Lecture Notes on Real Analysis

Xiaojing Ye

Contents

1 Preliminaries **3**
1.1 Basics of sets 3
1.2 Functions 5
1.3 Cardinality of sets 5
1.4 Topology of metric spaces 8

2 Measure and Measurable Sets **13**
2.1 σ-algebra 13
2.2 Outer measure 13
2.3 Measurable sets and Lebesgue measure 14
2.4 Non-measurable sets 19

3 Measurable Functions **21**
3.1 Extended real numbers 21
3.2 Simple functions 24
3.3 Convergence almost everywhere 25
3.4 Convergence in measure 26
3.5 Measurable functions and continuous functions 28
3.6 Measurability of composite functions 29

4 Lebesgue Integrals **30**
4.1 Integral of simple nonnegative functions 30
4.2 Integral of general nonnegative functions 31
4.3 Integral of general functions 33
4.4 Relation between Riemann and Lebesgue integrals 39
4.5 Iterated integrals 41
4.6 Convolution 44

5 Signed Measures and Differentiations **45**
5.1 Signed measure and decomposition 45
5.2 Radon-Nikodym theorem 47
5.3 Differentiation 49
5.4 Functions of bounded variation 53
5.5 Absolute continuity 57

6 L^p Spaces **60**
6.1 Important inequalities 60
6.2 L^p space 63
6.3 L^2 space and inner product 65
6.4 Dual space of L^p 68

7 Probability Theory **71**
7.1 Basic concepts 71
7.2 The law of large numbers 73
7.3 Central limit theorem 75

These notes outline the materials covered in class. Detailed derivations and explanations are given in lectures and/or the referenced books. The notes will be continuously updated with additional content and corrections. Questions and comments can be addressed to `xye@gsu.edu`.

1 Preliminaries

1.1 Basics of sets

A *set* A is a collection of elements with certain properties P, commonly written as $A = \{x : x \text{ satisfies } P\}$ (e.g., $A = \big\{x \in \mathbb{R} : x^2 > 1\big\}$). Recall the following definitions: subset $A \subset B$, union $A \cup B$, intersection $A \cap B$, complement A^c, empty set $\emptyset$, equal sets $A = B$, set minus $A \setminus B = A \cap B^c$.

Example 1.1. The following statements hold:

- $A \cap B \subset A \subset A \cup B$ for any A, B.
- $A \subset B$ iff $B^c \subset A^c$.
- $A \cap B = \emptyset$ iff $A \subset B^c$.
- Let $A, B \subset X$. If $E \cap A = E \cup B$ for any $E \subset X$, then $A = X$ and $B = \emptyset$.

We frequently consider a set of sets, and call it a family (or collection) of sets: $\mathcal{F} = \{A_\alpha : \alpha \in I\}$, where I is the index set. Here I can be finite $\{1, \dots, n\}$, countably infinite $\mathbb{N}$, or uncountably infinite. We also work with union and intersection of multiple (often infinitely many) sets:

$$\bigcup_{\alpha \in I} A_\alpha = \{x : \exists\, \alpha \in I,\ \text{s.t. } x \in A_\alpha\} \quad \text{and} \quad \bigcap_{\alpha \in I} A_\alpha = \{x : x \in A_\alpha,\ \forall\, \alpha \in I\}$$

The union and intersection satisfy the *distributive law*:

$$A \cap \Big(\bigcup_{\alpha \in I} B_\alpha\Big) = \bigcup_{\alpha \in I} (A \cap B_\alpha) \quad \text{and} \quad A \cup \Big(\bigcap_{\alpha \in I} B_\alpha\Big) = \bigcap_{\alpha \in I} (A \cup B_\alpha)$$

Example 1.2. Let $A_k = [a + \frac{1}{k}, b]$, then $\cup_{k=1}^{\infty} A_k = (a, b]$. Let $A_k = (a, b + \frac{1}{k})$, then $\cap_{k=1}^{\infty} A_k = (a, b]$.

Example 1.3. Let $A_\alpha = [0, -\log \alpha)$ where $\alpha \in I = (0, 1] \subset \mathbb{R}$, then $\cup_{\alpha \in I} A_\alpha = [0, \infty)$ and $\cap_{\alpha \in I} A_\alpha = \{0\}$.

Example 1.4. Suppose $f : [a, b] \to \mathbb{R}$. Show that $\{x \in [a, b] : |f(x)| > 0\} = \cup_{n=1}^{\infty} \{x \in [a, b] : |f(x)| > \frac{1}{n}\}$

Theorem 1.5 (De Morgan's law)**.** $(\cap_{\alpha \in I} A_\alpha)^c = \cup_{\alpha \in I} A_\alpha^c$ *and* $(\cup_{\alpha \in I} A_\alpha)^c = \cap_{\alpha \in I} A_\alpha^c$.

Example 1.6. Some basic tricks in proofs.

- Use of Venn diagram. For example, define the *symmetric difference of* A *and* B by $A \triangle B = (A \setminus B) \cup (B \setminus A)$, show $A \triangle B = (A \cup B) \setminus (A \cap B)$.
- $A \subset B$ iff $x \in A \Rightarrow x \in B$.
- $A = B$ iff $A \subset B$ and $B \subset A$.

Definition 1.7 (Limit of a sequence of monotone sets)**.** Suppose $A_1 \supset A_2 \supset \cdots A_k \supset \cdots$, then we say $\{A_k\}$ is *non-increasing* or simply *decreasing* (to be distinguished from *strictly decreasing* where $A_{k+1} \subsetneq A_k$ for all k), and $\cap_{k=1}^{\infty} A_k$ is called the *limit* of $\{A_k\}$, denoted by $\lim_{k \to \infty} A_k$ or simply $\lim_k A_k$. Similarly, suppose $A_1 \subset \cdots A_k \subset \cdots$, then $\{A_k\}$ is *non-decreasing* or simply *increasing*, and $\cup_{k=1}^{\infty} A_k$ is the limit of $\{A_k\}$, also denoted by $\lim_k A_k$.

Example 1.8. Let $A_k = [k, \infty) \subset \mathbb{R}$ for $k = 1, \ldots,$, then $\lim_k A_k = \emptyset$.

Example 1.9. Suppose $\{f_k\}$ is a sequence of real-valued functions defined on $\mathbb{R}$, and $f_1(x) \leq f_2(x) \leq \cdots \leq f_k(x) \leq \cdots$ and $f_k(x) \to f(x)$ as $k \to \infty$ for every $x \in \mathbb{R}$. For any $t \in \mathbb{R}$, define $A_k = \{x \in \mathbb{R} : f_k(x) > t\}$. Show that $\{A_k\}$ is increasing, and $\lim_k A_k = \{x \in \mathbb{R} : f(x) > t\}$.

Proof. It is clear that A_k is increasing and $\lim_k A_k \subset A := \{x \in \mathbb{R} : f(x) > t\}$. For every $x \in A$, there are $f(x) > t$, and $f_k(x) \uparrow f(x)$ as $k \to \infty$. Hence let $\epsilon = (f(x) - t)/2 > 0$, then there exists k' such that $f_{k'}(x) > f(x) - \epsilon = (f(x) + t)/2 > t$, and therefore $x \in A_{k'} \subset \cup_{k=1}^{\infty} A_k = \lim_k A_k$. $\square$

Definition 1.10 (Upper and lower limit of a sequence of sets)**.** Suppose $\{A_k\}$ is a sequence of sets. Denote $B_j = \cup_{k \geq j} A_k$, then $\{B_j\}$ is non-increasing. The *upper limit* of $\{A_k\}$ is denoted by

$$\limsup_{k\to\infty} A_k = \lim_{k\to\infty} B_k = \bigcap_{j=1}^{\infty} B_j = \bigcap_{j=1}^{\infty} \bigcup_{k=j}^{\infty} A_k$$

Similarly, the *lower limit* of $\{A_k\}$ is denoted by

$$\liminf_{k\to\infty} A_k = \bigcup_{j=1}^{\infty} \bigcap_{k=j}^{\infty} A_k$$

Note that $x \in \limsup_{k\to\infty} A_k = \cap_{j=1}^{\infty} \cup_{k=j}^{\infty} A_k$ means that: $\forall\, j \geq 1$, $\exists\, k \geq j$, such that $x \in A_k$. Similar for the lower limit.

Example 1.11. Show that $\liminf_{k\to\infty} A_k \subset \limsup_{k\to\infty} A_k$.

Proof. If $x \in \liminf_{k\to\infty} A_k$, then there exists $j \geq 1$, such that $x \in A_k$ for all $k \geq j$, which obviously implies that $x \in \limsup_{k\to\infty} A_k$. $\square$

Example 1.12. Suppose $f_n, f : \mathbb{R} \to \mathbb{R}$. Show that

$$\{x \in \mathbb{R} : \lim_{n\to\infty} f_n(x) \neq f(x)\} = \bigcup_{k=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \left\{x \in \mathbb{R} : |f_n(x) - f(x)| \geq \frac{1}{k}\right\}.$$

Proof. Note that $f_n(x) \nrightarrow f(x)$ at x means that there exists $\epsilon_0 > 0$ (or $k \in \mathbb{N}$ such that $\frac{1}{k} \leq \epsilon_0$), such that for any $N \in \mathbb{N}$, there is $|f_n(x) - f(x)| \geq \epsilon_0 \geq \frac{1}{k}$ for some $n \geq N$. Therefore, "there exists $k \geq 1$ ($\cup_{k=1}^{\infty}$), such that for any $N \geq 1$ ($\cap_{N=1}^{\infty}$), there exists an $n \geq N$ ($\cup_{n=N}^{\infty}$) for which $|f_n(x) - f(x)| \geq \frac{1}{k}$." $\square$

Example 1.13. Suppose $f_n(x) \to f(x)$ for every $x \in \mathbb{R}$. Show that, for any $t \in \mathbb{R}$, there is

$$\{x \in \mathbb{R} : f(x) \leq t\} = \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{x \in \mathbb{R} : f_n(x) \leq t + \frac{1}{k}\right\}.$$

Definition 1.14 (Cartesian product)**.** The *Cartesian product* of A and B is $A \times B = \{(a, b) : a \in A,\ b \in B\}$.

Definition 1.15. A few examples of Cartesian product:

- $A = \{1, 2, 3\}$, $B = \{4, 5\}$, then $A \times B = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}$.
- $[0, 1] \times [0, 1] = \{(x, y) : 0 \leq x, y \leq 1\}$.

1.2 Functions

Definition 1.16. We have a series of defintions regarding functions:

- Let X and Y be two sets. $f : X \to Y$ is called a *function* (or *mapping*, or *transformation*) if f assigns every $x \in X$ to one element $y \in Y$.
- Let $A \subset X$, then $f(A) = \{y \in Y : y = f(x) \text{ for some } x \in X\}$ is called the *image* of A under f. Let $B \subset Y$, then $f^{-1}(B) = \{x \in A : f(x) \in B\}$ is called the *inverse image* (or *pre-image*) of B under f.
- X is called the *domain* of f. $f(X)$ is the *range* of f.
- If $f(X) = Y$ then f is called a mapping from X *onto* Y (or f is *surjective*). If $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$ then f is called *one-to-one* (or f is *injective*).
- If f is both injective and surjective, then f is called *bijective*, or a *one-to-one correspondence* between X and Y. In this case f^{-1} exists and is also a one-to-one correspondence.
- Suppose $f : X \to Y$ and $g : Y \to Z$, then $g \circ f : X \to Z$ is called the *composition* of f and g, defined by $(g \circ f)(x) = g(f(x))$.

Example 1.17 (Characteristic function)**.** Suppose $A \subset X$. Define the *characteristic function* of A by

$$\chi_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \in A^c. \end{cases}$$

Then one can verify the following statements for every $x \in X$:

- $A \subset B \Rightarrow \chi_A(x) \leq \chi_B(x)$
- $\chi_{A \cup B}(x) = \chi_A(x) + \chi_B(x) - \chi_{A \cap B}(x)$
- $\chi_{A \cap B}(x) = \chi_A(x)\chi_B(x)$
- $\chi_{A \setminus B}(x) = \chi_A(x)(1 - \chi_B(x))$
- $\chi_{A \triangle B}(x) = |\chi_A(x) - \chi_B(x)|$

1.3 Cardinality of sets

We denote $|A|$ the *cardinal number* of A (informally the "number of elements in A"). This is clear if A is finite. However it is not obvious if A is infinite. We need the help of functions to "count" $|A|$.

Definition 1.18. X and Y is said to have the same *cardinal number* if there exists a one-to-one correspondence $f : X \to Y$. In this case, we denote $X \sim Y$. Then it is obvious that $\sim$ represents *an equivalence relation*: (i) $A \sim A$; (ii) $A \sim B$ iff $B \sim A$; (iii) If $A \sim B$ and $B \sim C$, then $A \sim C$.

Example 1.19. Sets with the same cardinality.

- $\mathbb{N} \sim \mathbb{Z} \sim \{0, 1, \dots\} \sim \{2n : n \in \mathbb{N}\}$.
- $\mathbb{N} \times \mathbb{N} \sim \mathbb{N}$ by setting $f((i, j)) = 2^{i-1} \cdot (2j - 1)$ (since every integer n can be uniquely represented by $n = 2^p \cdot q$ for some nonnegative integer p and odd integer q).

- $\mathbb{Q} \sim \mathbb{N}$.
- $(-1,1) \sim \mathbb{R}$ by setting $f(x) = \frac{x}{1-x^2}$ for $x \in (-1,1)$. [Or $f(x) = \tan(\frac{\pi}{2}x)$.]

Lemma 1.20 (Decomposition of sets by functions)**.** *Suppose $f : X \to Y$ and $g : Y \to X$. Then there exist $A_1, A_2 \subset X$ and $B_1, B_2 \subset Y$, such that $f(A_1) = B_1$, $g(B_2) = A_2$, $A_1 \cap A_2 = \emptyset$, $B_1 \cap B_2 = \emptyset$, $A_1 \cup A_2 = X$, and $B_1 \cup B_2 = Y$.*

Proof. Define $\Gamma := \{E \subset X : E \cap g(Y \setminus f(E)) = \emptyset\}$, $A_1 := \cup_{E \subset \Gamma} E$, $B_1 := f(A_1)$, $B_2 := Y \setminus B_1 = Y \setminus f(A_1)$, and $A_2 := g(B_2)$. Then it remains to show that $A_1 \cap A_2 = \emptyset$ and $A_1 \cup A_2 = X$.

For any $E \in \Gamma$, we know $E \subset A_1$ and hence $E \cap g(Y \setminus f(A_1)) \subset E \cap g(Y \setminus f(E)) = \emptyset$. Therefore $A_1 \cap g(Y \setminus f(A_1)) = \cup_{E \in \Gamma}(E \cap g(Y \setminus f(A_1))) = \emptyset$, i.e., $A_1 \in \Gamma$. Hence $A_1 \cap A_2 = \emptyset$.

If there exists $x_0 \in X \setminus (A_1 \cup A_2)$, then define $A = A_1 \cup \{x_0\}$ and hence there is $B_1 = f(A_1) \subset f(A)$. This implies that $Y \setminus f(A) \subset B_2$, and hence $g(Y \setminus f(A)) \subset g(B_2) = A_2$ and $A \cap g(Y \setminus f(A)) = \emptyset$ which means $A \in \Gamma$. This contradicts to $A_1 = \cup_{E \in \Gamma} E$. $\square$

Theorem 1.21 (Cantor-Bernstein)**.** *If $U \subsetneq X$ and $V \subsetneq Y$, and $X \sim V$ and $U \sim Y$, then $X \sim Y$.*

Proof. Let $f : X \to V$ and $g : Y \to U$ be one-to-one correspondences. By Lemma 1.20, there exist $A_1, A_2 \subset X$ and $B_1, B_2 \subset Y$, such that $A_1 \cap A_2 = \emptyset$, $B_1 \cap B_2 = \emptyset$, $A_1 \cup A_2 = X$, $B_1 \cup B_2 = Y$, $f(A_1) = B_1$ and $g(B_2) = A_2$ (which are still one-to-one correspondences as they are restrictions of f and g on A_1 and B_2 respectively). Define

$$h(x) = \begin{cases} f(x), & \text{if } x \in A_1 \\ g^{-1}(x), & \text{if } x \in A_2 \end{cases}$$

Then it is clear that h is a one-to-one correspondence between X and Y, so $X \sim Y$. $\square$

Corollary 1.22. *If $C \subset A \subset B$ and $C \sim B$, then $C \sim A \sim B$.*

Proof. If $C = A$ or $A = B$ then trivial. Otherwise $C \subsetneq A \subsetneq B$, then setting $X = A$, $Y = B$, $U = C$ and $V = A$ in Theorem 1.21 yields $A \sim B$. $\square$

Example 1.23. $(-1,1) \sim (-1,1] \sim [-1,1] \sim \mathbb{R}$.

Definition 1.24 (Cardinality of $\mathbb{N}$)**.** $\mathbb{N}$ is said to have cardinality $\aleph_0$ (pronounced as "aleph zero"). An infinite set of cardinality $\aleph_0$ is called *countable*; otherwise called *uncountable*.

Theorem 1.25 ($\aleph_0$ is the smallest cardinality of infinite sets)**.** *Every infinite set contains a countable set.*

Proof. Suppose E is infinite. Then we can pick $a_1, a_2 \dots,$ one by one from E, such that $a_{n+1} \in E \setminus \{a_1, \dots, a_n\} \neq \emptyset$, to get $\{a_k : k \in \mathbb{N}\} \subset E$. $\square$

Example 1.26. A few examples of sets of cardinality $\aleph_0$.

- If $A \sim \mathbb{N}$ and $B \sim \mathbb{N}$ then $A \cup B \sim \mathbb{N}$.
- If $A_n \sim \mathbb{N}$ for every $n \geq 1$, then $\cup_{n=1}^{\infty} A_n \sim \mathbb{N}$.
- $\mathbb{Q} \sim \mathbb{N}$. (Note that this only means that it is possible to list the elements of $\mathbb{Q}$ in *some* order, but not necessarily by their values.)

Example 1.27. The set of mutually disjoint open intervals in $\mathbb{R}$ is at most countable.

Proof. Define the function f that maps each interval to a rational number r in that interval. Then f is injective to $\mathbb{Q}$. $\square$

Example 1.28. If $f : \mathbb{R} \to \mathbb{R}$ is monotone, then $\{x \in \mathbb{R} : \lim_{y\to x^-} f(y) \neq \lim_{y\to x^+} f(y)\}$ is at most countable.

Proof. WLOG, assume non-decreasing. Then for each point in the set above, there exists $r_x \in \mathbb{Q}$ such that $\lim_{y\to x^-} f(y) < r_x < \lim_{y\to x^+} f(y)$. Define $g : x \mapsto r_x$, then g is injective. $\square$

Example 1.29. If E is a countable subset of $\mathbb{R}$, then $\exists\, x_0 \in \mathbb{R}$ such that $E \cap (E + \{x_0\}) = \emptyset$. [Hint: consider $A = \{r_n - r_m : r_n, r_m \in E, n \neq m\}$ which is countable, hence $\exists\, x_0 \in \mathbb{R} \setminus A$.]

Theorem 1.30. *If A is an infinite set and B is at most countable, then $A \sim A \cup B$.*

Proof. Suppose $B = \{b_1, b_2, \ldots\}$. Extract a countable set $A_1 = \{a_1, a_2, \ldots\}$ from A, and denote $A_2 = A \setminus A_1$. Then define

$$f(x) = \begin{cases} a_{2i-1} & \text{if } x = b_i \in B \\ a_{2i} & \text{if } x = a_i \in A_1 \\ a & \text{if } x = a \in A_2 \end{cases}$$

Hence $f : A \cup B \to A$ is a one-to-one correspondence. $\square$

Theorem 1.31. *X is infinite iff $X \sim A$ for some $A \subsetneq X$.*

Proof. The necessity is obvious. Extract a finite set B from X and define $A = X \setminus B$, then A is inifite, and $A \sim A \cup B = X$. $\square$

Definition 1.32 (Cardinality of $\mathbb{R}$)**.** $\mathbb{R}$ is said to have cardinality $\aleph_1$, also called *cardinality of the continuum* $\mathfrak{c} = \aleph_1 = 2^{\aleph_0}$.

We consider the cardinality of $(0, 1] \sim \mathbb{R}$. For every $x \in (0, 1]$, it can be written as $x = \sum_{n=1}^{\infty} \frac{a_n}{2^n}$ for $a_n \in \{0, 1\}$ and infinitely many a_n's being 1. To see this, note that every irrational number is a limit point of rational numbers, and if $a_k = 0$ for all $k > n$ then we can instead set $a_n = 0$ and $a_k = 1$ for all $k > n$. We can show that $A = \{(a_1, a_2, \ldots) : a_n \in \{0, 1\}\}$ is an uncountable set, and $(0, 1] \sim A$ (we only removed a subset of A, consisting of those with finitely many 1's, which correspond to some rational numbers that are collectively at most countable). We can interprete $|A| = 2^{\aleph_0}$ as A is the set of binary sequences.

Example 1.33. The following statements hold:

- If $|A_n| = \aleph_1$ for all $n \geq 1$, then $|\cup_{n=1}^{\infty} A_n| = \aleph_1$. [Hint: $A_k \sim (k, k+1]$.]
- $|\mathbb{R}^n| = |\mathbb{R}| = \aleph_1$. [Hint: $A_k = \mathbb{R}$ for $k = 1, \ldots, n$.]

Theorem 1.34 (There is no "cap" on cardinal number)**.** *Suppse $A \neq \emptyset$, then $A \nsim 2^A := \{E : E \subset A\}$.*

Proof. If not, then there exists a one-to-one correspondence $f : A \to 2^A$. Let $B = \{x \in A : x \notin f(x)\}$. Since $B \in 2^A$, there exists $y \in A$ such that $f(y) = B$. If $y \in B$, then $y \notin f(y) = B$; if $y \notin B = f(y)$, then $y \in B$. Both yield contraditions. □

1.4 Topology of metric spaces

Definition 1.35 (Euclidean space and norm)**.** We denote $\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_i \in \mathbb{R}, \forall\, i\}$ the n-dimensional *Euclidean space*. The *norm* of $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is defined by $|x| = (x_1^2 + \cdots + x_n^2)^{1/2}$.

One can verify the following properties of norms:

- $|x| \geq 0$; $|x| = 0$ iff $x = (0, \ldots, 0)$.
- $|ax| = |a||x|$ for any $a \in \mathbb{R}$.
- $|x + y| \leq |x| + |y|$. [Use the Cauchy-Schwarz inequality below.]

Theorem 1.36 (Cauchy-Schwarz)**.** *Let $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, then there is $(\sum_{i=1}^n x_i y_i) \leq (\sum_{i=1}^n x_i^2)^{1/2} \cdot (\sum_{i=1}^n y_i^2)^{1/2}$. In addition, the equality holds iff $x = ay$ or $y = ax$ for some $a \in \mathbb{R}_+$.*

Proof. Note that $\lambda^2 + b\lambda + c \geq 0$ for all λ iff $b^2 \leq 4c$. Use this fact and that the quadratic function $f(\lambda) = \sum_{i=1}^n (x_i + \lambda y_i)^2 \geq 0$ for all λ. □

Definition 1.37 (Metric space)**.** Let X be a set. Then $d : X \times X \to \mathbb{R}$ is called a *distance* (or a *metric*) on X if the followings hold for all $x, y, z \in X$:

- $d(x, y) \geq 0$ for all $x, y \in X$; and $d(x, y) = 0$ iff $x = y$.
- $d(x, y) = d(y, x)$.
- $d(x, y) \leq d(x, z) + d(y, z)$.

A set X with a distance d is called a *metric space*, denoted by (X, d) or simply X. Throughout this class, we set $d(x, y) = |x - y|$ for $x, y \in \mathbb{R}^n$ by default.

Definition 1.38. There are a series of definitions given (X, d):

- $\mathrm{diam}(E) := \sup\{d(x, y) : x, y \in E\}$ is the *diameter* of E. E is said to be *bounded* if $\mathrm{diam}(E) < \infty$.
- For any $x \in X$ and $\delta > 0$, $B(x, \delta) := \{y \in X : d(x, y) < \delta\}$ is called the *open ball* with center x and radius δ. $\overline{B(x, \delta)} = \{y \in X : d(x, y) \leq \delta\}$ is the *closed ball*.
- x is called an *interior point* of E if there exists an open ball $B(x, \delta) \subset E$ (i.e., there exists $\delta > 0$ such that $B(x, \delta) \subset E$).
- E is called *open* if every point of E is an interior point. E is called *closed* if E^c is open. [It is easy to show that an open ball $B(x, \delta)$ is literally open by definition, and a closed ball is closed.]

- (Only in $\mathbb{R}^n$) Suppose $a_i < b_i$ for $i = 1, \ldots, n$, then $I = (a_1, b_1) \times \cdots \times (a_n, b_n)$ is called an *open box* in $\mathbb{R}^n$. The volume of I is denoted by $|I| = \prod_{i=1}^n (b_i - a_i)$.
- A sequence $\{x_k\}$ in X is said to *converge* to x if $\lim_{k\to\infty} d(x_k, x) = 0$ (or simply denoted by $x_k \to x$).
- A sequence $\{x_k\}$ in X is said to be *Cauchy* if for any $\epsilon > 0$, there exists N, such that $d(x_n, x_m) < \epsilon$ for all $n, m \geq N$.
- Let E be an infinite subset of X. If there exists a sequence of distinct points $\{x_k\}$ such that $x_k \to x$, then x is called a *limit point* (or *accumulation point*) of E. [Note that a limit point of E needs not be in E.]
- The set of limit points of E is denoted by E'. The union $\bar{E} := E \cup E'$ is called the *clousure* of E. [$\bar{E}$ is a closed set; see below.]
- If $A \subset B$ and $\bar{A} = B$, then A is called *dense in* B, or A is a *dense subset of* B.
- If $x \in E$ and x is not a limit point of E, then x is called an *isolated point* of E (i.e., $\exists\, \delta > 0$, such that $B(x, \delta) \cap E = \{x\}$).
- If G_α is open for every $\alpha \in I$ and $E \subset \cup_{\alpha\in I} G_\alpha$, then $\{G_\alpha : \alpha \in I\}$ is called an *open cover* of E.
- E is called *compact* if every open cover of E contains a finite subcover. [In $\mathbb{R}^n$, E is compact iff E is closed and bounded; see below.]

