forall E \in \mathcal{M},\ \mu(E) = 0 \Rightarrow \lambda(E) = 0$.*

The following proposition gives an equivalent characterization of absolute continuity in $\epsilon - \delta$ language.

Proposition 5.3. *Let μ and λ respectively be positive and complex measures on a measurable space $(X, \mathcal{M})$, then $\lambda \ll \mu$ if and only if $\forall \epsilon > 0,\ \exists \delta > 0,\ \forall E \in \mathcal{M},\ \mu(E) < \delta \Rightarrow |\lambda|(E) < \epsilon$.*

Definition 5.4 (Measure support)**.** *Let λ be a complex measure on a measurable space $(X, \mathcal{M})$ and $A \in \mathcal{M}$, we say λ is **supported in** A if $\forall E \in \mathcal{M},\ \lambda(E) = \lambda(E \cap A)$.*

Definition 5.5 (Singularity)**.** *Let μ and λ respectively be positive and complex measures on a measurable space $(X, \mathcal{M})$. λ is said to be **singular** w.r.t μ, denoted by $\lambda \perp \mu$, if there is $A \in \mathcal{M}$ such that λ is supported on A and $\mu(A) = 0$.*

Definition 5.6 (Mutual singularity)**.** *Let λ_1, λ_2 be complex measures on a measurable space $(X, \mathcal{M})$, we say λ_1 and λ_2 are **mutually singular**, denoted by $\lambda_1 \perp \lambda_2$, if there exist two disjoint measurable sets A and B such that λ_1 and λ_2 are respectively supported on A and B.*

The following proposition shows that the definitions of singularity and mutual singularity are consistent.

Proposition 5.4. *Let μ and λ respectively be positive and complex measures on a measurable space $(X, \mathcal{M})$ then λ is singular w.r.t μ if and only if λ and μ are mutually singular.*

Proof. If λ is singular w.r.t. μ, then we can find A such that λ is supported in A and $\mu(A) = 0$, as a result, for every $E \in \mathcal{M}$, we have

$$\mu(E \cap A^c) = \mu(E) - \mu(E \cap A) = \mu(E). \tag{5.6}$$

Hence, μ is supported in A^c, which is a measurable set disjoint from A. Hence, μ and λ are mutually singular. For the other direction, we assume that λ and μ are mutually singular, then we have two disjoint measurable sets A and B such that λ and μ are respectively supported in A and B. As a result,

$$\mu(A) = \mu(A \cap B) = \mu(\emptyset) = 0. \tag{5.7}$$

Hence, λ is singular w.r.t. μ. ☐

Proposition 5.5. *Let μ be a (complex/real/positive) measure on a measurable space $(X, \mathcal{M})$ that is supported in $A \in \mathcal{M}$, and B be a measurable set with $A \subset B$, then λ is also supported on B.*

Proof. For each $E \in \mathcal{M}$, we have $\mu(E \cap A^c) = \mu(E) - \mu(E \cap A) = 0$. Thus $\mu(E \cap (B \backslash A)) = \mu(E \cap (B \backslash A) \cap A^c) = 0$. As a result, $\mu(E \cap B) = \mu(E \cap A) + \mu(E \cap (B \backslash A)) = \mu(E \cap A) = \mu(E)$. Hence, μ is supported on B. ☐

Proposition 5.6. *Let μ be a positive measure, and λ_1, λ_2 be complex measures on a measurable space $(X, \mathcal{M})$. Then,*

1. *if λ_1, λ_2 are supported in $A \in \mathcal{M}$, so is $\alpha_1\lambda_1 + \alpha_2\lambda_2$ for each $\alpha_1, \alpha_2 \in \mathbb{C}$;*
2. *if $\lambda_1 \ll \mu$ and $\lambda_2 \ll \mu$, then $\alpha_1\lambda_1 + \alpha_2\lambda_2 \ll \mu$ for each $\alpha_1, \alpha_2 \in \mathbb{C}$;*
3. *if $\lambda_1 \perp \mu$ and $\lambda_2 \perp \mu$, then $\alpha_1\lambda_1 + \alpha_2\lambda_2 \perp \mu$ for each $\alpha_1, \alpha_2 \in \mathbb{C}$;*

Proof.

1. For each $E \in \mathcal{M}$, we have

$$(\alpha_1\lambda_1 + \alpha_2\lambda_2)(E \cap A) = \alpha_1\lambda_1(E \cap A) + \alpha_2\lambda_2(E \cap A) = \alpha_1\lambda_1(E) + \alpha_2\lambda_2(E) = (\alpha_1\lambda_1 + \alpha_2\lambda_2)(E) \tag{5.8}$$

Hence, $\alpha_1\lambda_1 + \alpha_2\lambda_2$ is supported in A.

2. For each E with $\mu(E) = 0$, we have

$$(\alpha_1\lambda_1 + \alpha_2\lambda_2)(E) = \alpha_1\lambda_1(E) + \alpha_2\lambda_2(E) = \alpha_1 \cdot 0 + \alpha_2 \cdot 0 = 0. \tag{5.9}$$

Hence, $\alpha_1\lambda_1 + \alpha_2\lambda_2 \ll \mu$.

3. We can find A_1 and A_2 respectively for λ_1 and λ_2 such that λ_1 and λ_2 are respectively supported in A_1 and A_2, and $\mu(A_1) = 0$, and $\mu(A_2) = 0$. Then λ_1 and λ_2 are both supported in $A_1 \cup A_2$, and $\mu(A_1 \cup A_2) = 0$. In addition, by the first statement, we know $\alpha_1\lambda_1 + \alpha_2\lambda_2$ is supported in $A_1 \cup A_2$.

☐

Proposition 5.7. *Let λ be a complex measure on a measurable space $(X, \mathcal{M})$, then $\lambda \ll |\lambda|$.*

Proof. For each $E \in \mathcal{M}$ with $|\lambda|(\mu) = 0$, by definition of $|\lambda|$, we have

$$|\lambda(E)| \leq |\lambda|(E) = 0. \tag{5.10}$$

Hence $\lambda(E) = 0$. Therefore, $\lambda \ll |\lambda|$. ☐

Proposition 5.8. *Let μ and λ be a respective positive and complex measure on a measurable space $(X, \mathcal{M})$, then*

1. *if λ is supported in $A \in \mathcal{M}$, so is $|\lambda|$;*
2. *if $\lambda \ll \mu$, so is $|\lambda|$;*
3. *if $\lambda \perp \mu$, so is $|\lambda|$.*

Proof. 1. If λ is supported in A, then we have for each $E \in \mathcal{M}$,

$$|\lambda|(E) = \sup_{\mathcal{E}} \sum_{E\in\mathcal{E}} |\lambda(E_i)| = \sup_{\mathcal{E}} \sum_{E\in\mathcal{E}} |\lambda(E\cap A)| = |\lambda|(E\cap A). \tag{5.11}$$

Here $\mathcal{E}$ represents a countable partition of E, and thus $\{E\cap A | E \in \mathcal{E}\}$ is a countable partition of $E\cap A$. Hence, we can conclude that $|\lambda|$ is supported in A.

2. Assume $\lambda \ll \mu$. For each $A \in \mathcal{M}$ with $\mu(A) = 0$, we have $\mu(E) = 0$ and thus $\lambda(E) = 0$ for every measurable set $E \subset A$. Hence, for each countable partition $\mathcal{E}$ of A, we have

$$\sum_{E\in\mathcal{E}} |\lambda(E)| = 0. \tag{5.12}$$

By definition, we have $|\lambda|(A) = 0$. Hence, $|\lambda| \ll \mu$.

3. Assume $\lambda \perp \mu$. Then we can find $A \in \mathcal{M}$ such that λ is supported in A, and $\mu(A) = 0$. By the first statement, we know that $|\lambda|$ is also supported on A. Hence, $|\lambda| \perp \mu$.

□

Corollary 5.1. *Let λ_1 and λ_2 be complex measures on a measurable space $(X, \mathcal{M})$ with $\lambda_1 \perp \lambda_2$, then $|\lambda_1| \perp |\lambda_2|$.*

Proof. Since $\lambda_1 \perp \lambda_2$, we can find two disjoint measurable sets A and B such that λ_1, λ_2 are respectively supported in A and B. Then $|\lambda_1|$ and $|\lambda_2|$ are also respectively supported in A and B, thus $|\lambda_1| \perp |\lambda_2|$. □

Proposition 5.9. *Let μ be a positive measure and λ_1, λ_2 be complex measures on a measurable space $(X, \mathcal{M})$, with $\lambda_1 \ll \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 \perp \lambda_2$. In particular, if $\lambda_1 = \lambda_2 = \lambda$, then $\lambda = 0$.*

Proof. Since $\lambda_2 \perp \mu$, there are two disjoint measurable sets A and B such that μ and λ_2 are respectively supported in A and B. As a result, λ_1 is supported in A, thus $\lambda_1 \perp \lambda_2$. If $\lambda_1 = \lambda_2 = \lambda$, then for each $E \in \mathcal{M}$, we have

$$\lambda(E) = \lambda(E\cap A) = \lambda(E\cap A\cap B) = \lambda(\emptyset) = 0. \tag{5.13}$$

Hence $\lambda = 0$. □

5.2.2 Lebesgue Decomposition Theorem and Radon-Nikodym Theorem

Proposition 5.10. *Let μ be a positive measure on a measurable space $(X, \mathcal{M})$, and $h \in L^1(\mu)$, define $\lambda : \mathcal{M} \to \mathbb{C}$ by*

$$\lambda(E) = \int_E h d\mu,$$

then λ is absolutely continuous w.r.t. μ, i.e. $\lambda \ll \mu$.

Proof. This immediately follows from the fact that integration on a null set is zero. □

Lemma 5.1. *Let $(X, \mathcal{M}, \mu)$ be a σ-finite positive measure space, then there exists $f \in L^1(\mu)$ such that $0 < f < 1$.*

Proof. Since μ is σ-finite, we can write $X = \bigcup_{n=1}^{\infty} E_n$ with $\mu(E_n) < +\infty$. Let

$$g_n = \frac{2^{-n}}{1+\mu(E_n)}\chi_{E_n} \tag{5.14}$$

Let $f = \sum_{n=1}^{\infty} g_n$. We can see that for each $x \in X$,

$$0 < f(x) < \sum_{n=1}^{\infty} 2^{-n} = 1. \tag{5.15}$$

And by monotone convergence theorem,

$$\int_X f d\mu = \sum_{i=1}^{\infty} \int_X g_n d\mu = \sum_{i=1}^{\infty} \frac{2^{-n}}{1+\mu(E_n)}\mu(E_n) \leq \sum_{i=1}^{\infty} 2^{-n} = 1. \tag{5.16}$$

Hence, $f \in L^1(\mu)$. □

Theorem 5.2 (Lebesgue Decomposition Theorem)**.** *Let μ be a σ-finite positive measure on a measurable space $(X, \mathcal{M})$, and λ be a σ-finite complex measure on $(X, \mathcal{M})$, then there are unique complex measures λ_a and λ_s such that $\lambda = \lambda_a + \lambda_s$, $\lambda_a \ll \mu$, and $\lambda_s \perp \mu$.*

Theorem 5.3 (Radon-Nikodym Theorem)**.** *Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space, and λ be a finite measure that is absolutely continuous w.r.t. μ, then there is $h \in L^1(\mu)$ such that $d\lambda = hd\mu$, i.e.*

$$\forall E \in \mathcal{M},\ \lambda(E) = \int_E h d\mu.$$

Here, h is called the ***Radon-Nikodym derivative*** *of λ w.r.t μ, and is written as $h = d\lambda/d\mu$.*

These two theorems are proved together in the following.

Proof. First of all, we assume λ be positive and finite. The conclusion will be generalized later.

By previous lemma, we can choose $w \in L^1(\mu)$ with $0 < w < 1$. Define $\phi : \mathcal{M} \to [0, +\infty]$ by

$$\varphi(E) = \lambda(E) + \int_E w d\mu. \tag{5.17}$$

We can write this as $d\varphi = d\lambda + wd\mu$ for short. It is easy to verify that φ is a positive and finite measure on $(X, \mathcal{M})$, and $\lambda \leq \varphi$. Note that $L^2(X, \mathcal{M}, \varphi)$ is a Hilbert space. Consider the functional given by $f \mapsto \int_X f d\lambda$. Then we have

$$\left|\int_X f d\lambda\right| \leq \int_X |f| d\lambda \leq \int_X |f| d\varphi = \langle |f|, 1\rangle \leq ||f||_2 \cdot ||1||_2 = ||f||_2 (\varphi(X))^{1/2}. \tag{5.18}$$

Since $\varphi(X) < +\infty$, thus the functional defined above is a bounded linear functional. By Riesz representation theorem, we there exists a unique $\bar{g} \in L^2(\varphi)$ such that

$$\int_X f d\lambda = \langle f, \bar{g}\rangle = \int_X f g d\varphi, \quad \forall f \in L^2(\varphi). \tag{5.19}$$

This can be written as $d\lambda = g d\varphi$. Hence, for each measurable set E, we have

$$\lambda(E) = \int_X g d\varphi, \tag{5.20}$$

and thus the average of g on E has

$$A_E(g) = \frac{1}{\varphi(E)} \int_X g d\varphi = \frac{\lambda(E)}{\varphi(E)} \in [0, 1]. \tag{5.21}$$

Hence, we can conclude that $g(x) \in [0, 1],\ a.e.[\varphi]$. Without losing generality, we can assume that $0 \leq gle1$. Moreover, we have

$$\int_X f d\lambda = \int_X f g d\varphi = \int_X f g d\lambda + \int_X f g w d\mu. \tag{5.22}$$

This can be rewritten as

$$\int_X f(1-g) d\lambda = \int_X f g w d\mu. \tag{5.23}$$

Let $A = g^{-1}[0, 1)$, and $B = g^{-1}\{1\}$, then $X = A \sqcup B$ (due to $0 \leq gle1$). Define $\lambda_a, \lambda_s : \mathcal{M} \to [0, +\infty]$ by

$$\lambda_a(E) := \lambda(E \cap A) \quad \text{and} \quad \lambda_s(E) := \lambda(E \cap B). \tag{5.24}$$

It is easy to verify that λ_a and λ_s are both positive measures and $\lambda = \lambda_a + \lambda_s$. From Eq.(5.23), we have

$$\int_B w d\mu = \int_B g w d\mu \int_X \chi_B g w d\mu = \int_X \chi_B (1-g) d\mu = \int_B 0 d\mu = 0. \tag{5.25}$$

Since $0 < w < 1$, from this we can conclude that $\mu(B) = 0$; while it is obvious that λ_s is supported in B, thus $\lambda_s \perp \mu$. For λ_a, we let $f_n = (1 + \sum_{i=1}^n g^i)\chi_E$. Again, by Eq.(5.23), we have

$$\int_E (1 - g^{n+1}) d\lambda = \int_X f_n (1-g) d\lambda = \int_X f_n g w d\lambda = \int_E g(1 + g + \cdots + g^n) d\lambda. \tag{5.26}$$

As $n \to \infty$, we have $1 - g^{n+1} \uparrow chi_A$ and $g(1 + g + \cdots + g^n)w \uparrow h$ where h is some non-negative measurable function. By monotone convergence theorem, we have $\chi_A d\lambda = hd\mu$.

$$\lambda_a(E) = \lambda(E \cap A) = \int_E \chi_A d\lambda = \int_E hd\mu. \tag{5.27}$$

Hence, $\lambda_a \ll \mu$. Recall our assumption that λ is a finite measure, hence $h \in L^1(\mu)$.

Note that up to now we have prove Radon-Nikodym theorem for the case with λ being a positive measure. If λ is a finite complex measure, we can decompose it into real and imaginary, positive and negative parts (Hahn-Jordan decomposition), and find the Radon-Nikodym derivatives for each of this part.

For Lebesgue decomposition theorem, we can decompose the space into disjoint union of finite measure sets. And perform the decomposition within each set respectively.

In the following we show that the Lebesgue decomposition is unique. Suppose $\lambda = \lambda_a + \lambda_s = \lambda'_a + \lambda'_s$ with $\lambda_a, \lambda'_a \ll \mu$ and $\lambda_s, \lambda'_s \perp \mu$. Then, we have

$$\lambda'_a - \lambda_a = \lambda_s - \lambda'_s. \tag{5.28}$$

While $\lambda'_a - \lambda_a \ll \mu$ and $\lambda_s - \lambda' s \perp \mu$, thus $\lambda'_a = \lambda_a$ and $\lambda'_s = \lambda_s$.

The uniqueness of the Radon-Nikodym derivative (up to almost everywhere equality) follows readily from the fact that

$$\int_E hd\mu = \int_E h' d\mu \;\forall E \in \mathcal{M} \Rightarrow h = h' \;\; a.e.[\mu]. \tag{5.29}$$

□

Theorem 5.4 (Polar decomposition)**.** *Let μ be a complex measure on a measurable space $(X, \mathcal{M})$, then there exists a complex measurable function h with $|h(x)| = 1 \;\forall x \in X$, such that $d\mu = hd|\mu|$, i.e.*

$$\mu(E) = \int_E hd|\mu|, \quad \forall E \in \mathcal{M}.$$

Proof. Since $\mu \ll |\mu|$, by Radon-Nikodym theorem, there exists a complex measurable function h such that $\mu = hd|\mu|$. In the following, we show that $|h(x)| = 1$ almost everywhere w.r.t. $|\mu|$. Let $A_r = \{x \in X : |h(x)| < r\}$. Let $\{E_{r,j}\}_j$ be partition of A_r.

$$\sum_{j=1}^{\infty} |\mu(E_{r,j})| = \sum_{j=1}^{\infty} \left| \int_{E_{r,j}} hd|\mu| \right| \leq r \sum_{j=1}^{\infty} |\mu|(E_{r,j}) = r\mu(A_r). \tag{5.30}$$

This follows that $|\mu|(A_r) \leq r|\mu|(A_r)$, implying that $|\mu|(A_r) = 0$. As this holds for any $0 \leq r < 1$, we know that $|h(x)| \geq 1, \; a.e.[\mu]$. For any $E \in \mathcal{M}$ with $|\mu|(E) > 0$, we have

$$|A_E(h)| = \left| \frac{1}{|\mu|(E)} \int_E hd|\mu| \right| = \frac{|\mu(E)|}{|\mu|(E)} \leq 1. \tag{5.31}$$

Therefore, $|h(x)| \leq 1, \; a.e.[\mu]$. Then, we can conclude that $|h(x)| = 1, \; a.e.[\mu]$. Choose an $h' = h, \; a.e.[\mu]$ and $h'(x) = 1, \; \forall x \in X$, then we are done. □

Part II

Fundamentals of Functional Analysis

Chapter 6

Review of Linear Algebra

6.1 Vector Spaces

First, we review several key concepts in linear algebra, including vector space, linear functional, linear transform, subspaces, and quotient spaces. It is important to note that the definitions of some concepts are given in a generalized way that is different from the one in an elementary linear algebra textbook, such as linear span, basis, etc.

Definition 6.1 (Vector space)**.** *A **vector space** is a set X over a field $\mathbb{F}$ is a set X together with an addtion $+ : X \times X \to X$ and a scalar multiplication $\cdot : \mathbb{F} \times X \to X$, which satisfy the following axioms*

1. *(commutativity of addition)* $\forall x, y \in X \ x + y = y + x$;
2. *(associativity of addition)* $\forall x, y, z \in X \ (x + y) + z = x + (y + z)$;
3. *(identity element of addition)* $\exists 0 \in X \ \forall x \in X \ x + 0 = x$;
4. *(inverse element of addition)* $\forall x \in X \ \exists(-x) \ x + (-x) = 0$;
5. *(identity of scalar multiplication)* $\forall x \ 1x = x$;
6. *(distributivity of scalar multiplication over vector addition)* $\forall \alpha \in \mathbb{F}, x, y \in X \ \alpha(x + y) = \alpha x + \alpha y$;
7. *(distributivity of scalar multiplication over field addition)* $\forall \alpha, \beta \in \mathbb{F}, x \in X \ (\alpha + \beta)x = \alpha x + \beta x$;
8. *(compatibility of scalar multiplication and field multiplication)* $\forall \alpha, \beta \in \mathbb{F}, x \in X \ \alpha(\beta x) = (\alpha\beta)x$.

*For conciseness of notation, one typically writes $x - y$ to represent $x + (-y)$. When $\mathbb{F}$ is $\mathbb{R}$, it is called a **real vector space**; when $\mathbb{F}$ is $\mathbb{C}$, it is called a **complex vector space**.*

Definition 6.2 (Linear independence)**.** *Let S be a subset of a vector space X over field $\mathbb{F}$, if there exists a finite subset $x_1, \ldots, x_n \in S$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ which are not all zeros, such that $\sum_{i=1}^n \alpha_i x_i = 0$, then S is called **linearly dependent**; otherwise it is called **linear independent**.*

Definition 6.3 (Linear span)**.** *Let S be a subset of a vector space X, the **linear span** of X, denoted by $\operatorname{span} S$ is defined to be the set of all linear combinations of **finitely many** vectors in S,*

$$\operatorname{span} S = \left\{ \sum_{i=1}^n \alpha_i x_i \,\middle|\, x_1, \ldots, x_n \in S, \ \alpha_1, \ldots, \alpha_n \in \mathbb{F} \right\}$$

*If $\operatorname{span} S = X$, we say that S **spans** X.*

Definition 6.4 (Basis)**.** *A **basis** of a vector space X is a linearly independent set that spans X.*

Definition 6.5 (Dimension)**.** *If a vector space X admits a finite basis, then all basis in X is finite, and they have the same cardinality called the **dimension** of X, denoted by $\dim X$. Then such a vector space is called a **finite dimensional vector space**. Otherwise (X admits no finite basis), then it is called an **infinite dimensional vector space**, and we write $\dim X = +\infty$ in this case.*

Note that the notion of basis introduced above works well in finite dimensional spaces, but there are several subtly different ways in generalizing it to infinite dimensional spaces. We will discuss a generalized notion of basis later (when introducing normed spaces).

6.2 Subspaces

Definition 6.6 (Subspace)**.** *Let X be a vector space over the field $\mathbb{F}$, and $Y \subset X$, if Y together with the addition and scalar multiplication inherited from X forms a vector space over $\mathbb{F}$, then Y is called a **subspace** of X*

Proposition 6.1. *Let Y be a subset of a vector space X over field $\mathbb{F}$, then Y is a subspace of X if and only if $0 \in Y$ and Y is closed under linear operations i.e. $x, y \in Y \Rightarrow x + y \in Y$ and $x \in Y, \alpha \in \mathbb{F} \Rightarrow \alpha x \in Y$.*

Proposition 6.2. *Intersection of subspaces of a vector space is also a subspace.*

Proof. Let $\mathcal{U}$ be a collection of subspaces, denote $V = \bigcap_{U \in \mathcal{U}} U$.

1. Since every $U \in \mathcal{U}$ is a subspace, $0 \in U,\ \forall U \in \mathcal{U}$, thus $0 \in V$.
2. Let $x, y \in V$ and $\alpha, \beta \in \mathbb{F}$, then $x, y \in U,\ \forall U \in \mathcal{U}$, thus $\alpha x + \beta y \in U,\ \forall U \in \mathcal{U}$, which means that $\alpha x + \beta y \in V$. Hence, V is closed under linear operations.

To conclude, V is a subspace. □

Proposition 6.3 (Direct sum of subapces)**.** *Let X be a vector space over the field $\mathbb{F}$, U and V be subspaces of X, then denote*

$$U + V = \{u + v | u \in U,\ v \in V\},$$

*then $U + V$ is also a subspace of X. In particular, $U \cap V = \{0\}$ if and only if every $x \in U + V$ can be uniquely expressed as $x = u + v$ with $u \in U$ and $v \in V$. In this case, $U + V$ is called a **direct sum** of U and V, denoted by $U \oplus V$.*

Proof. We first show that $U + V$ is a subspace of X. Let $x_1, x_2 \in U + V$, then we can write $x_1 = u_1 + v_1$ and $x_2 = u_2 + v_2$, with $u_1, u_2 \in U$ and $v_1, v_2 \in V$. Hence, for any $\alpha_1, \alpha_2 \in \mathbb{F}$, we have

$$\alpha_1 x_1 + \alpha_2 x_2 = \alpha_1(u_1 + v_1) + \alpha_2(u_2 + v_2) = (\alpha_1 u_1 + \alpha_2 u_2) + (\alpha_1 v_1 + \alpha_2 v_2) \in U + V. \tag{6.1}$$

It follows that $U + V$ is a subspace of X, so is $U \oplus V$.

Then, we show that $U \cap V = \{0\}$ if and only if the decomposition of x into $u + v$ with $u \in U$ and $v \in V$ is unique. For one direction, we assume that $U \cap V = \{0\}$, and then show the uniqueness of the expression. Suppose $x = u + v = u' + v'$ with $u, u' \in U$ and $v, v' \in V$. Hence, $u - u' = v - v'$ which are in both U and V, thus $u - u' = v - v' = 0$, and therefore $u = u'$ and $v = v'$. For the other direction, we assume the uniqueness and prove that $U \cap V = \{0\}$. If there is non-zero vector $x \in U \cap V$, then $-x \in U \cap V$, and we can write $0 = 0 + 0$ and $0 = x + (-x)$, contradicting the uniqueness of expression. □

Proposition 6.4. *Let X be a vector space and U be its subspace, then there is a subspace V of X such that $X = U \oplus V$. (Note that V is generally not unique).*

Proposition 6.5. *Let U and V be subspaces of vector X over the field $\mathbb{F}$, then*

$$\dim(U + V) + \dim(U \cap V) = \dim(U) + \dim(V),$$

In particular,

$$\dim(U \oplus V) = \dim(U) + \dim(V).$$

6.3 Linear Maps

Definition 6.7 (Linear map)**.** *Let X and Y be vector spaces over field $\mathbb{F}$, a map $T : X \to Y$ is called a **linear map** if it preserves linear operations, i.e.*

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y),\ \forall x, y \in X,\ \alpha, \beta \in \mathbb{F}.$$

*A linear map is also called a **linear transform**, or a **linear operator**. We often write Tx instead of $T(x)$.*

Definition 6.8 (Domain and Range of linear map)**.** *Let $T : X \to Y$ be a linear map, the set $\{T(x)|x \in X\}$ is called the **range** or the **image** of T, denoted by $\operatorname{Im} T$. And X is called the **domain** of T, denoted by $\operatorname{Dom} T$.*

Proposition 6.6. *Let $T : X \to Y$ be a linear map, then $\operatorname{Im} T$ is subspace of Y.*

Proof. Let $y_1, y_2 \in \operatorname{Im} T$, then there are $x_1, x_2 \in X$ such that $y_1 = Tx_1$ and $y_2 = Tx_2$. Hence, for any $\alpha_1, \alpha_2 \in \mathbb{F}$, we have

$$\alpha_1 y_1 + \alpha_2 y_2 = \alpha_1 Tx_1 + \alpha_2 Tx_2 = T(\alpha_1 x_1 + \alpha_2 x_2) \in \operatorname{Im} T. \tag{6.2}$$

Hence $\operatorname{Im} T$ is a subspace of Y. □

Definition 6.9 (Kernel(Null space))**.** *Let $T : X \to Y$ be a linear map, the set $\{x \in X|T(x) = 0\}$, denoted by $\ker T$, is called the **kernel** or **null space** of f.*

Proposition 6.7. *Let $T : X \to Y$ be a linear map, then $\ker T$ is a subspace of X.*

Proof. Let $x_1, x_2 \in \ker T$, then for any $\alpha_1, \alpha_2 \in \mathbb{F}$, we have $T(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 Tx_1 + \alpha_2 Tx_2 = 0$, which follows that $\alpha_1 x_1 + \alpha_2 x_2 \in \ker T$. Hence, $\ker T$ is a subspace of X. □

Proposition 6.8. *Let $T : X \to Y$ be a bijective linear map, then the inverse map T^{-1} is also linear.*

Proof. Let $y_1, y_2 \in Y$, then since T is bijective, there exist unique $x_1, x_2 \in X$ such that $Tx_1 = y_1$ and $Tx_2 = y_2$, i.e. $x_1 = T^{-1}y_1$ and $x_2 = T^{-1}y_2$. Then, for any $\alpha_1, \alpha_2 \in \mathbb{F}$, we have $T(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 y_1 + \alpha_2 y_2$, thus $T^{-1}(\alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 x_1 + \alpha_2 x_2 = \alpha_1 T^{-1} x_1 + \alpha_2 T^{-1} x_2$. It means that T^{-1} is linear. □

Definition 6.10 (Linear isomorphism)**.** *Let X and Y be two vector spaces, a bijective linear map $T : X \to Y$ (there exists an inverse map) is called a **linear isomorphism**. X and Y are said to be **linearly isomorphic** if there is a linear isomorphism between them.*

Proposition 6.9. *Let $T : X \to Y$ be a linear map, then*

1. *T is injective if and only if $Tx = 0 \Rightarrow x = 0$, i.e. $\ker T = \{0\}$;*
2. *T is surjective if and only if $\operatorname{Im} T = Y$;*
3. *T is bijective (isomorphism) if and only if $\ker T = \{0\}$ and $\operatorname{Im} T = Y$.*

Proof.

1. First, we assume that T is injective. Then there is a unique x such that $Tx = 0$. And we know $T0 = 0$, thus $Tx = 0 \Rightarrow x = 0$. For the other direction, we assume that $Tx = 0 \Rightarrow x = 0$. Then for $x_1, x_2 \in X$ with $x_1 \neq x_2$, we have $Tx_1 - Tx_2 = T(x_1 - x_2) \neq 0$, i.e. $Tx_1 \neq Tx_2$, thus T is injective.
2. This is just a re-statement of the definition of a surjective map.
3. This immediately follows from the above two points.

□

The following proposition tells us that the characteristics of a linear map can be determined from its behavior on the basis.

Proposition 6.10. *Let $\mathcal{B}$ be a basis of a vector space X, and $T : X \to Y$ be a linear map, and $T\mathcal{B} = \{Te|e \in \mathcal{B}\}$,*

1. *if T is injective, then $T\mathcal{B}$ is linearly independent;*
2. *if T is surjective, then $T\mathcal{B}$ spans Y;*

3. *if T is bijective (isomorphism), then $T\mathcal{B}$ is a basis of Y.*

Proof. We only need to prove the first two points, while the third one is an immediate corollary of them.

1. Suppose T is injective and $T\mathcal{B}$ is linearly dependent, then there exists $e_1, \dots, e_n \in \mathcal{B}$, and $\alpha_1, \dots, \alpha_n \in \mathbb{F}$ such that $\sum_{i=1}^n \alpha_i T e_i = 0$, thus $T\left(\sum_{i=1}^n \alpha_i e_i\right) = 0$. Because T is injective, $\sum_{i=1}^n \alpha_i e_i = 0$, contradicting the assumption that $\mathcal{B}$ is a basis.

2. When T is surjective, for each $y \in Y$, we have $y = Tx$ for some $x \in X$. Since $\mathcal{B}$ is a basis of X, we can write $x = \sum_{i=1}^n \alpha_i e_i$ with $e_i \in \mathcal{B}$ for each i. As a result, $y = \sum_{i=1}^n \alpha_i T e_i$, thus $y \in \operatorname{span}(T\mathcal{B})$.

□

Theorem 6.1. *Let $T : X \to Y$ be a linear transform, then*

$$\dim(X) = \dim(\operatorname{Im} T) + \dim(\ker T).$$

Proof. Choose a subspace V of X such that $X = \ker T \oplus V$, then $\dim(X) = \dim(\ker T) + \dim(V)$. It suffices to show that V and $\operatorname{Im} T$ is isomorphic. Let $T' : V \to \operatorname{Im} T$ be the restriction of T, we are going to show that T' is an isomorphism.

Given $x \in V$ with $T'x = 0$, then $Tx = 0$, thus $x \in \ker T$. Note that $V \cap \ker T = \{0\}$, thus $x = 0$. Hence, T' is injective. For each $y \in \operatorname{Im} T$, then there exists $x \in X$ such that $y = Tx$. On the other hand, $X = \ker T \oplus V$, hence there is $u \in \ker T$ and $v \in V$ such that $x = u + v$. Therefore, $y = Tu + Tv$, and $Tu = 0$, thus $y = Tv = T'v$. Hence, T' is surjective. To conclude, T' is an linear isomorphism, thus $\dim(\operatorname{Im} T) = \dim(V)$. □

Definition 6.11 (Rank of linear transform)**.** *Let T be a linear transform, then* $\dim(\operatorname{Im} T)$ *is called the* ***rank*** *of T, denoted by* $\operatorname{rank} T$*.*

6.4 Quotient Spaces

Proposition 6.11. *Let X be a vector space and E be its subspace, then the relation $x \sim y$ defined by $x - y \in E$ is an equivalence relation, and the equivalence class containing x is given by $[x] = x + E = \{x + v | v \in E\}$, which is called a* ***coset*** *of E.*

Definition 6.12 (Quotient space)**.** *Let X be a vector space over field $\mathbb{F}$ and E be its subspace, then the equivalence relation $x \sim y$ defined by $x - y \in E$ induces the quotient set $Q = \{[x] | x \in E\}$, on which we can define addition as*

$$[x] + [y] = [x + y].$$

and scalar multiplication as

$$\alpha[x] = [\alpha x].$$

These operations are well defined (independent from the choice of representatives). Hence, Q together with the these linear operations forms a vector space over field $\mathbb{F}$, called the ***quotient space****, denoted by X/E.*

The following proof verifies that the linear operations defined on the quotient space are well defined.

Proof. For addition, we are to show $x \sim x'$, $y \sim y' \Rightarrow x+y \sim x'+y'$, which can be seen from $(x'+y') - (x+y) = (x' - x) + (y' - y) \in E$. For scalar multiplication, we are to show $x \sim x' \Rightarrow \alpha x \sim \alpha x'$, $\forall \alpha \in \mathbb{F}$, which can be seen from $\alpha x' - \alpha x = \alpha(x' - x) \in E$. □

Proposition 6.12. *Let X be a vector space, and E be its subspace, let $Y = X/E$, and define $T : X \to Y$ by $Tx = [x]$, then T is a linear map, and $\ker T = E$.*

Proof. Let $x, y \in X$, then for any $\alpha, \beta \in \mathbb{F}$, we have

$$T(\alpha x + \beta y) = [\alpha x + \beta y] = [\alpha x] + [\beta y] = \alpha[x] + \beta[y] = \alpha Tx + \beta Ty \tag{6.3}$$

Hence, T is linear. And

$$Tx = 0 \Leftrightarrow [x] = 0 \Leftrightarrow x \sim 0 \Leftrightarrow x \in E. \tag{6.4}$$

This implies that $\ker T = E$. □

Theorem 6.2. *Let $T : X \to Y$ be a linear map then $X/\ker T$ is a isomorphic to $\operatorname{Im} T$, in particular, the map $T' : X/\ker T \to \operatorname{Im} T$ given by $T'[x] = Tx$ is a well defined linear isomorphism.*

Proof. To show that T' is well defined, it is equivalent to showing that given $x, y \in X$ with $x \sim y$, we have $Tx = Ty$, this can be seen from

$$Tx - Ty = T(x - y) = 0, \tag{6.5}$$

the last equality is due to that $x - y \in \ker T$. Then, we show that T' is an isomorphism. First, given x with $T'[x] = 0$, then $Tx = 0$, thus $x \in \ker T$, i.e. $[x] = [0]$. Hence, T is injective. Second, for each $y \in \operatorname{Im} T$, we can find $x \in X$ with $Tx = y$, and thus $T'[x] = y$. Hence, T is surjective. □

Proposition 6.13 (Codimension)**.** *Let X be a vector space over field $\mathbb{F}$, and E be a subspace of X, then the dimension of the quotient space $\dim(X/E)$ is called the **codimension** of E in X, denoted by $\operatorname{codim} E$.*

Proposition 6.14. *Let X be a vector space, U and V be its subspaces with $X = U \oplus V$, then X/U is isomorphic to V, in particular, $T : V \to X/U$ defined by $v \mapsto [v]$ to V is an isomorphism.*

Proof. Suppose $Tv = [0]$, then $v \in U$, since $U \cap V = \{0\}$, thus $v = 0$. Hence, T is injective. Given any $[x] \in X/U$, we can write $x = u + v$, thus $x - v \in U$, i.e. $x \sim v$. As a result, $Tv = [v] = [x]$. It follows that T is surjective. □

Corollary 6.1. *Let X be a vector space over field $\mathbb{F}$, and E be a subspace of X, then*

$$\dim(E) + \operatorname{codim}(E) = \dim(E) + \dim(X/E) = \dim(X).$$

Chapter 7

Normed and Banach Spaces

7.1 Metric Spaces

7.1.1 Metric and Metric Spaces

Definition 7.1 (Metric (distance))**.** *Let X be a set, then $d : X \times X \to \mathbb{R}$ is called a **metric** or a **distance** on X if it satisfies*

1. *(positiveness) $d(x, y) \geq 0,\ \forall x, y \in X$, and $d(x, y) = 0 \Leftrightarrow x = y$;*
2. *(symmetry) $d(x, y) = d(y, x),\ \forall x, y \in X$;*
3. *(triangle inequality) $d(x, z) \leq d(x, y) + d(y, z),\ \forall x, y, z \in X$.*

Definition 7.2 (Metric space)**.** *A set X together with a metric d defined thereon is called a **metric space**, denoted by (X, d).*

Proposition 7.1. *Each metric space induces a topology, which is generated by the topological basis $\{B(x, r)|x \in X,\ r > 0\}$, where $B(x, r) = \{y \in X|d(y, x) < r\}$. Being a topological space with the induced topology, a metric space is Hausdorff.*

Definition 7.3 (Open ball, Closed ball, and Sphere)**.** *Let (X, d) be a metric space, then*

1. *$B(x_0, r) = \{x \in X|d(x, x_0) < r\}$ is called the **open ball** centered at x_0 with radius r;*
2. *$\bar{B}(x_0, r) = \{x \in X|d(x, x_0) \leq r\}$ is called the **closed ball** centered at x_0 with radius r;*
3. *$S(x_0, r) = \{x \in X|d(x, x_0) = r\}$ is called the **sphere** centered at x_0 with radius r.*

It is easy to see that open balls are open sets, closed balls and spheres are closed sets.