Theorem 1.39. *$x \in E'$ iff for any $\delta > 0$ there is $(B(x, \delta) \setminus \{x\}) \cap E \neq \emptyset$.*

Proof. Necessity is clear. Let $\delta_1 = 1$ and select $x_1 \in (B(x, \delta_1) \setminus \{x\}) \cap E$. Then for any $k \geq 1$, let $\delta_{k+1} = \frac{1}{2} d(x_k, x)$ and select $x_{k+1} \in (B(x, \delta_{k+1}) \setminus \{x\}) \cap E$, then we obtain a sequence $\{x_k\}$ which are distinct and $x_k \to x$, i.e., $x \in E'$. $\square$

Theorem 1.40. *E is closed iff $E' \subset E$.*

Proof. Suppose E is closed, then E^c is open. If $x \in E' \setminus E$, then $x \in E^c$ and there exists $\{x_k\} \subset E$ and $x_k \to x$. But this is a contradiction since x is an interior point of E^c.

Suppose $E' \subset E$. For any $x \in E^c$, we know $x \notin E'$, i.e., there exists $\delta > 0$ such that $B(x, \delta) \cap E = \emptyset$. Hence $B(x, \delta) \subset E^c$, i.e., x is an interior point of E^c. As x is arbitrary, we know E^c is open, and hence E is closed. $\square$

Theorem 1.41. *$\bar{E}$ is closed.*

Proof. For any $x \notin \bar{E} = E \cup E'$, there exists $\delta > 0$ such that $B(x, \delta) \cap E = \emptyset$. If $\exists\, y \in E'$ such that $y \in B(x, \delta)$, then there exists $\delta' > 0$ and $x' \in B(y, \delta') \subset B(x, \delta)$, contradiction. Hence $B(x, \delta) \cap E' = \emptyset$. Therefore $B(x, \delta) \subset (E \cup E')^c$, implying that $(E \cup E')^c$ is open. $\square$

Example 1.42. A few examples of limit points.
- Let $E = \{\frac{1}{n} : n \in \mathbb{N}\}$. Then $E' = \{0\}$. All points in E are isolated points.
- Let $E = \{\sqrt{m} - \sqrt{n} : m, n \in \mathbb{N}\}$. Then $E' = \mathbb{R}$. [Hint: for any $x \in \mathbb{R}$, let $x_n = \sqrt{\lfloor (x+n)^2 \rfloor} - \sqrt{n^2}$ (where $\lfloor x \rfloor := \max\{n \in \mathbb{Z} : n \leq x\}$). Then $\sqrt{(x+n)^2 - 1} - n < x_n < x$ and $x_n \to x$.]

Theorem 1.43. *Let $E_1, E_2 \subset \mathbb{R}^n$. Then $(E_1 \cup E_2)' = E_1' \cup E_2'$.*

Proof. It is clear that $E_j' \subset (E_1 \cup E_2)'$ for $j = 1, 2$. If $x \in (E_1 \cup E_2)'$, then there exists a sequence of distinct points $\{x_k\} \subset E_1 \cup E_2$, such that $x_k \to x$. Then at least one of E_1 and E_2 contains a subsequence of $\{x_k\}$ which also converges to x. Hence $(E_1 \cup E_2)' \subset E_1' \cup E_2'$. $\square$

Theorem 1.44 (Bolzano-Weierstrass). *Every bounded infinite set E of $\mathbb{R}^n$ has at least one limit point.*

Proof. Let E be contained in $\prod_{i=1}^n [a_i, b_i]$, then by focusing on the first components of the points in E we can extract a convergent sequence in $[a_1, b_1]$ (by Weierstrass theorem on $\mathbb{R}$) with limit c_1; then we focus on the second components of this sequence and extract a convergent subsequence with limit c_2, and so on, until we finish the n-th component with c_n. Then the sequence have distinct points and its limit is $c = (c_1, \ldots, c_n)$. $\square$

Theorem 1.45. *$f \in C(\mathbb{R}^n)$ iff for every $t \in \mathbb{R}$ the sets $E_1 = \{x \in \mathbb{R}^n : f(x) > t\}$ and $E_2 = \{x \in \mathbb{R}^n : f(x) < t\}$ are open.*

Proof. The necessity is clear. To show the sufficiency, suppose that for every t both E_1^c and E_2^c are closed. If f is not continuous at x_0, then there exists $\epsilon_0 > 0$ and a sequence $x_k \to x_0$ such that $|f(x_k) - f(x_0)| \geq \epsilon_0$. WLOG, suppose $f(x_k) \leq f(x_0) - \epsilon_0$ for all k. Then set $t = f(x_0) - \epsilon_0$, we know $E_1^c = \{x \in \mathbb{R}^n : f(x) \leq t\}$ is closed, which is a contradiction since $x_k \to x_0$ and $\{x_k\} \subset E_1^c$ but $x \notin E_1^c$. $\square$

Theorem 1.46 (Operations on open and closed sets). *Union of (finitely or infinitely many) open sets is open; Intersection of finitely many open sets is open. Contrary for closed sets. Namely,*

- *If F_α is closed and G_α is open for every $\alpha \in I$, then $\cap_{\alpha \in I} F_\alpha$ is closed, and $\cup_{\alpha \in I} G_\alpha$ is open.*
- *If F_k is closed and G_k is open for $k = 1, \ldots, n$, then $\cap_{k=1}^n F_k$ is closed, and $\cup_{k=1}^n G_k$ is open.*

Proof. We only show this for open sets. Then applying De Morgan's law implies those for closed sets.

For every $x \in \cup_{\alpha \in I} G_\alpha$, there exists $\alpha' \in I$ such that $x \in G_{\alpha'}$, and hence $\exists\, B(x, \delta) \subset G_{\alpha'} \subset \cup_{\alpha \in I} G_\alpha$.

For every $x \in \cap_{k=1}^n G_k$, there exist $\delta_k > 0$ such that $B(x, \delta_k) \subset G_k$ for every $k = 1, \ldots, n$. Let $\delta = \min\{\delta_k : k = 1, \ldots, n\} > 0$ (require finiteness!) then $B(x, \delta) \subset G_k$ for all k. $\square$

Definition 1.47 (G_δ-set and F_σ-set). We call $H = \cap_{k=1}^\infty G_k$ a *G_δ-set* if G_k is open for all k, and $K = \cup_{k=1}^\infty F_k$ an *F_σ-set* if F_k is closed for all k.

Theorem 1.48 (Compact sets are closed). *If K is a compact set, then K is closed.*

Proof. It suffices to show that K^c is open, i.e., every point of K^c is an interior point. For every $y \in K^c$ and $x \in K$, denote $\delta_x = \frac{1}{2}d(x,y) > 0$. Then $\{B(x,\delta_x) : x \in K\}$ is an open cover of K, and hence has a finite subcover $\{B(x_i,\delta_{x_i}) : 1 \le i \le k\}$ of K. Let $\delta := \min_{1\le i\le k} \delta_{x_i}$ (which is > 0), then $B(y,\delta) \cap B(x_i,\delta_{x_i}) = \emptyset$ for all i. Hence $B(y,\delta) \subset K^c$, i.e., y is an interior point of K^c. □

Theorem 1.49 (Closed subsets of a compact set are compact)**.** *If $F \subset K$, F is closed and K is compact, then F is compact.*

Proof. Let $\{G_\alpha : \alpha \in I\}$ be any open cover of F, then $\{G_\alpha, F^c : \alpha \in I\}$ is an open cover of X and hence also of K. As K is compact, there is a finite subcover $\{G_{\alpha_i}, F^c : 1 \le i \le k\}$ of K. Therefore $\{G_{\alpha_i} : 1 \le i \le k\}$ is a finite subcover of F by noting that F^c does not help covering F. □

Theorem 1.50 (Cantor)**.** *Suppose $F_k \neq \emptyset$ is compact for every k, and $F_1 \supset F_2 \supset \cdots \supset F_k \supset \dots$, then $\cap_{k=1}^{\infty} F_k \neq \emptyset$.*

Proof. Supose not, then $F_1 \subset X = (\cap_{k=1}^{\infty} F_k)^c = \cup_{k=1}^{\infty} F_k^c$. Therefore $\{F_k^c\}$ is an open cover of F_1. Hence there exists a finite subcover $\{F_{k_i}^c : 1 \le i \le l\}$ of F_1, i.e, $F_1 \subset \cup_{i=1}^{l} F_{k_i}^c$. Therefore $\cap_{i=1}^{l} F_{k_i} \subset F_1^c$. But $\cap_{i=1}^{l} F_{k_i} \subset F_1$. Hence $\cap_{i=1}^{l} F_{k_i} = \emptyset$, which is a contradiction since $F_k \neq \emptyset$ for all k. □

Lemma 1.51. *$I = [a_1,b_1] \times \cdots \times [a_n,b_n]$ is compact in $\mathbb{R}^n$.*

Proof. Suppose not, then I has an infinite open cover $\mathcal{A} = \{G_\alpha : \alpha\}$ which does not have a finite subcover. Perform bisection of each side of I, we obtain 2^n closed boxes, and at least one of them cannot be covered by finitely many open sets in $\mathcal{A}$. Hence we perform bisection of this box again, and obtain a smaller closed box that does not have a finite subcover either, and so on, which never ends. It is clear that the size of the box shrinks to 0 and converges to a point $x \in I$, which must be an interior point of some $G_{\alpha'}$, i.e., $\exists\,\delta > 0$ such that $B(x,\delta) \subset G_{\alpha'}$. Then we should have stopped within finitely many iterations of the bisection once the box is in $B(x,\delta)$ which is covered by $G_{\alpha'}$, a contradiction. □

Theorem 1.52 (Heire-Borel)**.** *Bounded closed sets in $\mathbb{R}^n$ are compact.*

Proof. Let F be closed and bounded. Then there exists a bounded closed box I such that $F \subset I$. Since I is compact and F is closed, we know F is compact. □

Example 1.53. Suppose $F \subset \mathbb{R}^n$ is closed and bounded, and $G \subset \mathbb{R}^n$ is open, and $F \subset G$. Then $\exists\,\delta > 0$ such that for every $x \in B(0,\delta)$, there is $F + \{x\} := \{y + x : y \in F\} \subset G$.

Proof. Since every $y \in F$ is an interior point of G, we know $\exists\,\delta_y > 0$ such that $B(y,\delta_y) \subset G$. Also $\{B(y,\frac{\delta_y}{2}) : y \in F\}$ is an open cover of F, and hence has a finite subcover $\{B(y_i,\frac{\delta_{y_i}}{2}) : 1 \le i \le k\}$. Namely, for every $y \in F$, we know $y \in B(y_i,\frac{\delta_{y_i}}{2})$ for some $i \in \{1,\dots,k\}$. Set $\delta = \min_{1\le i\le k} \frac{\delta_{y_i}}{2}$. Then for any

$x \in B(0,\delta)$ and $y \in F$, $x+y \in B(y_i, \delta_{y_i})$ for some $i \in \{1, \ldots, k\}$, and hence $x+y \in G$. $\square$

Before closing this section, we consider several examples of distances between sets.

Definition 1.54 (Distance between sets)**.** $d(x,E) := \inf\{d(x,y) : y \in E\}$ and $d(E_1, E_2) := \inf\{d(x,y) : x \in E_1,\ y \in E_2\}$.

Example 1.55. Suppose $E_1 = \{x = (x_1, x_2) \in \mathbb{R}^2 : x_2 = 0\}$ and $E_2 = \{x = (x_1, x_2) \in \mathbb{R}^2 : x_1 x_2 = 1\}$. Then $d(E_1, E_2) = 0$.

Theorem 1.56. *If $F \subset \mathbb{R}^n$ is nonempty and closed, and $x_0 \in \mathbb{R}^n$, then $\exists\, y \in F$ such that $d(x_0, F) = d(x_0, y)$.*

Proof. Choose $\delta > 0$ large enough such that $K = \overline{B(x_0, \delta)} \cap F$ is nonempty. Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) = d(x_0, x)$ for any $x \in \mathbb{R}^n$. Then f is continuous. Since K is compact, we know f attains its minimum on K at some $y \in K$. $\square$

Theorem 1.57. *Suppose $E \subset \mathbb{R}^n$ is nonempty. Then $f(x) : \mathbb{R}^n \to \mathbb{R}$ defined by $f(x) = d(x, E)$ is uniformly continuous.*

Proof. We can even show that $f(x)$ is Lipschitz continuous on $\mathbb{R}^n$, which implies uniform continuity. To this end, for any $x, y \in \mathbb{R}^n$ and $\epsilon > 0$, there exists $z \in E$, such that $d(y,z) - \epsilon < d(y,E) = f(y) \leq d(y,z)$. Hence $f(x) - f(y) < d(x,z) - (d(y,z) - \epsilon) = d(x,z) - d(y,z) + \epsilon \leq d(x,y) + \epsilon$. Since $\epsilon > 0$ is arbitrary, we know $f(x) - f(y) \leq d(x,y)$. Similiarly $f(y) - f(x) \leq d(x,y)$. Hence $|f(x) - f(y)| \leq d(x,y)$, i.e., f is 1-Lipschitz. $\square$

Corollary 1.58. *If F_1, F_2 are nonempty and closed, and at least one of them is bounded, then there exist $x_1 \in F_1$ and $x_2 \in F_2$, such that $d(x_1, x_2) = d(F_1, F_2)$.*

Proof. If $F_1 \cap F_2 \neq \emptyset$, then trivial. Otherwise, WLOG, suppose F_1 is bounded and hence is compact. Define $f(x) := d(x, F_2)$, then $f : \mathbb{R}^n \to \mathbb{R}$ is continuous and hence attains minimum over F_1 at some $x_1 \in F_1$. Note that there exists $\delta > 0$ such that $K = \overline{B(x_1, \delta)} \cap F_2$ is nonempty. Since K is compact (since K is closed and bounded) and $g(x) : \mathbb{R}^n \to \mathbb{R}$ defined by $g : x \mapsto d(x_1, x)$ is continuous, we know g attains minimum over K at some $x_2 \in K$. Hence $d(x_1, x_2) = g(x_2) = d(x_1, K) = d(x_1, F_2) = f(x_1) = d(F_1, F_2)$. $\square$

Example 1.59. Suppose $F_1, F_2 \subset \mathbb{R}^n$ are nonempty, closed, and disjoint. Then there exists a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ such that $0 \leq f \leq 1$, $F_1 = \{x : f(x) = 1\}$ and $F_2 = \{x : f(x) = 0\}$. [Hint: $f(x) = \frac{d(x,F_2)}{d(x,F_1)+d(x,F_2)}$.]

2 Measure and Measurable Sets

2.1 σ-algebra

Definition 2.1 (σ-algebra)**.** Let X be a nonempty set. Then $\Gamma \subset 2^X$ (i.e., Γ is a collection of subsets of X) is called a *σ-algebra of X* if:
1. $\emptyset \in \Gamma$;
2. If $A \in \Gamma$ then $A^c \in \Gamma$;
3. If $A_n \in \Gamma$ for $n = 1, 2, \dots$, then $\cup_{n=1}^{\infty} A_n \in \Gamma$.

Given the definition above, it is also easy to verify that the following statements hold if Γ is a σ-algebra of X:
1. $X \in \Gamma$;
2. If $A, B \in \Gamma$, then $A \setminus B \in \Gamma$;
3. If $A_k \in \Gamma$ for $k = 1, \dots, n$, then $\cup_{k=1}^{n} A_k \in \Gamma$;
4. If $A_n \in \Gamma$ for $k = 1, 2, \dots$, then $\cap_{k=1}^{\infty} A_k, \limsup_k A_k, \liminf_k A_k \in \Gamma$.

Definition 2.2 (Generated σ-algebra)**.** Suppose $\Sigma \subset 2^X$, and consider $\mathcal{A} = \{\Gamma : \Sigma \in \Gamma, \text{ and } \Gamma \text{ is a } \sigma\text{-algebra of } X\}$ (obviously $2^X \in \mathcal{A}$ and hence $\mathcal{A} \neq \emptyset$). Then $\Gamma(\Sigma) := \cap_{\Gamma \in \mathcal{A}} \Gamma$ is called *the σ-algebra of X generated by* Σ.

Remarks. It is easy to verify that $\Gamma(\Sigma)$ is a σ-algebra of X (i.e., $\Gamma(\Sigma) \in \mathcal{A}$): for example, if $A \in \Gamma(\Sigma)$, then $A \in \Gamma$ for all $\Gamma \in \mathcal{A}$. Since every Γ is a σ-algebra, $A^c \in \Gamma$. Therefore $A^c \in \Gamma(\Sigma) = \cap_{\Gamma \in \mathcal{A}} \Gamma$. Similar for the other two conditions. Hence $\Gamma(\Sigma)$ is the "smallest" σ-algebra of X containing Σ.

Definition 2.3 (Borel σ-algebra of $\mathbb{R}^n$)**.** Let $\Sigma = \{G \subset \mathbb{R}^n : G \text{ is open}\}$. Then the σ-algebra $\Gamma(\Sigma)$ generated by Σ, formally denoted by $\mathcal{B}(\mathbb{R}^n)$ or simply $\mathcal{B}$, is called the *Borel σ-algebra of* $\mathbb{R}^n$. A set $B \in \mathcal{B}$ is called a *Borel set*.

2.2 Outer measure

Definition 2.4 (Outer measure)**.** Let $\{I_k : k \in \mathbb{N}\}$ be a countable set of open boxes in $\mathbb{R}^n$. We call $\{I_k\}_k$ an open-box-cover of $E \subset \mathbb{R}^n$ if $E \subset \cup_{k=1}^{\infty} I_k$. Then the *outer measure* of E is defined by

$$\mu^*(E) = \inf \Big\{ \sum_{k=1}^{\infty} |I_k| : \{I_k\}_k \text{ is an open-box-cover of } E \Big\}$$

Remarks. If $\sum_k |I_k| = \infty$ for every open-box-cover of E, then we define $\mu^*(E) = \infty$; otherwise $\mu^*(E) < \infty$. Note that $\mu^*(E) \geq 0$ for any E.

Example 2.5. For any $x \in \mathbb{R}^n$, $\mu^*(\{x\}) = 0$. For any $t \in \mathbb{R}$ and $E = \{x = (x_1, \dots, x_n) : x_i = t, x_j \in \mathbb{R}, \ \forall\, j \neq i\}$, there is $\mu^*(E) = 0$.

Theorem 2.6. *Let I be an open box, then $\mu^*(\bar{I}) = |I|$. [Hint: WLOG consider $|I| < \infty$. For any $\epsilon > 0$, there exists an open box J such that $I \subset \bar{I} \subset J$ and $|J| < |I| + \epsilon$.]*

Theorem 2.7 (Properties of outer measure). *The outer measure μ^* in $\mathbb{R}^n$ has the following properties:*

1. $\mu^*(E) \geq 0$; $\mu^*(\emptyset) = 0$ *[Note that $\mu^*(E) = 0$ dose not imply $E = \emptyset$.]*
2. *If $E_1 \subset E_2$ then $\mu^*(E_1) \leq \mu^*(E_2)$.*
3. *(Sub-additivity) $\mu^*(\cup_{k=1}^{\infty} E_k) \leq \sum_{k=1}^{\infty} \mu^*(E_k)$.*

Proof. The first two are trivial. We only show the sub-additivity. For any $\epsilon > 0$ and any E_k, there exists an open-box-cover $\{I_{k,l} : l \in \mathbb{N}\}$ of E_k, such that

$$\sum_{l=1}^{\infty} |I_{k,l}| < \mu^*(E_k) + \frac{\epsilon}{2^k}.$$

Then $\{I_{k,l} : k, l \in \mathbb{N}\}$ is an open-box-cover of $E := \cup_{k=1}^{\infty} E_k$, and hence

$$\mu^*(E) \leq \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} |I_{k,l}| \leq \sum_{k=1}^{\infty} \Big(\mu^*(E_k) + \frac{\epsilon}{2^k}\Big) = \sum_{k=1}^{\infty} \mu^*(E_k) + \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we have $\mu^*(E) \leq \sum_{k=1}^{\infty} \mu^*(E_k)$. □

Example 2.8 (Countable sets have measure 0). If $E = \{x_k \in \mathbb{R}^n : k \in \mathbb{N}\}$, then $\mu^*(E) = 0$.

Example 2.9. Suppose $E \subset [a, b] \subset \mathbb{R}$ and $\mu^*(E) > 0$. Then for any $t \in (0, \mu^*(E))$, there exists $A \subset E$ such that $\mu^*(A) = t$. [Hint: Define $f(x) = \mu^*([a, x) \cap E)$ for every $x \in \mathbb{R}$. Then show that f is 1-Lipschitz on $\mathbb{R}$, i.e., $|f(x+\Delta x)-f(x)| \leq |\Delta x|$ for all $x, \Delta x \in \mathbb{R}$. Then apply the Intermediate Value Theorem of Continuous Functions to f.]

Theorem 2.10 (Outer measure is invariant of shifting). *For any $x \in \mathbb{R}^n$ and $E \subset \mathbb{R}^n$, there is $\mu^*(E) = \mu^*(E + x)$.*

Example 2.11. For any $\lambda \in \mathbb{R}$ and $E \subset \mathbb{R}$, there is $\mu^*(\lambda E) = |\lambda|\mu^*(E)$.

Example 2.12. If $\mu^*(A) = 0$, then $\mu^*(A \cup B) = \mu^*(B) = \mu^*(B \setminus A)$.

Proof. It follows from $\mu^*(B) \leq \mu^*(A \cup B) \leq \mu^*(A) + \mu^*(B) = \mu^*(B)$ and $\mu^*(B \setminus A) \leq \mu^*(B) \leq \mu^*(B \setminus A) + \mu^*(A) = \mu^*(B \setminus A)$. □

2.3 Measurable sets and Lebesgue measure

The problem with the outer measure is that there exist mutually disjoint E_k for $k = 1, 2, \ldots$, but $\mu^*(\cup_{k=1}^{\infty} E_k) < \sum_{k=1}^{\infty} \mu^*(E_k)$. We will restrict μ^* to the so-called measurable sets, such that $\mu^*(\cup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} \mu^*(E_k)$, i.e., μ^* is *countably additive*. Note that countable additivity implies finite additivity.

Definition 2.13 (Measurable set). A set $E \subset \mathbb{R}^n$ is called a *measurable set*, or simply that E is *measurable*, if for any $T \subset \mathbb{R}^n$, there is

$$\mu^*(T) = \mu^*(T \cap E) + \mu^*(T \cap E^c).$$

We call T a *test set* (note that it can be any set). The collection of all measurable sets in $\mathbb{R}^n$ is denoted by $\mathcal{M}(\mathbb{R}^n)$, or simply $\mathcal{M}$ in no danger of confusion.

Remarks. In order to show $E \in \mathcal{M}$, it suffices to show $\mu^*(T) \geq \mu^*(T \cap E) + \mu^*(T \cap E^c)$ for any test set T, since $\mu^*(T) \leq \mu^*(T \cap E) + \mu^*(T \cap E^c)$ is already implied by the sub-additivity of μ^*.

Example 2.14. If $\mu^*(E) = 0$, then $E \in \mathcal{M}$. [Hint: $\mu^*(T) \geq \mu^*(T \cap E^c) = \mu^*(T \cap E) + \mu^*(T \cap E^c)$ as $0 \leq \mu^*(T \cap E) \leq \mu^*(E) = 0$.]

Example 2.15. Let $E_1, E_2 \subset \mathbb{R}^n$ (not necessarily measurable). If there exists $S \in \mathcal{M}$ such that $E_1 \subset S$ and $E_2 \subset S^c$, then $\mu^*(E_1 \cup E_2) = \mu^*(E_1) + \mu^*(E_2)$.

Proof. Let $E_1 \cup E_2$ be the test set for S, then $\mu^*(E_1 \cup E_2) = \mu^*((E_1 \cup E_2) \cap S) + \mu^*((E_1 \cup E_2) \cap S^c) = \mu^*(E_1) + \mu^*(E_2)$. □

Theorem 2.16 (Properties of $\mathcal{M}$)**.** *The following statements hold for $\mathcal{M}$:*

1. *$\emptyset \in \mathcal{M}$.*
2. *If $E \in \mathcal{M}$, then $E^c \in \mathcal{M}$.*
3. *If $E_1, E_2 \in \mathcal{M}$, then $E_1 \cup E_2, E_1 \cap E_2, E_1 \setminus E_2 \in \mathcal{M}$.*
4. *If $E_k \in \mathcal{M}$ for all $k = 1, 2, \ldots$, then $\cup_{k=1}^{\infty} E_k \in \mathcal{M}$. If in addition that all E_k are mutually disjoint, then $\mu^*(\cup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} \mu^*(E_k)$.*

Proof. Items 1 and 2 are trivial. For Item 3, we only need to show $E_1 \cup E_2 \in \mathcal{M}$, since $E_1 \cap E_2 = (E_1^c \cup E_2^c)^c$ and $E_1 \setminus E_2 = E_1 \cap E_2^c$. For any test set T, consider $T \cap (E_1 \cup E_2)$, which can be partitioned into three mutually disjoint sets: $T \cap E_1 \cap E_2$, $T \cap E_1 \cap E_2^c$, and $T \cap E_1^c \cap E_2$ (use Venn diagram). Also note that $T \cap (E_1 \cup E_2)^c = T \cap E_1^c \cap E_2^c$. Therefore

$$\begin{aligned}
\mu^*(T) &= \mu^*(T \cap E_1) + \mu^*(T \cap E_1^c) \\
&= [\mu^*(T \cap E_1 \cap E_2) + \mu^*(T \cap E_1 \cap E_2^c)] + [\mu^*(T \cap E_1^c \cap E_2) \\
&\quad + \mu^*(T \cap E_1^c \cap E_2^c)] \\
&\geq \mu^*(T \cap (E_1 \cup E_2)) + \mu^*(T \cap E_1^c \cap E_2^c)
\end{aligned}$$

where the first equality is due to $E_1 \in \mathcal{M}$ with T as test set, the second equality is due to $E_2 \in \mathcal{M}$ and $T \cap E_1$ and $T \cap E_1^c$ as test sets, and the inequality is due to the sub-additivity of μ^* applied to the first three terms. Note that it is easy to show that $\sum_{i=1}^{k} E_k \in \mathcal{M}$ and $\mu^*(\cup_{i=1}^{k} E_k) = \sum_{i=1}^{k} \mu^*(E_k)$.