7.1.2 Convergence in Metric Spaces

Definition 7.4 (Convergence in metric space)**.** *Let (X, d) be a metric space, a sequence $(x_n)_{n=1}^{\infty}$ is said to **converges** to $x \in X$, if $d(x_n, x) \to 0$, and x is called the limit of (x_n), denoted by $\lim_{n\to\infty} x_n = x$ or $x_n \to x$.*

Definition 7.5 (Bounded set)**.** *A subset A of a metric space (X, d) is said to be **bounded**, if there exists $M > 0$ such that $d(x, y) \leq M,\ \forall x, y \in A$.*

Definition 7.6 (Cauchy sequence)**.** *Let (X, d) be a metric space, a sequence $(x_n)_{n=1}^{\infty}$ is called a **Cauchy sequence** for every $\epsilon > 0$, there exists N such that $d(x_m, x_n) < \epsilon,\ \forall m, n > N$.*

The relations between convergence, boundedness and Cauchy sequence are summarized by the following proposition.

Proposition 7.2. *Let (X, d) be a metric space, and $(x_n)_{n=1}^{\infty}$ be a sequence in X, then*

1. *if (x_n) is a Cauchy sequence, then (x_n) is bounded;*

2. *if (x_n) is convergent, then (x_n) is a Cauchy sequence;*

3. *if (x_n) is a Cauchy sequence and has a convergent subsequence that converges to x, then (x_n) converges to x;*

4. *if every subsequence of (x_n) has a subsequence that converges to the same limit x, then (x_n) converges to x.*

Proof. 1. Since (x_n) is Cauchy, we can find N, such that $d(x_N, x_m) < \epsilon$ for some $\epsilon > 0$ and every $m > N$. Let $D = \max d(x_1, x_N), \ldots, d(x_{N-1}, x_N), \epsilon$, then for any $i, j \in \mathbb{N}^+$, we have

$$d(x_i, x_j) \leq d(x_i, x_N) + d(x_j, x_N) \leq 2D. \tag{7.1}$$

Hence, (x_n) is bounded.

2. Let (x_n) converge to x, then for any $\epsilon > 0$, there exists N such that $d(x_n, x) < \epsilon/2$ for any $n > N$. Hence, for any $m, n > N$, we have $d(x_m, x_n) \leq d(x_m, x) + d(x_n, x) < \epsilon$, which means that (x_n) is Cauchy.

3. Let (x_n) be a Cauchy sequence with a convergent subsequence (x_{n_k}) that converges to x. Then, given $\epsilon > 0$, we can find N such that $d(x_{n_k}, x) < \epsilon/2$ when $n_k > N$, and $d(x_n, x_m) < \epsilon/2$ when $n, m > N$, thus for every $n > N$, we can find a $n_K > n$, and $d(x_n, x) \leq d(x_n, x_{n_K}) + d(x_{n_K}, x) < \epsilon$. Hence (x_n) converges to x.

4. Suppose (x_n) does not converge to x, we can find a subsequence of (x_{n_k}) such that $d(x_{n_k}, x) > \epsilon, \forall k \in \mathbb{N}^+$ for some $\epsilon > 0$. It is easy to see that every subsequence of (x_{n_k}) does not converge to x, violating the assumption that every subsequence has a subsequence that converges to x.

□

Definition 7.7 (Dense set)**.** *Let X be a topological space, and $A \subset X$ is called **dense** in X, if its closure $\bar{A} = X$.*

There are several equivalent statements in testing whether a set is dense.

Proposition 7.3. *Let X be a metric space, A be a subset of X then the following statements are equivalent*

1. *A is dense;*

2. *for every $x \in X$, there is a sequence $(a_n)_{n=1}^{\infty}$ in A such that $a_n \to x$;*

3. *every open ball in X contains some point in A.*

Proof. 1. (1) ⇒ (2). This immediately follows from the fact in topology that every point of $\bar{A}$ is either in A or a limit point of A;

2. (2) ⇒ (3). Given $x \in X$, we find (a_n) in A that converges to x. For any $\epsilon > 0$, there exists $N \in \mathbb{N}^+$ such that $d(a_n, x) < \epsilon$ i.e. $a_n \in B(x, \epsilon)$ for every $n > N$. Hence, $B(x, \epsilon) \cap A \neq \emptyset$.

3. (3) ⇒ (1). If statment (3) does not hold, then we can find a non-empty open ball U with $U \cap A = \emptyset$, which implies that $A \subset U^c$. Since U^c is closed. $\bar{A} \subset U^c$, thus $\bar{A} \neq X$.

□

Definition 7.8 (Separable metric space)**.** *A metric space X is said to be **separable** if there exists a countable dense set in X.*

Definition 7.9 (Complete metric space)**.** *A metric space X is said to be **complete** if every Cauchy sequence in X is convergent.*

7.2 Normed Spaces

7.2.1 Norm and Normed Spaces

In the following discusssion, the field $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$.

Definition 7.10. *Let X be a vector space over the field $\mathbb{F}$, a **norm** on X is a real-valued function $||\cdot||$ on X that satisfies*

1. $\forall x \in X$ $||x|| \geq 0$ *and* $||x|| = 0 \Leftrightarrow x = 0$*;*
2. $\forall x \in X, \alpha \in \mathbb{F}$ $||\alpha x|| = |\alpha|||x||$*;*
3. $\forall x, y \in X,\ ||x+y|| \leq ||x|| + ||y||$.

Definition 7.11 (Normed space)**.** *A (real or complex) vector space X together with a norm $||\cdot||$ defined thereon is called a **normed space**, denoted by $(X, ||\cdot||)$.*

Proposition 7.4. *Let $(X, ||\cdot||)$ be a normed space, then $d : X \times X \to \mathbb{R}$ defined by $d(x,y) = ||x-y||$ is a metric on X, which is called the **induced metric**. Hence, every normed space is considered as a metric space with with the induced normed.*

Proposition 7.5. *Let $(X, ||\cdot||)$ be a normed space, then the norm $x \mapsto ||x||$ is a continuous function.*

Proof. When $x_n \to x$, by definition of convergence in metric space, it has $||x_n - x|| \to 0$, hence $|||x_n|| - ||x||| \leq ||x_n - x|| \to 0$, which implies that $||\cdot||$ is continuous. □

Proposition 7.6. *The addition $(x,y) \mapsto x+y$ and scalar multiplication $(\alpha, x) \mapsto \alpha x$ in a normed space is continuous.*

Proof. Let $(x_n) \to x$, $(y_n) \to y$, and $(\alpha_n) \to \alpha$. Then,

$$||(x+y) - (x_n + y_n)|| = ||(x - x_n) + (y - y_n)|| = ||x - x_n|| + ||y - y_n|| \to 0 \tag{7.2}$$

And, since (x_n) is convergent, thus it is bounded, and there is M such that $||x_n|| \leq M$ for every n, and thus

$$||\alpha_n x_n - \alpha x|| \leq ||\alpha_n x_n - \alpha x_n|| + ||\alpha x_n - \alpha x|| \leq M||\alpha_n - \alpha|| + ||\alpha|| \cdot ||x_n - x|| \to 0. \tag{7.3}$$

The two equations above respectively show the continuity of addition and scalar multiplication. □

Definition 7.12 (Equivalent norms)**.** *Let X be a vector space, $||\cdot||_1$ and $||\cdot||_2$ be two norms defined on X if there are $a, b > 0$ such that*

$$a||x||_1 \leq ||x||_2 \leq b||x||_1,\ \forall x \in X,$$

*then $||\cdot||_1$ and $||\cdot||_2$ are said to be **equivalent norms**. It is easy to verify that that this is an equivalence relation between norms on X.*

Theorem 7.1. *Let X be a vector space, $||\cdot||_1$ and $||\cdot||_2$ be two norms on X, then they are equivalent norms if and only if they induce the same topology of X.*

Proof. First we show that equivalent norms induce the same topology. Suppose there are $a, b > 0$ such that $a||x||_1 \leq ||x||_2 \leq b||x||_1$, $\forall x \in X$. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be the topologies respectively induced by $||\cdot||_1$ and $||\cdot||_2$. Given $A \in \mathcal{T}_1$, then for every $a \in A$, we can find an $r > 0$ such that $\{x|||x-a||_1 < r\} \subset A$, and hence $\{x|||x-a||_2 < br\} \subset A$, thus $A \in \mathcal{T}_2$. Likewise, we can show $A \in \mathcal{T}_2 \Rightarrow A \in \mathcal{T}_1$. As a result, $\mathcal{T}_1 = \mathcal{T}_2$.

Then we show that the norms that induce the same topology are equivalent. Suppose both $||\cdot||_1$ and $||\cdot||_2$ induce the topology $\mathcal{T}$, then within $\{x|||x||_1 < 1\}$, we can find $\{x|||x||_2 \leq a\}$ for some $a > 0$, and in $\{x|||x||_2 < 1\}$ we can find $C = \{x|||x||_1 < b\}$ for some $b > 0$, hence we have $||x||_2 = a \Rightarrow ||x||_1 < 1$ and $||x||_1 = b \Rightarrow ||x||_2 < 1$, therefore, $(1/b)||x||_1 \leq ||x||_2 \leq a||x||_1$ for every x, which implies that $||\cdot||_1$ and $||\cdot||_2$ are equivalent norms. □

Theorem 7.2. *All norms defined on a finite dimensional vector space are equivalent.*

Proof. Let X be an n-dimensional vector spaces with a basis $\{e_1, \dots, e_n\}$. Then any vector $x \in X$ can be uniquely written as $x = \sum_{i=1}^n \alpha_i e_i$. We define $||x||_1 = \sum_{i=1}^n |\alpha_i|$, then it is easy to verify that $||\cdot||_1$ is a norm on X. It suffices to show that every norm on X is equivalent to $||\cdot||_1$.

Let $||\cdot||$ be a norm on X. Then for every $x = \sum_{i=1}^n \alpha_i e_i \in X$, we have

$$||x|| \leq \sum_{i=1}^n |\alpha_i|||e_i|| \leq \left(\max_{i=1}^n ||e_i||\right) \sum_{i=1}^n |\alpha_i| = \left(\max_{i=1}^n ||e_i||\right) ||x||_1. \tag{7.4}$$

Let $S = \{x|||x|| = 1\}$, which is a closed set due to the continuity of norm, thus $||\cdot||_1$ attains the minimum value on S, i.e. there exists $x^* \in S$ with $||x^*||_1 = \min_{x\in S} ||x||_1$. Since $x^* \neq 0$, thus $||x^*||_1 > 0$. Hence, for every $x \neq 0$ we have

$$||x||_1 = ||x|| \cdot ||(x/||x||)||_1 \geq ||x|| \cdot ||x^*||_1. \tag{7.5}$$

Combining the results above, we can conclude that $||\cdot||$ is equivalent to $||\cdot||_1$. □

7.2.2 Seminorm and Norm of Quotient Spaces

Definition 7.13 (Seminorm)**.** *Let X be a vector space over the field $\mathbb{F}$, a* ***seminorm*** *on X is a real-valued function ρ on X that satisfies*

1. $\forall x \in X \;\; \rho(x) \geq 0$;
2. $\forall x \in X, \alpha \in \mathbb{F} \;\; \rho(\alpha x) = |\alpha|\rho(x)$;
3. $\forall x, y \in X, \;\; \rho(x + y) \leq \rho(x) + \rho(y)$.

In particular, a seminorm ρ is a norm if it satisfies $\rho(x) = 0 \Rightarrow x = 0$.

Theorem 7.3. *Let E be a subspace of a vector space X, and $\rho : X \to \mathbb{R}$ be a seminorm on X, then the function $\rho' : X/E \to \mathbb{R}$ given by $\rho'([x]) = \inf_{y\in E} \rho(x - y)$ is a well defined seminorm. In particular, if ρ is a norm on X, and E is closed (w.r.t the topology induced by ρ), then ρ' is a norm on X/E.*

Proof.

1. First of all, we need to show that the given construction is well defined, i.e. when $x \sim x'$, we have $\inf_{y\in X_0} \rho(x - y) = \inf_{y'\in X_0} \rho(x' - y')$.

   Given $x, x' \in X$ with $x \sim x'$, i.e. $x - x' \in E$. Let $d = \inf_{y\in X_0} \rho(x-y)$, we can find $(y_n)_{n=1}^\infty$ in E such that $\rho(x - y_n) \to d$. Then for x', we let $y'_n = x' - x + y_n$, then $y'_n \in E$ for each n, and $\rho(x' - y'_n) = \rho(x - y_n)$, as a result, $\rho(x' - y'_n) \to d$. Hence,

   $$\inf_{y'\in E} \rho(x' - y') \leq \inf_n \rho(x' - y'_n) = d = \inf_{y\in E} \rho(x - y). \tag{7.6}$$

   Likewise, we have $\inf_{y\in E} \rho(x - y) \leq \inf_{y'\in E} \rho(x' - y')$. Hence, the equality holds, implying that the construction is well defined.

2. Then, we need to show that the defined function is a seminorm.

   (a) $\rho([x])$ is non-negative for each $x \in X$, which directly inherits from the non-negativity of ρ. In particular, when $[x] = [0]$, since $x \in E$, we have $\rho'([x]) \leq rho(x - x) = \rho(0) = 0$.

   (b) Consider $\rho(\alpha[x])$. If $\alpha = 0$,

   $$\rho'(\alpha[x]) = \rho'([\alpha x]) = \rho'([0]) = 0 = |\alpha|\rho'([x]). \tag{7.7}$$

   If $\alpha \neq 0$, let (y_n) be a sequence in E that has $||x - y_n|| \to \rho'([x])$. Then (αy_n) is also a sequence in E, and it has $\rho(\alpha x - \alpha y_n) = |\alpha|\rho(x - y_n)$. Hence by definition, we have

   $$\rho'(\alpha[x]) = \rho'([\alpha x]) = \inf_{y\in X_0} \rho(\alpha x - y) \leq \inf_n \rho(\alpha x - \alpha y_n) = |\alpha| \cdot \inf_n \rho(x - y_n) = |\alpha| \cdot \rho'([x]). \tag{7.8}$$

   On the other hand, we note that $x = \alpha^{-1}\alpha x$. Applying the above conclusion, we have

   $$\rho'([x]) = \rho'(\alpha^{-1}[\alpha x]) \leq |\alpha|^{-1} \cdot \rho'([\alpha x]). \tag{7.9}$$

   which implies that $\rho'([\alpha x]) \geq |\alpha|\rho'([x])$. Combining the above two results, we obtain $\rho'(\alpha[x]) = |\alpha| \cdot \rho'([x])$.

(c) Let $x, y \in X$, then there exists sequences $(u_n)_{n=1}^{\infty}$ and $(v_n)_{n=1}^{\infty}$ in E such that $\rho(x - u_n) \to \rho'([x])$ and $\rho(y - v_n) \to \rho'([y])$, let $w_n = u_n + v_n$, then (w_n) is also a sequence in E, and we have

$$\begin{aligned}\rho'([x] + [y]) = \rho'([x + y]) &= \inf_{w \in X_0} \rho(x + y - w) \leq \inf_n \rho(x + y - w_n) \\ &= \inf_n \rho((x - u_n) + (y - v_n)) \leq \lim_{n \to \infty} \rho(x - u_n) + \rho(y - v_n) = \rho'([x]) + \rho'([y]).\end{aligned} \tag{7.10}$$

The triangle inequality is thus established.

Therefore, the function defined by $\rho([x]) = \inf_{y \in X_0} \rho(x - y)$ is a well-defined seminorm on X/X_0.

3. Finally, we show that when ρ is a norm, ρ' is a norm on X/E. Suppose $\rho'([x]) = 0$, then there exists a sequence (y_n) in E such that $\rho(x, y_n) \to 0$, which means that $y_n \to x$. Since E is closed, it contains all its limit points, thus $x \in E$, i.e. $x \sim 0$ or equivalently $[x] = [0]$. Hence ρ' is a norm.

□

Theorem 7.4. *Let X be a (complex or real) vector space, and ρ be a seminorm on X, then $E = \{x | \rho(x) = 0\}$ is a subspace of X. On the quotient space X/E, the function given by $||[x]|| = \inf_{y \in E} \rho(x - y)$ is a norm, and it has $||[x]|| = \rho(x)$.*

Proof. We first show that E is a closed subspace. Let $x, y \in E$, and $\alpha, \beta \in \mathbb{F}$, then by sub-additivity, we have $\rho(\alpha x + \beta y) \leq \alpha\rho(x) + \beta\rho(y) = 0$, thus $\alpha x + \beta y \in E$.

By theorem 7.3, we know that $|| \cdot ||$ as defined above is a seminorm. To show that it is a norm, it suffices to show that $||[x]|| = 0 \Rightarrow x \in E$. By definition, there is a sequence (y_n) in E such that $\rho(x, y_n) \to 0$. By sub-additivity of ρ, we have $\rho(x) \leq \rho(x, y_n) + \rho(y_n) = \rho(x, y_n)$, which can be arbitrarily small. Hence, $\rho(x) = 0$, thus $x \in E$. Then, we can conclude that $|| \cdot ||$ is a norm on $X\ E$.

Finally, we show the equality $||[x]|| = \rho(x)$. For each $y \in E$, we have $\rho(x) \leq \rho(x, y) + \rho(y) = \rho(x, y)$, hence $\rho(x) \leq \inf_{y \in E} \rho(x, y) = ||[x]||$. On the other hand, $\rho(x) = \rho(x - 0) \geq \inf_{y \in E} \rho(x, y)$. Together, we have $||[x]|| = \rho(x)$. □

7.3 Banach Spaces

Recall that a complete metric space is a space in which every Cauchy sequence converges.