For Item 4, WLOG, we assume E_k are mutually disjoint; otherwise replace E_k by $F_k := E_k \setminus (\cup_{i=1}^{k-1} E_i)$ for $k \geq 2$ which are also in $\mathcal{M}$. Denote $S_k = \cup_{i=1}^{k} E_i$ and $S = \cup_{i=1}^{\infty} E_i$. Note that $S_k \in \mathcal{M}$ due to Item 3, and $S^c \subset S_k^c$. Therefore, for any T, there is

$$\mu^*(T) = \mu^*(T \cap S_k) + \mu^*(T \cap S_k^c) \geq \sum_{i=1}^{k} \mu^*(T \cap E_i) + \mu^*(T \cap S^c)$$

Letting $k \to \infty$, we obtain $\mu^*(T) \geq \sum_{i=1}^{\infty} \mu^*(T \cap E_i) + \mu^*(T \cap S^c) \geq \mu^*(T \cap S) + \mu^*(T \cap S^c)$. Hence $S \in \mathcal{M}$. Letting $T = S$ yields $\mu^*(S) \geq \sum_{k=1}^{\infty} \mu^*(S \cap E_k) = \sum_{k=1}^{\infty} \mu^*(E_k)$. □

The theorem above implies that $\mathcal{M}$, the collection of measurable sets in $\mathbb{R}^n$, is a σ-algebra of $\mathbb{R}^n$. This is formally called the Lebesgue measure of $\mathbb{R}^n$.

Definition 2.17 (Lebesgue measure)**.** For any $E \in \mathcal{M}$, the *Lesbesgue measure* μ of E is defined by $\mu(E) = \mu^*(E)$. The space $\mathbb{R}^n$, the set $\mathcal{M}$, and μ, constitute the *Lebesgue measure space* $(\mathbb{R}^n, \mathcal{M}, \mu)$ on $\mathbb{R}^n$.

Remarks. In general, for any set X and its σ-algebra $\mathcal{A}$, and an extended real-valued $\mu : \mathcal{A} \to \mathbb{R} \cup \{\infty\}$, we call $(X, \mathcal{A}, \mu)$ a *measurable space* if (i) $0 \le \mu(E) \le \infty$ for any $E \in \mathcal{A}$; (ii) $\mu(\emptyset) = 0$; and (iii) μ is countably additive, i.e., $\mu(\cup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} \mu(E_k)$ for any countable collection of mutually disjoint sets $\{E_k \in \mathcal{A} : k \in \mathbb{N}\}$. If $\mu(X) = \infty$, but $X = \cup_{k=1}^{\infty} E_k$ where $E_k \in \mathcal{M}$ and $\mu(E_k) < \infty$ for all k, then μ is called *σ-finite*.

Theorem 2.18 (Continuity of μ from below)**.** *Suppose $E_k \in \mathcal{M}$ for all $k = 1, 2, \dots,$ and $E_1 \subset E_2 \subset \dots,$ then there is $\mu(\lim_k E_k) = \lim_k \mu(E_k)$.*

Proof. If $\lim_k \mu(E_k) = \infty$ then trivial. Denote $E_0 = \emptyset$ and $D_k = E_k - E_{k-1} \in \mathcal{M}$ for all $k = 1, 2, \dots$. Then $\lim_k E_k = \cup_{k=1}^{\infty} E_k = \cup_{k=1}^{\infty} D_k$. Hence,

$$\mu\left(\lim_{k\to\infty} E_k\right) = \mu\left(\bigcup_{k=1}^{\infty} D_k\right) = \sum_{k=1}^{\infty} \mu(D_k) = \lim_k \sum_{j=1}^{k} \mu(D_j) = \lim_k \mu(E_k)$$

where we used the countable additivity of measures in the second equality. $\square$

Corollary 2.19 (Continuity of μ from above)**.** *Suppose $E_k \in \mathcal{M}$ for all $k = 1, 2, \dots,$ $\mu(E_1) < \infty$, and $E_1 \supset E_2 \supset \dots,$ then $\mu(\lim_k E_k) = \lim_k \mu(E_k)$.*

Proof. Denote $F_k = E_1 \setminus E_k$ for all k. Then $\emptyset = F_1 \subset F_2 \subset \cdots$. Therefore,

$$\begin{aligned}\mu(E_1) - \lim_{k\to\infty} \mu(E_k) = \lim_{k\to\infty} (\mu(E_1) - \mu(E_k)) = \lim_{k\to\infty} \mu(F_k) = \mu\left(\lim_{k\to\infty} F_k\right) \\ = \mu\left(\bigcup_{k=1}^{\infty} F_k\right) = \mu\left(\bigcup_{k=1}^{\infty} (E_1 \cap E_k^c)\right) = \mu\left(E_1 \cap (\bigcup_{k=1}^{\infty} E_k^c)\right) \\ = \mu\left(E_1 \cap (\bigcap_{k=1}^{\infty} E_k)^c\right) = \mu(E_1) - \mu\left(\lim_{k\to\infty} E_k\right).\end{aligned}$$

Hence $\mu(\lim_k E_k) = \lim_k \mu(E_k)$. $\square$

Remarks. We need the boundedness in Corollary 2.19: for example, let $E_k = [k, \infty)$ for every k, then $\lim_k E_k = \emptyset$ and $\mu(\lim_k E_k) = 0 \ne \infty = \lim_k \mu(E_k)$. So we need E_k to be bounded starting from some k (WLOG, $k = 1$ in Corollary 2.19). This boundedness is not required in Theorem 2.18 since E_k can grow unbounded and we will just get ∞ both sides.

Example 2.20. Suppose $E_k \in \mathcal{M}$ for all k and $\sum_{k=1}^{\infty} \mu(E_k) < \infty$, then $\mu(\limsup_k E_k) = 0$.

Proof. Let $B_k = \cup_{j=k}^{\infty} E_j$, then B_k is non-increasing and $\mu(B_k) \le \sum_{j=k}^{\infty} \mu(E_j) \to 0$ as $k \to \infty$. Hence there is $\mu(\limsup_k E_k) = \mu(\lim_k B_k) = \lim_k \mu(B_k) = 0$. $\square$

Corollary 2.21 (Fatou's lemma for measures). *Suppose $E_k \in \mathcal{M}$ for all k. Then there is $\mu(\liminf_k E_k) \le \liminf_k \mu(E_k)$.*

Proof. Let $B_k = \cap_{j\ge k} E_j$ for every $k \in \mathbb{N}$. Then B_k is non-decreasing, $B_k \subset E_k$ for all k, and hence $\mu(\liminf_k E_k) = \mu(\lim_k B_k) = \lim_k \mu(B_k) \le \liminf_k \mu(E_k)$. $\square$

Remarks. We have two remarks regarding the Fatou's lemma for measures.

- In general we do not have $\mu(\liminf_k E_k) = \liminf_k \mu(E_k)$. For example, let $E_k = [0, \frac{1}{2}]$ if k is odd and $(\frac{1}{2}, 1]$ if even, then $\liminf_k E_k = \emptyset$, and $\mu(\liminf_k E_k) = 0 < \frac{1}{2} = \liminf_k \mu(E_k)$ for all k.
- If $E_k \subset E$ for all k and $\mu(E) < \infty$, then we also have $\mu(\limsup_k E_k) \ge \limsup_k \mu(E_k)$ by substituting E_k with E_k^c for every k and observing that $(\limsup_k E_k)^c = \liminf_k E_k^c$.

Lemma 2.22 (Carathéodory). *Suppose G is an open proper subset of $\mathbb{R}^n$, $E \subset G$ (E needs not be in $\mathcal{M}$). Let $E_k = \{x \in E : d(x, G^c) \ge \frac{1}{k}\}$ for every $k \in \mathbb{N}$. Then $\lim_k \mu^*(E_k) = \mu^*(E)$.*

Proof. It is clear that $E_1 \subset E_2 \subset \cdots \subset E_k \subset \cdots \subset E$. Hence $\mu^*(E_k) \le \mu^*(E)$ for all k. For any $x \in E \subset G$, x is an interior point of G, and hence there exists $\delta > 0$ such that $B(x, \delta) \subset G$, i.e., $d(x, G^c) \ge \delta$. Hence $E = \cup_{k=1}^\infty E_k$. WLOG, assume $\mu^*(E) < \infty$.

Let $D_k = E_{k+1} \setminus E_k$ for $k = 1, 2, \dots$, then we have

$$\infty > \mu^*(E) \ge \mu^*(E_{2k}) \ge \mu^*(\bigcup_{j=1}^{k} D_{2j}) = \sum_{j=1}^{k} \mu^*(D_{2j}), \quad \forall\, k \in \mathbb{N},$$

where we used the fact that $d(D_{2i}, D_{2j}) > 0$ for all $i < j \le k$ to obtain the equality (μ^* is additive if two sets are separated with a positive distance). Hence $\sum_{j=k+1}^\infty \mu^*(D_{2j}) \to 0$ as $k \to \infty$. Similarly, $\sum_{j=k}^\infty \mu^*(D_{2j+1}) \to 0$ as $k \to \infty$. Therefore, for any $\epsilon > 0$, there exists k large enough, such that $\sum_{j=k+1}^\infty \mu^*(D_{2j}) < \frac{\epsilon}{2}$ and $\sum_{j=k}^\infty \mu^*(D_{2j+1}) < \frac{\epsilon}{2}$. On the other hand, note that $E = E_{2k} \cup (\cup_{j=k+1}^\infty D_{2j}) \cup (\cup_{j=k}^\infty D_{2j+1})$, we know that

$$\begin{aligned}\mu^*(E) &\le \mu^*(E_{2k}) + \sum_{j=k+1}^\infty \mu^*(D_{2j}) + \sum_{j=k}^\infty \mu^*(D_{2j+1}) \\ &< \mu^*(E_{2k}) + \epsilon \le \lim_{k\to\infty} \mu^*(E_k) + \epsilon.\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we have $\mu^*(E) \le \lim_k \mu^*(E_k)$. $\square$

Theorem 2.23 (Closed sets are measurable). *If $F \subset \mathbb{R}^n$ is closed, then $F \in \mathcal{M}$.*

Proof. For any test set $T \subset \mathbb{R}^n$, consider $T \cap F^c$, which is a subset of the open set F^c. Denote $F_k = \{x \in T \cap F^c : d(x, F) \ge \frac{1}{k}\}$. Then by Lemma 2.22, $\lim_k \mu^*(F_k) = \mu^*(T \cap F^c)$. Therefore

$$\begin{aligned}\mu^*(T) &= \mu^*\big((T \cap F) \cup (T \cap F^c)\big) \ge \mu^*((T \cap F) \cup F_k) \\ &= \mu^*(T \cap F) + \mu^*(F_k) \to \mu^*(T \cap F) + \mu^*(T \cap F^c)\end{aligned}$$

as $k \to \infty$, where we used the fact that $d(T \cap F, F_k) \geq \frac{1}{k} > 0$ to obtain the second equality above. Hence $\mu^*(T) \geq \mu^*(T \cap F) + \mu^*(T \cap F^c)$. So $F \in \mathcal{M}$. $\square$

Corollary 2.24 (Open sets are measurable). *If $G \subset \mathbb{R}^n$ is open, then $G \in \mathcal{M}$.*

Corollary 2.25 (Borel sets are measurable). *If $A \in \mathcal{B}(\mathbb{R}^n)$, then $A \in \mathcal{M}$.*

Proof. Since $\mathcal{B}(\mathbb{R}^n)$ is the smallest σ-algebra generated by the family of open sets, and $\mathcal{M}$ is a σ-algebra also containing all open sets due to Corollary 2.24, we know $\mathcal{B}(\mathbb{R}^n) \subset \mathcal{M}$. $\square$

Theorem 2.26. *Suppose $E \in \mathcal{M}$. For any $\epsilon > 0$, there exist an open set G such that $E \subset G$ and $\mu(G \setminus E) < \epsilon$, and a closed set F such that $F \subset E$ and $\mu(E \setminus F) < \epsilon$.*

Proof. First assume $\mu(E) < \infty$. Then there exists an open-box-cover $\{I_k\}$ of E such that $G = \cup_{k=1}^{\infty} I_k$ and $\mu(G) < \mu(E) + \epsilon$. Hence $0 \leq \mu(G \setminus E) = \mu(G) - \mu(E) < \epsilon$.

If $\mu(E) = \infty$, then denote $E_k = E \cup B(0, k)$ for every $k \in \mathbb{N}$. Then $\mu(E_k) < \infty$ for all k, and $E = \cup_{k=1}^{\infty} E_k$. For every k, there exists G_k such that $E_k \subset G_k$ and $\mu(G_k \setminus E_k) < \frac{\epsilon}{2^k}$. Now let $G = \cup_{k=1}^{\infty} G_k$, then $E \subset G$ and $G \setminus E = \cup_{k=1}^{\infty}(G_k \setminus E)$. Hence

$$\mu(G \setminus E) \leq \sum_{k=1}^{\infty} \mu(G_k \setminus E) \leq \sum_{k=1}^{\infty} \mu(G_k \setminus E_k) < \sum_{k=1}^{\infty} \frac{\epsilon}{2^k} = \epsilon.$$

For any $E \in \mathcal{M}$, we know $E^c \in \mathcal{M}$. Hence there exists an open set G, and therefore a closed set $F = G^c$, such that $\mu(E \setminus F) = \mu(E \cap F^c) = \mu(E \cap G) = \mu(G \setminus E^c) < \epsilon$. $\square$

Theorem 2.27. *Suppose $E \in \mathcal{M}$. Then there eixsts a G_δ-set H such that $E \subset H$ and $\mu(H \setminus E) = 0$. Similarly, there exists an F_σ-set K such that $K \subset E$ and $\mu(E \setminus K) = 0$.*

Proof. There exists an open set G_k such that $E \subset G_k$ and $\mu(G_k \setminus E) < \frac{1}{k}$ for every $k \in \mathbb{N}$. Let $H = \cap_{k=1}^{\infty} G_k$ which is a G_σ-set, then $E \subset H$ and $\mu(H \setminus E) \leq \mu(G_k \setminus E) < \frac{1}{k}$ for all k. Hence $\mu(H \setminus E) = 0$. $\square$

Theorem 2.28. *For any $E \subset \mathbb{R}^n$ (needs not be in $\mathcal{M}$), there exists a G_δ-set H such that $E \subset H$ and $\mu(H) = \mu^*(E)$.*

Proof. There exists an open G_k such that $E \subset G_k$ and $\mu(G_k) < \mu^*(E) + \frac{1}{k}$ for every $k \in \mathbb{N}$. Let $H = \cap_{k=1}^{\infty} G_k$ then $E \subset H$ and $\mu^*(E) \leq \mu(H) \leq \mu^*(E) + \frac{1}{k}$ for all k. Hence $\mu(H) = \mu^*(E)$. $\square$

Remarks. However, there may not exist an F_σ-set K such that $K \subset E$ and $\mu(K) = \mu^*(E)$ if $E \notin \mathcal{M}$. See Example 2.35 below.

Theorem 2.29. *Suppose $E_k \subset \mathbb{R}^n$ but need not be in $\mathcal{M}$ for all $k \in \mathbb{N}$. Then $\mu^*(\liminf_k E_k) \leq \liminf_k \mu^*(E_k)$.*

Proof. By Theorem 2.28, for every E_k, there exists a G_σ-set H_k (hence $H_k \in \mathcal{M}$), such that $E_k \subset H_k$ and $\mu(H_k) = \mu^*(E_k)$. Hence $\mu^*(\liminf_k E_k) \le \mu(\liminf_k H_k) \le \liminf_k \mu(H_k) = \liminf_k \mu^*(E_k)$. □

Corollary 2.30. *Suppose $E_k \subset \mathbb{R}^n$ but need not be in $\mathcal{M}$, and $E_1 \subset E_2 \subset \cdots \subset E_k \subset \cdots$, then $\mu^*(\lim_k E_k) = \lim_k \mu^*(E_k)$.*

Proof. Note that $\mu^*(\lim_k E_k) \le \lim_k \mu^*(E_k)$ by Theorem 2.29. The converse is obvious because $E_k \subset \lim_k E_k$ for all k. □

Example 2.31 (Cantor set)**.** Consider the closed interval $[0, 1] \subset \mathbb{R}$. Let $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ by removing the open middle third $(\frac{1}{3}, \frac{2}{3})$ from $[0, 1]$. Then let $C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$, by removing the middle thirds from both of $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$, and so on. Then C_k is compact for every k, and $C_1 \supset C_2 \supset \cdots$. Then $C = \cap_{k=1}^\infty C_k$ is called the *Cantor set*. In addition, C has the following properties: (i) C is compact, nowhere dense, totally disconnected, and has no isolated point; (ii) $\mu(C) = 0$; and (iii) $|C| = \mathfrak{c}$.

Proof. (i) Note that C is closed and bounded, hence compact. The remaining three are easy to check; (ii) The total length of intervals removed is $\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots = \sum_{k=1}^\infty \frac{2^{k-1}}{3^k} = 1$; (iii) It can be shown that $C = \{\sum_{k=1}^\infty \frac{a_k}{3^k} : a_k = \{0, 2\}\}$, hence $|C| = \mathfrak{c}$ (see comment below Definition 1.32). □

2.4 Non-measurable sets

We provide example of *non-measurable* sets. We have shown that $E \in \mathcal{M}$ if $\mu^*(E) = 0$. Hence a non-measurable set must have positive outer measure. To show that there are non-measurable sets, we consider $\mathbb{R}$ in this subsection, and recall that $\lambda + E = \{\lambda + x : x \in E\}$ is the translation set of $E \subset \mathbb{R}$ by $\lambda \in \mathbb{R}$. Note that $\mu^*(\lambda + E) = \mu^*(E)$ since (outer) measure is translation invariant.

Lemma 2.32. *Suppose $E \in \mathcal{M}$ and $\mu(E) < \infty$. If there exists a bounded countably infinite set $\Lambda \subset \mathbb{R}$ such that the collection of translation sets $\{\lambda + E : \lambda \in \Lambda\}$ are disjoint, then $\mu(E) = 0$.*

Proof. Since $\{\lambda + E : \lambda \in \Lambda\}$ is a collection of countably many disjoint sets, by the countable additivity of measures, we know $\mu(\cup_{\lambda\in\Lambda}(\lambda + E)) = \sum_{\lambda\in\Lambda} \mu(\lambda + E) = \sum_{\lambda\in\Lambda} \mu(E)$. Since both Λ and E are bounded, we know $\cup_{\lambda\in\Lambda}(\lambda + E)$ is bounded and hence $\mu(\cup_{\lambda\in\Lambda}(\lambda + E)) < \infty$. If $\mu(E) > 0$, then $\mu(\cup_{\lambda\in\Lambda}(\lambda + E)) = \sum_{\lambda\in\Lambda} \mu(E) = \infty$, which is a contradiction. Hence $\mu(E) = 0$. □

For any $E \subset \mathbb{R}$, we define the equivalence relation $\sim$ for any $x, y \in E$: $x \sim y$ iff $x - y \in \mathbb{Q}$ (it is easy to verify that $\sim$ is an equivalence relation by checking that it is reflexive, symmetric, and transitive). Then E is a disjoint union of its equivalence classes. Let $C_E \subset E$ be the set that contains *exactly one point* from each equivalence class in E. Then we know the two properties below hold:

1. For any $x, y \in C_E$, $x - y \notin \mathbb{Q}$ since x, y belong to different equivalence classes of E.

2. For any $x \in E$, there exists $c \in C_E$ and $q \in \mathbb{Q}$ such that $x = c + q$, since C_E contains a point from the same equivalence class which x belongs to.

Note that the first property also implies that $\{\lambda + C_E : \lambda \in \Lambda\}$ is a collection of disjoint sets provided that $\Lambda \subset \mathbb{Q}$. Now the following theorem shows that non-measurable sets exist.

Theorem 2.33 (Vitali: non-measurable sets exist)**.** *Any set E with $\mu^*(E) > 0$ contains non-measurable subset.*

Proof. WLOG, we assume $0 < \mu^*(E) < \infty$ and $E \subset [-b, b] \subset \mathbb{R}$ (otherwise we take $B(0, b) \cap E$ for some $b \in \mathbb{N}$). Let $\Lambda = [-2b, 2b] \cap \mathbb{Q}$. Let C_E be the subset of E containing exactly one point from each equivalence class, then $\{\lambda + C_E : \lambda \in \Lambda\}$ is a collection of countably many disjoint sets.

If $C_E \in \mathcal{M}$, then by Lemma 2.32, we know $\mu(C_E) = 0$. Now for any $x \in E$, there exists $c \in C_E$ and $q \in \mathbb{Q}$, such that $q = x - c$. Since $C_E \subset E \subset [-b, b]$, we know $q \in [-2b, 2b]$ and hence $q \in \Lambda$. Therefore $x \in \cup_{\lambda\in\Lambda}(\lambda + C_E)$. As x is arbitrary, we know $E \subset \cup_{\lambda\in\Lambda}(\lambda + C_E)$. However $0 < \mu^*(E) \le \mu(\cup_{\lambda\in\Lambda}(\lambda + C_E)) = \sum_{\lambda\in\Lambda} \mu(\lambda + C_E) = \sum_{\lambda\in\Lambda} \mu(C_E) = 0$, which is a contradiction. Hence $C_E \notin \mathcal{M}$. □

Theorem 2.34. *There exist disjoint sets $A, B \subset \mathbb{R}$ such that $\mu^*(A \cup B) < \mu^*(A) + \mu^*(B)$.*

Proof. If not, then for any $E \notin \mathcal{M}$ and $T \subset \mathbb{R}$, there is $\mu^*(T) \ge \mu^*(T \cap E) + \mu^*(T \cap E^c)$ since $T \cap E$ and $T \cap E^c$ are disjoint. This implies $E \in \mathcal{M}$ by definition, a contradiction. □

Example 2.35. Let $E = [0, 1) \subset \mathbb{R}$ and C_E be the subset of E containing exactly one point from each equivalence class in E as above. Let $D_\lambda = \{x + \lambda \pmod 1 : x \in C_E\}$ for every $\lambda \in \Lambda := [0, 1) \cap \mathbb{Q}$. Then $E = \cup_{\lambda\in\Lambda} D_\lambda$ is a countable disjoint union, $\mu^*(D_\lambda) = \mu^*(C_E) > 0$ for all $\lambda \in \Lambda$, and

$$1 = \mu^*([0, 1)) = \mu^*(E) = \mu^*(\bigcup_{\lambda\in\Lambda} D_\lambda) < \sum_{\lambda\in\Lambda} \mu^*(D_\lambda) = \infty.$$

Moreover, for any measurable subset $K \subset C_E$, we know $\{\lambda + K : \lambda \in \Lambda\}$ is a family of measurable and disjoint sets, $\mu(\lambda+K) = \mu(K)$, and $\infty > \mu^*(\cup_{\lambda\in\Lambda}(\lambda+C_E)) \ge \mu(\cup_{\lambda\in\Lambda}(\lambda + K)) = \sum_{\lambda\in\Lambda} \mu(\lambda + K)$. Hence $\mu(K) = 0 < \mu^*(C_E)$ (see remark after Theorem 2.28).

3 Measurable Functions

3.1 Extended real numbers

Recall the following properties of the extended real numbers $\bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$:

1. For any $x \in \mathbb{R}$, there is $-\infty < x < \infty$.
2. For any $x \in \mathbb{R}$, there are

$$
\begin{aligned}
x + (+\infty) &= (+\infty) + x = (+\infty) + (+\infty) = +\infty \\
x - (+\infty) &= (-\infty) + x = (-\infty) + (-\infty) = -\infty \\
x - (+\infty) &= (-\infty) - (+\infty) = -\infty \\
x - (-\infty) &= (+\infty) - (-\infty) = +\infty \\
\pm(+\infty) &= \pm\infty \\
\pm(-\infty) &= \mp\infty \\
\pm(\mp\infty) &= -\infty \\
|\pm\infty| &= +\infty
\end{aligned}
$$

3. For any $x \in \mathbb{R}$ and $x \neq 0$, we denote $\operatorname{sign}(x) = 1$ if $x > 0$ and -1 if $x < 0$. Then there are $x \cdot (\pm\infty) = \pm\operatorname{sign}(x) \cdot \infty$, and

$$
(\pm\infty) \cdot (\pm\infty) = +\infty, \quad (\pm\infty) \cdot (\mp\infty) = -\infty
$$

4. $(\pm\infty) - (\pm\infty)$ and $\pm(\infty) + (\mp\infty)$ are not defined. We define $0 \cdot (\pm\infty) = 0$ in this course.

Definition 3.1 (Measureable function)**.** Suppose $E \subset \mathbb{R}^n$. We call the function $f : E \to \bar{\mathbb{R}}$ *measurable*, or f is a *measurable function on* E, if $\{x \in E : f(x) > t\} \in \mathcal{M}$ for any $t \in \mathbb{R}$.

It turns out that we only need to show $\{x \in E : f(x) > r\} \in \mathcal{M}$ for all r in a dense set D of $\mathbb{R}$, for example $D = \mathbb{Q}$.

Theorem 3.2. *Suppose $f : E \to \bar{\mathbb{R}}$ and D is a dense subset of $\mathbb{R}$. If $\{x \in E : f(x) > r\} \in \mathcal{M}$ for any $r \in D$, then f is measurable.*

Proof. For any $t \in \mathbb{R}$, there exist $\{r_k\} \subset D$ such that $r_k \downarrow t$. Hence $\{x \in E : f(x) > t\} = \cup_{k=1}^{\infty} \{x \in E : f(x) > r_k\} \in \mathcal{M}$. □

Example 3.3. If $f : [a, b] \to \bar{\mathbb{R}}$ is monotone, then f is measurable.

Proof. WLOG, assume f is non-decreasing. Then for any $t \in \mathbb{R}$, $\{x \in [a, b] : f(x) > t\}$ is either the empty set, or a single point, or a subinterval of $[a, b]$, which are all in $\mathcal{M}$. □

Theorem 3.4. *If $f : E \to \bar{\mathbb{R}}$ is measurable, then the sets on the left hand side below are all measurable sets for any $t \in \mathbb{R}$:*

1. $\{x \in E : f(x) \le t\} = E \setminus \{x \in E : f(x) > t\}$.
2. $\{x \in E : f(x) \ge t\} = \cap_{k=1}^{\infty} \{x \in E : f(x) > t - \frac{1}{k}\}$.

3. $\{x \in E : f(x) < t\} = E \setminus \{x \in E : f(x) \geq t\}$.

4. $\{x \in E : f(x) = t\} = \{x \in E : f(x) \geq t\} \cap \{x \in E : f(x) \leq t\}$.

5. $\{x \in E : f(x) < \infty\} = \cup_{k=1}^{\infty} \{x \in E : f(x) < k\}$.

6. $\{x \in E : f(x) = +\infty\} = E \setminus \{x \in E : f(x) < +\infty\}$.

7. $\{x \in E : f(x) > -\infty\} = \cup_{k=1}^{\infty} \{x \in E : f(x) > -k\}$.

8. $\{x \in E : f(x) = -\infty\} = E \setminus \{x \in E : f(x) > -\infty\}$.

Remarks. We can use any of the first three as the definition of measurable sets.

Theorem 3.5. *If $f : E_1 \cup E_2 \to \bar{\mathbb{R}}$, and f is measurable on E_1 and E_2, then f is measurable on $E_1 \cup E_2$.*

Proof. $\{x \in E_1 \cup E_2 : f(x) > t\} = \{x \in E_1 : f(x) > t\} \cup \{x \in E_2 : f(x) > t\} \in \mathcal{M}$. □

Theorem 3.6. *If $f : E \to \bar{\mathbb{R}}$, and $A \subset E$ is measurable, then f is measurable on A.*

Proof. $\{x \in A : f(x) > t\} = A \cap \{x \in E : f(x) > t\} \in \mathcal{M}$. □

Example 3.7. Suppose $E \in \mathcal{M}$. Then $\chi_E : \mathbb{R}^n \to \bar{\mathbb{R}}$ is measurable.

Proof. Note that all of the three sets below are measurable:

$$\{x \in \mathbb{R}^n : \chi_E(x) > t\} = \begin{cases} \mathbb{R}^n & \text{if } t < 0 \\ E & \text{if } 0 \leq t < 1 \\ \emptyset & \text{if } t \geq 1 \end{cases}$$

□

Theorem 3.8. *Suppose $f, g : E \to \mathbb{R}$ are measurable. Then cf, $f + g$, $f \cdot g$ are measurable for any $c \in \mathbb{R}$.*

Proof. If $c = 0$ then trivial. If $c > 0$ then $\{x \in E : cf(x) > t\} = \{x \in E : f(x) > t/c\} \in \mathcal{M}$. Similarly for $c < 0$.

Let $\mathbb{Q} = \{r_k\}$, then $\{x : f + g > t\} = \cup_{k=1}^{\infty} \{x : f > r_k\} \cap \{x : g > t - r_k\} \in \mathcal{M}$.

We first can show f^2 is measurable: $\{x : f^2 > t\} = \{x : f > \sqrt{t}\} \cup \{x : f < -\sqrt{t}\}$ if $t \geq 0$ and $\{x : f^2 > t\} = E$ if $t < 0$. In either case, $\{x : f^2 > t\} \in \mathcal{M}$, hence f^2 is measurable. Then note that $f \cdot g = \frac{1}{4}((f + g)^2 - (f - g)^2)$. □

The result above can be generalized to functions $f : E \to \bar{\mathbb{R}}$.

Theorem 3.9. *If $f_k : E \to \bar{\mathbb{R}}$ is measurable for every k, then $\sup_k f_k(x)$, $\inf_k f_k(x)$, $\limsup_k f_k(x)$, $\liminf_k f_k(x)$ are all measurable functions.*

Proof. Denote $f = \sup_k f_k$. Then for any $t \in \mathbb{R}$, $\{x : f > t\} = \cup_{k=1}^{\infty} \{x : f_k > t\} \in \mathcal{M}$. The other three can be verified by noting that $\inf_k f_k = -\sup_k(-f_k)$, $\limsup_k f_k = \lim_j \sup_{k \geq j} f_k = \inf_j \sup_{k \geq j} f_k$ and $\liminf_k f_k = -\limsup_k(-f_k)$. □

Corollary 3.10. *If $f_k, f : E \to \bar{\mathbb{R}}$, f_k is measurable, and $f_k \to f$ for every $x \in E$, then f is measurable.*

Example 3.11. Suppose that $f : E \to \bar{\mathbb{R}}$ is measurable. Define $f^+(x) = \max(f(x), 0)$ and $f^-(x) = \max(-f(x), 0)$ for all $x \in E$. Then f^+, f^- are measurable.

It is clear that f is measurable iff f^+, f^- are measurable.

Example 3.12. If $f : E \to \bar{\mathbb{R}}$ is measurable, then $|f|$ is measurable. However the converse may not be true. [Hint: $f(x) = 2(\chi_E(x) - \frac{1}{2})$ for some $E \notin \mathcal{M}$.]

Example 3.13. Suppse $f : \mathbb{R}^2 \to \mathbb{R}$. $f(x, y)$ is continuous in y for every x and measurable in x for every y. Then f is measurable on $\mathbb{R}^2$.

Proof. For every $k \in \mathbb{N}$, define $f_k(x, y) = f(x, \frac{j}{k})$ for $\frac{j-1}{k} < y \leq \frac{j}{k}$. Then for every $t \in \mathbb{R}$, there is

$$\{(x, y) : f_k(x, y) > t\} = \bigcup_{j=-\infty}^{\infty} \left\{x : f_k\left(x, \frac{j}{k}\right) > t\right\} \times \left(\frac{j-1}{k}, \frac{j}{k}\right]$$

which implies that f_k is measurable. Then $\lim_k f(x, y) = f(x, y)$ for all y implies that $f(x, y)$ is measurable. $\square$

Example 3.14 (Continuous functions are measurable)**.** If $E \in \mathcal{M}$ and $f \in C(E)$, then f is measurable. [Hint: for every t, $\{x : f(x) > t\}$ is open.]

Definition 3.15. Let $E \subset \mathbb{R}^n$. We say a property P holds *almost everywhere (a.e. in short)* if there exists $A \subset E$ such that $\mu(E \setminus A) = 0$ and $P(x)$ holds for every $x \in A$.

Example 3.16. If $f, g : E \to \bar{\mathbb{R}}$ and $\mu(\{x \in E : f(x) \neq g(x)\}) = 0$, then $f = g$ a.e. E.

Definition 3.17. If $f : E \to \bar{\mathbb{R}}$ and $\mu(\{x \in E : |f(x)| = +\infty\}) = 0$, then $|f| < \infty$ a.e. E, or f is finite a.e. E.

Remarks. Note that this is different from "bounded by M a.e. E", which is $\mu(\{x \in E : |f(x)| > M\}) = 0$.

Theorem 3.18. *If $f, g : E \to \bar{\mathbb{R}}$, $f = g$ a.e. E, and f is measurable, then g is measurable.*

Proof. Let $A = \{x : f = g\}$, then $\mu(E \setminus A) = 0$ and $A \in \mathcal{M}$. Therefore $\{x \in E : g(x) > t\} = \{x \in A : g(x) > t\} \cup \{x \in A^c : g(x) > t\} = \{x \in A : f(x) > t\} \cup \{x \in A^c : g(x) > t\} \in \mathcal{M}$ since both sets are measurable. $\square$

Example 3.19. Suppose $E \in \mathcal{M}$ and $0 < \mu(E) < \infty$. If $0 < f < \infty$ a.e. E, then for any $\delta > 0$, there exits $E_\delta \subset E$ and $K > 0$, such that $\mu(E_\delta) < \delta$ and $\frac{1}{K} \leq f(x) \leq K$ for all $x \in E \setminus E_\delta$.

Proof. Define $A_k = \{x \in E : \frac{1}{k} \le f(x) \le k\}$. Then it is clear that A_k is non-decreasing. Let $A = \cup_{k=1}^{\infty} A_k$. Then $E = A \cup Z_0 \cup Z_1$ where $Z_0 = \{x \in E : f(x) = 0\}$ and $Z_1 = \{x \in E : f(x) = \infty\}$. Note that $\mu(Z_0) = \mu(Z_1) = 0$. Hence $\mu(E) = \mu(E \setminus (Z_0 \cup Z_1)) = \mu(A) = \mu(\cup_{k=1}^{\infty} A_k) = \lim_k \mu(A_k)$. Thus there exists K, such that for $E_\delta = Z_0 \cup Z_1 \cup A_K^c$, there are $\mu(E_\delta) < \delta$ and $\frac{1}{K} \le f(x) \le K$ for all $x \in A_K = E \setminus E_\delta = E \setminus (Z_0 \cup Z_1 \cup A_K^c)$. □

3.2 Simple functions

Definition 3.20 (Simple function)**.** A function $f : E \to \bar{\mathbb{R}}$ is called *simple* if the range set $f(E)$ is finite. That is, E can be partitioned into p mutually disjoint sets $E_1, \dots, E_p$, such that $f(x) = \sum_{i=1}^{p} c_i \chi_{E_i}(x)$ for p distinct values $c_1, \dots, c_p \in \bar{\mathbb{R}}$.

It is clear that, if $E_i \in \mathcal{M}$ for all i, then the simple function f is measurable.

Theorem 3.21 (Pointwise convergence to a nonnegative measurable function by a monotone sequence of simple functions)**.** *If $f : E \to \bar{\mathbb{R}}$, f is measurable, and $f \ge 0$, then there exists a sequence of non-decreasing simple functions $\{f_k\}$ such that $\lim_k f_k(x) = f(x)$ for every $x \in E$.*

Proof. For every $k \in \mathbb{N}$, we partition $[0, \infty]$ into $[0, k)$ and $[k, \infty]$, and further partition $[0, k)$ into $k \cdot 2^k$ segments of length $\frac{1}{2^k}$. Then define $E_{k,j} = \{x \in E : \frac{j-1}{2^k} \le f(x) < \frac{j}{2^k}\}$ for $j = 1, \dots, k \cdot 2^k$ and $E_k = \{x \in E : f(x) \ge k\}$, and a simple function f_k as follows:

$$f_k(x) = \begin{cases} \frac{j-1}{2^k}, & \text{if } x \in E_{k,j},\ j = 1, \dots, k \cdot 2^k, \\ k, & \text{if } x \in E_k. \end{cases}$$

Then it is clear that $f_k \le f$ and f_k is non-decreasing. For any $x \in E$, if $f(x) < \infty$, then there exists an integer $K > f(x)$, and hence $0 \le f(x) - f_k(x) \le \frac{1}{2^k}$ for all $k \ge K$, and hence $\lim_k f_k(x) = f(x)$; if $f(x) = \infty$, then $f_k(x) = k$ and hence $\lim_k f_k(x) = \infty = f(x)$. □

Theorem 3.22 (Pointwise (uniform) convergence to a (uniformly bounded) measurable function by a sequence of simple functions)**.** *If $f : E \to \bar{\mathbb{R}}$ is measurable, then there exists a sequence of simple functions $\{f_k\}$ such that $|f_k| \le f$ and $f_k \to f$. If in addition $|f| \le M$, then $f_k \rightrightarrows f$.*

Proof. The first statement can be verified by noting that $f = f^+ - f^-$. If $|f| \le M$, then for any $\epsilon > 0$, by the same construction above, there exists an integer $K \ge M$, such that $|f(x) - f_k(x)| \le \frac{1}{2^k} \le \frac{1}{2^K} < \epsilon$ for all $x \in E$ and $k \ge K$. □

Definition 3.23 (Support of a function)**.** Suppose $f : E \to \bar{\mathbb{R}}$. Then the *support of* f, denoted by $\text{supp}(f)$, is defined by the closure of $\{x \in E : f(x) \ne 0\}$ (hence a support is always closed). If $\text{supp}(f)$ is bounded, then f is said to have a *compact support*.

Example 3.24. Let $g_k(x) = f_k(x)\chi_{B(0,k)}(x)$ where f_k is the simple function in the previous theorem, then g_k is also simple and has compact support, and $g_k(x) \to f(x)$ for all $x \in E$.

3.3 Convergence almost everywhere

Definition 3.25 (Convergence almost everywhere)**.** Suppose $f_k, f : E \to \bar{\mathbb{R}}$. Then f_k is said to *converge to f almost everywhere on E*, denoted by $f_k \to f$ a.e. E, if there exists $Z \subset E$, such that $\mu(Z) = 0$ and $\lim_k f_k(x) = f(x)$ for all $x \in E \setminus Z$.

It is clear that if f_k is measurable for every k and $f_k \to f$ a.e. E, then f is measurable.

Lemma 3.26. *Suppose $f_k, f : E \to \bar{\mathbb{R}}$ where $\mu(E) < \infty$, and $f_k \to f$ a.e. E. Then for any $\epsilon > 0$, define the set $E_k(\epsilon) = \{x \in E : |f_k(x) - f(x)| \geq \epsilon\}$, there is $\lim_k \mu(\cup_{j=k}^\infty E_j(\epsilon)) = \mu(\limsup_k E_k(\epsilon)) = 0$.*

Proof. Note that $f_k \to f$ a.e. E means that $\mu(\cup_{i=1}^\infty \cap_{k=1}^\infty \cup_{j=k}^\infty E_j(\frac{1}{i})) = 0$ (To see this, recall that $\lim_k f_k(x) \neq f(x)$ means that there exists $i \in \mathbb{N}$, such that for any $k \in \mathbb{N}$ there exists $j \geq k$ and $|f_j(x) - f(x)| \geq \frac{1}{i}$. But the set of such "non-convergent" points has measure 0). Also note that $E_k(\epsilon_1) \subset E_k(\epsilon_2)$ for any $0 < \epsilon_2 < \epsilon_1$. Hence for any $\epsilon > 0$, there exists $i_0 \in \mathbb{N}$, such that $\frac{1}{i_0} < \epsilon$ and $\mu(\cap_{k=1}^\infty \cup_{j=k}^\infty E_j(\epsilon)) \leq \mu(\cap_{k=1}^\infty \cup_{j=k}^\infty E_j(\frac{1}{i_0})) \leq \mu(\cup_{i=1}^\infty \cap_{k=1}^\infty \cup_{j=k}^\infty E_j(\frac{1}{i})) = 0$. Note that $\limsup_k E_k(\epsilon) = \cap_{k=1}^\infty \cup_{j=k}^\infty E_j(\epsilon)$ and $\cup_{j=k}^\infty E_j(\epsilon)$ is non-increasing in k, we obtain that $\lim_k \mu(\cup_{j=k}^\infty E_j(\epsilon)) = \mu(\lim_k \cup_{j=k}^\infty E_j(\epsilon)) = \mu(\limsup_k E_k(\epsilon)) = 0$, where we needed $\mu(E) < \infty$ to obtain the first equality. □

Theorem 3.27 (Egorov: almost everywhere convergence + bounded domain ⇒ "nearly" uniform convergence)**.** *Suppose $f_k, f : E \to \bar{\mathbb{R}}$ where $\mu(E) < \infty$. If f_k and f are finite and $f_k \to f$ a.e. E, then for any $\delta > 0$, there exists $E_\delta \subset E$, such that $\mu(E_\delta) < \delta$ and $f_k \rightrightarrows f$ on $E \setminus E_\delta$.*

Proof. For any $\delta > 0$ and $i \in \mathbb{N}$, denote $E_k(\frac{1}{i}) = \{x \in E : |f_k(x) - f(x)| \geq \frac{1}{i}\}$, then by the lemma above, we know that there exists $j_i \in \mathbb{N}$ such that $\mu(\cup_{k=j_i}^\infty E_k(\frac{1}{i})) < \frac{\delta}{2^i}$.

Now consider $E_\delta = \cup_{i=1}^\infty \cup_{k=j_i}^\infty E_k(\frac{1}{i})$, there is $\mu(E_\delta) \leq \sum_{i=1}^\infty \mu(\cup_{k=j_i}^\infty E_k(\frac{1}{i})) < \sum_{i=1}^\infty \frac{\delta}{2^i} = \delta$. Moreover, there is

$$E \setminus E_\delta = \bigcap_{i=1}^\infty \bigcap_{k=j_i}^\infty \left(E_k\left(\frac{1}{i}\right)\right)^c = \bigcap_{i=1}^\infty \bigcap_{k=j_i}^\infty \left\{x \in E : |f_k(x) - f(x)| < \frac{1}{i}\right\}$$

Therefore, for any i, there exists j_i, such that $|f_k(x) - f(x)| < \frac{1}{i}$ for all $x \in E \setminus E_\delta$ and $k \geq j_i$. This means $f_k \rightrightarrows f$ on $E \setminus E_\delta$. □

Remarks. A few remarks are in place:

1. The boundedness of E is necessary. [Hint: consider $f_k(x) = \chi_{[0,k]}(x)$ and $f(x) = 1$ for all $x \in \mathbb{R}$. Or consider $f_k(x) = \frac{x}{k}$ and $f(x) = 0$ for $x \in \mathbb{R}$.]

2. If $\mu(E) = \infty$, we can still show that for any $M > 0$, there exists $E_M \subset E$ such that $\mu(E_M) > M$ and $f_k \rightrightarrows f$ on E_M. [Hint: choose any set $F_M \subset \mathbb{R}^n$ with $\mu(F_M) = M + \delta$, and $E_M = F_M \setminus E_\delta$ by applying Theorem 3.27 (Egorov) to $E = F_M$.]
3. There exists a sequence of sets $\{E_j\}$ with non-decreasing measure to $\mu(E)$, such that $f_k \rightrightarrows f$ on E_j for every j, and $\mu(E \setminus (\cup_{j=1}^{\infty} E_j)) = 0$.
4. We can choose E_δ such that $E \setminus E_\delta$ is also closed. To this end, just choose $E_{\delta/2}$ at the first place such that $\mu(E_{\delta/2}) < \delta/2$ and $f_k \rightrightarrows f$ on $E \setminus E_{\delta/2}$, and choose a closed set $F \subset E \setminus E_{\delta/2}$ such that $\mu((E \setminus E_{\delta/2}) \setminus F) < \delta/2$ (by Theorem 2.26), then $\mu(E \setminus F) < \delta$, and $f_k \rightrightarrows f$ on F.

Example 3.28. $f_k(x) = x^k$ on $[0, 1]$. Then for any $\delta > 0$ we can show $f_k \rightrightarrows f$ on $[0, 1 - \delta]$.

3.4 Convergence in measure

Definition 3.29 (Convergence in measure)**.** Suppose $f_k, f : E \to \bar{\mathbb{R}}$ and f_k, f finite a.e. E. We say that f_k *converges to* f *in measure*, denoted by $f_k \xrightarrow{\mu} f$, if for any $\epsilon > 0$, there is $\lim_{k\to\infty} \mu(\{x \in E : |f_k - f| \geq \epsilon\}) = 0$. [Using the $E_k(\epsilon)$ notation above, this definition can be stated as: $f_k \xrightarrow{\mu} f$ on E if for any $\epsilon > 0$ there is $\lim_k \mu(E_k(\epsilon)) = 0$.]

Note that $\mu(\{x \in E : |f_k| = \infty\}) = 0$ for all k, so it does not affect the convergence in measure.

Theorem 3.30 (Convergence in measure $\Rightarrow$ unique limit in the sense of a.e.)**.** *If $f_k \xrightarrow{\mu} f$ and $f_k \xrightarrow{\mu} g$, then $f = g$ a.e.*

Proof. For any $\epsilon > 0$, denote $E_k(\epsilon) = \{x : |f_k(x) - f(x)| \geq \epsilon\}$ and $F_k(\epsilon) = \{x : |f_k(x) - g(x)| \geq \epsilon\}$. Note that $|f(x) - g(x)| \leq |f_k(x) - f(x)| + |f_k(x) - g(x)|$. Hence, there is $\{x : |f - g| \geq \epsilon\} \subset E_k(\frac{\epsilon}{2}) \cup F_k(\frac{\epsilon}{2})$. Since $\mu(E_k(\frac{\epsilon}{2})), \mu(F_k(\frac{\epsilon}{2})) \to 0$ as $k \to \infty$, we know $\mu(\{x : |f - g| \geq \epsilon\}) = 0$. Setting $\epsilon = \frac{1}{n}$ and taking countable union for $n \in \mathbb{N}$ yield that $\mu(\{x : |f - g| > 0\}) = 0$. $\square$

Theorem 3.31 (Convergence a.e. + bounded domain $\Rightarrow$ convergence in μ)**.** *Suppose $f_k, f : E \to \bar{\mathbb{R}}$ where $\mu(E) < \infty$, f_k, f finite a.e. E, and $f_k \to f$ a.e. E, then $f_k \xrightarrow{\mu} f$.*

Proof. Since $f_k \to f$ a.e. E, we know that for any $\epsilon > 0$ there is $\mu(\cap_{j=1}^{\infty} \cup_{k=j}^{\infty} E_k(\epsilon)) = 0$ where $E_k(\epsilon) = \{x \in E : |f_k(x) - f(x)| \geq \epsilon\}$. Note that $A_k(\epsilon) = \cup_{j=k}^{\infty} E_j(\epsilon)$ is non-increasing in k and $E_k \subset A_k$, we have $\lim_k \mu(E_k(\epsilon)) \leq \lim_k \mu(A_k(\epsilon)) = \mu(\lim_k A_k(\epsilon)) = 0$ (the first equality requires $\mu(E) < \infty$). Hence $f_k \xrightarrow{\mu} f$ on E. $\square$

Remarks. We have two remarks regarding Theorem 3.31.

- An alternative proof: If $f_k \to f$ a.e. E, then for any $\epsilon > 0$, there is $\mu(\limsup_k E_k(\epsilon)) = 0$. By the Fatou's lemma for sets and $\mu(E) < \infty$, we have $0 = \mu(\limsup_k E_k(\epsilon)) \geq \limsup_k \mu(E_k(\epsilon)) \geq 0$. Hence $\lim_k \mu(E_k(\epsilon)) = 0$.

- The boundedness of E in Theorem 3.31 is again necessary: consider $f_k(x) = \chi_{[0,k]}(x)$ and $f(x) = 1$ for all $x \in \mathbb{R}$. Or consider $f_k(x) = \frac{x}{k}$ and $f(x) = 0$ for $x \in \mathbb{R}$.

Theorem 3.32 (Almost uniform convergence $\Rightarrow$ convergence in μ)**.** *Suppose $f_k, f : E \to \bar{\mathbb{R}}$, f_k, f finite a.e. E. If for any $\delta > 0$, there exists $E_\delta \subset E$ such that $\mu(E_\delta) < \delta$ and $f_k \rightrightarrows f$ on $E \setminus E_\delta$, then $f_k \xrightarrow{\mu} f$.*

Proof. For any $\delta > 0$, there is $E_\delta \subset E$, $\mu(E_\delta) < \delta$, and $f_k \rightrightarrows f$ on $E \setminus E_\delta$. Hence for any $\epsilon > 0$, there exists an integer K depending on ϵ and δ, such that $|f_k(x) - f(x)| < \epsilon$ for all $x \in E \setminus E_\delta$ and $k \geq K$. Therefore $E_k(\epsilon) = \{x \in E : |f_k(x) - f(x)| \geq \epsilon\} \subset E_\delta$, i.e., $\mu(E_k(\epsilon)) \leq \mu(E_\delta) < \delta$ for all $k \geq K$. Therefore $\lim_k \mu(E_k(\epsilon)) = 0$, i.e., $f_k \xrightarrow{\mu} f$ on E. □

Definition 3.33 (Cauchy in measure)**.** Suppose $f_k : E \to \bar{\mathbb{R}}$. We say $\{f_k\}$ is *Cauchy in measure* if for any $\epsilon > 0$, there is $\mu(\{x \in E : |f_k(x) - f_j(x)| \geq \epsilon\}) \to 0$ as $k, j \to \infty$. In other words, for any $\epsilon, \delta > 0$, there exists $K \in \mathbb{N}$, such that $\mu(\{x \in E : |f_k(x) - f_j(x)| \geq \epsilon\}) < \delta$ for all $k, j \geq K$.

Theorem 3.34 (Cauchy in measure $\Rightarrow$ convergence in measure)**.** *Suppose $\{f_k\}$ is Cauchy in measure on E. Then there exists $f : E \to \bar{\mathbb{R}}$ finite a.e. E such that $f_k \xrightarrow{\mu} f$ on E.*

Proof. We first show that there exists a subsequence of $\{f_k\}$ that converges to f in measure; then we show that the entire sequence $f_k \xrightarrow{\mu} f$.

Since $\{f_k\}$ is Cauchy in measure, we know that for every $i \in \mathbb{N}$ (and let $\epsilon = \delta = \frac{1}{2^i}$), there exists k_i, such that $\mu(\{x \in E : |f_l(x) - f_j(x)| \geq \frac{1}{2^i}\}) < \frac{1}{2^i}$ for all $l, j \geq k_i$. WLOG, we assume $k_i < k_{i+1}$, hence we have a subsequence $\{f_{k_i}\}$, denoted by $\{g_i\}$ for short, of $\{f_k\}$, such that $\mu(E_i) < \frac{1}{2^i}$ where $E_i = \{x \in E : |g_i(x) - g_{i+1}(x)| \geq \frac{1}{2^i}\}$. Now consider $S = \cap_{l=1}^\infty \cup_{i=l}^\infty E_i$. Since $\cup_{i=l}^\infty E_i$ is decreasing in l, we have

$$\mu(S) = \mu(\lim_{l\to\infty} \bigcup_{i=l}^{\infty} E_i) = \lim_{l\to\infty} \mu(\bigcup_{i=l}^{\infty} E_i) \leq \lim_{l\to\infty} \sum_{i=l}^{\infty} \mu(E_i) < \lim_{l\to\infty} \frac{1}{2^{l-1}} = 0.$$

Also, for any $x \in S^c = \cup_{l=1}^\infty \cap_{i=l}^\infty E_i^c$, we know $\sum_{i=l}^\infty |g_i(x) - g_{i+1}(x)| \leq \sum_{i=l}^\infty \frac{1}{2^i} = \frac{1}{2^{l-1}} \to 0$ as $l \to \infty$, which means $\{g_i(x) - g_{i+1}(x)\}$ is absolutely convergent and $\{g_i(x)\}$ is Cauchy. Let $f(x)$ be the limit of $\{g_i(x)\}$ for every $x \in S^c$ and arbitrary in S (it does not matter as $\mu(S) = 0$). Hence $g_i \to f$ on S^c, and $g_i \to f$ a.e. E.

Now we show $g_i \xrightarrow{\mu} f$ on E (note that we cannot directly get this from $g_i \to f$ a.e. E using Theorem 3.31 since E may be unbounded). Note that $|g_l(x) - f(x)| \leq \sum_{i=l}^\infty |g_i(x) - g_{i+1}(x)|$ a.e. E. Denote $F_l(\epsilon) = \{x \in E : |g_l(x) - f(x)| \geq \epsilon\}$. Then for any $\epsilon, \delta > 0$, there exists l large enough, such that $\frac{1}{2^{l-1}} < \min(\epsilon, \delta)$, and $\mu(F_l(\epsilon)) \leq \mu(F_l(\frac{1}{2^{l-1}})) \leq \mu(\cup_{i=l}^\infty E_i) \leq \sum_{i=l}^\infty \mu(E_i) < \frac{1}{2^{l-1}} < \delta$. Hence $g_i \xrightarrow{\mu} f$.

Finally, we will show $f_k \xrightarrow{\mu} f$. To this end, for any $\epsilon > 0$, consider

$$\{x : |f_k(x) - f(x)| \geq \epsilon\} \subset \left\{x : |f_k(x) - f_{k_i}(x)| \geq \frac{\epsilon}{2}\right\} \cup \left\{x : |f_{k_i}(x) - f(x)| \geq \frac{\epsilon}{2}\right\}$$

Note that the two sets on the right hand side have measure approaching 0 as $k, i \to \infty$. Hence $\mu(\{x : |f_k(x) - f(x)| \geq \epsilon\}) \to 0$ as $k \to \infty$. □

Remarks. It is easy to show the converse of Theorem 3.34. Hence $\{f_k\}$ is Cauchy in measure iff $f_k \xrightarrow{\mu} f$.