Definition 7.14 (Banach space)**.** *A complete normed space is called a **Banach space**.*

Proposition 7.7. *A subspace Y of a Banach space X is complete if and only if it is closed.*

Proof. First, suppose Y is complete. Given a sequence in Y that converges to some point $x \in X$, then it is Cauchy, and by completeness of Y, it converges in Y, thus $x \in Y$, which implies that Y contains all its limit points, thus Y is closed. Second, suppose Y is closed. Given a Cauchy sequence in Y, it must converge to some point $x \in X$ due to completeness of X, and thus by closedness of Y, we have $x \in Y$, which implies that Y is complete. □

Definition 7.15 (Isometry)**.** *Let X and Y be normed spaces, then $T : X \to Y$ is called a **isometry** if it is an isomorphism and preserves norm, i.e. $||Tx|| = ||x||$, $\forall x \in X$. X and Y are said to be **isometric** if there is an isometry between them.*

Theorem 7.5 (Existence of completion of normed space)**.** *Let E be a normed vector space, then there exists a complete normed space $\tilde{E}$ and a linear transform T such that $||Tx|| = x$, $\forall x \in E$, and TE is dense in $\tilde{E}$. Here, $\tilde{E}$ is called a **completion** of E.*

Proof. The basic idea is that we start from the vector space of Cauchy sequences and then derive a quotient space by combining the ones that are equivalent (we will define the equivalence later during the construction). Let $\mathcal{C}(E)$ be the set of all Cauchy sequences of E.

Claim 1 $\mathcal{C}(E)$ *is a vector space, i.e. it is closed under addition and scalar multiplication.*

Proof of Claim 1. Let $x, y \in \mathcal{C}(E)$, then given $\epsilon > 0$, there exists sufficiently large N such that $||x_m - x_n|| < \epsilon/2$ and $||y_m - y_n|| < \epsilon/2$ when $m, n > N$. Hence,

$$||(x+y)_m - (x+y)_n|| = ||(x_m - x_n) + (y_m - y_n)|| \leq ||x_m - x_n|| + ||y_m - y_n|| < \epsilon. \tag{7.11}$$

Thus, $x+y \in \mathcal{C}(E)$. Given $\alpha \in \mathbb{C}$, if $\alpha = 0$, then for any sequence x, αx is the zero sequence, which is obviously Cauchy and thus in $\mathcal{C}(E)$. If $\alpha \neq 0$, given $x \in \mathcal{C}(E)$ and $\epsilon > 0$, there exists M such that $||x_m - x_n|| < \epsilon/|\alpha|$. As a result,

$$||(\alpha x)_m - (\alpha x)_n|| = ||\alpha(x_m - x_n)|| = |\alpha| \cdot ||x_m - x_n|| < \epsilon. \tag{7.12}$$

Hence, $\alpha x \in \mathcal{C}(E)$. Then, we can conclude that $\mathcal{C}(E)$ is a vector space. □

Given $x, y \in \mathcal{C}(E)$, we define the following relation

$$x \sim y \Leftrightarrow \lim_{n\to\infty} ||x_n - y_n|| = 0.$$

Claim 2 *The relation defined above is an equivalence relation.*

Proof of Claim 2. First, $x \sim x$ is obvious, since $||x_n - x_n|| = 0$, $\forall n$, its limit must be zero. Second, $x \sim y \Rightarrow y \sim x$ follows from the fact that $||x_n - y_n|| = ||y_n - x_n||$. Third, if $x \sim y$ and $y \sim z$, then by triangle inequality, $||x_n - z_n|| \leq ||x_n - y_n|| + ||y_n - z_n|| \to 0$, thus $x \sim z$. □

In this sense, we call x and y *equivalent Cauchy sequences* if $x \sim y$, and we use $[x]$ to denote all the Cauchy sequences that are equivalent to x.

Claim 3 *Based on the equivalence relation defined above,* $[0]$ *is a subspace of* $\mathcal{C}(E)$*.*

Proof of Claim 3. Given $x, y \in \mathcal{C}(E)$ with $x \sim 0$ and $y \sim 0$, and $\alpha, \beta \in \mathbb{C}$, then

$$\lim_{n\to\infty} ||\alpha x_n + \beta y_n|| \leq \lim_{n\to\infty} (|\alpha| \cdot ||x_n|| + |\beta| \cdot ||y_n||) = |\alpha| \lim_{n\to\infty} ||x_n|| + |\beta| \lim_{n\to\infty} ||y_n|| = 0. \tag{7.13}$$

Hence, $\alpha x_n + \beta y_n \sim 0$, implying that $[0]$ is a subspace. □

Hence, we can define the quotient space $\tilde{E} = \mathcal{C}(E)/[0]$ with $[x] + [y] = [x + y]$ and $[\alpha x] = \alpha[x]$. (It is a well known result in linear algebra that the operations in quotient space are well defined).

Claim 4 *The vector space* $\tilde{E}$ *defined above can be equipped with a norm* $||\cdot||$ *defined by* $||[x]|| = \lim_{n\to\infty} ||x_n||$*.*

Proof of Claim 4. First of all, we need to verify that such a definition is well defined, i.e. $x \sim y \Rightarrow \lim_{n\to\infty} ||x_n|| = \lim_{n\to\infty} ||y_n||$. This is briefly shown below.

$$\lim_{n\to\infty} ||x_n|| = lim_{n\to\infty} ||y_n + (x_n - y_n)|| \leq \lim_{n\to\infty} ||y_n|| + \lim_{n\to\infty} ||x_n - y_n|| = \lim_{n\to\infty} ||y_n||, \tag{7.14}$$

likewise, we can get $\lim_{n\to\infty} ||y_n|| \leq \lim_{n\to\infty} ||x_n||$. Then, the equality is established. In the following, we show that this real valued function is a norm on $\tilde{E}$.

1. $||[x]||$ is non-negative, which follows from the fact that $||x_n||$ is non-negative. And, when $[x] = [0]$, i.e. $x \sim 0$, then by definition of the equivalence relation, we have

$$||[x]|| = 0 \Leftrightarrow \lim_{n\to\infty} ||x_n|| = 0 \leftrightarrow x \sim 0 \Leftrightarrow [x] = [0]. \tag{7.15}$$

2. Given $x \in \mathcal{C}(E)$ and $\alpha \in \mathbb{C}$, we have

$$||\alpha \cdot [x]|| = ||[\alpha \cdot x]|| = \lim_{n\to\infty} ||\alpha x_n|| = \lim_{n\to\infty} |\alpha| \cdot ||x_n|| = |\alpha| \cdot \lim_{n\to\infty} ||x_n|| = |\alpha| \cdot ||[x]||. \tag{7.16}$$

3. Given $x, y \in \mathcal{C}(E)$, we have

$$||[x]+[y]|| = ||[x+y]|| = \lim_{n\to\infty} ||x_n+y_n|| \leq \lim_{n\to\infty} (||x_n||+||y_n||) = \lim_{n\to\infty} ||x_n||+\lim_{n\to\infty} ||y_n|| = ||[x]||+||[y]||. \tag{7.17}$$

Hence, we can conclude that $||\cdot||$ is a well-defined norm on $\tilde{E}$. $\square$

Claim 5 *The normed space $(\tilde{E}, ||\cdot||)$ defined above is complete, i.e. it is a Banach space.*

Proof of Claim 5. Let $([x_1],[x_2],\ldots)$ be a Cauchy sequence in $\tilde{E}$. Note that each element in this sequence is in itself an equivalence class of Cauchy sequences. We use x_{ij} to denote the j-th element in x_i. To show that $\tilde{E}$ is complete, it suffices to show that there is $[y] \in \tilde{E}$ such that $\lim_{n\to\infty} ||[x_n - y]|| = 0$.

We define y by $y_j = x_{jj}$ (take the j-th vector in the sequence x_j to be the j-th vector in the sequence y). First, we need to show that $y \in \mathcal{C}(E)$, i.e. it is a Cauchy sequence. We do this as follows.

Since $([x_n])$ is Cauchy, given $\epsilon > 0$, we can find N_1 such that $||[x_m]-[x_n]|| = \lim_{k\to\infty} ||x_{mk} - x_{nk}|| < \epsilon/3$ when $m,n > N_1$. Then, we can choose N_2, such that when $k > N_2$, $||x_{mk} - x_{nk}|| < \epsilon/3$. Hence, when $m,n,q > \max(N_1,N_2)$, we have $||x_{mm} - x_{qm}|| < \epsilon/3$, and $||x_{qn} - x_{nn}|| < \epsilon/3$. On the other hand, we fix some $q > \max(N_1,N_2)$, since x_q is Cauchy, we can choose N_3 such that when $m,n > N_3$, we have $||x_{qm} - x_{qn}|| < \epsilon/3$. Together, for every $m,n > \max(N_1,N_2,N_3)$, we have

$$||y_m - y_n|| = ||x_{mm} - x_{nn}|| \leq ||x_{mm} - x_{qm}|| + ||x_{qm} - x_{qn}|| + ||x_{qn} - x_{nn}|| < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon. \quad (7.18)$$

It follows that y is a Cauchy sequence.

Then, we show that $([x_n])$ converges to $[y]$. Given $\epsilon > 0$, we can choose M, such that $||[x_n - x_k]|| < \epsilon$ for each $n,k < \epsilon$, meaning $\lim_{i\to\infty} ||x_{ni} - xki|| < \epsilon$, so when k is sufficiently large $||x_{nk} - x_{kk}|| < \epsilon$, which follows that $\lim_{k\to\infty} ||x_{nk} - x_{kk}|| < \epsilon$. By definition, it means that $||[x_n]-[y]|| < \epsilon$, $\forall n > M$. As we can find such M for arbitrarily small $\epsilon > 0$, we can conclude that $\lim_{n\to\infty} ||[x_n]-[y]|| = 0$, in other words, $[x_n] \to [y]$. Therefore, $\tilde{E}$ is complete. $\square$

We use $\tilde{x} = (x,x,\ldots)$ to denote a constant sequence (which is clearly Cauchy), and define $T : E \to \tilde{E}$ by

$$Tx = [\tilde{x}].$$

Claim 6 *The T defined above is an injective linear map with $||Tx|| = ||x||$, $\forall x \in E$.*

Proof of Claim 6. The linearity of T can be seen from below. Given $x,y \in E$ and $\alpha,\beta \in \mathbb{C}$, we have

$$T(\alpha x + \beta y) = [\widetilde{\alpha x + \beta y}] = [\alpha\tilde{x} + \beta\tilde{y}] = \alpha[\tilde{x}] + \beta[\tilde{y}] = \alpha Tx + \beta Ty. \quad (7.19)$$

Then we show that it preserves norm as follows.

$$||Tx|| = ||[\tilde{x}]|| = \lim_{n\to\infty} ||\tilde{x}_n|| = ||x||. \quad (7.20)$$

Since T preserves norm, $Tx = 0 \Rightarrow ||x|| = ||Tx|| = 0 \Rightarrow x = 0$, thus T is injective. $\square$

Claim 7 *TE is dense in $\tilde{E}$, i.e. $\overline{TE} = \tilde{E}$.*

Proof of Claim 7. It suffices to prove that given any $[y] \in \tilde{E}$, there is a sequence of $(x_n)_{n=1}^{\infty}$ in E such that $Tx_n \to [y]$.

Given $[y] \in \tilde{E}$, we define $x_n = y_n$, then since y is a Cauchy sequence.

$$\lim_{n\to\infty} ||[y] - Tx_n|| = \lim_{n\to\infty} ||[y]-[\tilde{x}_n]|| = \lim_{n\to\infty}\left(\lim_{k\to\infty} ||y_k - \tilde{x}_{nk}||\right) = \lim_{n\to\infty}\lim_{k\to\infty} ||y_k - y_n|| = 0. \quad (7.21)$$

It means that $Tx_n \to y$, and therefore, we can conclude that TE is dense in $\tilde{E}$. $\square$

The entire proof of existence is completed. $\square$

Theorem 7.6 (Uniqueness of completion of normed space)**.** *Let $\tilde{E}_1$ and $\tilde{E}_2$ be two completion of E, then they are isometric. Formally, if $\tilde{E}_1$ and $\tilde{E}_2$ are complete normed spaces, and there are linear maps $T_1 : E \to \tilde{E}_1$ and $T_2 : E \to \tilde{E}_2$ which satisfy $\overline{T_iE} = \tilde{E}$ and $||T_ix|| = ||x||$ $\forall x \in E$ for each $i = 1,2$, then there is an isometry $\tilde{T} : \tilde{E}_1 \to \tilde{E}_2$.*

Proof. We have shown above that T_1 and T_2 are both injective. By restricting their target spaces, we get $T_1' : E \to T_1E$ and $T_2' : E \to T_2E$, which are both isomorphisms. Define $T = T_2' \circ T_1'^{-1} : T_1E \to T_2E$. Then T is an isomorphism between T_1E and T_2E.

Based on T, we define $\tilde{T} : \tilde{E}_1 \to \tilde{E}_2$ as follows. Given $y_1 \in \tilde{E}_1$, since T_1E is dense in $\tilde{E}_1$, we can choose a sequence (x_n) in E such that (T_1x_n) converges to y_1. Because T_1 preserves norm, we know that $||x_m - x_n|| = ||T_1x_m - T_1x_n||$, and note that (T_1x_n) is Cauchy (due to convergence), hence (x_n) is Cauchy. And, since T_2 also preserves norm, we can see that the sequence (T_2x_n) in $\tilde{E}_2$ is also a Cauchy sequence. By completeness of $\tilde{E}_2$, T_2x_n converges to a unique element $y_2 \in \tilde{E}_2$. The map $\tilde{T}$ is defined to be $y_1 \mapsto y_2$, where y_2 is found as described above.

Claim 1 *The map $\tilde{T}$ is well-defined.*

Proof of Claim 1. We need to prove that y_2 is independent from the choice of the intermediate Cauchy sequence (x_n) in E. Let x and x' be two sequences in E such that $T_1x_n \to y_1$ and $T_1x_n' \to y_1$, hence $T_1(x_n - x_n') \to 0$, which follows that $||x_n - x_n'|| = ||T_1(x_n - x_n')|| \to 0$. Note that T_2 also preserves norm, thus $||T_2x_n - T_2x_n'|| \to 0$, implying that T_2x_n and T_2x_n' converges to the same limit. Therefore, y_2 chosen by the described process is unique for each y_1, meaning that $\tilde{T}$ is well defined. □

Claim 2 *$\tilde{T}$ is a linear map with $||\tilde{T}x|| = ||x||$ for every $x \in \tilde{E}_1$.*

Proof of Claim 2. Let $x_1, y_1 \in \tilde{E}_1$ and $\alpha, \beta \in \mathbb{C}$, we choose two sequences (u_n) and (v_n) in E with $T_1u_n \to x_1$ and $T_1v_n \to y_1$. Let $x_2 = \tilde{T}x_1 = \lim_{n\to\infty} T_2u_n$ and $y_2 = \tilde{T}y_1 = \lim_{n\to\infty} T_2v_n$. (We have shown above that x_2 and y_2 are well defined and independent from the choice of (u_n) and (v_n)). Then, it is easy to see that $T_1(\alpha u_n + \beta v_n) \to (\alpha x_1 + \beta y_1)$, hence

$$\tilde{T}(\alpha x_1 + \beta y_1) = \lim_{n\to\infty} T_2(\alpha u_n + \beta v_n) = \alpha \lim_{n\to\infty} T_2u_n + \beta \lim_{n\to\infty} T_2v_n = \alpha\tilde{T}x_1 + \beta\tilde{T}x_2. \tag{7.22}$$

Hence, $\tilde{T}$ is a linear map.

In addition, by continuity of norm and the assumption that T_1 and T_2 preserve norms, we have

$$||\tilde{T}x_1|| = ||\lim_{n\to\infty} T_2u_n|| = \lim_{n\to\infty} ||T_2u_n|| = \lim_{n\to\infty} ||u_n|| = \lim_{n\to\infty} ||T_1u_n|| = ||\lim_{n\to\infty} T_1u_n|| = ||x_1||. \tag{7.23}$$

Hence, $\tilde{T}$ preserves norm. □

Claim 3 *$\tilde{T}$ is a linear isomorphism.*

Proof of Claim 3. It suffices to show that $\tilde{T}$ is bijective. First, since $\tilde{T}$ preserves norm, thus

$$\tilde{T}x = 0 \Rightarrow ||\tilde{T}x|| = 0 \Rightarrow ||x|| = 0 \Rightarrow x = 0. \tag{7.24}$$

Hence, $\tilde{T}$ is injective. Then we show that it is also surjective. Given $y_2 \in \tilde{E}_2$, we can find a sequence (x_n) in E with $T_2x_n \to y_2$, and hence x_n is a Cauchy sequence, thus T_1x_n converges to an element in $\tilde{E}_1$, denoted by y_1. According to the construction process described above, we know $\tilde{T}y_1 = y_2$. Therefore, we can conclude that $\tilde{T}$ is a linear isomorphism. □

To sum up, $\tilde{T}$ is an isometry. The proof of the theorem of uniqueness is completed. □

Proposition 7.8. *Every finite dimensional normed space is a Banach space.*

Proof. Let X be a normed space with a basis $\{e_1, \ldots, e_n\}$. Given a Cachy sequence $(x_i)_{i=1}^{\infty}$ in X. Let $x_i = \sum_{j=1}^{n} \alpha_{ij}e_j$. Note that all norms are equivalent in finite dimensional space, thus (x_i) is Cachy in terms of the norm defined by $||x||_1 = \sum_{j=1}^{n} |\alpha_j|$ where $x = \sum_{j=1}^{n} \alpha_j e_j$. From this, we can readily see that $(\alpha_{ij})_{i=1}^{\infty}$ is Cachy for each j, hence it is convergent (due to completeness of $\mathbb{R}$ and $\mathbb{C}$) Define $\beta_j = \lim_{i\to\infty} a_{ij}$, and $x = \sum_{j=1}^{n} \beta_j e_j$. Then

$$||x_i - x|| \le \sum_{j=1}^{n} |\alpha_{ij} - \beta_j| \cdot ||e_j|| \to 0. \tag{7.25}$$

Hence $x_i \to x$ by definition. Therefore, we can conclude that X is a Banach space. □

Corollary 7.1. *Every finite dimensional subspace of a normed space is closed.*

Chapter 8

Hilbert Spaces

8.1 Inner Product and Inner Product Spaces

Definition 8.1 (Inner product). *Let H be a complex vector space, then a bineary operation $\langle\cdot,\cdot\rangle : H\times H \to \mathbb{C}$ is called an **inner product** if it satisfies*

1. *(linearity w.r.t the first argument) $\forall x, y, z \in H, \alpha, \beta \in \mathbb{C}$ $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$;*
2. *(conjugate symmetry) $\forall x, y \in H$ $\langle x, y\rangle = \overline{\langle y, x\rangle}$;*
3. *(positiveness) $\forall x \in H$ $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0 \Leftrightarrow x = 0$.*

Definition 8.2 (Inner product space). *A vector space H together with an inner product defined thereon is called an **inner product space**.*

Theorem 8.1 (Cauchy-Schwartz inequality). *Let H be an inner product space and $x, y \in H$, then*

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

and the equality holds if and only if x and y are linearly dependent.