Theorem 3.35 (Reisz: Convergence in $\mu \Rightarrow \exists$ subsequence converges a.e.)**.** *If $f_k \xrightarrow{\mu} f$ on E, then there exists a subsequence $f_{k_i} \to f$ a.e. E.*

Proof. From the proof of Theorem 3.34, there exists a subsequence $\{f_{k_i}\}$ of $\{f_k\}$, such that $f_{k_i} \to g$ a.e. E for some g and $f_{k_i} \xrightarrow{\mu} g$. On the other hand, $f_{k_i} \xrightarrow{\mu} f$. Hence $f = g$ a.e. E. Therefore $f_{k_i} \to f$ a.e. E. □

3.5 Measurable functions and continuous functions

Theorem 3.36 (Lusin: finite a.e. $\Rightarrow$ "nearly" continuous)**.** *Suppose $f : E \to \bar{\mathbb{R}}$ is finite a.e. E, then for any $\delta > 0$, there exists $F \subset E$ where F is closed and $\mu(E \setminus F) < \delta$, such that f is continuous on F.*

Proof. WLOG, assume that $f : E \to \mathbb{R}$ (since $\mu(\{x : |f| = \infty\}) = 0$).

We first prove the case where f is simple, i.e., $f(x) = \sum_{i=1}^{p} c_i \chi_{E_i}(x)$ for mutually disjoint $E_1, \ldots, E_p$. To this end, for every E_i, there exists a closed subset $F_i \subset E_i$ and $\mu(E_i \setminus F_i) < \frac{\delta}{p}$ (by Theorem 2.26). Let $F = \cup_{i=1}^{p} F_i$. Then F is closed, and $\mu(E \setminus F) < \delta$. Moreover $F_1, \ldots, F_p$ are also mutually disjoint, and f is constant on each F_i. Hence f is continuous on F.

Next we prove the case where f is a general measurable function. WLOG, assume $|f| \leq 1$ (note that the transformation $g(x) = f(x)/(1 + |f(x)|)$ and its inverse $f(x) = g(x)/(1 - |g(x)|)$). Then there exists a sequence of non-decreasing simple functions $\{f_k\}$, such that $f_k \to f$ a.e. E. Then for every k, there exists a closed subset F_k of E with $\mu(E \setminus F_k) < \frac{\delta}{2^k}$, such that f_k is continuous on F_k. Let $F = \cap_{k=1}^{\infty} F_k$, then F is closed, and $\mu(E \setminus F) = \mu(\cup_{k=1}^{\infty}(E \setminus F_k)) < \delta$. Hence $f_k \rightrightarrows f$ on F. Since f_k is continuous, we know f is continuous on F. □

Corollary 3.37. *Suppose $f : E \to \bar{\mathbb{R}}$ is finite a.e. E. Then for any $\delta > 0$, there exists a continuous function $g : E \to \bar{\mathbb{R}}$ such that $\mu(\{x : f(x) \neq g(x)\}) < \delta$.*

Corollary 3.38. *Suppose $f : E \to \bar{\mathbb{R}}$ is finite a.e. E. Then there exists a sequence of continuous functions $\{g_k\}$ such that $g_k \to f$ a.e. E.*

Proof. Consider sequences $\epsilon_k, \delta_k \downarrow 0$. Then for every k, there exists g_k such that $\mu(\{x : |f(x) - g_k(x)| \geq \epsilon_k\}) < \delta_k$. Hence $g_k \xrightarrow{\mu} f$. By Reisz theorem above, there exists a subsequence of $\{g_k\}$, still denoted by $\{g_k\}$, such that $g_k \to f$ a.e. E. □

3.6 Measurability of composite functions

Lemma 3.39. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$. Then f is measurable iff $f^{-1}(G) \in \mathcal{M}$ for any open set $G \subset \mathbb{R}$.*

Proof. Necessity is obvious. Now suppose f is measurable. Then for any $(a, b) \subset \mathbb{R}$, we know $f^{-1}((a, b)) = f^{-1}((a, \infty)) \setminus f^{-1}([b, \infty)) \in \mathcal{M}$. Recall that for any open set $G \subset \mathbb{R}$, it can be written as the union of countably many disjoint open intervals as $G = \cup_{k=1}^{\infty}(a_k, b_k)$. Hence $f^{-1}(G) = \cup_{k=1}^{\infty} f^{-1}((a_k, b_k)) \in \mathcal{M}$. □

Theorem 3.40. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is continuous and $g : \mathbb{R}^n \to \mathbb{R}$ is measurable, then $f \circ g : \mathbb{R}^n \to \mathbb{R}$ is measurable.*

Proof. Since f is continuous, we know that for any open G, $f^{-1}(G)$ is open. Hence $(f \circ g)^{-1}(G) = g^{-1}(f^{-1}(G)) \in \mathcal{M}$. □

Remarks. Note that if f is measurable and g is continuous, $f \circ g$ is not necessarily measurable.

Theorem 3.41. *Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is continuous, and $\mu(T^{-1}(Z)) = 0$ for any $Z \subset \mathbb{R}^n$ with $\mu(Z) = 0$, and $f : \mathbb{R}^n \to \mathbb{R}$ is measurable, then $f \circ T$ is measurable.*

Proof. For any open set G in $\mathbb{R}$, $f^{-1}(G) \in \mathcal{M}$. Hence there exist a G_δ-set H and a measure zero set Z such that $f^{-1}(G) = H \setminus Z$. Therefore $T^{-1}(f^{-1}(G)) = T^{-1}(H) \setminus T^{-1}(Z) \in \mathcal{M}$ since $T^{-1}(H), T^{-1}(Z) \in \mathcal{M}$. □

Corollary 3.42. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is measurable and $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear nondegenerate transformation, then $f \circ T$ is measurable.*

Example 3.43. Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is measurable. Show that $g(x, y) = f(x - y) : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is measurable.

Proof. Define $h(x, y) = f(x)$. Then for any $t \in \mathbb{R}$, $\{(x, y) : h(x, y) > t\} = \{x : f(x) > t\} \times \mathbb{R} \in \mathcal{M}(\mathbb{R}^2)$. Hence h is measurable. Let $T(x, y) = (x - y, x + y)$ then T is a linear nondegenerate transformation. Therefore $g(x, y) = h(T(x, y))$ is measurable. □

4 Lebesgue Integrals

4.1 Integral of simple nonnegative functions

Definition 4.1 (Integral of simple nonnegative function)**.** Suppose $f : \mathbb{R}^n \to \bar{\mathbb{R}}_+$ is a simple measurable function such that $f(x) = \sum_{i=1}^p c_i \chi_{A_i}(x)$, where $\{A_i\}$ are mutually disjoint and $\mathbb{R}^n = \cup_{i=1}^p A_i$. Then for any $E \in \mathcal{M}$, the *integral of f on E* is defined by

$$\int_E f(x)\,\mathrm{d}\mu(x) = \sum_{i=1}^p c_i\,\mu(E \cap A_i).$$

Recall that we define $0 \cdot \infty = 0$. Hence the integral is not affected if $c_i = 0$ or $\mu(E \cap A_i) = 0$. For notation simplicity, we write the integral of f over E in either of the ways below:

$$\int_E f(x)\,\mathrm{d}\mu(x) = \int_X f(x)\chi_E(x)\,\mathrm{d}\mu(x) = \int_E f\,\mathrm{d}\mu = \int_E f, \quad \text{and} \quad \int_X f = \int f.$$

Example 4.2. Let $f = \sum_{i=1}^p c_i \chi_{A_i}$ be a simple nonnegative function. Show that $\mu(A_i) < \infty$ for all i if $\int f < \infty$.

Example 4.3. Consider the function $f(x) = \chi_{\mathbb{Q}}(x) = 1$ if $x \in \mathbb{Q}$ and 0 otherwise. Then $\int_0^1 f = 0$.

Theorem 4.4 (Linearity of integral)**.** *Suppose $f, g : \mathbb{R}^n \to \bar{\mathbb{R}}$ are simple nonnegative measurable functions. Then for any $E \in \mathcal{M}$ and $c \in \mathbb{R}$, there are $\int_E cf = c\int_E f$ and $\int_E (f+g) = \int_E f + \int_E g$.*

Proof. The first one is trivial to show. Now suppose $f(x) = \sum_{i=1}^p a_i \chi_{A_i}(x)$ and $g(x) = \sum_{j=1}^q b_j \chi_{B_j}(x)$. Then it is clear that $f(x) + g(x) = \sum_{i=1}^p \sum_{j=1}^q (a_i + b_j)\chi_{A_i \cap B_j}(x)$. Hence

$$\begin{aligned}\int_E (f+g) &= \sum_{i=1}^p \sum_{j=1}^q (a_i + b_j)\mu(E \cap A_i \cap B_j)\\ &= \sum_{i=1}^p a_i \mu(E \cap A_i) + \sum_{j=1}^q b_j \mu(E \cap B_j)\\ &= \int_E f + \int_E g,\end{aligned}$$

which proves the linearity. □

Theorem 4.5. *Suppose $E_k \in \mathcal{M}$ and is increasing in k. Let $E = \cup_{k=1}^\infty E_k$. If $f : \mathbb{R}^n \to \mathbb{R}$ is a simple nonnegative measurable function, then $\int_E f = \lim_k \int_{E_k} f$.*

Proof. Suppose $f(x) = \sum_{i=1}^{p} a_i \chi_{A_i}(x)$. Then there is

$$\begin{aligned}\lim_{k\to\infty}\int_{E_k} f = \lim_{k\to\infty}\sum_{i=1}^{p} a_i\mu(E_k\cap A_i) &= \sum_{i=1}^{p} a_i\mu\left(\lim_{k\to\infty}(E_k\cap A_i)\right)\\ &= \sum_{i=1}^{p} a_i\mu\left(\lim_{k\to\infty}(E_k)\cap A_i\right) = \sum_{i=1}^{p} a_i\mu(E\cap A_i) = \int_E f,\end{aligned}$$

which completes the proof. $\square$

4.2 Integral of general nonnegative functions

Definition 4.6. Suppose $f : E \to \bar{\mathbb{R}}$ is nonnegative measurable function. The integral of f on E is defined by

$$\int_E f = \sup\left\{\int_E h : h \text{ is simple nonnegative, } h(x) \le f(x),\ \forall\, x \in E\right\}$$

We call f *integrable* if $\int_E f < \infty$.

Theorem 4.7 (Properties of integral)**.** *The following statements hold:*

1. *If $f, g : E \to \mathbb{R}_+$ and $f \le g$, then $\int_E f \le \int_E g$. Hence if g is integrable then f is integrable. If $f \le M$ and $\mu(E) < \infty$, then f is integrable.*
2. *If $A \subset E$ is measurable, then $\int_A f = \int_E f\chi_A$.*
3. *If $f = 0$ a.e. E, then $\int_E f = 0$.*

Proof. Item 1 is trivial. Item 2 can be verified by

$$\begin{aligned}\int_A f &= \sup\left\{\int_E h : h \text{ is simple nonnegative, } h \le f\right\}\\ &= \sup\left\{\int_E h\chi_A : h \text{ is simple nonnegative, } h\chi_A \le f\chi_A\right\}\\ &= \int_E f\chi_A\end{aligned}$$

For Item 3, let $E_k = \{x \in E : f(x) \ge \frac{1}{k}\}$. Then E_k is non-decreasing. Moreover, $\frac{1}{k}\mu(E_k) \le \int_{E_k} f \le \int_E f = 0$, which implies that $\mu(E_k) = 0$ for all k. Moreover $\{x \in E : f(x) > 0\} = \cup_{k=1}^{\infty} E_k$, therefore $\mu(\{x \in E : f(x) > 0\}) \le \sum_{k=1}^{\infty}\mu(E_k) = 0$. $\square$

Theorem 4.8 (Integrable functions are finite a.e.)**.** *Suppose $f : E \to \bar{\mathbb{R}}_+$ is integrable. Then f is finite a.e. E.*

Proof. Let $E_k = \{x : f(x) > k\}$. Then E_k is decreasing and $A := \{x : f(x) = \infty\} = \cap_{k=1}^{\infty} E_k$. Note that $k \cdot \mu(E_k) \le \int_{E_k} f \le \int_E f < \infty$ and $\mu(E_k) < \infty$ for all k, we know $\mu(A) = \lim_k \mu(E_k) \le \frac{1}{k}\int_E f \to 0$ as $k \to \infty$. Hence $\mu(A) = 0$. $\square$

Theorem 4.9 (Beppo Levi: $f_k \uparrow f \Rightarrow \int f_k \uparrow \int f$)**.** *Suppose $f_k : E \to \bar{\mathbb{R}}_+$, $f_k(x) \le f_{k+1}(x)$ for every $x \in E$ and k, and $\lim_k f_k(x) = f(x)$ for all $x \in E$. Then $\lim_k \int_E f_k = \int_E f$.*

Proof. It is clear that $\int_E f$ is well defined. Since $f_k \uparrow f$, we know $\int_E f_k \uparrow$ and $\lim_k \int_E f_k \le \int_E f$.

For any simple function $h \le f$ and any $c \in (0,1)$, consider $E_k = \{x : f_k(x) \ge ch(x)\}$. Then $E_k \uparrow E$. Hence

$$\lim_{k\to\infty} \int_E f_k \ge \lim_{k\to\infty} \int_{E_k} f_k \ge \lim_{k\to\infty} \int_{E_k} ch = \int_E ch = c \int_E h.$$

Letting $c \to 1$, we have $\lim_k \int_E f_k \ge \int_E h$. Hence $\lim_k \int_E f_k \ge \int_E f$. $\square$

Remarks. If $\int_E f_1 < \infty$ and $f_k \downarrow f \ge 0$, then $\int_E f_k \downarrow \int_E f$.

Theorem 4.10 (Linearity of integral)**.** *Suppose $f, g : E \to \bar{\mathbb{R}}_+$ and $\alpha, \beta \ge 0$. Then $\int_E (\alpha f + \beta g) = \alpha \int_E f + \beta \int_E g$.*

Proof. Only need to show this for $\alpha = \beta = 1$. Let $f_k \uparrow f$ and $g_k \uparrow g$, then $f_k + g_k \uparrow f + g$. Hence

$$\int_E (f+g) = \lim_{k\to\infty} \int_E (f_k + g_k) = \lim_{k\to\infty} \left\{ \int_E f_k + \int_E g_k \right\} = \int_E f + \int_E g.$$

$\square$

Example 4.11. Suppose $f, g : E \to \bar{\mathbb{R}}_+$ and $f = g$ a.e. E. Then $\int_E f = \int_E g$. [Hint: $f = f\chi_{E_1} + f\chi_{E_2}$ where $E_1 = \{x : f = g\}$ and $E_2 = E \setminus E_1$.]

Theorem 4.12. *Suppose $f_k : E \to \bar{\mathbb{R}}_+$. Then $\int_E \sum_{k=1}^\infty f_k = \sum_{k=1}^\infty \int_E f_k$.*

Proof. Let $s_k(x) = \sum_{i=1}^k f_i(x)$ and $s(x) = \sum_{k=1}^\infty f_k(x)$ for every x and k. Then $s_k \uparrow s$. Hence $\lim_k \int_E s_k = \int_E s$. $\square$

Corollary 4.13. *Suppose $\{E_k\}_k \subset \mathcal{M}$ are mutually disjoint. If f is integrable on $E = \cup_{k=1}^\infty E_k$, then $\int_E f = \sum_{k=1}^\infty \int_{E_k} f$.*

Proof. Let $f_k = f\chi_{E_k}$. Then

$$\sum_{k=1}^\infty \int_{E_k} f = \sum_{k=1}^\infty \int_E f\chi_{E_k} = \int_E \Big(\sum_{k=1}^\infty f\chi_{E_k} \Big) = \int_E f$$

by the theorem above. $\square$

Remarks. If $f \equiv 1$, then the corollary above reduces to $\mu(E) = \sum_{k=1}^\infty \mu(E_k)$.

Example 4.14. If $E_1, \dots, E_p$ are subsets of $[0,1]$, and every point on $[0,1]$ is covered by at least k of $E_1, \dots, E_p$ where $k \le p$. Then $\mu(E_i) \ge \frac{k}{p}$ for some i.

Proof. $\sum_{i=1}^p \mu(E_i) = \int_0^1 \sum_{i=1}^p \chi_{E_i}(x) \ge \int_0^1 k = k$. $\square$

Lemma 4.15 (Fatou's lemma for integrals)**.** *Suppose* $f_k : E \to \bar{\mathbb{R}}_+$. *Then* $\int_E \liminf_k f_k \le \liminf_k \int_E f_k$.

Proof. Define $g_k(x) = \inf_{j\ge k} f_j(x)$ for every $x \in E$ and k. Then g_k is non-decreasing and nonnegative. Hence

$$\int_E \liminf_{k\to\infty} f_k = \int_E \lim_{k\to\infty} g_k = \lim_{k\to\infty} \int_E g_k \le \liminf_{k\to\infty} \int_E f_k,$$

since $g_k(x) \le f_k(x)$ on E. □

Remarks. If $\int_E f_k \le M$ for all k, then $\int_E \liminf_{k\to\infty} f_k \le M$.

Example 4.16. We may have "<" hold in Lemma 4.15 in some cases: let

$$f_k(x) = \begin{cases} 0, & \text{if } x = 0 \\ k, & \text{if } 0 < x < \frac{1}{k} \\ 0, & \text{if } \frac{1}{k} \le x \le 1 \end{cases}$$

for every k. Then $f_k \to 0$ on $[0, 1]$. But $\int_0^1 \lim_k f_k = 0 < 1 = \lim_k \int_0^1 f_k$.

Theorem 4.17. *Suppose* $f : E \to \bar{\mathbb{R}}_+$ *is finite a.e.* E *and* $\mu(E) < \infty$. *If* $[0, \infty)$ *is partitioned such that* $0 = y_0 < y_1 < \cdots$ *and* $y_{k+1} - y_k < \delta$ *for all* k, *and define* $E_k = \{x \in E : y_k \le f(x) < y_{k+1}\}$. *Then* f *is integrable iff* $\sum_{k=1}^{\infty} y_k \mu(E_k) < \infty$. *Moreover* $\lim_{\delta\to 0} \sum_{k=1}^{\infty} y_k \mu(E_k) = \int_E f$.

Proof. For every k, there is $y_k\mu(E_k) \le \int_{E_k} f \le y_{k+1}\mu(E_k) < \delta\mu(E_k)+y_k\mu(E_k) \le \delta\mu(E) + y_k\mu(E_k)$. By the squeeze theorem, f is integrable iff $\sum_{k=1}^{\infty} y_k\mu(E_k) < \infty$. Taking $\delta \to 0$ completes the proof. □

4.3 Integral of general functions

Definition 4.18. Suppose $f : E \to \bar{\mathbb{R}}$ is measurable. If at least one of $\int_E f^+$ and $\int_E f^-$ is finite, then the *integral of* f is defined by $\int_E f = \int_E f^+ - \int_E f^-$. If both are finite, then f is called *integrable*, denoted by $f \in L(E)$.

Remarks. Note that $\int_E |f| = \int_E f^+ + \int_E f^-$. Hence $f \in L(E)$ iff $|f| \in L(E)$. In addition, $|\int_E f| \le \int_E |f|$.

Example 4.19. If $f : E \to \mathbb{R}$ is bounded, $\mu(E) < \infty$, then $f \in L(E)$. [Hint: There exists $M > 0$ such that $|f| \le M$. Hence $\int_E |f| \le M \cdot \mu(E) < \infty$.]

Theorem 4.20 (Properties of integral)**.** *The following statements hold:*

1. *If* $f \in L(E)$, *then* $|f|$ *is finite a.e.* E.
2. *If* $E \in \mathcal{M}$ *and* $f = 0$ *a.e.* E, *then* $\int_E f = 0$.
3. *If* $f : E \to \mathbb{R}$ *is measurable,* $g \in L(E)$, *and* $|f| \le g$, *then* $f \in L(E)$.
4. *If* $f \in L(\mathbb{R}^n)$, *then* $\lim_k \int_{\mathbb{R}^n \setminus B(0,k)} f = 0$.

Proof. Item 1 follows Theorem 4.8. Item 2 follows from $f^{\pm} = 0$ a.e. E. Item 3 follows from $0 \le f^{\pm} \le g$ due to $|f| \le g$. Item 4 follows from $\int f^{\pm} \le \int |f| < \infty$ and $f_k^{\pm} \downarrow 0$ where $f_k^{\pm} := f^{\pm}\chi_{\mathbb{R}^n \setminus B(0,k)}$. □

Theorem 4.21 (Linearity of integral)**.** *For any $c \in \mathbb{R}$, $\int_E cf = c\int_E f$ and $\int_E (f+g) = \int_E f + \int_E g$.*

Example 4.22. Suppose $f : [0,1] \to \mathbb{R}$ is measurable. If $\int_0^1 |f(x)| \log(1 + |f(x)|) < \infty$, then $f \in L([0,1])$.

Proof. Let $E_1 = \{x : |f(x)| \ge e-1\}$, then $|f(x)| \le |f(x)|\log(1+|f(x)|)$ for all $x \in E_1$. Let $E_2 = E \setminus E_1$, then $|f(x)| \le e-1$. Hence $\int_0^1 |f| = \int_{E_1} |f| \log(1 + |f|) + \int_{E_2} (e-1) < \infty$. □

Example 4.23 (Jensen's inequality)**.** Suppose $w : E \to \mathbb{R}_+$ and $\int_E w = 1$. If $\phi : [a,b] \to \mathbb{R}$ is convex, $f : E \to [a,b]$ is measurable and $f \in L(E)$. Then $\phi(\int_E fw) \le \int_E \phi(f)w$.

Proof. Denote $y_0 = \int_E fw$. Then $a \le y_0 \le b$. Since ϕ is convex, there exists $z \in \mathbb{R}$ such that $\phi(y) \ge \phi(y_0) + z \cdot (y - y_0)$ for all y. Hence by setting $y = f(x)$, multiplying w on both sides, and taking integral over E, we obtain

$$\int_E \phi(f(x))w(x) \ge \Big(\int_E w(x)\Big)\,\phi(y_0) + z \cdot \Big(\int_E fw - y_0\Big) = \phi(y_0),$$

which is the claimed inequality. □

Theorem 4.24. *Suppose $E_k \in \mathcal{M}$ are mutually disjoint. If $f \in L(E)$ where $E = \cup_{k=1}^{\infty} E_k$, then $\int_E f = \sum_{k=1}^{\infty} \int_{E_k} f$.*

Proof. Note that $\sum_{k=1}^{\infty} \int_{E_k} f^{\pm} = \int_E f^{\pm} \le \int_E |f| < \infty$. Hence $\sum_{k=1}^{\infty} \int_E f = \sum_{k=1}^{\infty} \int_E f^+ - \sum_{k=1}^{\infty} \int_E f^- = \int_E f$. □

Theorem 4.25 (Absolute continuity of integral)**.** *Suppose $f \in L(E)$. Then for any $\epsilon > 0$, there exists $\delta > 0$, such that for any $E_\delta \subset E$ satisfying $\mu(E_\delta) < \delta$, there is $|\int_{E_\delta} f| \le \int_{E_\delta} |f| < \epsilon$.*

Proof. WLOG, assume $f : E \to \mathbb{R}_+$. Then there exists a simple function g such that $0 \le g \le f$ and $0 \le \int_E f - \int_E g < \frac{\epsilon}{2}$. Let M be the bound of g and $\delta = \frac{\epsilon}{2M}$. Then for any E_δ satisfying $\mu(E_\delta) < \delta = \frac{\epsilon}{2M}$, there is

$$\int_{E_\delta} f = \int_{E_\delta} (f-g) + \int_{E_\delta} g \le \int_E (f-g) + M \cdot \mu(E_\delta) < \epsilon,$$

which completes the proof. □

Example 4.26 (Intermediate value theorem of integrals)**.** Suppose $f \in L(E)$ where $E \subset \mathbb{R}$ and $0 < C := \int_E f < \infty$. Then for any $c \in (0, C)$, there exists $t \in \mathbb{R}$ such that $\int_{E \cap (-\infty, t]} f = c$.

Proof. Define $g(t) = \int_{E\cap(-\infty,t]} f(x)$. From the theorem above, we know that for any $\epsilon > 0$, there exists $\delta > 0$, such that for any $|\Delta t| < \delta$, there is

$$|g(t+\Delta t) - g(t)| \leq \int_{E\cap[t,t+\Delta t)} |f(x)| < \epsilon.$$

Hence $g(t)$ is continuous. Since $g(-\infty) = 0$ and $g(+\infty) = C$, by the intermediate value theorem of continuous functions, there exists t such that $g(t) = c \in (0, C)$. □

Theorem 4.27. *If $f \in L(\mathbb{R}^n)$, then for any $y_0 \in \mathbb{R}^n$, there is $\int_{\mathbb{R}^n} f(x+y_0) = \int_{\mathbb{R}^n} f(x)$.*

Theorem 4.28 (Lebesgue dominated convergence theorem)**.** *Suppose $f_k \in L(E)$ and $f_k \to f$ a.e. E. If there exists $g \in L(E)$ such that $|f_k| \leq g$, then $\int_E f_k \to \int_E f$.*

Proof. Define $h_k = |f_k - f|$, then $h_k \to 0$ a.e. E and $0 \leq h_k \leq 2g$ for all k. Hence $h_k, 2g \in L(E)$. Moreover, by Fatou's Lemma 4.15,

$$\int_E 2g = \int_E \lim_{k\to\infty}(2g - 2h_k) \leq \liminf_{k\to\infty} \int_E (2g - 2h_k) = \int_E 2g - \limsup_{k\to\infty} \int_E 2h_k$$

Hence $\limsup_k \int_E h_k \leq 0$ and therefore $|\int_E f_k - \int_E f| \leq \int_E |f_k - f| = \int_E h_k \to 0$ as $k \to \infty$. □

Remarks. The Lebesgue dominated convergence theorem actually implies a stronger result: $\int_E |f_k - f| \to 0$ as $k \to \infty$.

Example 4.29. In general, $\int_E f_k \to \int_E f$ does *not* imply $\int_E |f_k - f| \to 0$. For example, let $f_k(x) = 1$ if $x \in [\frac{j}{2^k}, \frac{j+1}{2^k})$ and j is odd, and 0 otherwise. Let $f(x) = \frac{1}{2}$. Then $\int_E f_k = \int_E f$, but $\int_E |f_k - f| = \frac{1}{2}$.

Example 4.30. In general, $\int_E |f_k - f| \to 0$ does *not* imply $f_k \to f$ a.e. For example, let $f_k(x) = 1$ if $x \in [\frac{i}{2^j}, \frac{i+1}{2^j})$ and 0 elsewhere for $k = 2^j + i$, where $j \geq 1$ and $i = 0, \ldots, 2^j - 1$. Then $f_k \xrightarrow{\mu} f$, but we do not have $f_k \to f$ a.e.

Theorem 4.31. *If $\int_E |f_k - f| \to 0$, then $f_k \xrightarrow{\mu} f$. Moreover, there exists a subsequence $f_{k_j} \to f$ a.e.*

Proof. For any $\epsilon > 0$, denote $E_k(\epsilon) = \{x \in E : |f_k(x) - f(x)| \geq \epsilon\}$, there is

$$\epsilon\,\mu(E_k(\epsilon)) = \int_{E_k(\epsilon)} \epsilon \leq \int_{E_k(\epsilon)} |f_k - f| \leq \epsilon \int_E |f_k - f| \to 0.$$

Hence $\mu(E_k(\epsilon)) \to 0$ as $k \to \infty$, i.e., $f_k \xrightarrow{\mu} f$. □

Remarks. The converse is not true in general. See Example 4.16.

Theorem 4.32 (Bounded convergence theorem)**.** *If $|f_k| \le M$, $f_k \to f$ a.e. E, and $\mu(E) < \infty$, then $\int_E f_k \to \int_E f$.*

Proof. Set $g(x) = M$ and note that $g \in L(E)$ since $\mu(E) < \infty$. Then by Dominated Convergence Theorem, $\int_E f_k \to \int_E f$. $\square$

Theorem 4.33 (Dominated convergence theorem for $f_k \xrightarrow{\mu} f$)**.** *Suppose $f_k \in L(\mathbb{R}^n)$, $f_k \xrightarrow{\mu} f$, and there exists $g \in L(\mathbb{R}^n)$ such that $|f_k| \le g$. Then $f \in L(\mathbb{R}^n)$ and $\int_{\mathbb{R}^n} f_k \to \int_{\mathbb{R}^n} f$.*

Proof. Since $f_k \xrightarrow{\mu} f$, there exists a subsequence $f_{k_j} \to f$ a.e. and $\int_E f_{k_j} \to \int_E f$. Hence $f \in L(\mathbb{R}^n)$, and $|f| \le g$ a.e. It remains to show that $\int_{\mathbb{R}^n} f_k \to \int_{\mathbb{R}^n} f$.

For any $\epsilon > 0$, we first choose $R > 0$ large enough, and denote $B = B(0, R)$ for short, such that $2\int_{\mathbb{R}^n \setminus B} g < \frac{\epsilon}{3}$ (by Theorem 4.20 Item 4). Then

$$\int_{\mathbb{R}^n \setminus B} |f_k - f| \le 2\int_{\mathbb{R}^n \setminus B} g < \frac{\epsilon}{3}.$$

Now we work on B which is bounded. We choose $\delta > 0$ small enough, such that for any $E_\delta \subset B$ satisfying $\mu(E_\delta) < \delta$, there is $2\int_{E_\delta} g < \frac{\epsilon}{3}$ (by Theorem 4.25). Hence $\int_{E_\delta} |f_k - f| \le 2\int_{E_\delta} g < \frac{\epsilon}{3}$. Since $f_k \xrightarrow{\mu} f$, there exists an integer K large enough, such that $\mu(C_k) < \delta$ for all $k \ge K$, where $C_k := \{x \in B : |f_k - f| \ge \frac{\epsilon}{3m}\} \subset B$ and $m := \mu(B)$. Then there is

$$\int |f_k - f| = \int_{\mathbb{R}^n \setminus B} |f_k - f| + \int_{C_k} |f_k - f| + \int_{B \setminus C_k} |f_k - f| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + m\frac{\epsilon}{3m} = \epsilon$$

for all $k \ge K$. Hence $\int |f_k - f| \to 0$. $\square$

Example 4.34. Suppose $f \in C([0, \infty))$, and $f(x) \to l$ as $x \to \infty$, then for any $A > 0$, there is $\lim_k \int_0^A f(kx) = Al$.

Proof. Since $f(x) \to l$, we know there exists $X > 0$ such that $|f(x)| < |l| + 1$ for all $x \ge X$. Since f is continuous, there exists $m = \max_{0 \le x \le X} |f(x)| < \infty$. Then $|f(x)| \le M$ for all $x \ge 0$ where $M = \max(m, |l| + 1)$. Define $f_k(x) = f(kx)$, then $f_k(x) \le M$ on $[0, A]$. By the bounded convergence theorem, there is

$$\lim_{k\to\infty} \int_0^A f(kx) = \lim_{k\to\infty} \int_0^A f_k(x) = \int_0^A \lim_{k\to\infty} f_k(x) = \int_0^A \lim_{k\to\infty} f(kx) = Al.$$

$\square$

Example 4.35. For any $\alpha > 1$, show that $\int_0^1 \frac{kx \sin x}{1+(kx)^\alpha}\,\mathrm{d}x \to 0$ as $k \to 0$.

Proof. Denote $f_k(x) = \frac{kx \sin x}{1+(kx)^\alpha}$. Then $|f_k(x)| \le \frac{kx}{1+(kx)^\alpha} \le \frac{1}{(kx)^{\alpha-1}} \to 0$. By DCT Theorem 4.28, $\lim_k \int_0^1 f = \int_0^1 \lim_k f_k = 0$. $\square$

Example 4.36. For any $a > 0$, show that $\int_a^\infty \frac{k^2 x e^{-k^2x^2}}{1+x^2} \to 0$ as $k \to \infty$.

Proof. Consider change of variable $u = kx$ and then

$$\int_a^\infty \frac{k^2xe^{-k^2x^2}}{1+x^2}\,\mathrm{d}x = \int_{ka}^\infty \frac{ue^{-u^2}}{1+(u/k)^2}\,\mathrm{d}u = \int_0^\infty \chi_{[ka,\infty)}(u)\frac{ue^{-u^2}}{1+(u/k)^2}\,\mathrm{d}u.$$

Denote $f_k(u) = \chi_{[ka,\infty)}(u)\frac{ue^{-u^2}}{1+(u/k)^2}$, then it is clear that $f_k(u) \to 0$ a.e. and $|f_k(u)| \le ue^{-u^2} \in L(\mathbb{R})$. Hence by DCT there is $\lim_k \int_0^\infty f = 0$. □

Corollary 4.37. *Let $f_k \in L(E)$. If $\sum_{k=1}^\infty \int_E |f_k| < \infty$, then $\sum_{k=1}^\infty f_k(x)$ converges a.e. E. Define $f(x) = \sum_{k=1}^\infty f_k(x)$ for every $x \in E$. Then $f \in L(E)$ and $\int_E f = \sum_{k=1}^\infty \int_E f_k$.*

Proof. Let $s_k(x) = \sum_{i=1}^k |f_i(x)|$ and $s(x) = \sum_{k=1}^\infty |f_k(x)|$ for every $x \in E$. Then $\int_E s = \lim_k \int_E s_k = \sum_{k=1}^\infty \int_E |f_k| < \infty$ by Theorem 4.12. Hence $s \in L(E)$ and s is finite a.e. E. Define $g_k(x) = \sum_{i=1}^k f_i(x)$, then $|g_k(x)| \le s(x)$. By Theorem 4.28 (DCT), there is $\int_E f = \int_E \lim_k g_k = \lim_k \int_E g_k = \sum_{k=1}^\infty \int_E f_k$. □

Theorem 4.38 (Interchange derivative and integral)**.** *Suppose $f : E \times (a,b) \to \mathbb{R}$, $f(\cdot, y) \in L(E)$ for every $y \in (a,b)$, and $f(x,\cdot)$ is differentiable on (a,b) for every $x \in E$. If there exists $g \in L(E)$ such that $|\partial_y f(x,y)| \le g(x)$ for any $(x,y) \in E \times (a,b)$, then for any $y \in (a,b)$, there is*

$$\frac{\mathrm{d}}{\mathrm{d}y}\Big(\int_E f(x,y)\,\mathrm{d}\mu(x)\Big) = \int_E \Big(\partial_y f(x,y)\Big)\,\mathrm{d}\mu(x).$$

Proof. For any $y \in (a,b)$, consider any nonzero sequence $e_k \to 0$, and define $f_k(x) = \frac{1}{e_k}(f(x,y+e_k) - f(x,y))$ for every $x \in E$. By the mean value theorem of derivatives, there exists $\xi_k \in (y, y+e_k)$, such that $|f_k(x)| = |\partial_y f(x,\xi_k)| \le g(x) \in L(E)$. Note that $\lim_k f_k(x) = \partial_y f(x,y)$. Hence by Theorem 4.28 (DCT), there is

$$\frac{\mathrm{d}}{\mathrm{d}y}\Big(\int_E f(x,y)\,\mathrm{d}\mu(x)\Big) = \lim_{k\to\infty}\int_E f_k = \int_E \lim_{k\to\infty} f_k = \int_E \Big(\partial_y f(x,y)\Big)\,\mathrm{d}\mu(x),$$

which completes the proof. □

Example 4.39. Let $f \in L(\mathbb{R})$ and $\sum_{k=1}^\infty \frac{1}{a_k} < \infty$ where $a_k > 0$ for all k. Then $\lim_k f(a_k x) = 0$ a.e.

Proof. Denote $f_k(x) = f(a_k x)$. Then $\sum_{k=1}^\infty \int |f_k| = \sum_{k=1}^\infty \int \frac{|f(a_k x)|}{a_k}\,\mathrm{d}(a_k x) = \sum_{k=1}^\infty \int \frac{|f(x)|}{a_k}\,\mathrm{d}x < \infty$. Hence $\sum_{k=1}^\infty |f_k|$ is finite a.e., which implies $f_k \to 0$ a.e. □

Theorem 4.40. *If $f \in L(E)$, then for any $\epsilon > 0$, there exists $g \in C(\mathbb{R}^n)$ with compact support, such that $\int_E |f - g| < \epsilon$.*

Proof. For any $\epsilon > 0$, there exists a compact set K such that $\int_{E\setminus K} |f| < \epsilon/4$ and a simple function $\tilde{h}$ with $\text{supp}(\tilde{h}) \subset K$ such that $\int_K |f - \tilde{h}| < \epsilon/4$. Define $h = \tilde{h}\chi_K$, then

$$\int_E |f - h| = \int_E |f - \tilde{h}\chi_K| = \int_K |f - \tilde{h}| + \int_{E\setminus K} |f| < \frac{\epsilon}{2}.$$

Let $M > 0$ be such that $|h| \le M$. By Theorem 3.36 (Lusin), for $\delta = \epsilon/(4M)$, there exists (closed) $F \subset K$ and a continuous function g where $|g| \le M$ and $\text{supp}(g) \subset K$, such that $\mu(K \setminus F) < \delta = \epsilon/(4M)$ and $h|_F = g|_F$. Then

$$\int_E |h - g| = \int_K |h - g| = \int_{K\setminus F} |h - g| \le 2M\mu(K \setminus F) < 2M \cdot \frac{\epsilon}{4M} = \frac{\epsilon}{2}.$$

Hence $\int_E |f - g| \le \int_E |f - h| + \int_E |h - g| < \epsilon$. $\square$

Corollary 4.41. *Suppose $f \in L(E)$. Then there exists a sequence of continuous functions $\{g_k\}$ with bounded support for every k, such that $\int_E |g_k - f| \to 0$ and $g_k \to f$ a.e. E.*

Proof. The first claim follows the theorem above. Since $\int_E |g_k - f| \to 0$, we know $g_k \xrightarrow{\mu} f$ and there exists a subsequence of $\{g_k\}$, still denoted by $\{g_k\}$, such that $g_k \to f$ a.e. E. $\square$

Example 4.42. Suppose $f \in L(\mathbb{R}^n)$. If $\int_{\mathbb{R}^n} f\phi = 0$ for any continuous function ϕ with compact support, then $f = 0$ a.e.

Proof. Suppose not, i.e., $\mu(\{x : f(x) \neq 0\}) > 0$. WLOG, assume $\mu(\{x : f(x) > 0\}) > 0$, then there exists $c > 0$ and $E \subset \mathbb{R}^n$ such that $\mu(E) > 0$ and $f(x) \ge c$ on E (because $\{x : f(x) > 0\} = \cup_{k=1}^\infty E_k$ where $E_k = \{x : f(x) \ge \frac{1}{k}\}$ and $0 < \mu(\cup_{k=1}^\infty E_k) \le \sum_{k=1}^\infty \mu(E_k)$, which means $\mu(E_k) > 0$ for some k). Then there exists a sequence of continuous functions $\{\phi_k\}$ with compact supports such that $\phi_k \uparrow \chi_E$ and $\int |\chi_E - \phi_k| \to 0$. Note that $|f\phi_k| \le |f\chi_E| \le |f| \in L(\mathbb{R}^n)$, we have

$$0 < c\mu(E) \le \int f\chi_E = \int \lim_{k\to\infty} f\phi_k = \lim_{k\to\infty} \int f\phi_k = 0,$$

where we applied Theorem 4.28 (DCT) to obtain the second equality. Contradiction. $\square$

Example 4.43. Suppose $f \in L([a, b])$. If $\int_a^b f\phi' = 0$ for any differentiable function ϕ with support $\text{supp}(\phi) \subset (a, b)$, then $f \equiv c$ a.e. for some c.

Proof. For any continuous function g with $\text{supp}(g) \subset (a, b)$ and continuous function h with $\text{supp}(h) \subset (a, b)$ and $\int_a^b h = 1$, define

$$\phi(x) = \int_a^x g(t)\,\mathrm{d}t - \int_a^b g(t)\,\mathrm{d}t \cdot \int_a^x h(t)\,\mathrm{d}t.$$

Note that $\mathrm{supp}(\phi) \subset (a, b)$. Then $\phi'(x) = g(x) - \int_a^b g(t)\,\mathrm{d}t \cdot h(x)$ and $\int_a^b f\phi' = \int_a^b fg - \int_a^b fh \int_a^b g = \int_a^b (f - \int_a^b fh) g(x)\,\mathrm{d}x$. Since g is arbitrary, we have $f(x) = \int_a^b fh$ a.e. for all continuous h and $\int_a^b h = 1$. $\square$

Theorem 4.44. *If $f \in L(\mathbb{R}^n)$, then $\lim_{h\to 0} \int_{\mathbb{R}^n} |f(x+h) - f(x)|\,\mathrm{d}x = 0$.*

Proof. By Theorem 4.40 above, we can decompose $f = f_1 + f_2$, where f_1 is continuous and has compact support, and f_2 is such that $\int_{\mathbb{R}^n} |f_2| < \epsilon/4$. Since f_1 is continuous on a compact set K, f_1 is uniformly continuous, there exists $\delta > 0$ such that $|f_1(x+h) - f_1(x)| < \epsilon/(2\mu(K))$ for any $h \in (0, \delta)$. Therefore

$$\int_{\mathbb{R}^n} |f_1(x+h) - f_1(x)|\,\mathrm{d}x = \int_K |f_1(x+h) - f_1(x)|\,\mathrm{d}x < \mu(K) \cdot \frac{\epsilon}{2\mu(K)} = \frac{\epsilon}{2}.$$

Therefore, we obtain that

$$\int_{\mathbb{R}^n} |f(x+h) - f(x)| \leq \int_{\mathbb{R}^n} |f_1(x+h) - f_1(x)| + \int_{\mathbb{R}^n} |f_2(x+h) - f_2(x)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

where we used the fact that $\int_{\mathbb{R}^n} |f_2(x+h) - f_2(x)| \leq 2\int_{\mathbb{R}^n} |f_2| < \epsilon/2$. $\square$

Corollary 4.45. *If $f \in L(E)$, then there exists a sequence of simple functions $\{\phi_k\}$ where $\mathrm{supp}(\phi_k)$ is compact for every k, such that $\phi_k \to f$ a.e. E and $\int_E |\phi_k - f| \to 0$.*

Proof. For any $\epsilon > 0$, there exists continuous g with bounded $\mathrm{supp}(g)$ such that $\int_E |f - g| < \frac{\epsilon}{2}$. For g, there exists simple function ϕ such that $\int |g - \phi| < \frac{\epsilon}{2}$. Hence $\int |f - \phi| < \epsilon$. Let $\epsilon_k = \frac{1}{k}$, then there exists ϕ_k such that $\int |f - \phi_k| \to 0$. Hence $\phi_k \xrightarrow{\mu} f$ and there exists a subsequence of $\{\phi_k\}$, still denoted by $\{\phi_k\}$, such that $\phi_k \to f$ a.e. $\square$

4.4 Relation between Riemann and Lebesgue integrals

We consider the integrals of bounded functions defined on $I = [a, b]$ only. Recall the definition of Riemann integral: consider a partition $\Delta^{(n)} : a = x_0^{(n)} < x_1^{(n)} < \cdots < x_{k_n}^{(n)} = b$ of I into k_n segments. Denote $|\Delta^{(n)}| = \max_{1\leq i\leq k_n} |x_i^{(n)} - x_{i-1}^{(n)}|$. For such a partition $\Delta^{(n)}$, denote

$$M_i^{(n)} = \sup\{f(x) : x_{i-1}^{(n)} \leq x \leq x_i^{(n)}\}, \quad m_i^{(n)} = \inf\{f(x) : x_{i-1}^{(n)} \leq x \leq x_i^{(n)}\}$$

Then the Darboux upper and lower integrals are defined by the two limits below as $|\Delta^{(n)}| \to 0$ and $n \to \infty$:

$$\overline{\int_a^b} f = \lim_{|\Delta^{(n)}|\to 0} \sum_{i=1}^{k_n} M_i^{(n)} (x_i^{(n)} - x_{i-1}^{(n)}), \quad \underline{\int_a^b} f = \lim_{|\Delta^{(n)}|\to 0} \sum_{i=1}^{k_n} m_i^{(n)} (x_i^{(n)} - x_{i-1}^{(n)})$$

Definition 4.46 (Riemann integral)**.** We call a function $f : I \to \mathbb{R}$ *Riemann integrable on I* if $\overline{\int_a^b} f = \underline{\int_a^b} f$.

For a sequence of partitions $\{\Delta^{(n)} : n \in \mathbb{N}\}$ where $\Delta^{(n+1)}$ is a refinement of $\Delta^{(n)}$ for every n (i.e., $\Delta^{(n+1)}$ retains all the partition points in $\Delta^{(n)}$ and may add new points). Define $w_n(x) = \sum_{i=1}^{k_n}(M_i^{(n)} - m_i^{(n)})\chi_{[x_i^{(n)},x_{i-1}^{(n)})}(x) \geq 0$, then $w_{n+1}(x) \leq w_n(x)$ for all n and x. Suppose $|f| \leq M$ for some $M > 0$, then $|w_n(x)| \leq 2M(b-a)$.

We also define the *oscillation* of the function f at every point $x \in I$ by

$$w_f(x) = \lim_{\delta\to 0} \sup\{|f(x') - f(x'')| : x', x'' \in B(x, \delta)\}$$

It is easy to verify that f is continuous at x iff $w_f(x) = 0$: Sufficiency is trivial; For necessity, for any $\epsilon > 0$, there exists $\delta_0 > 0$ such that $|f(x') - f(x)| < \epsilon/4$ for all $x' \in B(x, \delta_0)$. Hence $|f(x') - f(x'')| < \epsilon/2$ for all $x', x'' \in B(x, \delta_0)$. Denote $W_f(x, \delta) := \sup\{|f(x') - f(x'')| : x', x'' \in B(x, \delta)\}$, then for all $\delta \in (0, \delta_0)$, there is $W_f(x, \delta) \leq W_f(x, \delta_0) \leq \epsilon/2 < \epsilon$. Hence $w_f(x) = \lim_{\delta\to 0} W_f(x, \delta) = 0$.

Lemma 4.47. *The oscillation $w_f : (a, b) \to \mathbb{R}$ is a measurable function.*

Proof. It suffices to show that $E_t := w_f^{-1}((-\infty, t))$ is measurable (we actually can show it is open) for any $t \in \mathbb{R}$. For any $x \in E_t$, we know $w_f(x) < t$, and hence there exists $\delta_0 > 0$ such that $W(x, \delta_0) < t$. For any $y \in B(x, \delta_0)$, there exists $\delta_y > 0$ such that $B(y, \delta_y) \subset B(x, \delta_0)$, and hence $W(y, \delta) \leq W(y, \delta_y) \leq W(x, \delta_0) < t$ for all $\delta < \delta_y$. This implies that $w_f(y) = \lim_{\delta\to 0} W(y, \delta) < t$, i.e., $y \in E_t$. As y is arbitrary, we know $B(x, \delta_0) \subset E_t$, which means E_t is open. $\square$

It is trivial to extend the domain of w_f to $I = [a, b]$ and keep its measurability. We hereafter consider its domain on I. Moreover, it is easy to verify that $w_n \to w_f$ as $|\Delta^{(n)}| \to 0$ and $n \to \infty$. Since $|w_n| \leq 2M(b-a)$ and $\mu(I) < \infty$, we know $\int w_n \to \int w_f$ by DCT. Due to the definition of the Darboux upper and lower integrals, we also have $\int w_f = \overline{\int_a^b} f - \underline{\int_a^b} f$.

Theorem 4.48. *Suppose $f : [a, b] \to \mathbb{R}$ is bounded. Then f is Riemann integrable iff $\mu(\{x \in [a, b] : f(x)$ not continuous at $x\}) = 0$*

Proof. Necessity. If f is Riemann integrable, then $\int w_f = 0$. Hence $w_f = 0$ a.e.

Sufficiency. Suppose f is continuous a.e. Then $w_f = 0$ a.e. Then $0 = \int w_f = \overline{\int_a^b} f - \underline{\int_a^b} f$. Hence f is Riemann integrable. $\square$

Theorem 4.49. *If f is Riemann integrable on $[a, b]$, then $f \in L([a, b])$.*

Proof. From the theorem above, f is continuous a.e. $[a, b]$. For any partition $\Delta : a = x_0 < x_1 < \cdots < x_n = b$, there is $\int_I f = \sum_{i=1}^n \int_{x_{i-1}}^{x_i} f$. Hence $\sum_{i=1}^n m_i(x_i - x_{i-1}) \leq \int_a^b f \leq \sum_{i=1}^n M_i(x_i - x_{i-1})$. Since f is Riemann integrable by taking limit $|\Delta| \to 0$, we obtain $\overline{\int_a^b} f = \int_a^b f = \underline{\int_a^b} f$ $\square$

4.5 Iterated integrals

We denote $\mathcal{F}$ the set of all nonnegative measurable functions $f : \mathbb{R}^n = \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}_+$ that satisfy the following three properties:

1. For a.e. $x \in \mathbb{R}^p$, $f(x, \cdot) \geq 0$ is measurable on $\mathbb{R}^q$.
2. Let $F_f(x) := \int_{\mathbb{R}^q} f(x, y) \,\mathrm{d}y$. Then F_f is measurable and $F_f \geq 0$ a.e. $\mathbb{R}^p$.
3. There is $\int_{\mathbb{R}^p} F_f \,\mathrm{d}x = \int_{\mathbb{R}^p} \left(\int_{\mathbb{R}^q} f(x, y) \,\mathrm{d}y\right) \mathrm{d}x = \int_{\mathbb{R}^n} f \,\mathrm{d}x \,\mathrm{d}y$.

Lemma 4.50. *The following statements hold for the set $\mathcal{F}$:*

(i) If $f \in \mathcal{F}$ and $a \geq 0$, then $af \in \mathcal{F}$.
(ii) If $f_1, f_2 \in \mathcal{F}$, then $f_1 + f_2 \in \mathcal{F}$.
(iii) If $f, g \in \mathcal{F}$, $f - g \geq 0$, and $g \in L(\mathbb{R}^n)$, then $f - g \in \mathcal{F}$.
(iv) If $f_k \in \mathcal{F}$, $f_k \leq f_{k+1}$ for all k, and $\lim_k f_k = f$, then $f \in \mathcal{F}$.

Proof. It is trivial to verify (i) and (ii). For (iii), since $g \in \mathcal{F}$ and $g \in L(\mathbb{R}^n)$, we know F_g is finite a.e. $\mathbb{R}^p$. For every $x \in \mathbb{R}^p$ where $F_g(x) < \infty$, we know $g(x, \cdot) \in L(\mathbb{R}^q)$ and hence $g(x, \cdot)$ is finite a.e. $\mathbb{R}^q$. Hence $g(x, y)$ is finite a.e. $\mathbb{R}^n$. Then it is easy to verify the three properties of $f - g$, which implies $f - g \in \mathcal{F}$.

For (iv), it is easy to verify Property 1 of $\mathcal{F}$ for f. By Theorem 4.9 (Beppo-Levi), we know

$$F_f(x) = \int_{\mathbb{R}^q} f(x, y) \,\mathrm{d}y = \lim_{k \to \infty} \int_{\mathbb{R}^q} f_k(x, y) \,\mathrm{d}y = \lim_{k \to \infty} F_{f_k}(x)$$

which implies that $F_f \geq 0$ is measurable (as the limit of a sequence of measurable functions). Moreover, as $F_{f_k} \uparrow F_f$, we know

$$\int_{\mathbb{R}^p} F_f(x) \,\mathrm{d}x = \lim_{k \to \infty} \int_{\mathbb{R}^p} F_{f_k}(x) \,\mathrm{d}x = \lim_{k \to \infty} \int_{\mathbb{R}^n} f_k \,\mathrm{d}x \,\mathrm{d}y = \int_{\mathbb{R}^n} f \,\mathrm{d}x \,\mathrm{d}y$$

where we used Theorem 4.9 (Beppo-Levi) to obtain the first equality, the Property 3 of $f_k \in \mathcal{F}$ for the second equality, and Beppo-Levi again for the last equality. This verifies Property 3 of f. Hence $f \in \mathcal{F}$. $\square$

Theorem 4.51 (Tonelli)**.** *If $f : \mathbb{R}^n = \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}_+$ is measurable, then $f \in \mathcal{F}$.*

Proof. By Lemma 4.50(iv) and that every measurable function is a limit of a sequence of simple functions, it suffices to prove the case where f is simple. Due to Lemma 4.50(ii), it suffices to prove $\chi_E \in \mathcal{F}$ where E is measurable. As E can be written as a disjoint union of an F_σ-set K and measure zero set Z, we prove the claim in the following steps.

Firstly, it is easy to verify that $\chi_E \in \mathcal{F}$ if E is a (possibly) half-open half-closed box (an open box plus some of its $2n$ facets) in $\mathbb{R}^n$.