Proof. Without losing generality, we assume $y \neq 0$ through the entire proof.

$$0 \leq \langle x - \lambda y, x - \lambda y\rangle = \langle x, x\rangle + |\lambda|^2\langle y, y\rangle - 2\mathrm{Re}(\bar{\lambda}\langle x, y\rangle) \tag{8.1}$$

Let $\lambda = \langle x, y\rangle/\langle y, y\rangle$, then we have

$$\langle x, x\rangle + \frac{|\langle x, y\rangle|^2}{\langle y, y\rangle} - 2\frac{|\langle x, y\rangle|^2}{\langle y, y\rangle} \geq 0. \tag{8.2}$$

It follows that

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle. \tag{8.3}$$

The inequality is derived. Suppose x and y are linearly independent, we can write $x = \lambda y$. then

$$|\langle x, y\rangle|^2 = |\langle \lambda y, y\rangle|^2 = |\lambda|^2\langle y, y\rangle^2 \tag{8.4}$$

and

$$\langle x, x\rangle\langle y, y\rangle = \langle \lambda y, \lambda y\rangle\langle y, y\rangle = |\lambda|^2\langle y, y\rangle^2 \tag{8.5}$$

Combining them, we get the equality. For the other direction, we assume that the equality holds, then

$$\langle x, x\rangle + \frac{|\langle x, y\rangle|^2}{\langle y, y\rangle} - 2\frac{|\langle x, y\rangle|^2}{\langle y, y\rangle} = 0, \tag{8.6}$$

by letting $\lambda = \langle x, y\rangle/\langle y, y\rangle$, the equality can be written into

$$\frac{|\langle x, y\rangle|^2}{\langle y, y\rangle} - 2\frac{|\langle x, y\rangle|^2}{\langle y, y\rangle} = \langle x - \lambda y, x - \lambda y\rangle = 0. \tag{8.7}$$

It implies that $x - \lambda y = 0$. ∎

Proposition 8.1. *Let H be an inner product space and, then $||\cdot|| : H \to \mathbb{R}$ given by $||x|| = \langle x, x\rangle^{1/2}$ defines a norm on H, hence an inner product space can be considered as a normed space with the induced norm.*

Proof. The positiveness of $||\cdot||$ follows from the positiveness of inner product, while the $||\alpha x|| = |\alpha|\cdot||x||$ follows from the linearity and conjuate symmetry. It remains to show the triangle inequality.

$$||x+y||^2 = \langle x+y, x+y\rangle^2 = \langle x,x\rangle + \langle y,y\rangle + 2\mathrm{Re}\langle x,y\rangle = ||x||^2 + ||y||^2 + 2\mathrm{Re}\langle x,y\rangle. \tag{8.8}$$

By Cauchy-Schwartz's inequality

$$\mathrm{Re}\langle x,y\rangle \le |\langle x,y\rangle| \le \langle x,x\rangle^{1/2}\langle y,y\rangle^{1/2} = ||x||\cdot||y||. \tag{8.9}$$

Hence,

$$||x+y||^2 \le ||x||^2 + ||y||^2 + 2||x||\cdot||y|| = (||x|| + ||y||)^2. \tag{8.10}$$

The triangle inequality is derived. □

The following are some typical examples of Hilbert space.

1. $\mathbb{C}^n$: $\langle x,y\rangle = \sum_{i=1}^n x_i\bar{y}_i$;
2. l^2: $\langle x,y\rangle = \sum_{i=1}^\infty x_i\bar{y}_i$;
3. $L^2(X,\mu)$: $\langle f,g\rangle = \int_X f\bar{g}d\mu$.

Proposition 8.2. *The inner product is a continuous function of both arguments.*

Proof. Consider an inner product space H, let $(x_n)_{n=1}^\infty$ be a sequence converging to $x \in H$, and $(y_n)_{n=1}^\infty$ be a sequence converging to $y \in H$, it suffices to show that $\langle x_n, y_n\rangle \to \langle x,y\rangle$. This can be seen from

$$\begin{aligned} |\langle x_n, y_n\rangle - \langle x,y\rangle| &\le |\langle x_n,y_n\rangle - \langle x,y_n\rangle| + |\langle x,y_n\rangle - \langle x,y\rangle| \\ &= |\langle x_n - x, y_n\rangle| + |\langle x, \langle\rangle, y_n - y\rangle| \le ||x_n - x||\cdot||y_n|| + ||x||\cdot||y_n - y|| \to 0. \end{aligned} \tag{8.11}$$

Note that in the above deduction, we made use of the fact that $||y_n||$ is bounded. □

We have seen that every inner product space is a normed space (every inner product induces a norm), but the converse is in general not true, namely not every normed space can be an inner product space (not every norm can be induced by an inner product). The following theorem specifies the condition when the converse can be true.

Theorem 8.2. *Let $(X, ||\cdot||)$ be a complex normed space, then $||\cdot||$ can be induced by an inner product if and only if it satisfies the parallelogram law as follows*

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

under such condition, the inner product satisfies the following polarization identity.

$$\langle x,y\rangle = \frac{1}{4}\left(||x+y||^2 - ||x-y||^2 + i||x+iy||^2 - i||x-iy||^2\right).$$

Proof. For one direction, we assume that $||\cdot||$ is induced by an inner product, i.e. $||x|| = \langle x,x\rangle^{1/2}$ for every $x \in X$, and then prove the parallelogram law as follows.

$$\begin{aligned} ||x+y||^2 + ||x-y||^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\ &= ||x||^2 + ||y||^2 + 2\mathrm{Re}\langle x,y\rangle + ||x||^2 + ||y||^2 - 2\mathrm{Re}\langle x,y\rangle \\ &= 2(||x||^2 + ||y||^2). \end{aligned} \tag{8.12}$$

In addition, we have

$$||x+y||^2 - ||x-y||^2 = (||x||^2 + ||y||^2 + 2\mathrm{Re}\langle x,y\rangle) - (||x||^2 + ||y||^2 - 2\mathrm{Re}\langle x,y\rangle) = 4\mathrm{Re}\langle x,y\rangle, \tag{8.13}$$

and

$$(||x+iy||^2 - ||x-iy||^2) = (||x||^2 + ||y||^2 + 2\text{Re}\langle x, iy\rangle) - (||x||^2 + ||y||^2 - 2\text{Re}\langle x, iy\rangle) = 4\text{Im}\langle x, y\rangle. \tag{8.14}$$

Hence, the polarization identity is established.

For the other direction, the parallelogram law is assumed, we construct an inner product that induces the given norm. We define $\langle \cdot, \cdot \rangle$ by

$$\langle x, y\rangle = \frac{1}{4}\left(||x+y||^2 - ||x-y||^2 + i||x+iy||^2 - i||x-iy||^2\right). \tag{8.15}$$

One can verify that such a construction is an inner product. $\square$

Corollary 8.1. *An isometry T between Hilbert spaces preserves inner product, i.e. $\langle Tx, Ty\rangle = \langle x, y\rangle$.*

Proposition 8.3. *From theorem 8.2, we know that the inner product is uniquely determined by the norm, since an isometry preserves the norm, it also preserves the inner product.*

8.2 Orthogonality

Definition 8.3 (Orthogonality). *Let H be an inner product space and $x, y \in H$, then x and y are said to be **orthogonal** if $\langle x, y\rangle = 0$, denoted by $x \perp y$.*

We notation is often used for sets. Let A, B be subsets of H, and $x \in H$, then $x \perp A$ means $x \perp y$, $\forall y \in A$, and $A \perp B$ means $x \perp y$, $\forall x \in A$ $y \in B$.

Definition 8.4 (Orthonormal set). *Let H be an inner product space and $S \subset H$, then S is called an **orthogonal set** if the elements in S are mutually orthogonal, i.e. $x \perp y$, $\forall x, y \in S$, $x \neq y$. Particularly, if $||x|| = 1$, $\forall x \in S$, S is called an **orthonormal set**, or an **orthonormal system**.*

Theorem 8.3 (Pythagorean theorem). *Let H be an inner product space, and $x_1, \ldots, x_n \in H$ be mutually orthogonal, then*

$$||x_1 + \cdots + x_n||^2 = ||x_1||^2 + \cdots + ||x_n||^2.$$

Proof. The proof is simply done by expanding the inner product in the left hand side, and then applying the orthogonal condition to eliminate the mutual terms. $\square$

Corollary 8.2. *Let H be an inner product space, and $\{e_1, \ldots, e_n\}$ be an orthonormal set in H, then for any $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$, we have*

$$\left|\left|\sum_{i=1}^n \alpha_i e_i\right|\right|^2 = \sum_{i=1}^n |\alpha_i|^2.$$

Proposition 8.4. *A orthogonal set of non-zero vectors is linearly independent.*

Proof. Let H be an inner product space, and $S \subset H$ be an orthogonal set. Then it is obvious that any finite subset $\{x_1, \ldots, x_n\} \subset S$ is orthogonal. Given $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ with $\sum_{i=1}^n \alpha_i x_i = 0$, then by Pythagorean, we have

$$\left|\left|\sum_{i=1}^n \alpha_i x_i\right|\right| = \sum_{i=1}^n |\alpha_i|^2 ||x_i||^2 = 0, \tag{8.16}$$

implying that $\alpha_i = 0$ for each $i = 1, \ldots, n$. Hence, $x_1, \ldots, x_n$ are linearly independent. As the linear independence holds for any finite subset of S, S is linearly independent. $\square$

Theorem 8.4 (Bessel's Inequality). *Let $(e_n)_{n=1}^\infty$ be an orthonormal sequence in an inner product space H, then for any $x \in H$, we have*

$$\sum_{n=1}^\infty |\langle x, e_n\rangle|^2 \leq ||x||^2. \tag{8.17}$$

(Remark: this inequality also holds for finite sequence).

Proof. Let $y_n = \sum_{i=1}^n \langle x, e_i\rangle e_i$, and $z_n = x - y_n$. Then $||y||^2 = \sum_{i=1}^n |\langle x, e_i\rangle|^2$. In addition, we for each $i \le n$ we have

$$\langle z_n, e_i\rangle = \langle x, e_i\rangle - \sum_{j=1}^n (\langle x, e_j\rangle) e_j e_i = \langle x, e_i\rangle - \langle x, e_j\rangle = 0, \tag{8.18}$$

thus $e_i \perp z_n$, which follows that $y_n \perp z_n$, and hence by Pythagorean, we have

$$\sum_{i=1}^n |\langle x, e_i\rangle|^2 = ||y_n||^2 = ||x||^2 - ||z_n||^2 \le ||x||^2. \tag{8.19}$$

As this holds for any $n \in \mathbb{N}^+$, we can conclude that

$$\lim_{i\to\infty} \sum_{i=1}^\infty |\langle x, e_i\rangle|^2 \le ||x||^2. \tag{8.20}$$

□

Theorem 8.5 (Gram-Schmidt orthonormalization). *Let $\{x_n\}_{n=1}^\infty$ be a linearly independent subset of an inner product space H, then we let*

$$\begin{aligned} e_1 &= \frac{x_1}{||x_1||} \\ r_n &= x_n - \sum_{i=1}^{n-1} \langle x_n, e_i\rangle e_i, \ \forall n > 1 \\ e_n &= \frac{r_n}{||r_n||} \end{aligned}$$

Then $\{e_n\}_{n=1}^\infty$ be an orthonormal set, and for each n, $\operatorname{span}\{e_i\}_{i=1}^n = \operatorname{span}\{x_i\}_{i=1}^n$, and $\operatorname{span}\{e_i\}_{i=1}^\infty = \operatorname{span}\{x_i\}_{i=1}^\infty$. In particular if $\{x_n\}_{n=1}^\infty$ is a basis of H, then $\{e_i\}_{i=1}^\infty$ is also a basis of H.

Proof. We first need to show that the procedure is well defined, i.e. $r_n \neq 0$ for each $n > 1$, which directly follows the linear independence assumption. And, $||e_n|| = 1$, $\forall n$ directly follows from the normalization steps. To show that $\{e_n\}$ is orthogonal, it suffices to show that $\langle r_n, e_i\rangle = 0$ for every $i < n$. This can be seen from

$$\langle r_n, e_i\rangle = \langle x_n, e_i\rangle - \sum_{j=1}^{n-1} \langle x_n, e_i\rangle \langle e_j, e_i\rangle = \langle x_n, e_i\rangle - \langle x_n, e_i\rangle = 0. \tag{8.21}$$

Subsequently, we show that $\operatorname{span}\{e_i\}_{i=1}^n = \operatorname{span}\{x_i\}_{i=1}^n$. This can be easily shown by induction and the construction rules. When this is proved, $\operatorname{span}\{e_i\}_{i=1}^\infty = \operatorname{span}\{x_i\}_{i=1}^\infty$ immediately follows due to the definition of span. □

8.3 Hilbert Spaces and Closed Subspaces

Definition 8.5 (Hilbert space). *A complete inner product space is called a **Hilbert space**.*

Theorem 8.6 (Projection theorem). *Let H be a Hilbert space, and E be a closed convex subset of H, then for every $x \in H$, there exists a unique $y \in E$ such that*

$$||x - y|| = \inf_{z\in E} ||x - z||.$$

*Here, y is called the **projection** of x onto E, denoted by $y = P_E x$. In particular, if C is a closed subspace of H (C is obviously convex), then $x - P_E x \perp E$. Conversely, if there is a point $y \in E$ such that $x - y \perp E$, then y is unique and is given by $y = P_E x$.*

Proof. Given x and E, let $d = \inf_{z\in E} ||x - z||$. Then, we can find a sequence $(z_n)_{n=1}^\infty$ such that $||x - z_n|| \to d$. We claim that (z_n) is a Cauchy sequence, which is shown as follows. Given $\epsilon > 0$, we can find N such that

$||x - z_n||^2 < d^2 - \epsilon/2$ for any $n > N$. Let $m, n > N$ and $v = (z_m + z_n)/2$, by convexity of E, $v \in E$, thus $||x - v|| \geq d$. According to parallelogram law, we have

$$||z_m - z_n||^2 = 2(||x - z_m||^2 + ||x - z_n||^2) - 4||x - v||^2 < 4d^2 + \epsilon - 4d^2 = \epsilon. \tag{8.22}$$

It implies that (z_n) is a Cauchy sequence, and since E is closed in a complete space, z_n converges to some $y \in E$, thus $||x - y|| = d$.

In the following, we show that such y is unique. Let $y_1, y_2 \in E$ with $||x - y_1|| = ||x - y_2|| = d$, we have $v = (y_1 + y_2)/2 \in E$ by convexity of E. Hence, by parallelogram, we have

$$||y_1 - y_2||^2 = 2(||x - y_1||^2 + ||x - y_2||^2) - 4||x - v||^2 \leq 4d^2 - 4d^2 = 0. \tag{8.23}$$

Hence $y_1 = y_2$. The uniqueness is shown.

Let E be a closed subspace, and $x \in E$ and $y = P_E x$, if $x - y \perp E$ does NOT hold, then we can find $v \in E$ such that $\langle x - y, v\rangle \neq 0$. Choose $\alpha = \frac{\langle x-y,v\rangle}{||v||^2}$, then we have

$$||x - (y + \alpha v)||^2 = ||x - y||^2 - |\alpha|^2|\langle x - y, v\rangle|^2 < ||x - y||^2. \tag{8.24}$$

Note that $y + \alpha v \in E$, this result contradicts the assumption that $||x - y||$ attains the minimum. Hence, we can conclude that $x - y \perp E$.

Finally, we prove that the $y \in E$ that satisfies $x - y \perp E$ must have $y = P_E x$. For such a y we on one hand have $||x - y|| \geq ||x - P_E x||$, and on the other hand, by Pythagorean

$$||x - y||^2 = ||x - P_E x||^2 - ||P_E x - y||^2, \tag{8.25}$$

hence the only possibility is that $||P_E x - y|| = 0$, thus $y = P_E x$. □

Proposition 8.5. *Let E be a closed subspace of a Hilbert space H, then the following conditions are equivalent*

1. $x \in E$;
2. $x = P_E x$;
3. $||x||^2 = \langle x, P_E x\rangle$;
4. $||x|| = ||P_E x||$.

Proof. 1. (1) ⇔ (2). Since $x \in E$, it is obvious that $\inf_{z\in E} ||x - z||$ attains minium at $z = x$, thus $P_E x = x$. Since $P_E x \in E$, the other direction is obvious.

2. (2) ⇒ (3).
$$\langle x, x\rangle = \langle x, P_E x + (x - P_E x)\rangle = \langle x, P_E x\rangle + \langle x, x - P_E x\rangle = \langle x, P_E x\rangle. \tag{8.26}$$

3. (3) ⇒ (4).
$$||x - P_E x||^2 = ||x||^2 + ||P_E x||^2 - 2\mathrm{Re}\langle x, P_E x\rangle = ||x||^2 + ||P_E x||^2 - 2||x||^2 = ||P_E x||^2 - ||x||^2 \geq 0. \tag{8.27}$$
Hence $||x||^2 = ||P_E x||^2$.

4. (4) ⇒ (1), (2). By Pythagorean and projection theorem, we have
$$||x||^2 = ||P_E x||^2 + ||x - P_E x||^2, \tag{8.28}$$
hence $||x - P_E x|| = 0$, which implies (2), and equivalently (1).

□

Proposition 8.6. *Let H be an inner product space, and $S \subset H$ be any subset, define $S^\perp = \{y | y \perp x, \forall x \in S\}$. Then $S^\perp$ is a closed subspace. And $S \subseteq (S^\perp)^\perp$.*

Proof. Let $x, y \in S^\perp$, $\alpha, \beta \in \mathbb{C}$, then for every $z \in S$, we have

$$\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle = 0, \tag{8.29}$$

which implies that $\alpha x + \beta y \in S^\perp$, hence $S^\perp$ is a subspace of H. In addition, let (x_n) be a sequence in $S^\perp$ that converges to $x \in H$, then for every $y \in S$, by continuity of inner product,

$$\langle x, y\rangle = \lim_{n\to\infty} \langle x_n, y\rangle = \lim_{n\to\infty} 0 = 0, \tag{8.30}$$

which follows that $S^\perp$ is closed. On the other hand,

$$x \in S \Rightarrow x \perp y \ \forall y \in S^\perp \Rightarrow x \in (S^\perp)^\perp, \tag{8.31}$$

thus $S \subseteq (S^\perp)^\perp$. □

Theorem 8.7 (Orthogonal complement)**.** *Let E be a closed subspace of a Hilbert space H, and $E^\perp = \{y | y \perp x \ \forall x \in E\}$, then*

1. *$H = E \oplus E^\perp$;*
2. *$(E^\perp)^\perp = E$.*

*Here, $E^\perp$ is called the **orthogonal complement** of E in H.*

Proof. 1. ($H = E \oplus E^\perp$). Given $x \in E \cap E^\perp$, by definition, $x \perp x$, hence $x = 0$, which implies that $E \cap E^\perp = \{0\}$. Let $x \in H$, then by projection theorem, we can write $x = P_E x + (x - P_E x)$, in which $P_E x \in E$ and $x - P_E x \in E^\perp$, hence we can conclude that $H = E \oplus E^\perp$.