Secondly, for any open $G \subset \mathbb{R}^n$, we can rewrite G as a disjoint union $G = \cup_{k=1}^\infty I_k$ where each I_k is a half-open half-closed box. Let $E_k = \cup_{j=1}^k I_j$, then we know that $\chi_{E_k} \in \mathcal{F}$ from Lemma 4.50(ii), and then $\chi_E \in \mathcal{F}$ from $\chi_{E_k} \uparrow \chi_G$ and Lemma 4.50(iv).

Thirdly, we show that $\chi_F \in \mathcal{F}$ if $F \subset \mathbb{R}^n$ is bounded and closed. To this end, we first know $F \subset G_1 := B(0, k)$ for some $k \in \mathbb{N}$. Hence $G_2 = G_1 \setminus F = G_1 \cap F^c$

is open. Since $\chi_{G_1} - \chi_{G_2} \geq 0$ and $\chi_{G_2} \in L(\mathbb{R}^n)$, we know by Lemma 4.50(iii) that $\chi_F = \chi_{G_1} - \chi_{G_2} \in \mathcal{F}$.

Fourthly, we show that if $E_k \downarrow E$ and $\mu(E_1) < \infty$, $\chi_{E_k} \in \mathcal{F}$ for all k, then $\chi_E \in \mathcal{F}$. To this end, let $D_k = E_1 \setminus E_k$ and $D = E_1 \setminus E$, then $0 \leq \chi_{D_k} \uparrow \chi_D$ and $\chi_D \in L(\mathbb{R}^n)$ (since $\mu(E_1 \setminus E) \leq \mu(E_1) < \infty$). Hence by Lemma 4.50(iv) and Theorem 4.28 (DCT) we can see that $\chi_D \in \mathcal{F}$. Hence $\chi_E = \chi_{E_1} - \chi_D$ by Lemma 4.50(iii).

Fifthly, we show that if $\mu(E) = 0$ then $\chi_E \in \mathcal{F}$. To this end, consider a sequence of non-increasing open sets G_k such that $E \subset G_k$ and $\mu(G_k) \to 0$. Let $H = \cap_{k=1}^{\infty} G_k$, then $E \subset H$ and $\mu(H) = 0$. From the second and fourth points above, we know $\chi_H \in \mathcal{F}$ and hence $\chi_H(x, \cdot) = 0$ and $\int_{\mathbb{R}^n} \chi_H = 0$. As $0 \leq \chi_E \leq \chi_H$, we know χ_E satisfies all three properties of $\mathcal{F}$ and hence $\chi_E \in \mathcal{F}$.

Sixthly, we show that if K is an F_σ-set and $\mu(K) < \infty$, then $\chi_K \in \mathcal{F}$. Suppose $K = \cup_{k=1}^{\infty} F_k$ where F_k is closed and bounded for all k. Let $D_k = S_k \setminus S_{k-1}$ where $S_k := \cup_{j=0}^{k} F_k$ (assume $F_0 = \emptyset$). Note that both F_k and S_{k-1} are closed bounded sets, and hence by Lemma 4.50(iii) and the third point above, we know that $\chi_{D_k} = \chi_{S_k} - \chi_{S_{k-1}} \in \mathcal{F}$. Therefore, by Lemma 4.50(ii), $\chi_{\cup_{j=1}^{k} D_j} = \chi_{S_k} \in \mathcal{F}$. Since $S_k \uparrow K$, we know $\chi_K \in \mathcal{F}$ by Lemma 4.50(iv).

Finally, let $E = K \cup Z$ where K is an F_σ-set and Z is a measure zero set, and $K \cap Z = \emptyset$. Then $\chi_E = \chi_K + \chi_Z \in \mathcal{F}$. $\square$

Theorem 4.52 (Fubini)**.** *If $f \in L(\mathbb{R}^n)$, then the following statements hold:*

1. *For a.e. $x \in \mathbb{R}^p$, $f(x, \cdot) \in L(\mathbb{R}^q)$.*
2. *Let $F_f(x) = \int_{\mathbb{R}^q} f(x, y)\,\mathrm{d}y$, then $F_f(x) \in L(\mathbb{R}^p)$.*
3. *There is $\int_{\mathbb{R}^n} f = \int_{\mathbb{R}^p} \left(\int_{\mathbb{R}^q} f(x, y)\,\mathrm{d}y\right)\mathrm{d}x = \int_{\mathbb{R}^q} \left(\int_{\mathbb{R}^p} f(x, y)\,\mathrm{d}x\right)\mathrm{d}y$.*

Proof. Let $f = f^+ - f^-$. Since $f \in L(\mathbb{R}^n)$, we know $f^\pm \in L(\mathbb{R}^n)$. From Theorem 4.51 we know $f^\pm \in \mathcal{F}$, which implies the claims as all integrals are finite. $\square$

Example 4.53. Suppose $f \in L([0, \infty))$ and $a > 0$, then

$$\int_0^\infty \Big(\int_0^\infty f(y) e^{-xy}\,\mathrm{d}y\Big) \sin(ax)\,\mathrm{d}x = a \int_0^\infty \frac{f(y)}{a^2 + y^2}\,\mathrm{d}y.$$

Proof. Note that for any fixed $y > 0$, there is $\int_0^\infty e^{-xy} \sin(ax)\,\mathrm{d}x = \frac{a}{a^2+y^2}$ (apply integration by parts twice). Hence we just need to show the condition in Theorem 4.52 (Fubini) holds, i.e., $\sin(ax) f(y) e^{-xy} \in L([0, \infty)^2)$.

Note that $|\sin(ax) f(y) e^{-xy}| \leq |f(y)| \in L([\delta, X] \times (0, \infty))$ for any $0 < \delta < X < \infty$ (the integrand is bounded by $\int_0^\infty |f(y)|\,\mathrm{d}y$). Hence, by Theorem 4.52 (Fubini) on $[\delta, X] \times (0, \infty)$, there is

$$\int_\delta^X \Big(\int_0^\infty f(y) e^{-xy}\,\mathrm{d}y\Big) \sin(ax)\,\mathrm{d}x = \int_\delta^X \int_0^\infty \sin(ax) f(y) e^{-xy}\,\mathrm{d}y\,\mathrm{d}x$$

For any fixed $y > 0$, e^{-xy} is decreasing in x and hence the second mean value theorem for integrals implies that there exists $c \in (\delta, X)$ such that

$$\int_\delta^X e^{-xy} \sin(ax)\,\mathrm{d}x = e^{-\delta y} \int_\delta^c \sin(ax)\,\mathrm{d}x + e^{-Xy} \int_c^X \sin(ax)\,\mathrm{d}x.$$

Therefore, we can show that

$$\Big|\int_\delta^X e^{-xy} \sin(ax)\,\mathrm{d}x\Big| \le \Big|\int_\delta^c \sin(ax)\,\mathrm{d}x\Big| + \Big|\int_c^X \sin(ax)\,\mathrm{d}x\Big| \le \frac{4}{a}$$

for all $0 < \delta < X < \infty$. Let $\delta_k \to 0$ and $X_k \to \infty$ as $k \to \infty$, and denote $G_k(y) := \int_{\delta_k}^{X_k} e^{-xy} \sin(ax) f(y)\,\mathrm{d}x$, then we know $G_k(y) \to \int_0^\infty e^{-xy} \sin(ax) f(y)\,\mathrm{d}x$ as $k \to \infty$, and $|G_k(y)| \le 4|f(y)|/a \in L((0,\infty))$. Then we have

$$\begin{aligned}
\int_0^\infty \Big(\int_0^\infty f(y) e^{-xy}\,\mathrm{d}y\Big) \sin(ax)\,\mathrm{d}x &= \lim_{k\to\infty} \int_{\delta_k}^{X_k} \Big(\int_0^\infty f(y) e^{-xy}\,\mathrm{d}y\Big) \sin(ax)\,\mathrm{d}x \\
&= \lim_{k\to\infty} \int_0^\infty \Big(\int_{\delta_k}^{X_k} f(y) e^{-xy} \sin(ax)\,\mathrm{d}x\Big)\,\mathrm{d}y \\
&= \lim_{k\to\infty} \int_0^\infty G_k(y)\,\mathrm{d}y \\
&= \int_0^\infty \Big(\int_0^\infty e^{-xy} \sin(ax) f(y)\,\mathrm{d}x\Big)\,\mathrm{d}y
\end{aligned}$$

where the second equality is due to Theorem 4.52 (Fubini) on $[\delta_k, X_k] \times (0,\infty)$, the third equality is due to the definition of G_k, and the last equality is due to Theorem 4.28 applied to the sequence $G_k(y)$. $\square$

Example 4.54. Show that $\int_0^\infty e^{-x^2}\,\mathrm{d}x = \frac{\sqrt{\pi}}{2}$.

Proof. Consider $\int_0^\infty \int_0^\infty y e^{-y^2x^2} e^{-y^2}\,\mathrm{d}x\,\mathrm{d}y$. Then by Theorem 4.51 (Tonelli), we have

$$\begin{aligned}
\int_0^\infty \int_0^\infty y e^{-y^2x^2} e^{-y^2}\,\mathrm{d}x\,\mathrm{d}y &= \int_0^\infty e^{-y^2} \Big(\int_0^\infty y e^{-y^2x^2}\,\mathrm{d}x\Big)\,\mathrm{d}y \\
&= \Big(\int_0^\infty e^{-y^2}\,\mathrm{d}y\Big) \cdot \Big(\int_0^\infty e^{-u^2}\,\mathrm{d}u\Big) \\
&= \Big(\int_0^\infty e^{-y^2}\,\mathrm{d}y\Big)^2
\end{aligned}$$

where we applied the change of variable $u = yx$. On the other hand,

$$\int_0^\infty \int_0^\infty y e^{-y^2x^2} e^{-y^2}\,\mathrm{d}y\,\mathrm{d}x = \int_0^\infty \frac{1}{2(x^2+1)}\,\mathrm{d}x = \frac{1}{2} \arctan(x)\big|_0^\infty = \frac{\pi}{4}.$$

Hence $\int_0^\infty e^{-y^2}\,\mathrm{d}y = \frac{\sqrt{\pi}}{2}$. $\square$

Remarks. An alternative proof is based on polar coordinate: $(\int_0^\infty e^{-y^2}\,\mathrm{d}y)^2 = \int_0^\infty \int_0^\infty e^{-x^2} e^{-y^2}\,\mathrm{d}x\,\mathrm{d}y = \int_0^{\pi/2} \int_0^\infty e^{-\rho^2} \rho\,\mathrm{d}\rho\,\mathrm{d}\theta = \frac{\pi}{4}$.

4.6 Convolution

Definition 4.55 (Convolution)**.** Suppose f, g are measurable on $E \subset \mathbb{R}^n$. We call $\int_{\mathbb{R}^n} f(x-y)g(y)\,\mathrm{d}y$, a function of x, the *convolution of f and g*, denoted by $f * g$ or $(f*g)(x)$, if the integral exists for every $x \in E$. Note that $f * g = g * f$.

Theorem 4.56. *If $f, g \in L(\mathbb{R}^n)$, then $f * g$ is finite a.e., and*

$$\int_{\mathbb{R}^n} |f * g| \leq \left(\int_{\mathbb{R}^n} |f|\right) \cdot \left(\int_{\mathbb{R}^n} |g|\right).$$

Proof. We first consider the case where $f, g \geq 0$. By Theorem 4.51 (Tonelli), there is

$$\begin{aligned}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(x-y)g(y)\,\mathrm{d}y\,\mathrm{d}x &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} f(x-y)\,\mathrm{d}x\right) g(y)\,\mathrm{d}y \\ &= \left(\int_{\mathbb{R}^n} f(x)\,\mathrm{d}x\right)\cdot\left(\int_{\mathbb{R}^n} g(y)\,\mathrm{d}y\right) < \infty.\end{aligned}$$

For general functions f, g, note that $|f * g| \leq |f| * |g|$. $\square$

Example 4.57 (Convolution is continuous)**.** Suppose $f \in L(\mathbb{R}^n)$, and g is measurable and uniformly bounded a.e., then $F(x) = (f * g)(x)$ is uniformly continuous on $\mathbb{R}^n$.

Proof. Suppose $M > 0$ is such that $|g| \leq M$ a.e. Since $f \in L(\mathbb{R}^n)$, by Theorem 4.44, we know for any $\epsilon > 0$, there exists $\delta > 0$ such that $\int_{\mathbb{R}^n} |f(x+h) - f(x)|\,\mathrm{d}x < \epsilon/M$ for all $|h| \leq \delta$. Hence,

$$\begin{aligned}|F(x+h) - F(x)| &\leq \int_{\mathbb{R}^n} |f(x+h-y) - f(x-y)||g(y)|\,\mathrm{d}y \\ &\leq M\int_{\mathbb{R}^n} |f(z+h) - f(z)|\,\mathrm{d}z < M \cdot \frac{\epsilon}{M} = \epsilon,\end{aligned}$$

for all $|h| \leq \delta$. Hence F is uniformly continuous on $\mathbb{R}^n$. $\square$

5 Signed Measures and Differentiations

5.1 Signed measure and decomposition

Definition 5.1 (Signed measure)**.** Suppose $(X, \mathcal{M})$ is a measurable space. A function $\nu : \mathcal{M} \to \bar{\mathbb{R}}$ is called a *signed measure* if (i) $\nu(\emptyset) = 0$; (ii) ν can take ∞ or $-\infty$ but not both; (iii) countable additivity: $\nu(\cup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} \nu(E_k)$ for any countable family of mutually disjoint sets $\{E_k\}$.

Note that if in addition $\nu(E) \geq 0$ for any $E \in \mathcal{M}$, then ν is called a *positive measure*, or simply *measure*.

Theorem 5.2 (Continuity of signed measure)**.** *If E_k is increasing and $E = \cup_{k=1}^{\infty} E_k$, then $\nu(E_k) \to \nu(E)$. If E_k is decreasing, $\nu(E_1) < \infty$, and $E = \cap_{k=1}^{\infty} E_k$, then $\nu(E_k) \to \nu(E)$.*

Proof. If E_k is increasing, then denote $D_k = E_k \setminus \cup_{i=1}^{k-1} E_i$. By countable additivity, there is

$$\nu(E) = \nu(\bigcup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} \nu(D_k) = \lim_{k\to\infty} \sum_{i=1}^{k} \nu(D_i) = \lim_{k\to\infty} \sum_{i=1}^{k} \nu(E_i) = \sum_{k=1}^{\infty} \nu(E_k).$$

If E_k is decreasing, then consider $F_k = E_1 \setminus E_k$ which is increasing, the rest follows similarly. $\square$

Definition 5.3 (ν-positive/negative/null set)**.** Suppose ν is a signed measure on $(X, \mathcal{M})$. Then a set E is called *ν-positive (resp. ν-negative, ν-null)* if $\nu(F) \geq 0$ (resp. ≤ 0, $= 0$) for any $F \subset E$.

Lemma 5.4. *If $\{P_k\}$ are ν-positive, then $\cup_{k=1}^{\infty} P_k$ is ν-positive.*

Proof. Denote $Q_k = P_k \setminus \cup_{i=1}^{k-1} P_i$, then for any $E \subset \cup_{k=1}^{\infty} P_k$, there is

$$\nu(E) = \nu(E \cap (\bigcup_{k=1}^{\infty} Q_k)) = \sum_{k=1}^{\infty} \nu(E \cap Q_k) \geq 0,$$

since $E \cap Q_k \subset P_k$. $\square$

Theorem 5.5 (Hahn Decomposition Theorem)**.** *If ν is a signed measure on $(X, \mathcal{M})$, then there exists positive P and negative N such that $X = P \cup N$, $P \cap N = \emptyset$. If P' and N' is another such pair, then $P \triangle P'$ and $N \triangle N'$ are null. The pair (P, N) is called a* Hahn decomposition *of ν.*

Proof. (i) WLOG, assume that $\nu : \mathcal{M} \to \mathbb{R} \cup \{-\infty\}$. Consider the family of ν-positive sets: $\mathcal{P} = \{P \in \mathcal{M} : P \text{ is } \nu\text{-positive}\}$. Define $m = \sup\{\nu(P) : P \in \mathcal{P}\}$, then there exists a sequence $\{P_k\} \subset \mathcal{P}$ such that $\lim_k \nu(P_k) = m < \infty$. Define $P = \cup_{k=1}^{\infty} P_k$ and $N = X \setminus P$. Then, by Lemma 5.4, P is ν-positive, and $\nu(P) = m$.

(ii) Now we only need to show that N is ν-negative. First of all, if $E \subset N$ and $\nu(E) > 0$, then E cannot be ν-positive: otherwise $E \cap P \subset N \cap P = \emptyset$ and $E \cup P$ is ν-positive, and hence $\nu(E \cup P) = \nu(E) + \nu(P) > m$, contradiction.

Next, for any $E \subset N$ with $\nu(E) > 0$, there exists $B \subset E$ such that $\nu(B) < 0$ (since E is not ν-positive), then let $A = E \setminus B$, we have $A \subset E$ and $\nu(A) = \nu(E) - \nu(B) > \nu(E)$.

Now we are ready to show that N is ν-negative. If not, then there exists $A_0 \subset N$ such that $\nu(A_0) > 0$. Then choose the smallest $n_1 \in \mathbb{N}$ such that there exists $A_1 \subset A_0$ that satisfies $\nu(A_1) \geq \nu(A_0) + \frac{1}{n_1} > \nu(A_0)$; then choose the smallest $n_2 \in \mathbb{N}$ such that there exists $A_2 \subset A_1$ and $\nu(A_2) \geq \nu(A_1) + \frac{1}{n_2} > \nu(A_1)$; and so on. Note that A_k is decreasing, $\nu(A_k)$ is increasing, and $\nu(A_k) \geq \nu(A_{k-1}) + \frac{1}{n_k}$ for all k. Let $A = \cap_{k=1}^{\infty} A_k = \lim_k A_k$. Then $\infty > \nu(A) = \lim_k \nu(A_k) \geq \nu(A_0) \sum_k \frac{1}{n_k} > 0$. Therefore $n_k \to \infty$ as $k \to \infty$. Since $A \subset N$ and $\nu(A) > 0$, there again exists $B \subset A$ and $n \in \mathbb{N}$ such that $\nu(B) \geq \nu(A) + \frac{1}{n} > \nu(A)$. Therefore, there exists k large enough, such that $n_k > n$, but $B \subset A \subset A_{k-1}$, and $\nu(B) \geq \nu(A) + \frac{1}{n} \geq \nu(A_{k-1}) + \frac{1}{n}$, which contradicts to the construction of n_k (smallest integer) and A_k. Hence N must be ν-negative.

(iii) If P' and N' is another such pair, then $P \setminus P' \subset P$ and $P \setminus P' \subset N'$. Hence $P \setminus P'$ is both ν-positive and ν-negative, and therefore is ν-null. Note that $N \triangle N' = P \triangle P'$ is therefore also ν-null. □

Definition 5.6 (Mutually singular measures)**.** Two signed measures μ and ν are called *mutually singular*, denoted by $\mu \perp \nu$, if there exist $E \in \mathcal{M}$ such that E is ν-null and E^c is μ-null.

Theorem 5.7 (Jordan Decomposition Theorem)**.** *If ν is a signed measure on $\mathcal{M}$, then there exist unique positive measures ν^+ and ν^-, such that $\nu^+ \perp \nu^-$ and $\nu = \nu^+ - \nu^-$.*

Proof. Let $X = P \cup N$ be a Hahn decomposition of ν. Define $\nu^{\pm}$ such that $\nu^+(E) = \nu(E \cap P)$ and $\nu^-(E) = -\nu(E \cap N)$ for any $E \in \mathcal{M}$. Then it is easy to verify that both ν^+ and ν^- are positive measures on $\mathcal{M}$. Moreover, for any $E \subset N$, $\nu^+(E) = \nu(E \cap P) = \nu(\emptyset) = 0$; and for any $E \subset P$, $\nu^-(E) = \nu(E \cap N) = \nu(\emptyset) = 0$. Hence N is ν^+-null and $P = N^c$ is ν^--null, i.e., $\nu^+ \perp \nu^-$.

If there exists another Hahn decomposition $X = P' \cup N'$ with $\mu^{\pm}$ defined similarly, then $P \triangle P'$ is ν-null. Hence, for any $E \in \mathcal{M}$, there is $\mu^+(E) = \nu(E \cap P') = \nu(E \cap P) = \nu^+(E)$. Hence $\mu^+ = \nu^+$. Similarly $\mu^- = \nu^-$. □

Definition 5.8 (Total variation of signed measure)**.** $|\nu| = \nu^+ + \nu^-$ is called the *total variation of* ν. That is, $|\nu|(E) = \nu^+(E) + \nu^-(E) = \nu(E \cap P) - \nu(E \cap N)$ for any $E \in \mathcal{M}$.

Definition 5.9 (Integrable wrt. signed measure)**.** We call f *integrable with respect to* ν, where the integral is denoted by $\int f \,\mathrm{d}\nu = \int f \,\mathrm{d}\nu^+ - \int f \,\mathrm{d}\nu^-$, if f is integrable with respect to both ν^+ and ν^-.

5.2 Radon-Nikodym theorem

Definition 5.10 (Absolute continuity)**.** We say a signed measure ν is *absolutely continuous* with respect to a positive measure μ, denoted by $\nu \ll \mu$, if $\nu(E) = 0$ for any $E \in \mathcal{M}$ with $\mu(E) = 0$. Note that $\nu \ll \mu$ iff $\nu^{\pm} \ll \mu$ iff $|\nu| \ll \mu$.

Theorem 5.11. *If $\nu \perp \mu$ and $\nu \ll \mu$, then $\nu = 0$.*

Proof. Let E be such that E is μ-null and E^c is ν-null. Then E is also ν-null since $\nu \ll \mu$. Hence X is ν-null, i.e., $\nu = 0$. $\square$

Theorem 5.12. *Suppose ν is a signed measure and μ is a measure, then $\nu \ll \mu$ iff for any $\epsilon > 0$, there exists $\delta > 0$ such that $|\nu(E)| < \epsilon$ for all E satisfying $\mu(E) < \delta$.*

Proof. Since $\nu \ll \mu$ iff $|\nu| \ll \mu$, we only need to show this for positive measure ν. Sufficiency is trivial. To prove necessity, assume that there exists $\epsilon_0 > 0$, such that for any $k \in \mathbb{N}$ there are $|\nu(E_k)| > \epsilon_0$ and $\mu(E_k) < 2^{-k}$. Let $F_k = \cup_{i=k}^{\infty} E_i$ and $F = \cap_{k=1}^{\infty} F_k$. Then $\mu(F_k) \leq 2^{1-k}$ and $\mu(F) = \lim_k \mu(F_k) = 0$. However $\nu(F_k) \geq \nu(E_k) \geq \epsilon_0$ for all k, which implies $\nu(F) = \lim_k \nu(F_k) \geq \epsilon_0$ and contradicts to $\nu \ll \mu$. $\square$

Lemma 5.13. *Suppose ν and μ are finite measures on $\mathcal{M}$. Then either $\nu \perp \mu$ or there exist $\epsilon > 0$ and $E \in \mathcal{M}$ with $\mu(E) > 0$, such that E is $(\nu - \epsilon\mu)$-positive.*

Proof. Consider signed measures $\nu - k^{-1}\mu$ with a Hahn decomposition $X = P_k \cup N_k$, for any $k \in \mathbb{N}$. Let $P = \cup_{k=1}^{\infty} P_k$ and $N = \cap_{k=1}^{\infty} N_k$. Note that N is $(\nu - k^{-1}\mu)$-negative for all k, and hence $0 \leq \nu(N) \leq k^{-1}\mu(N) \to 0$. Hence N is ν-null.

If P is μ-null, then $\nu \perp \mu$ and done. Otherwise, $\mu(P) > 0$, and hence there exists k such that $\mu(P_k) > 0$, and P_k is $(\nu - k^{-1}\mu)$-positive. Taking $E = P_k$ and $\epsilon = k^{-1}$ completes the proof. $\square$

Theorem 5.14 (Radon-Nikodym)**.** *Suppose ν is a σ-finite signed measure and μ is a σ-finite measure. Then there exist unique σ-finite signed measures λ and ρ, such that $\lambda \perp \mu$, $\rho \ll \mu$, and $\nu = \lambda + \rho$.*

Proof. (i) We first consider the case where both μ and ν are finite positive measures. Define

$$\mathcal{F} = \Big\{ f : X \to [0, \infty] : \int_E f \, \mathrm{d}\mu \leq \nu(E), \ \ \forall E \in \mathcal{M} \Big\}$$

Note that $0 \in \mathcal{F}$ and hence $\mathcal{F}$ is nonempty. For any $f, g \in \mathcal{F}$, $h = \max\{f, g\} \in \mathcal{F}$: let $A = \{x : f(x) \geq g(x)\}$, then

$$\int_E h = \int_{E \cap A} f + \int_{E \cap A^c} g \leq \nu(E \cap A) + \nu(E \cap A^c) = \nu(E).$$

Let $m = \sup\{\int_X f : f \in \mathcal{F}\} \leq \nu(X) < \infty$, then there exists a sequence $\{f_k\} \subset \mathcal{F}$, such that $\int_X f_k \to m$. Define $g_k(x) = \max_{1 \leq i \leq k} f_i(x)$ for all $x \in X$

and $f(x) = \sup_k f_k(x)$. Note that $g_k \in \mathcal{F}$ and $g_k \uparrow f$. By Theorem 4.9 (Beppo Levi), we know $m \le \lim_k \int_X f_k \le \lim_k \int_X g_k = \int_X f \le m$ and hence $\int_X f = m$. Moreover, $\int_E f = \lim_k \int_E g_k \le \nu(E)$ for all $E \in \mathcal{M}$, and hence $f \in \mathcal{F}$.

Now we claim that λ, defined by $\lambda(E) = \nu(E) - \int_E f \,\mathrm{d}\mu$ for any $E \in \mathcal{M}$ (we write this as $\mathrm{d}\lambda = \mathrm{d}\nu - f\,\mathrm{d}\mu$ for short), satisfies $\lambda \perp \mu$. If not, then by Lemma 5.13, there exists $\epsilon > 0$ and A, such that $\mu(A) > 0$ and A is $(\lambda - \epsilon\mu)$-positive. Then for any $E \in \mathcal{M}$, there is

$$\int_E (f+\epsilon\chi_A)\,\mathrm{d}\mu = \int_{E\cap A^c} f\,\mathrm{d}\mu + \int_{E\cap A} (f+\epsilon\chi_A)\,\mathrm{d}\mu \le \nu(E\cap A^c) + \nu(E\cap A) = \nu(E),$$

where we used the fact

$$\begin{aligned} 0 \le (\lambda - \epsilon\mu)(E\cap A) &= \lambda(E\cap A) - \epsilon\mu(E\cap A) \\ &= \nu(E\cap A) - \int_{E\cap A} f\,\mathrm{d}\mu - \int_E \epsilon\chi_A\,\mathrm{d}\mu \\ &= \nu(E\cap A) - \int_{E\cap A} (f+\epsilon\chi_A)\,\mathrm{d}\mu \end{aligned}$$

to obtain the inequality above. Hence $f + \epsilon\chi_A \in \mathcal{F}$. However $\int_X (f+\epsilon\chi_A)\,\mathrm{d}\mu = m + \epsilon\mu(A) > m$, contradiction.