2. ($(E^\perp)^\perp = E$). From proposition 8.6, we know that $E \subseteq (E^\perp)^\perp$. For the other direction, let $x \in (E^\perp)^\perp$. Note that $x - P_E x \in E^\perp$, then $\langle x, x - P_E x\rangle = 0$, implying that $||x||^2 = \langle x, P_E x\rangle$, hence $x \in E$.

□

Proposition 8.7. *Let E_1 and E_2 be two subspaces of H, then $E_1 \subseteq E_2 \Rightarrow E_2^\perp \subseteq E_1^\perp$. In particular, when both E_1 and E_2 are closed, then the converse also holds, i.e. $E_1 \subseteq E_2 \Leftrightarrow E_2^\perp \subseteq E_1^\perp$.*

Proof. For one direction, if $E_1 \subseteq E_2$, then $y \perp x, \forall x \in E_2 \ \Rightarrow \ y \perp x, \forall x \in E_1$, then $E_2^\perp \subseteq E_1^\perp$. For the other direction, when E_1 and E_2 are both closed, we have $E_1 = E_1^\perp$ and $E_2 = E_2^\perp$, hence

$$E_2^\perp \subseteq E_1^\perp \Rightarrow (E_1^\perp)^\perp \subseteq (E_2^\perp)^\perp \Rightarrow E_1 \subseteq E_2. \tag{8.32}$$

□

8.4 Complete Orthonormal Systems

Definition 8.6 (Convergent series)**.** *Let X be a normed space and $(x_i)_{i=1}^\infty$ be a sequence in X, then the **partial sum** of (x_i) is defined by $s_n = \sum_{i=1}^n x_i$. If the sequence of partial sums (s_n) converges to $s \in X$, then we say that the **infinite series** $\sum_{i=1}^\infty x_i$ **converges to** s. This can be written as*

$$s = \sum_{i=1}^\infty x_i.$$

Definition 8.7 (Absolute convergence)**.** *Let X be a normed space and $(x_i)_{i=1}^\infty$ be a sequence in X, then the infinite series $\sum_{i=1}^\infty x_i$ is said to be **absolutely convergent** if the infinite series $\sum_{i=1}^\infty ||x_i||$ converges.*

The following theorem states the relation between absolute convergence and convergence of an infinite series.

Theorem 8.8. *Let X be a normed space and $(x_i)_{i=1}^\infty$ be a sequence in X, then if the infinite series $\sum_{i=1}^\infty x_i$ is absolutely convergent, then its partial sum is a Cauchy sequence. In paricular, if X is a Banach space, then it converges.*

Proof. Suppose $\sum_{i=1}^{\infty} x_i$ is absolutely convergent. Let $a_n = \sum_{i=1}^{n} ||x_i||$, then a_n is a bounded monotone sequence, and thus it converges to some $a \in [0, +\infty)$. Given $\epsilon > 0$, we can find N such that $a - a_n = \sum_{i=n+1}^{\infty} ||x_i|| < \epsilon$, $\forall n > N$. Let $s_n = \sum_{i=1}^{n} x_i$ be the partial sum of the given sequence. Then, for any $m > n > N$, we have

$$||s_m - s_n|| = \left|\left|\sum_{i=n+1}^{m} x_i\right|\right| \leq \sum_{i=n+1}^{m} ||x_i|| \leq a - a_n < \epsilon, \tag{8.33}$$

which implies that (s_n) is a Cachy sequence. When X is complete, it converges. □

Theorem 8.9. *Let H be a Hilbert product space, and $(e_i)_{i=1}^{\infty}$ be an orthonormal sequence in H, then*

$$\lim_{n\to\infty} \left|\left|\sum_{i=1}^{n} \alpha_i e_i\right|\right| = \left(\sum_{i=1}^{\infty} |a_i|^2\right)^{1/2},$$

if the right hand side is finite, then the infinite series $\sum_{i=1}^{\infty} \alpha_i e_i$ converges and

$$\left|\left|\sum_{i=1}^{\infty} \alpha_i e_i\right|\right| = \left(\sum_{i=1}^{\infty} |a_i|^2\right)^{1/2}.$$

Proof. By Pythagorean theorem, we have

$$\left|\left|\sum_{i=1}^{n} \alpha_i e_i\right|\right|^2 = \sum_{i=1}^{n} |\alpha_i|^2. \tag{8.34}$$

Both hand side are increasing sequences, thus

$$\lim_{n\to\infty} \left|\left|\sum_{i=1}^{n} \alpha_i e_i\right|\right|^2 = \lim_{n\to\infty} \sum_{i=1}^{n} |\alpha_i|^2. \tag{8.35}$$

By continuity $x \mapsto x^{1/2}$ for $x \geq 0$, we have

$$\lim_{n\to\infty} \left|\left|\sum_{i=1}^{n} \alpha_i e_i\right|\right| = \left(\lim_{n\to\infty} \sum_{i=1}^{n} |\alpha_i|^2\right)^{1/2} = \left(\sum_{i=1}^{\infty} |\alpha_i|^2\right)^{1/2}. \tag{8.36}$$

If the right hand side is finite, then the infinite series $\sum_{i=1}^{n} \alpha_i e_i$ is absolutely convergent, there is $s \in H$ with $s = \sum_{i=1}^{\infty} \alpha_i e_i$. By continuity of norm, we have

$$||s|| = ||\lim_{n\to\infty} s_n|| = \lim_{n\to\infty} ||s_n|| = \left(\sum_{i=1}^{\infty} |\alpha_i|^2\right)^{1/2}. \tag{8.37}$$

□

Definition 8.8 (Complete system)**.** *Let X be a normed space, then a subset S is called a* ***complete system*** *if $\overline{\operatorname{span} S} = X$.*

We give several examples of complete systems in Hilbert spaces

1. $\{x \mapsto x^n\}_{n=0}^{\infty}$ is a complete system in $(C[0,1], ||\cdot||_\infty)$ and in $L^2[0,1]$.
2. $\{x \mapsto 1\} \cup \{x \mapsto \cos(nx)\}_{n=1}^{\infty} cup \{x \mapsto \sin(nx)\}_{n=1}^{\infty}$ is a complete system in $L^2[-\pi, \pi]$.

Definition 8.9 (Complete orthonormal system)**.** *An orthonormal set in a Hilbert space that is also a complete system is called a* ***complete orthonormal system****.*

Theorem 8.10. *Let S be a subset of Hilbert space H, then S is a complete system if and only if $x \perp S$ (i.e. $x \perp y, \forall y \in S$) implies $x = 0$.*

Proof. For one direction, we assume that S is a complete system in H. and $x \perp y, \forall y \in S$. Since $\overline{\operatorname{span} S} = H$, we can find a sequence (x_n) in $\operatorname{span} S$ such that $x_n \to x$. By continuity of inner product and the fact that $x \perp s$ for every $s \in \operatorname{span} S$, then

$$\langle x, x \rangle = \lim_{n\to\infty} \langle x_n, x \rangle = \lim_{n\to\infty} 0 = 0. \tag{8.38}$$

It follows that $x = 0$.

For the other direction, let $E = \overline{\operatorname{span} S}$, it suffices to show that $H \subseteq E$. Let $x \in H$, then $x = P_E x + (x - P_E x)$. Clearly, $x - P_E x \perp S$, hence $x - P_E x = 0$, thus $x = P_E x \in E$, implying that $H \subseteq E$. □

In discussing infinite dimensional normed spaces, we use the following notion of basis, which is a generalization of the basis in finite dimensional space.

Definition 8.10 (Basis in normed space)**.** *Let X be a (infinite dimensional) normed space, then $\{e_i\}_{i=1}^\infty$ is called a* ***basis*** *of X if for each $x \in X$, there exists a unique sequence $(\alpha_i)_{i=1}^\infty$ in $\mathbb{C}$ such that such that $x = \sum_{i=1}^\infty \alpha_i e_i$.*

Theorem 8.11. *A complete orthonormal system of a Hilbert space is a basis. Concretely, let H be a Hilbert space with a complete orthonormal system $\{e_i\}_{i=1}^\infty$, then for each $x \in H$, there is a unique sequence $(\alpha_i)_{i=1}^\infty \in l^2$ such that $x = \sum_{i=1}^\infty \alpha_i e_i$, and $\alpha_i = \langle x, e_i \rangle$. The result can be written as*

$$x = \sum_{i=1}^\infty \langle x, e_i \rangle e_i.$$

Proof. From Bessel's inequality, we know that for each $x \in H$

$$\sum_{i=1}^\infty |\langle x, e_i \rangle|^2 \le ||x||^2, \tag{8.39}$$

hence the infinite series $\sum_{i=1}^\infty \langle x, e_i \rangle e_i$ is absolutely convergent and thus converges. Let $y = \sum_{i=1}^\infty \langle x, e_i \rangle e_i$, then $y - x \perp e_i$ for each e_i. Since $\{e_i\}_{i=1}^\infty$ is a complete system, by theorem 8.10, we have $y - x = 0$, thus $y = x$. This shows that every $x \in H$ can be expressed in form of the given infinite series.

In the following, we show that such an expression is unique. Suppose $x = \sum_{i=1}^\infty \alpha_i e_i$, then we have

$$z = \sum_{i=1}^\infty (\alpha_i - \langle x, e_i \rangle) e_i = 0. \tag{8.40}$$

and for each i,

$$\alpha_i - \langle x, e_i \rangle = \langle z, e_i \rangle. \tag{8.41}$$

Combining the above two leads to $\alpha_i = \langle x, e_i \rangle$ for each i. □

Theorem 8.12 (Parseval's identity)**.** *Let $\{e_i\}_{i=1}^n$ be an orthonormal set in a Hilbert space H, then it is complete if and only if the following Parseval's identity is satisfied for every $x \in H$*

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

Proof. If $\{e_i\}_{i=1}^\infty$ is complete, then it is a basis, and thus for each $x \in H$, we have

$$x = \sum_{i=1}^\infty \langle x, e_i \rangle e_i. \tag{8.42}$$

By theorem 8.9, we get

$$||x||^2 = \sum_{i=1}^\infty |\langle x, e_i \rangle|^2. \tag{8.43}$$

The parseval's identity is thus established. For the other direction, the parseval's identity is assumed. let $E = \overline{\operatorname{span}\{e_i\}_{i=1}^\infty}$, then given $x \in H$, by Pythagorean, we have

$$||x - P_E x|| = ||x||^2 - ||P_E x||^2 = ||x||^2 - ||x||^2 = 0. \tag{8.44}$$

It implies that $x = P_E x \in E$. Hence, the given set is a complete system. □

Up to now, we have derived several theorems in characterizing a complete orthonormal system. They are summarized by the following theorem.

Theorem 8.13 (Characterization of complete orthonormal system)**.** *Let $\{e_i\}_{i=1}^{\infty}$ be an orthonormal set in a Hilbert space H, then the following statements are equivalent:*

1. *$\{e_i\}_{i=1}^{\infty}$ is a complete system (i.e. $\overline{\{e_i\}_{i=1}^{\infty}} = H$);*
2. *$\langle x, e_i \rangle = 0 \ \forall i \Rightarrow x = 0$;*
3. *$\{e_i\}_{i=1}^{\infty}$ is a basis of H (i.e. $\forall x \in H \ x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$);*
4. *parseval's identity is satisfied (i.e. $\forall x \in H \ ||x||^2 = \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2$).*

Not every Hilbert space admits a complete orthonormal sequence, the following proposition gives an equivalent condition.

Proposition 8.8. *A Hibert space has a complete orthonormal sequence if and only if it is separable (i.e. it has a countable dense set).*

Proof. Let H be a Hilbert space, then if H has a complete orthonormal sequence $(e_i)_{i=1}^{\infty}$, we let $S_n = \{\sum_{i=1}^{n} \alpha_i e_i | \alpha_i \in \mathbb{Q}\}$, then it is easy to show that $\bigcup_{n=1}^{\infty} S_n$ is a dense set in H.

For the other direction, if H is separable, there is a countable dense, denoted by $S = s_{i}{}_{i=1}^{\infty}$, from which we can select a maximal linearly independent subset by induction, and by applying Gram-Schmidt orthonormalization to this subset, we can obtain a complete orthonormal system. □

Theorem 8.14. *Every infinite dimensional separable Hilbert spaces are isometric. In particular, every infinite dimensional separable Hilbert space is isometric to l^2.*

Proof. Let H be an infinite dimensional separable Hilbert space, we can choose an orthonormal basis $\{e_i\}_{i=1}^{\infty}$. Then we can define $T : H \to l^2$ by $Tx = (\langle x, e_i \rangle)_{i=1}^{\infty}$. By theorem 8.9(for showing onto) 8.11(for showing one-to-one) and parseval's identity (for showing norm-perserving), we can prove that T is an isometry. Hence H is isometric to l^2. Therefore, all infinite dimensional separable Hilbert spaces are isometric. □

Chapter 9

Linear Functionals

9.1 The Space of Linear Functionals

All vector spaces in this notes are real or complex vector spaces.

Definition 9.1 (Linear functional)**.** *Let X be a vector space over the field $\mathbb{F}$ ($\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$), then a linear map $f : X \to \mathbb{F}$ is called a **linear functional**.*

In other words, a linear functional is a linear map whose range is in $\mathbb{R}$ or $\mathbb{C}$.

Definition 9.2 (The space of linear functionals)**.** *All linear functionals defined on a vector space X forms a vector space, called **the space of linear functionals**, denoted by $X^{\#}$. The addition and scalar multiplication are defined as*

$$(f + g)(x) = f(x) + g(x) \qquad \text{and} \qquad (\alpha f)(x) = \alpha f(x).$$

The following are two examples of linear functionals:

1. Given $f \in C[0, 1]$, $F : C[0, 1] \to \mathbb{R}$ defined by $F(x) = \int_0^1 x(t)f(t)dt$ is a linear functional;
2. Let H be a inner product space, given $y \in H$, then $f : H \to \mathbb{C}$ defined by $f(x) = \langle x, y\rangle$ is a linear functional.

Lemma 9.1. *Let $f \neq 0$ be a linear functional on X, then and $x_0 \in X$ satisfy $f(x_0) \neq 0$. Then for each $x \in X$, it can be uniquely written into $x = y + \lambda x_0$, where $y \in \ker f$. In other words, $X = \ker f \oplus \operatorname{span}\{x_0\}$.*

Proof. Since $f \neq 0$, there is $v \in X$ such that $f(x_0) \neq 0$. Then, we can write

$$x = (x - f(x)x_0) + f(x)x_0 = \left(x - \frac{f(x)}{f(x_0)}x_0\right) + \frac{f(x)}{f(x_0)}x_0. \tag{9.1}$$

Note that

$$f\left(x - \frac{f(x)}{f(x_0)}x_0\right) = f(x) - \frac{f(x)}{f(x_0)}f(x_0) = 0. \tag{9.2}$$

For uniqueness, if there are $y_1, y_2 \in \ker f$ and $\lambda_1, \lambda_2 \in \mathbb{F}$ with $y_1 + \lambda_1 x_0 = y_2 + \lambda_2 x_0$, then $(\lambda_1 - \lambda_2)x_0 = y_2 - y_1 \in \ker f$, thus $f((\lambda_1 - \lambda_2)x_0) = 0$, since $f(x_0) \neq 0$, we have $\lambda_1 = \lambda_2$, as a result, $y_1 = y_2$. ☐

Corollary 9.1. *Let f be a non-zero linear functional on X, then*

$$\operatorname{codim}(\ker f) = 1.$$

Corollary 9.2. *Let f, g be non-zero linear functionals on X, then $\ker f = \ker g$ if and only if $f = \lambda g$ for some $\lambda \neq 0$.*

Proof. First, assume $f = \lambda g$ for some $\lambda \neq 0$. Then $f(x) = 0 \Leftrightarrow g(x) = 0$, it implies that $\ker f = \ker g$. For the other direction, assume $\ker f = \ker g = E$. We can choose $x_0 \in E$. By the lemma above, we know that for each $x \in X$, we can write $x = y + \gamma x_0$ with $y \in E$. Then, $f(x) = \gamma f(x_0)$ and $g(x) = \gamma g(x_0)$. Let $\lambda = f(x_0)/g(x_0)$, we know $f(x) = \lambda g(x)$. Note that λ is a constant independent of x. ☐

9.2 Dual Spaces

9.2.1 Bounded Linear Functionals

Definition 9.3 (Bounded linear functional)**.** *Let X be a normed space, then a linear functional f on X is called a **bounded linear functional** if there exists $c > 0$ such that $|f(x)| \le c||x||$, $\forall x \in X$.*

Lemma 9.2. *Let f be a linear functional on a normed space X, then f is continuous if and only if f is continuous at* 0.

Proof. The "only if" direction is trivial. For "if" part, let $x \in X$, for any sequence (x_n) that converges to x, we have $x_n - x$ converges to 0. Suppose f is continuous at 0, then $f(x_n) - f(x) = f(x_n - x) \to 0$, implying that $f(x_n) \to f(x)$, thus f is continuous on X. □

Theorem 9.1. *Let f be a linear functional on a normed space X, then f is continuous if and only if f is bounded.*

Proof. Suppose f is continuous, then there exists $\delta > 0$, such that $|f(x)| < 1$ whenever $||x|| \le \delta$. Then, for each $x \in X$ and $x \neq 0$, we let $u = \frac{\delta}{||x||}x$, then $||u|| = \delta$ and

$$|f(x)| = \frac{||x||}{\delta}|f(u)| \le \frac{1}{\delta}||x||. \tag{9.3}$$

Hence, f is bounded. For the other direction, assume there is $M > 0$ such that $|f(x)| \le M||x||$ for each $x \in X$. Let (x_n) be a sequence that converges to x, then $|f(x_n) - f(x)| = |f(x_n - x)| \le M||x_n - x|| \to 0$, which means that $f(x_n) \to f(x)$. □

Proposition 9.1. *Let f be a linear functional on a normed space X, then f is bounded if and only if* $\ker f$ *is closed.*

Proof. If f is bounded, then it is continuous, we immediately have $\ker f = f^{-1}\{0\}$ is closed, since $\{0\}$ is closed. On the other hand, if f is unbounded. Then we can find a sequence $(x_n)_{n=1}^{\infty}$ with $||x_n|| = 1$ and $f(x_n) > n$ for each n. Let $y_n = x_n/f(x_n)$, then $f(y_n) = 1$ and $||y_n|| < 1/n$. Hence $y_n \to 0$. Let $z_n = y_n - y_1$, then $f(z_n) = 0$, i.e. $z_n \in \ker f$, and $z_n \to -y_1$, where $f(-y_1) = -1$. It shows that there exists a sequence in $\ker f$ that converges to some point outside $\ker f$. Hence, $\ker f$ is not closed. □

Lemma 9.3. *Let E be a closed subspace of a normed space X and $x_0 \in X \cap E^c$ such that $X = E \oplus \operatorname{span}\{x_0\}$. Let $d = \inf_{y\in E} ||x_0 - y||$, then there exists a bounded linear functional f on E such that $\ker f = E$, $||f|| = 1$, and $f(x_0) = d$.*

Proof. Since E is closed, E^c is open. Hence, there is a open ball $B(x_0, \epsilon) \subset E^c$ for some $\epsilon > 0$, and thus $d > \epsilon > 0$. Since $X = E \oplus \operatorname{span}\{x_0\}$, each $x \in X$ can be written uniquely as $x = y + \alpha x_0$. Then we can define a functional f_0 on X by $f_0(x) := \alpha$. Due to the uniqueness of decomposition, f_0 is well-defined. The linearity of f_0 can be seen from

$$(\alpha x_1 + \alpha x_2) = \alpha_1(y_1 + \lambda_1 x_0) + \alpha_2(y_2 + \lambda_2 x_0) = (\alpha_1 y_1 + \alpha_2 y_2) + (\alpha_1\lambda_1 + \alpha_2\lambda_2)x_0 \tag{9.4}$$

It means that $f_0(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1\lambda_1 + \alpha_2\lambda_2 = \alpha_1 f_0(x_1) + \alpha_2 f_0(x_2)$. In addition, we can easily see that $f_0(x_0) = 1$ and $f_0(y) = 0$, $\forall y \in E$, implying that $\ker f = E$. In the following, we will show that f_0 is a bounded functional with $||f_0|| = 1/d$. For each $x \in X$ and $x \notin E$, there exists $y \in E$ and $\alpha \neq 0$ such that

$$x = y + \alpha x_0 = \alpha(x_0 - y'), \tag{9.5}$$

where $y' = -y/\alpha \in E$. Hence,

$$|f_0(x)| = |\alpha| \le \frac{||x||}{||x_0 - y'||} \le \frac{1}{d}||x||. \tag{9.6}$$

In addition, when $x \in E$, $f(x) = 0 \le (1/d)||x||$. The above shows that f_0 is bounded with $||f_0|| \le 1/d$. To complete the proof, it remains to show the other direction, namely $||f_0|| \ge 1/d$. By definition of d, we can choose a sequence (y_n) in E such that $||x_0 - y_n|| \to d$. And, it can be readily seen that $f_0(x_0 - y_n) = 1$ for each n. As a result,

$$||f_0|| \ge \sup_n \frac{|f_0(x_0 - y_n)|}{||x_0 - y_n||} \ge \lim_{n\to\infty} \frac{1}{||x_0 - y_n||} = \frac{1}{d}. \tag{9.7}$$

Take $f = df_0$, then we are done. □

Proposition 9.2. *Let E be a closed subspace of a normed space X with* $\operatorname{codim}(E) = 1$*, then there exists a bounded linear functional f with* $\ker f = E$*.*

Proof. Since $\operatorname{codim}(E) = 1$, there exists a subspace V with $\dim(V) = 1$ such that $X = E \oplus V$. Then, there exists $x_0 \in X \cap E^c$ such that $V = \operatorname{span}\{x_0\}$. By lemma 9.3, we know there is a bounded linear functional f on X such that $\ker f = E$. □

To sum up, for a linear functional f on a normed space, the following statements are equivalent:

1. f is bounded;
2. f is continuous;
3. $\ker f$ is closed.

9.2.2 Dual Space

Proposition 9.3. *All bounded linear functionals on X form a linear subspace of $X^{\#}$.*

Proof. Let f and g be bounded linear functionals on X, and $|f(x)| \leq A||x||$, $|g(x)| \leq B||x||$ for every $x \in X$. Then, we have for each x

$$|(f+g)(x)| = |f(x)+g(x)| \leq |f(x)| + |g(x)| \leq A||x|| + B||x|| = (A+B)||x|| \tag{9.8}$$

and

$$|(\alpha f)(x)| = |\alpha f(x)| \leq (|\alpha| A)||x|| \tag{9.9}$$

Hence, $f+g$ and αf are both bounded linear functionals. □

Proposition 9.4. *Let $||\cdot||$ be a real valued functions defined on all bounded linear functionals, which is given by*

$$||f|| = \sup_{x \neq 0} \frac{|f(x)|}{||x||} = \sup\{|f(x)| : ||x|| = 1\}.$$

Then $||\cdot||$ is a norm on the space of all bounded linear functionals.

Proof. The non-negativeness of $||\cdot||$ directly follows from the definition. And

$$||\alpha f|| = \sup\{|\alpha f(x)| : ||x|| = 1\} = \sup\{|\alpha| \cdot |f(x)| : ||x|| = 1\} = |\alpha| \sup\{|f(x)| : ||x|| = 1\} = |\alpha| \cdot ||f||. \tag{9.10}$$

For sub-additivity, we have

$$\begin{aligned} ||f+g|| &= \sup\{|f(x)+g(x)| : ||x|| = 1\} \leq \sup\{|f(x)| + |g(x)| : ||x|| = 1\} \\ &= \sup\{|f(x)| : ||x|| = 1\} + \sup\{|g(x)| : ||x|| = 1\} \\ &= ||f|| + ||g||. \end{aligned} \tag{9.11}$$

In addition, when $||f|| = 0$, we have $|f(x)| = 0$ for all $x \neq 0$, hence $f = 0$. □

Definition 9.4. *Let X be a normed vector space. The vector space of all bounded linear functionals on X endowed with the norm defined above is called the* ***dual space*** *of X, denoted by X^*.*

Proposition 9.5. *Let X be a normed space, and $f \in X^*$, then*

$$|f(x)| \leq ||f|| \cdot ||x||, \ \forall x \in X.$$

Proposition 9.6. *Let X be a normed space, and $f \in X^*$ with $f \neq 0$, then*

$$\inf\{||x|| : f(x) = 1\} = \frac{1}{||f||}.$$

Proof. For every x with $f(x) = 1$, we have $||f|| \cdot ||x|| \geq |f(x)| = 1$, thus $||x|| \geq 1/||f||$, hence

$$\inf\{||x|| : f(x) = 1\} \geq 1/||f||. \tag{9.12}$$

By definition of $||f||$, we can choose a sequence (x_n) with $||x_n|| = 1$, $\forall n$ and $|f(x_n)| \to ||f||$. Without losing generality, we assume $f(x_n) \neq 0$ for all n. Let $y_n = x_n/f(x_n)$, then $f(y_n) = 1$, $\forall n$. and $||y_n|| = ||x_n||/|f(x_n)| \to 1/||f||$. Hence

$$\inf\{||x|| : f(x) = 1\} \leq \inf_{n\in\mathbb{N}^+} ||y_n|| \leq \lim_{n\to\infty} ||y_n|| = 1/||f||. \tag{9.13}$$

Combining the two inequalities together, we can get the equality. □

Theorem 9.2. *Let X be a normed vector space, then its dual space X^* is a Banach space.*

Proof. Consider a Cauchy sequence (f_n) in X^*, then for each $x \in X$, we have $||f_n(x) - f_m(x)|| \leq ||f_n - f_m|| \cdot ||x||$, hence $(f_n(x))$ is a Cauchy sequence. By completeness of a real field or a complex field, the limit of $(f_n(x))$ exists. Then we can define a functional f by

$$f(x) := \lim_{n\to\infty} f_n(x).$$

It is easy to verify that f is linear (due to linearity of limit operation). In addition, since (f_n) is Cauchy, there is N such that $||f_n - f_m|| < \epsilon$ when $n, m > N$, thus for each x, and $n > N$

$$|f(x) - f_n(x)| = |\lim_{m\to\infty} f_m(x) - f_n(x)| = \lim_{m\to\infty} |f_m(x) - f_n(x)| \leq \lim_{m\to\infty} ||f_m - f_n|| \cdot ||x|| \leq \epsilon||x||. \tag{9.14}$$

Hence, $f - f_n \in X^*$, thus $f = (f - f_n) + f_n \in X^*$. Then, we can conclude that X^* is complete. □

Proposition 9.7. *Let X be a normed space, and (f_n) be a sequence in X^* that converges to $f \in X^*$, then for each $x \in X$, we have*

$$\lim_{n\to\infty} f_n(x) = f(x).$$

Proof. This can be seen from $|f(x) - f_n(x)| \leq ||f_n - f|| \cdot ||x|| \to 0$. □

Proposition 9.8. *Let X be a normed space, and (f_n) be a sequence in X^* that converges to $f \in X^*$, and (x_n) be a sequence in X that converges to $x \in X$, then*

$$\lim_{n\to\infty} f_n(x_n) = f(x).$$

Proof. Since (x_n) is convergent, it is bounded. Let $M = \sup_n ||x_n||$, then

$$|f(x) - f_n(x_n)| \leq |f(x) - f(x_n)| + |f(x_n) - f_n(x_n)| \leq ||f|| \cdot ||x - x_n|| + ||f|| \cdot M \to 0. \tag{9.15}$$

□

9.2.3 Riesz Representation Theorem on Hilbert Space

Theorem 9.3 (Riesz Representation Theorem)**.** *Let H be a Hilbert space, and f be a linear functional on H. Then $f \in H^*$ if and only if there exists a unique $y \in H$ such that $f(x) = \langle x, y\rangle$, $\forall x \in H$. In addition, $||f|| = ||y||$.*

Proof. We first show the "if" part. Let f be defined by $f(x) = \langle x, y\rangle$. By Cauchy-Swartz inequality, we have for each x, $|f(x)| \leq ||x|| \cdot ||y||$, which implies that $||f|| \leq ||y||$. Take $x = y$, then $||f|| \cdot ||y|| \geq |f(y)| = ||y||^2$, hence $||f|| \geq ||y||$. Therefore $||f|| = ||y||$. Obviously, $f \in H^*$.

For the "only if" part, let $f \in H^*$. If $f = 0$, then we can take $y = 0$. In the following, we assume $f \neq 0$. Let $L = \ker f$, then L is a closed subspace of H. with $\mathrm{codim}(L) = 1$. Since $H = L \oplus L^\perp$, we have $\dim(L^\perp) = 1$. Hence, $L^\perp = \mathrm{span}\{y_0\}$ for some $y_0 \in H$. It is easy to see that $f(y_0) \neq 0$, otherwise $y_0 \in L \cap L^\perp = \{0\}$. For each $x \in H$, we can write $x = z + \lambda y_0$ with $z \in L$. Then $f(x) = \lambda f(y_0)$. Take $y = y_0 f(y_0)/||y_0||^2$, then

$$\langle x, y_0 f(y_0)/||y_0||^2\rangle = \langle z + \lambda y_0, y_0 f(y_0)/||y_0||^2\rangle = \lambda\langle y_0, y_0 f(y_0)/||y_0||^2\rangle = \lambda f(y_0) = f(x). \tag{9.16}$$

Hence $f = \langle \cdot, y\rangle$.

Finally, we show the uniqueness. If $f(x) = \langle x, y_1\rangle = \langle x, y_2\rangle$ for each x, Then $\langle x, y_1 - y_2\rangle = 0$, $\forall x \in H$. Hence, $||y_1 - y_2|| = \langle y_1 - y_2, y_1 - y_2\rangle = 0$, thus $y_1 = y_2$. □

The following is a corollary derived by applying the Riesz representation theorem to Lebesgue measure theory.

Corollary 9.3. *Let $(X, \mathcal{M}, \mu)$ be a measure space, then $L^2(\mu)$ is a Hilbert space. Let $\lambda : L^2(\mu) \to \mathbb{R}$ be a bounded linear functional, then there exists $g \in L^2(\mu)$ such that*

$$\lambda(f) = \int_X f\bar{g}d\mu, \quad \forall f \in L^2(\mu)$$

and

$$||g||_2 = \int_X |g|^2 d\mu = \sup_{f\neq 0} \frac{|\lambda(f)|}{||f||_2}.$$

9.2.4 Hahn-Banach Theorem

Hahn-Banach theorem is one of the most important theorem in functional analysis. It states that *every bounded linear functional defined on a subspace has a norm-preserving extension.*

Theorem 9.4 (Hahn-Banach Theorem). *Let X be a normed space, and E be its subspace, then for each $f_0 \in E^*$, there exists $f \in X^*$ with $f|_E = f_0$ and $||f||_{X^*} = ||f_0||_{E^*}$.*

Proposition 9.9. *Let X be a normed space, for each $x_0 \in X$, there exists $f \in X^*$ such that $||f|| = 1$ and $f(x_0) = ||x_0||$.*

Proof. Define a linear functional one the subspace span$\{x_0\}$, by $f_0(\lambda x_0) := \lambda||x_0||$. It is easy to verify that this is a linear functional and has norm $||f_0|| = 1$. By Hahn-Banach theorem, there is an extension $f \in X^*$ with $||f|| = 1$ and $f(x_0) = f_0(x_0) = ||x_0||$. □

Proposition 9.10. *Let X be a normed space, then for each $x \in X$, we have*

$$||x|| = \sup_{f\in X^*, f\neq 0} \frac{|f(x)|}{||f||} = \sup\{|f(x)| : ||f|| = 1\}.$$

Proof. First, from $|f(x)| \leq ||f|| \cdot ||x||$, we know

$$\sup\{|f(x)| : ||f|| = 1\} \leq ||x||. \tag{9.17}$$

On the other hand, for each x, we can find f' with $||f'|| = 1$ and $f'(x) = ||x||$, hence

$$\sup\{|f(x)| : ||f|| = 1\} \geq |f'(x)| = ||x||. \tag{9.18}$$

Then the equality is established. □

Corollary 9.4. *Let X be a normed space, if $x \in X$ has $f(x) = 0$, $\forall f \in X^*$, then $x = 0$.*

Proof. We can choose $f \in X^*$ such that $f(x) = ||x||$, by assumption, we have $||x|| = 0$, thus $x = 0$. □

Corollary 9.5. *Let X be a normed space, and $x_1, x_2 \in X$ have $x_1 \neq x_2$, then there is $f \in X^*$ such that $f(x_1) \neq f(x_2)$.*

Proof. We can choose $f \in X^*$ such that $f(x_1 - x_2) = ||x_1 - x_2|| > 0$, then $f(x_1) \neq f(x_2)$. □

Proposition 9.11. *Let X be a normed vector space and E be a proper closed subspace of X, if $x_0 \in X \cap E^c$ and $d = \inf_{y\in E} ||x_0 - y||$ then there exists $f \in X^*$ such that $||f|| = 1$, $f(y) = 0$ on E, and $f(x_0) = d$.*

Proof. Since E is closed and $x_0 \notin E$, we know that $d > 0$. Otherwise, we can find a sequence in E that converges to x_0, contradicting the closedness of E. Consider the augmented space $E^+ = E \oplus \text{span}\{x_0\}$, then E is a subspace of E^+ with codim$(E) = 1$ (note that the codimension is w.r.t E^+). By lemma 9.3, there exists a bounded linear functional f_0 defined on E^+ such that $||f_0|| = 1$, $f_0 = 0$ on E and $f(x_0) = d$. By Hahn-Banach theorem, we can find a norm-preserving extension of f_0 on X, then we are done. □

Proposition 9.12. *Let X be a normed space, for each subspace L of X we denote*

$$L^\perp = \{f \in X^* | f(x) = 0,\ \forall x \in L\},$$

and for each subspace F of X^* we denote

$$F_\perp = \{x \in X | f(x) = 0,\ \forall f \in F\}.$$

Let L be a subspace of X, then

1. *$L^\perp$ is a closed subspace of X^*;*
2. *$L \subset (L^\perp)_\perp$;*
3. *If L is closed, then $L = (L^\perp)_\perp$.*

Proof. 1. It is easy to verify that $L^\perp$ is a subspace of X^*. Let (f_n) be a sequence in $L^\perp$ that converges to $f \in X^*$. To prove that $L^\perp$ is closed, It suffices to show that $f \in L^\perp$. This can be seen from

$$f(x) = \lim_{n\to\infty} f_n(x) = 0. \tag{9.19}$$

2. For each $x \in L$, we have $f(x) = 0$ for every $f \in L^\perp$, thus $x \in (L^\perp)_\perp$ by definition. It follows that $L \subset (L^\perp)_\perp$.

3. It suffices to show that $(L^\perp)_\perp \subset L$, which equivalent to $x \notin L \Rightarrow x \notin (L^\perp)_\perp$. Let $x \notin L$, then since L is closed, $d = \inf_{y\in L} ||x - y|| > 0$, then there exists $f \in L^\perp$ such that $f(x) = d$, thus $x \notin (L^\perp)_\perp$.

□

9.2.5 Second Dual

Definition 9.5 (Second dual)**.** *Let X be a normed space, then the dual space of X^* is called the* ***second dual space*** *of X, denoted by X^{**}.*

Proposition 9.13. *Let X be a normed vector space, for each $x \in X$, we can define a linear functional $\tilde{x} : X^* \to \mathbb{F}$ by $\tilde{x}(f) = f(x)$, then $\tilde{x}$ is a bounded linear functional (i.e. $g_x \in X^{**}$) with $||\tilde{x}|| = ||x||$. The map $x \mapsto \tilde{x}$ is called the* ***canonical embedding*** *of X into X^{**}, which is injective.*

Proof. 1. For each x, $\tilde{x}$ is a linear functional on X^*, which follows from

$$\tilde{x}(\alpha f + \beta g) = (\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x) = \alpha\tilde{x}(f) + \beta\tilde{x}(g) \tag{9.20}$$

2. We then show $||\tilde{x}|| = ||x||$ as follows (

$$||\tilde{x}|| = \sup\{\tilde{x}(f) : ||f|| = 1\} = \sup\{|f(x)| : ||f|| = 1\} = ||x||. \tag{9.21}$$

Note that we utilize proposition 9.10.

3. The map $x \mapsto \tilde{x}$ is linear. This can be seen from

$$(\widetilde{\alpha x + \beta y})(f) = f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) = \alpha\tilde{x}(f) + \beta\tilde{y}(f). \tag{9.22}$$

4. The map $x \mapsto \tilde{x}$ is injective. This can be seen from $\tilde{x} = 0 \Rightarrow f(x) = 0, \forall f \in X^* \Rightarrow x = 0$.

□

Definition 9.6 (Reflexivity)**.** *A normed space X is said to be* ***reflexive*** *if the canonical embedding of X into X^{**} is an isomorphism. In other words, for each $z \in X^{**}$ we can find $x \in X$ such that $z(f) = f(x)$, $\forall f \in X^*$.*

Proposition 9.14. *Every reflexive normed space is complete.*

Proof. If a normed space X is reflexive, then the canonical embedding is an isometry between X and X^{**}. We know that X^{**} is complete, since it is in itself a dual space. Hence, X is complete. □

Proposition 9.15 (Dual basis)**.** *Let X be an n-dimensional vector space with basis $\{e_1, \ldots, e_n\}$, then there exists a unique set of vectors $\{f_1, \ldots, f_n\}$ in X^* such that*

$$f_i(e_j) = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

And $\{f_1, \ldots, f_n\}$ forms a basis of X^*, which is called the ***dual basis*** *of $\{e_1, \ldots, e_n\}$.*

Proof. First, we show the existence of $\{f_1, \ldots, f_n\}$ by construction. Since $\{e_1, \ldots, e_n\}$ is a basis of X, for each x it can be uniquely written as $x = \sum_{i=1}^n \alpha_i e_i$ with $\alpha_i \in \mathbb{F}$. It is easy to check that α_i depends linearly on x. We define f_i by $f_i(x) := \alpha_i$. Then f_i is well-defined (due to uniqueness of basis expansion) and is a linear functional. In addition, since every norm is equivalent for a finite dimensional spaces. There is C, such that

$$\sum_{i=1}^n |\alpha_i| \leq C||x||. \tag{9.23}$$

From this, we can readily see that f_i is a bounded functional for each i.

Then we show that $\{f_1, \ldots, f_n\}$ is a basis. Suppose there exists $c_1, \ldots, c_n$ such that $\sum_{i=1}^n c_i f_i = 0$. Then for each e_k, we have

$$\sum_{i=1}^n c_i f_i(e_k) = \sum_{i=1}^n c_i \delta_{ik} = c_k = 0. \tag{9.24}$$

It follows that to make $\sum_{i=1}^n c_i f_i$ zero, all its coefficients must be zero. Hence $\{f_1, \ldots, f_n\}$ is linearly independent.

Given $f \in X^*$, for each $x \in X$, it can be written as $x = \sum_{i=1}^n \alpha_i e_i$, then we have

$$f(x) = f\left(\sum_{i=1}^n \alpha_i e_i\right) = \sum_{i=1}^n \alpha_i f(e_i). \tag{9.25}$$

And let $f' = \sum_{i=1}^n f(e_i) f_i$, then

$$f'(x) = \sum_{i=1}^n f(e_i) f_i(x) = \sum_{i=1}^n f(e_i)\alpha_i. \tag{9.26}$$

Clearly $f = f' = \sum_{i=1}^\infty \alpha_i f_i$. Hence, we can conclude that $\{f_1, \ldots, f_n\}$ is a basis of X^*. □

Proposition 9.16. *Every finite dimensional space is reflexive.*

Proof. Let X be an n-dimensional space, and $\mathcal{E} = \{e_1, \ldots, e_n\}$ be its basis. Then, we can get a dual basis $\{f_1, \ldots, f_n\}$ for X^*. To show that the canonical embedding of X into X^{**} is an isomorphism, it suffices to show that $\{\tilde{e}_1, \ldots, \tilde{e}_n\}$ is a dual basis of $\{f_1, \ldots, f_n\}$ for X^{**}. This can be seen from

$$\tilde{e}_i(f_j) = f_j(e_i) = \delta_{ij}, \quad \forall i, j = 1, \ldots, n \tag{9.27}$$

Then we are done. □

Chapter 10

Linear Operators

10.1 The Space of Linear Operators

In functional analysis, a linear map is also called a **linear operator**.

Definition 10.1 (Bouned linear operator)**.** *Let X and Y be normed vector spaces, a linear operator $T : X \to Y$ is called a **bounded linear operator** if there exists $M > 0$ such that*

$$||Tx|| \leq M||x||, \ \forall x \in X$$

The following proposition shows that all bounded linear operators between two given normed spaces constitute a vector space.

Proposition 10.1. *Let $T_1, T_2 : X \to Y$ be bounded linear operators, then for any $\alpha_1, \alpha_2 \in \mathbb{C}$, $\alpha_1 T_1 + \alpha_2 T_2$ is also a bounded linear operator.*

Proof. Since T_1 and T_2 are bounded, there are $M_1, M_2 > 0$ such that for each $x \in X$, $||T_1 x|| \leq M_1||x||$ and $||T_2 x|| \leq M_2||x||$, thus

$$||(\alpha_1 T_1 + \alpha_2 T_2)x|| = ||\alpha_1 T_1 x + \alpha_2 T_2 x|| \leq |\alpha_1| \cdot ||T_1 x|| + |\alpha_2| \cdot ||T_2 x|| \leq (|\alpha_1|M_1 + |\alpha_2|M_2)||x||. \tag{10.1}$$

Hence, $\alpha_1 T_1 + \alpha_2 T_2$ is also bounded. □

Definition 10.2 (The normed space of bounded linear operators)**.** *Let X and Y be normed vector spaces, then all bounded linear operators from X to Y form a vector space, denoted by $L(X, Y)$. In addition, this space is endowed with a norm $|| \cdot || : L(X, Y) \to \mathbb{R}$ defined by*

$$||T|| = \sup_{x \neq 0} \frac{||Tx||}{||x||} = \sup\{||Tx|| : ||x|| = 1\}.$$

It is easy to verify that the norm defined above is really a norm. In addition, by the definition of operator norm, we immediately have

Proposition 10.2. *Let $T : X \to Y$ be a bouned linear operator, then we have*

$$||Tx|| \leq ||T|| \cdot ||x||, \ \forall x \in X.$$

Proposition 10.3. *Let X and Y be normed vector spaces, and $T : X \to Y$ be a linear operator, then T is bounded if and only if T is continuous.*

Proof. First, assume T is bounded, i.e. there exists $M > 0$ such that $||Tx|| \leq M||x||$ for each $x \in X$. Then for any $x_0 \in X$ and $\epsilon > 0$, we can choose $\delta = \epsilon/M$ such that when $||x - x_0|| < \delta$, we have $||Tx - Tx_0|| \leq M||x - x_0|| < M\delta$, thus T is continuous.

For the other direction, we assume T is continuous. Then, there exists $\delta > 0$ such that when $||x|| \leq \delta$, $||Tx|| \leq 1$. Then for arbitrary $x \in X$, we have

$$||Tx|| = \left\| T \frac{\delta x}{||x||} \right\| \cdot \frac{||x||}{\delta} \leq \frac{1}{\delta}||x||. \tag{10.2}$$

Hence, T is bounded. □

Corollary 10.1. *Let $T : X \to Y$ be a bounded linear operator, then for any sequence (x_n) in X that converges to x, we have $Tx_n \to Tx$.*

Corollary 10.2. *Let $T : X \to Y$ be a linear operator between normed spaces X and Y, then T is bounded then $\ker T$ is closed.*

Proof. If T is bounded, thus T is continuous, hence $\ker T = T^{-1}\{0\}$ is closed. □

Proposition 10.4. *Let $A : X \to Y$ and $B : Y \to Z$ be bounded linear operators, then the operator $BA : X \to Z$ is bounded and has*

$$||BA|| \leq ||A|| \cdot ||B||.$$

Proof. For every $x \in X$ we have

$$||BAx|| \leq ||B|| \cdot ||Ax|| \leq ||B|| \cdot ||A|| \cdot ||x||. \tag{10.3}$$

□

If T is a linear operator from X to X, then we use the following notation T^n to denote the composition of n such operator.

Corollary 10.3. *Let $T : X \to X$ be a bounded linear operator, then*

$$||T^n|| \leq ||T||^n.$$

Theorem 10.1. *Let X be any normed space, and Y be a Banach space, then $L(X, Y)$ is a Banach space.*

Proof. What we need to show here is that $L(X, Y)$ is complete. Let (A_n) be any Cauchy sequence in $L(X, Y)$. Then given $\epsilon > 0$, there exists N such that $\forall m, n > N$, $||A_m - A_n|| < \epsilon$, thus $\forall m, n > N$, $||A_m x - A_n x|| \leq ||A_m - A_n|| \cdot ||x|| < \epsilon||x||$, thus $(A_n x)$ is a Cauchy sequence for each $x \in X$. Since Y is complete, $(A_n x)$ converges for each x. Then we can define a map $A : X \to Y$ by

$$Ax := \lim_{n\to\infty} A_n x. \tag{10.4}$$

It is trivial to verify that A is a linear map. Since (A_n) is a Cauchy sequence, it is bounded, i.e. $\sup_n ||A_n|| < +\infty$. Hence, we have

$$||Ax|| \leq \sup_n ||A_n x|| \leq \sup_n(||A_n|| \cdot ||x||) = (\sup_n ||A_n||)||x||. \tag{10.5}$$

This shows that $A \in L(X, Y)$. In addition, given $\epsilon > 0$, there is $N \in \mathbb{N}^+$ such that when $m, n > N$, $||A_n - A_m|| < \epsilon$, and thus

$$\forall x \; ||x|| \leq 1 \Rightarrow ||A_n x - A_m x|| \leq \epsilon. \tag{10.6}$$

Take $m \to +\infty$, by continuity of norm, we have for each $n > N$,

$$||A_n x - Ax|| = ||A_n x - \lim_{m\to\infty}(A_m x)|| = \lim_{m\to\infty} ||A_n x - A_m x|| \leq \epsilon. \tag{10.7}$$

It follows that $||A_n - A|| \to 0$, i.e. $A_n \to A$. □

10.2 Inverse Operators

Definition 10.3 (Inverse operator)**.** *Let $A : X \to Y$ be a linear operator, if there is a linear operator $B : Y \to X$ such that $AB = \mathrm{Id}_X$ and $BA = \mathrm{Id}_Y$ then B is called the* ***linear operator*** *of A, denoted by A^{-1}, and A is called* ***invertible****.*

Proposition 10.5. *A linear operator $A : X \to Y$ is invertible if and only if it is bijective (one-to-one and onto).*

Proposition 10.6. *Let $A : X \to Y$ and $B : Y \to Z$ be invertible linear operators, then $BA : X \to Z$ is invertible and $(BA)^{-1} = A^{-1}B^{-1}$.*

Proof. This can be readily seen from $(BA)(A^{-1}B^{-1}) = B\mathrm{Id}_Y B^{-1} = BB^{-1} = \mathrm{Id}_Z$ and $(A^{-1}B^{-1})(BA) = A^{-1}\mathrm{Id}_Y A = A^{-1}A = \mathrm{Id}_X$. □

Proposition 10.7. *Let $T : X \to Y$ be a linear operator onto Y such that there exists $b > 0$ and $||Tx|| \geq b||x||$ $\forall x \in X$, then T is invertible and T^{-1} is a bounded linear operator with $||T^{-1}|| \leq 1/b$.*

Proof. First, we can see T is injective from

$$Tx = 0 \Rightarrow 0 \geq b||x|| \Rightarrow ||x|| = 0 \Rightarrow x = 0. \tag{10.8}$$

Hence, T is bijective, and thus it is invertible. In addition, we note that $x = TT^{-1}x$, thus

$$||x|| = ||TT^{-1}x|| \geq b||T^{-1}x||, \ \forall x \in X \tag{10.9}$$

Thus $||T^{-1}x|| \leq (1/b)||x||$, which follows that T^{-1} is bounded and has $||T|| \leq 1/b$. □

Proposition 10.8. *Let $A : X \to X$ be a bounded linear operator with $||A|| < 1$, then $I - A$ is invertible, and*

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k, \quad \textit{with} \ \ ||(I - A)^{-1}|| \leq (1 - ||A||)^{-1}.$$

Proof. Let $B = \sum_{k=0}^{\infty} A^k$. First, since $||A|| < 1$,

$$\sum_{k=0}^{\infty} ||A||^k \leq (1 - ||A||)^{-1}.$$

It shows that as an infinite series, B converges, and B is a bounded operator with $||B|| \leq (1 - ||A||)^{-1}$. It remains to show that B is the inverse of $I - A$. We note that $I + AB = I + BA = B$, thus $I = B - AB = (I - A)B$ and $I = B - BA = B(I - A)$. □

Corollary 10.4. *Let A be a bounded linear operator that has a bounded inverse, and B be a linear operator that satisfies $||A - B|| < 1/||A^{-1}||$, then B is invertible.*

Proof. We note that

$$B = A - (A - B) = A(I - A^{-1}(A - B)). \tag{10.10}$$

And

$$||A^{-1}(A - B)|| < ||A^{-1}||(1/||A^{-1}||) = 1 \tag{10.11}$$

Hence, $(I - A^{-1}(A - B))$ is invertible, thus B is invertible. □

This corollary tells us that for each bounded linear operator with bounded inverse, there is a neighborhood in which the operators possess such properties as well.

In the following, we introduce the Open mapping theorm. It is an important theorem in functional analysis, which gives a useful sufficient condition to judge whether the inverse of a bounded operator is bounded.

Theorem 10.2 (Open mapping theorem)**.** *Let X and Y be Banach spaces, and $T : X \to Y$ be a bounded linear operator, then T is an open mapping, i.e. T sends every open set in X to an open set in Y.*

Corollary 10.5. *Let X and Y be Banach spaces, and $T : X \to Y$ be an invertible bounded linear operator, then T^{-1} is bounded.*

Proof. By open mapping theorem, T is an open mapping. Hence, $(T^{-1})^{-1}(U) = T(U)$ is open for each open set U in X, thus T^{-1} is continuous, thus it is bounded. □

10.3 Dual Operators

Definition 10.4 (Dual operator)**.** *Let $A : X \to Y$ be a bounded linear operator. Then for each $h \in Y^*$, the linear functional f on X defined by $f(x) := h(Ax)$ is a bounded, i.e. $f \in X^*$. This can be seen from $|f(x)| \leq ||h|| \cdot ||A|| \cdot ||x||$. Then, we can define an operator $A^* : Y^* \to X^*$ by*

$$(A^*h)(x) = h(Ax), \quad \forall h \in Y^*$$

*which is called the **dual operator** of A.*

Proposition 10.9. *Let $A : X \to Y$ be a bounded linear operator, then its dual operator $A^* : Y^* \to X^*$ is also a bounded linear operator, with $||A^*|| = ||A||$.*

Proof. We first show that A^* is linear. This can be seen from

$$\begin{aligned}(A^*(\alpha f + \beta g))(x) &= (\alpha f + \beta g)(Ax) = \alpha f(Ax) + \beta g(Ax) = \alpha(A^*f)(x) + \beta(A^*g)(x) \\ &= (\alpha A^*f + \beta A^*g)(x).\end{aligned} \tag{10.12}$$

In addition, we have for each $f \in Y^*$,

$$|(A^*f)(x)| = |f(Ax)| \leq ||f|| \cdot ||A|| \cdot ||x||. \tag{10.13}$$

Hence $||A^*f|| \leq ||A|| \cdot ||f||$, which implies that $||A^*|| \leq ||A||$. On the other hand, from the corollary of Hahn-Banach theorem, for every $x \in X$, there is $g_x \in Y^*$ with $||g_x|| = 1$, and $g_x(Ax) = ||Ax||$. Let $f_x = A^*g_x$, then

$$||Ax|| = g_x(Ax) = (A^*g_x)(x) \leq ||A^*g_x|| \cdot ||x|| \leq ||A^*|| \cdot ||g_x|| \cdot ||x|| = ||A^*|| \cdot ||x||. \tag{10.14}$$

Taking supremium of the left hand side, we get $||A|| \leq ||A^*||$. $\square$

Proposition 10.10. *Let $A, B \in L(X, Y)$, then $(\alpha A + \beta B)^* = \alpha A^* + \beta B^*$ for each $\alpha, \beta \in \mathbb{C}$.*

Proof. For each $f \in Y^*$ and $x \in X$, we have

$$\begin{aligned}((\alpha A + \beta B)^*f)(x) &= f((\alpha A + \beta B)(x)) = f(\alpha Ax + \beta Bx) = \alpha f(Ax) + \beta f(Bx) \\ &= \alpha(A^*f)(x) + \beta(B^*f)(x) = ((\alpha A^* + \beta B^*)(f))(x).\end{aligned} \tag{10.15}$$

$\square$

Proposition 10.11. *Let $A \in L(X, Y)$ be invertible, then A^* is also invertible, and $(A^*)^{-1} = (A^{-1})^*$.*

Proof. For each $f \in Y^*$ and $x \in X$, we have

$$((A^{-1})^*A^*f)(x) = (A^*f)(A^{-1}x) = f(AA^{-1}x) = f(x), \tag{10.16}$$

thus $(A^{-1})^*A^*f = f$, $\forall f \in Y^*$. Similarly, we can show $A^*(A^{-1})^*f = f$, $\forall f \in X^*$. $\square$

Proposition 10.12. *Let $A \in L(X, Y)$ and $B \in L(Y, Z)$ then $(BA)^* = A^*B^*$.*

Proof. For each $f \in Z^*$, and $x \in X$, we have

$$((BA)^*f)(x) = f(BAx) = (B^*f)(Ax) = (A^*(B^*f))(x), \tag{10.17}$$

hence $(BA)^* = A^*B^*$. $\square$

10.4 Convergence of Operators

There are several different notions of convergence in regard to linear operators.

Definition 10.5 (Convergence of operators)**.** *Let X and Y be normed vector spaces, (A_n) be a sequence of operators in $L(X, Y)$ and $A \in L(X, Y)$, then*

1. *we say* (A_n) ***converges uniformly*** *to* A *or* (A_n) ***converges in norm*** *to* A *if* $||A_n - A|| \to 0$*, denoted by* $A_n \to A$*;*

2. *we say* (A_n) ***converges strongly*** *to* A *if* $A_n x \to Ax,\ \forall x \in X$*, denoted by* $A_n \xrightarrow{s} A$*;*

3. *we say* (A_n) ***converges weakly*** *to* A *if* $f(A_n x) \to Ax,\ \forall x \in X, f \in Y^*$*, denoted by* $A_n \xrightarrow{w} A$*.*

Proof. Uniform convergence of operators implies strong convergence; strong convergence implies weak convergence. □

Proof. 1. $(A_n \to A \Rightarrow A_n \xrightarrow{s} A)$. Assume $A_n \to A$, for each $x \in X$, we have

$$||A_n x - Ax|| \leq ||A_n - A|| \cdot ||x|| \to 0 \tag{10.18}$$

Hence, $A_n \xrightarrow{s} A$.

2. $(A_n \xrightarrow{s} A \Rightarrow A_n \xrightarrow{w} A)$. This directly follows from the fact that when f is a bounded linear functional, then $x_n \to x \Rightarrow f(x_n) \to f(x)$.

□