If there exists f', λ' such that $\mathrm{d}\nu = \mathrm{d}\lambda' + f'\,\mathrm{d}\mu$, then $\mathrm{d}\lambda - \mathrm{d}\lambda' = f\,\mathrm{d}\mu - f'\,\mathrm{d}\mu = (f - f')\,\mathrm{d}\mu$. Hence $(\lambda - \lambda') \ll \mu$. Moreover, since $\lambda, \lambda' \perp \mu$, there exist E and E' such that E is λ-null, E' is λ'-null, and $E^c, (E')^c$ are μ-null. Hence $E \cap E'$ is $(\lambda - \lambda')$-null and $E^c \cup (E')^c$ is μ-null, which means $(\lambda - \lambda') \perp \mu$. Therefore, by Theorem 5.11, $\lambda - \lambda' = 0$, and hence $f - f' = 0$ μ-a.e.

(ii) Next we consider the case where both μ and ν are σ-finite measures. Since there exist $\{A_k\}$ and $\{B_k\}$ such that $X = \cup_k A_k = \cup_k B_k$, μ is finite on A_k and ν is finite on B_k. Then $\{A_i \cap B_j : i, j \in \mathbb{N}\}$ is countable. Denote this set by $\{C_k\}$ (WLOG assume they are disjoint, otherwise take $C_k \setminus (\cup_{j=1}^{k-1} C_j)$ for all k), then μ and ν are both finite on C_k for any k. Define $\mu_k(E) = \mu(E \cap C_k)$ and $\nu_k(E) = \nu(E \cap C_k)$ for any $E \in \mathcal{M}$ and k. Then applying (i) we know there exist unique λ_k, f_k such that $\mathrm{d}\lambda_k = \mathrm{d}\nu_k - f_k\,\mathrm{d}\mu_k$ on C_k. Let $\lambda = \sum_k \lambda_k$ and $f = \sum_k f_k$. Then it is easy to verify that $\lambda \perp \mu$ and $\nu = \lambda + f\,\mathrm{d}\mu$ on X.

(iii) Finally consider the general case where ν is σ-finite signed measure. Let $\nu = \nu^+ - \nu^-$ be the Jordan decomposition of ν, then applying (ii) to each of $\nu^\pm$ yields unique $\lambda^\pm, f^\pm$ such that $\lambda^\pm = \nu^\pm - f^\pm\,\mathrm{d}\mu$ and $\lambda^\pm \perp \mu$. Let $\lambda = \lambda^+ - \lambda^-$ and $f = f^+ - f^-$, then $\lambda \perp \mu$ and $\nu = \lambda + f\,\mathrm{d}\mu$, which completes the proof. □

Definition 5.15 (Lebesgue decomposition and Radon-Nikodym derivative)**.** We call $\mathrm{d}\nu = \mathrm{d}\lambda + f\,\mathrm{d}\mu$ from Theorem 5.14 the *Lebesgue decomposition* of ν with respect to μ. If $\nu \ll \mu$, then $\mathrm{d}\nu = f\,\mathrm{d}\mu$ and f is called the *Radon-Nikodym derivative* of ν with respect to μ, denoted by $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.

Theorem 5.16. *Suppose ν is σ-finite signed measure, μ, λ are σ-finite measures on $(X, \mathcal{M})$, and $\nu \ll \mu \ll \lambda$. Then the following statements hold:*

1. *(Change of variable) If g is ν-integrable, then $g \cdot \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ is μ-integrable, and $\int g \,\mathrm{d}\nu = \int g \cdot \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \,\mathrm{d}\mu$.*
2. *(Chain rule) $\frac{\mathrm{d}\nu}{\mathrm{d}\lambda} = \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \cdot \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ λ-a.e.*

Proof. 1. By considering $\nu^{\pm}$ separately, it suffices to prove the result for positive measure ν. We first verify the claim for $g = \chi_E$ where $E \in \mathcal{M}$:

$$\int g \,\mathrm{d}\nu = \nu(E) = \int_E \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \,\mathrm{d}\mu = \int \chi_E \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \,\mathrm{d}\mu.$$

(Note that we identify $\mathrm{d}\nu / \mathrm{d}\mu$ with f where $\mathrm{d}\nu = f \,\mathrm{d}\mu$.) Then it is easy to verify this for g being nonnegative simple functions by linearity, then general nonnegative functions, and finally for general function g.

2. By Item 1, we have $\int g \,\mathrm{d}\mu = \int g \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \,\mathrm{d}\lambda$ for all μ-integrable function g. Then for any $E \in \mathcal{M}$, we substitute g by $\chi_E \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$, then

$$\nu(E) = \int \chi_E \,\mathrm{d}\nu = \int \chi_E \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \,\mathrm{d}\mu = \int_E \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \,\mathrm{d}\lambda.$$

This implies that $\frac{\mathrm{d}\nu}{\mathrm{d}\lambda} = \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ a.e. λ.

$\square$

5.3 Differentiation

We focus on the case where $f : \mathbb{R} \to \mathbb{R}$ in the remainder of this chapter.

Definition 5.17 (Vitali cover)**.** The collection $\mathcal{F}$ of closed intervals is called a *Vitali cover* of E if for any $\epsilon > 0$ and any $x \in E$, there exists $I \in \mathcal{F}$ such that $\mu(I) < \epsilon$ and $x \in I$.

Example 5.18. Suppose $E = [a, b]$. Let $\{r_k\} = [a, b] \cap \mathbb{Q}$ and $I_{k,m} = [r_k - \frac{1}{m}, r_k + \frac{1}{m}]$ for $k, m \in \mathbb{N}$. Then $\mathcal{F} = \{I_{k,m} : k, m \in \mathbb{N}\}$ is a Vitali cover of E.

Lemma 5.19 (Vitali covering lemma)**.** *Suppose $E \subset \mathbb{R}$ and $\mu^*(E) < \infty$. If $\mathcal{F}$ is a Vitali cover of E, then for any $\epsilon > 0$ there exist a finite number of disjoint sets $\{I_j : 1 \le j \le k\} \subset \mathcal{F}$, such that $\mu^*(E \setminus \cup_{j=1}^k E_j) < \epsilon$.*

Proof. WLOG, we assume $\mathcal{F}$ only contains bounded closed intervals. Since $\mu^*(E) < \infty$, there exists an open set G such that $E \subset G$ and $\mu(G) < \infty$. Since $\mathcal{F}$ is a Vitali cover of E, WLOG we assume $I \subset G$ for any $I \in \mathcal{F}$.

Now we perform the following interval selection procedure: we first choose $I_1 \in \mathcal{F}$ arbitrarily. Inductively, suppose we have already chosen $I_1, \ldots, I_k \in \mathcal{F}$. If $E \subset \cup_{j=1}^k I_j$, then we can terminate because the claim is proved. Otherwise, we denote $\mathcal{F}_k := \{I \in \mathcal{F} : I \cap (\cup_{j=1}^k I_j) = \emptyset\}$ and $\delta_k := \sup\{|I| : I \in \mathcal{F}_k\}$ (here $|I|$ denotes the length of the interval I for short), and then choose $I_{k+1} \in \mathcal{F}$ such that $|I_{k+1}| > \frac{\delta_k}{2}$ (this is possible since δ_k is taken as the supremum over $\mathcal{F}_k$).

If this set selection procedure continues for infinitely many steps, then we obtain a sequence of intervals $\{I_j\}_{j=1}^{\infty}$. Since $\sum_{k=1}^{\infty}|I_k| = \mu(\cup_{k=1}^{\infty} I_k) \le \mu(G) < \infty$, we know $\sum_{j=k+1}^{\infty}|I_j| \to 0$ as $k \to \infty$.

Now let $\epsilon > 0$ be arbitrary and fixed and k large enough so that $\sum_{j=k+1}^{\infty}|I_j| < \frac{\epsilon}{5}$. Denote $S := E \setminus (\cup_{j=1}^{k} I_j)$. Then we want to show $\mu^*(S) < \epsilon$. To this end, let $x \in S$ be arbitrary, then $x \notin \cup_{j=1}^{k} I_j$. Notice that $\cup_{j=1}^{k} I_j$ is a closed set, we know there exists $I \in \mathcal{F}$ such that $x \in I$ and $I \cap (\cup_{j=1}^{k} I_j) = \emptyset$. Moreover, $|I| \le \delta_k < 2|I_{k+1}|$ due to the criterion to select I_{k+1}.

Furthermore, notice that $|I_j| \to 0$ as $j \to \infty$. In addition, $I \cap (\cup_{j=k+1}^{\infty} I_j) \ne \emptyset$ because otherwise we would have selected I over some I_j during the procedure (the former has a fixed width while the width of the latter tends to 0 as $j \to 0$). Let $k_0 \ge k+1$ be the smallest index such that $I \cap I_{k_0} \ne \emptyset$, then $|I| \le \delta_{k_0 - 1} < 2|I_{k_0}|$. Now for each $j \ge k+1$ we define I'_k to be the closed interval with the same center as I_k but 5 times larger radius, then $x \in I \subset I'_{k_0-1}$. Since $x \in S$ is arbitrary, we know $S \subset \cup_{j=k+1}^{\infty} I'_k$, and $\mu^*(S) \le \mu(\cup_{j=k+1}^{\infty} I'_j) \le 5\mu(\cup_{j=k+1}^{\infty} I_j) \le 5\sum_{j=k+1}^{\infty}|I_j| < \epsilon$. This completes the proof. □

Remarks. Vitali covering lemma can be extended to $\mathbb{R}^n$. It is easy to show that there exists a countable collection of sets $\{E_k\}$ such that $\mu^*(E \setminus (\cup_{k=1}^{\infty} E_k)) = 0$.

Definition 5.20 (Dini derivatives)**.** Suppose $f : \mathbb{R} \to \mathbb{R}$, define $D^{\pm}f(x)$ and $D_{\pm}f(x)$ at $x \in \mathbb{R}$ by

$$D^{\pm}f(x) = \limsup_{h\to 0^{\pm}} \frac{f(x+h) - f(x)}{h}, \quad D_{\pm}f(x) = \liminf_{h\to 0^{\pm}} \frac{f(x+h) - f(x)}{h}$$

Then $D^{\pm}$ is called the *upper right/left Dini derivative of f at x*. Similarly, $D_{\pm}$ is called the *lower right/left Dini derivative of f at x*.

Remarks. Here are several remarks regarding the four Dini derivatives:

- For any f and x, there are $D_+f(x) \le D^+f(x)$ and $D_-f(x) \le D^-f(x)$.
- $D^+(-f) = -D_+(f)$ and $D^-(-f) = -D_-(f)$.
- If $D_+f(x) = D^+f(x)$, then we say f has *right derivative at x*. Similarly, if $D_-f(x) = D^-f(x)$, then we say f has *left derivative at x*.
- If all four derivatives are equal, then f is called *differentiable at x*.

Example 5.21. Suppose $a < b$ and $a' < b'$, and define

$$f(x) = \begin{cases} ax\sin^2(\frac{1}{x}) + bx\cos^2(\frac{1}{x}), & x > 0, \\ 0, & x = 0, \\ a'x\sin^2(\frac{1}{x}) + b'x\cos^2(\frac{1}{x}), & x < 0. \end{cases}$$

Then we can show that

$$D^+f(0) = \limsup_{h\to 0^+} \frac{f(h) - f(0)}{h} = \limsup_{h\to 0^+} \left\{ a\sin^2\left(\frac{1}{h}\right) + b\cos^2\left(\frac{1}{h}\right) \right\} = b.$$

Similarly, $D_+f(0) = a$, $D^-f(0) = b'$, $D_-f(0) = a'$.

Example 5.22. If $f \in C([a,b])$, then there exist $x_0 \in (a,b)$ and $k \in \mathbb{R}$, such that $D_- f(x_0) \geq k \geq D^+ f(x_0)$ or $D^- f(x_0) \leq k \leq D_+ f(x_0)$.

Proof. Let $k = (f(b) - f(a))/(b-a)$. Consider $g(x) = f(x) - kx$. Then $g \in C([a,b])$. Note that $g(a) = f(a) - ka = (bf(a) - af(b))(b-a) = g(b)$. Hence there exists $x_0 \in C$ such that $g(x)$ attains max or min at $x_0 \in (a,b)$. If x_0 is a maximizer, then $D^+ g(x_0) = D^+ f(x_0) - k \leq 0$ and $D_- g(x_0) = D_- f(x_0) - k \geq 0$, which implies that $D_- f(x_0) \geq k \geq D^- f(x_0)$. Similarly, if x_0 is minimizer, then $D^- f(x_0) \leq k \leq D_+ f(x_0)$. ☐

Theorem 5.23 (Lebesgue)**.** *Suppose $f : [a,b] \to \mathbb{R}$ is non-decreasing, then f is differentiable a.e. $[a,b]$ and $\int_a^b f'(x)\,\mathrm{d}x \leq f(b) - f(a)$.*

Proof. (i) Note that if $D^+ f(x) \leq D_- f(x)$ and $D^- f(x) \leq D_+ f(x)$, then all four Dini derivatives are equal and f is differentiable at x. Hence, if f is not differentiable at x, then either $D^+ f(x) > D_- f(x)$ or $D^- f(x) > D_+ f(x)$. Let $E_1 = \{x : D^+ f(x) > D_- f(x)\}$ and $E_2 = \{x : D^- f(x) > D_+ f(x)\}$. We then need to show $\mu(E_1 \cup E_2) = 0$. To this end, it suffices to show that $\mu(E_1) = 0$, as $\mu(E_2) = 0$ can be proved similarly. Let $r, s \in \mathbb{Q}$ and $E_{r,s} = \{x : D^+ f(x) > r > s > D_- f(x)\}$, then $E_1 = \cup_{r,s \in \mathbb{Q}} E_{r,s}$. Hence it suffices to show that $\mu(E_{r,s}) = 0$ for all $r, s \in \mathbb{Q}$.

Now we denote $E = E_{r,s}$ for short. For any $\epsilon > 0$, consider an open set G such that $E \subset G$ and $\mu(G) < \mu^*(E) + \epsilon$ (such G exists due to the definition of outer measure), and define the collection of closed intervals:

$$\mathcal{G} = \{[x-h, x] \subset G : x \in [a,b], f(x) - f(x-h) < sh \text{ for some } h > 0\}$$

Thus $\mathcal{G}$ is a Vitali cover of E (since $x \in E$ implies that $D^- f(x) < s$). Hence there exist a finite number of disjoint intervals $[x_1 - h_1, x_1], \ldots, [x_p - h_p, x_p]$, such that $\mu^*(E) - \epsilon < \mu(E \cap (\cup_{i=1}^p [x_i - h_i, x_i]))$ and $\sum_{i=1}^p h_i \leq \mu(G) < \mu^*(E) + \epsilon$. Since $f(x_i) - f(x_i - h_i) < sh_i$, we have

$$\sum_{i=1}^p (f(x_i) - f(x_i - h_i)) < s \sum_{i=1}^p h_i < s(\mu^*(E) + \epsilon).$$

Now define $F = E \cap (\cup_{i=1}^p (x_i - h_i, x_i))$. Consider the collection of closed intervals

$$\mathcal{F} = \{[y, y+l] \subset F : f(y+l) - f(y) > rl \text{ for some } l > 0\}$$

Hence $\mathcal{F}$ is a Vitali cover of F, and there exist a finite number of disjoint intervals $[y_1, y_1 + l_1], \ldots, [y_q, y_q + l_q]$, such that $\sum_{j=1}^q l_j > \mu(F) - \epsilon > \mu^*(E) - 2\epsilon$. Hence $\sum_{j=1}^q (f(y_j + l_j) - f(y_j)) > r \sum_{j=1}^q l_j > r(\mu^*(E) - 2\epsilon)$. Since f is non-decreasing and $[y_j, y_j + l_j] \subset [x_i - h_i, x_i]$ for some i, we know $\sum_{i=1}^p (f(x_i) - f(x_i - h_i)) \geq \sum_{j=1}^q (f(y_j + l_j) - f(y_j))$. Hence $r(\mu^*(E) - 2\epsilon) < s(\mu^*(E) + \epsilon)$. Since $\epsilon > 0$ is arbitrary, we have $r\mu^*(E) \leq s\mu^*(E)$, which implies that $\mu^*(E) = 0$ since $r > s$.

(ii) Consider $f_k(x) = k(f(x+\frac{1}{k}) - f(x))$. Then $f_k \to f'$ a.e. and

$$\int_a^b f' = \int_a^b \lim_{k\to\infty} f_k \le \liminf_{k\to\infty} \int_a^b f_k = \liminf_{k\to\infty} \int_a^b k\Big(f(x+\frac{1}{k}) - f(x)\Big)$$
$$= \liminf_{k\to\infty} k\Big(\int_b^{b+\frac{1}{k}} f - \int_a^{a+\frac{1}{k}} f\Big) \le f(b) - f(a)$$

as $k \to \infty$, where we used Lemma 4.15 (Fatou) to obtain the first inequality and f is non-decreasing (and constant over $[b, b+\frac{1}{k}]$) to obtain the second inequality. $\square$

Remarks. In general we only have the inequality above. For example, let

$$f(x) = \begin{cases} 0, & 0 \le x < \frac{1}{2}, \\ 1, & \frac{1}{2} \le x \le 1. \end{cases}$$

Then $f' = 0$ a.e., but $\int_0^1 f' = 0 < 1 = f(1) - f(0)$.

Theorem 5.24. *Suppose $f_k : [a,b] \to \mathbb{R}$ is non-decreasing in x for all k, and $\sum_k f_k(x)$ converges for any $x \in [a,b]$, then $(\sum_k f_k(x))' = \sum_k f_k'(x)$ a.e. $[a,b]$.*

Proof. Since f_k is non-decreasing, f_k' exists and $f_k' \ge 0$ a.e. $[a,b]$ for all k. Denote $s_k(x) = \sum_{j=1}^k f_k(x)$ and $r_k(x) = \sum_{j=k+1}^\infty f_j(x)$. Then both s_k and r_k are nondecreasing and hence have derivatives a.e. $[a,b]$. Note that

$$\Big(\sum_{k=1}^\infty f_k\Big)' = (s_k + r_k)' = s_k' + r_k' = \sum_{j=1}^k f_k' + r_k.$$

Hence it suffices to show that $r_k' \to 0$ a.e. as $k \to \infty$. Note that $r_k' = f_{k+1}' + r_{k+1}' \ge r_{k+1}' \ge 0$ a.e. Hence $r_k' \downarrow \phi$ for some $\phi \ge 0$ a.e. $[a,b]$. Then

$$0 \le \int_a^b \phi = \int_a^b \lim_{k\to\infty} r_k' \le \liminf_{k\to\infty} \int_a^b r_k' \le \liminf_{k\to\infty} (r_k(b) - r_k(a)) = 0.$$

where we used Theorem 5.23 (Lebesgue) to obtain the last inequality and the fact that $r_k(x) = \sum_{j=k+1}^\infty f_j(x) \to 0$ as $k \to \infty$ for every x to obtain the last equality. Hence $\phi = 0$ a.e. $[a,b]$. $\square$

Example 5.25. Consider $\{r_k\} = [0,1] \cap \mathbb{Q}$. Define

$$f_k(x) = \begin{cases} 0, & 0 \le x < r_k \\ \frac{1}{2^k}, & r_k \le x \le 1 \end{cases}$$

and $s(x) = \sum_{k=1}^\infty f_k(x)$. It is then easy to verify that $s(x) < s(y)$ if $x < y$, and $s'(x) = \sum_k f_k'(x) = 0$ a.e. $[a,b]$. Namely, s is strictly increasing but $s' = 0$ a.e.

5.4 Functions of bounded variation

Definition 5.26 (Functions of bounded variation)**.** Suppose $f : [a, b] \to \mathbb{R}$ and there is a partition $\Delta : a = x_0 < x_1 < \cdots < x_n = b$ of $[a, b]$. Then the *variation of f by partition Δ* is defined

$$\mathrm{V}(f, [a, b], \Delta) = \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|.$$

The *total variation* of f on $[a, b]$ is defined by

$$\mathrm{TV}(f, [a, b]) = \sup\{\mathrm{V}(f, [a, b], \Delta) : \Delta \text{ is a partition of } [a, b]\}$$

and f is called a *function of bounded variation* if $\mathrm{TV}(f, [a, b]) < \infty$. The set of functions of bounded variation is denoted by $\mathrm{BV}([a, b])$. (We simply denote $\mathrm{TV}(f)$ if the interval $[a, b]$ is clear from the context.)

Example 5.27. If $f : [a, b] \to \mathbb{R}$ is monotone, then $f \in \mathrm{BV}([a, b])$.

Proof. WLOG, assume f is non-decreasing. Then for any $\Delta > 0$, there is

$$\mathrm{V}(f, [a, b], \Delta) = \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| = f(b) - f(a) < \infty.$$

Hence $\mathrm{TV}(f) = f(b) - f(a) < \infty$, and $f \in \mathrm{BV}([a, b])$. □

Example 5.28. If $f : [a, b] \to \mathbb{R}$, and f is differentiable, and $|f'| \leq M$ for all x. Then $f \in \mathrm{BV}([a, b])$.

Proof. For any partition Δ, there is

$$\mathrm{V}(f, [a, b], \Delta) = \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| \leq M \sum_{i=1}^{n} (x_i - x_{i-1}) = M(b - a).$$

Hence $\mathrm{TV}(f) \leq M(b - a) < \infty$. □

Example 5.29. Suppose $f : [a, b] \to \mathbb{R}$ is defined by $f(x) = x \sin(\pi/x)$ if $0 < x \leq 1$ and 0 if $x = 0$. Then $f \notin \mathrm{BV}([0, 1])$.

Proof. Consider partition $\Delta_k : 0 < \frac{2}{2k-1} < \frac{2}{2k-3} < \cdots < \frac{2}{3} < 1$. Then

$$\mathrm{V}(f, [0, 1], \Delta_k) = \frac{2}{2k-1} + \Big(\frac{2}{2k-1} + \frac{2}{2k-3}\Big) + \cdots + \frac{2}{3} = 2 \sum_{j=1}^{k} \frac{2}{2j-1} \to \infty$$

as $k \to \infty$. Hence $\mathrm{TV}(f) = \infty$. □

Theorem 5.30. *The following statements hold:*

1. *If $f \in \mathrm{BV}([a, b])$ then f is uniformly bounded.*
2. $\mathrm{BV}([a, b])$ *is a linear space.*

3. $\mathrm{TV}(f,[a,b]) = \mathrm{TV}(f,[a,c]) + \mathrm{TV}(f,[c,b])$ *for any* $c \in [a,b]$.
4. *If* $f \in \mathrm{BV}([a,b])$*, then* $|f| \in \mathrm{BV}([a,b])$.
5. *If* $f, g \in \mathrm{BV}([a,b])$*, then* $\max\{f,g\} \in \mathrm{BV}([a,b])$.

Proof. Items 1 2, and 4 are trivial to prove. For item 5, note that $\max\{f,g\} = \frac{f+g}{2} + \frac{|f-g|}{2}$ and hence it follows from item 4.

For item 3, consider any partition Δ of $[a,b]$, then $\Delta' = \Delta \cup \{c\}$ is also a partition of $[a,b]$. Moreover,

$$\begin{aligned} \mathrm{V}(f,[a,b],\Delta) &\le \mathrm{V}(f,[a,b],\Delta') \\ &= \mathrm{V}(f,[a,c],\Delta' \cap [a,c]) + \mathrm{V}(f,[c,b],\Delta' \cap [c,b]) \\ &\le \mathrm{TV}(f,[a,c]) + \mathrm{TV}(f,[c,b]). \end{aligned}$$

Hence $\mathrm{TV}(f,[a,b]) \le \mathrm{TV}(f,[a,c]) + \mathrm{TV}(f,[c,b])$.

On the other hand, for any $\epsilon > 0$, there exist partition Δ_1 of $[a,c]$ and Δ_2 of $[c,b]$, such that

$$\mathrm{TV}(f,[a,c]) - \frac{\epsilon}{2} < \mathrm{V}(f,[a,c],\Delta_1), \quad \mathrm{TV}(f,[c,b]) - \frac{\epsilon}{2} < \mathrm{V}(f,[c,b],\Delta_2)$$

Note that $\Delta = \Delta_1 \cup \Delta_2$ is a partition of $[a,b]$. Hence

$$\begin{aligned} \mathrm{TV}(f,[a,c]) + \mathrm{TV}(f,[c,b]) - \epsilon &< \mathrm{V}(f,[a,c],\Delta_1) + \mathrm{V}(f,[c,b],\Delta_2) \\ &= \mathrm{V}(f,[a,b],\Delta) \\ &\le \mathrm{TV}(f,[a,b]) \end{aligned}$$

As ϵ is arbitrary, we know $\mathrm{TV}(f,[a,c]) + \mathrm{TV}(f,[c,b]) \le \mathrm{TV}(f,[a,b])$. $\square$

For a partition $\Delta : a = x_0 < x_1 < \cdots < x_n = b$ of the interval $[a,b]$, we can obtain a set of $n+1$ points: $(x_0, f(x_0)), \dots, (x_n, f(x_n))$ in $\mathbb{R}^2$. We connect these $n+1$ points using straight line segments, and sum the lengths of these line segments to obtain the total length:

$$l_\Delta(f) = \sum_{i=1}^{n} \Big((x_i - x_{i-1})^2 + (f(x_i) - f(x_{i-1}))^2 \Big)^{1/2}.$$

Then we can take the supremum of $l_\Delta(f)$ over all partitions:

$$l(f) = \sup\{l_\Delta(f) : \Delta \text{ is a partition of } [a,b]\}.$$

The following theorem reveals the relation between $\mathrm{TV}(f)$ and $l(f)$:

Theorem 5.31. *Suppose* $f : [a,b] \to \mathbb{R}$*. Then* $\mathrm{TV}(f) < \infty$ *iff* $l(f) < \infty$.

Proof. For any partition Δ, there is $l_\Delta(f) \le \sum_{i=1}^n |x_i - x_{i-1}| + |f(x_i) - f(x_{i-1})|$ (because $(u^2+v^2)^{1/2} \le u + v$ for any $u, v \ge 0$). Hence $l_\Delta(f) \le (b-a) + \sum_{i=1}^n |f(x_i) - f(x_{i-1})|$. Therefore $\mathrm{V}(f,[a,b],\Delta) \le l_{\Delta(f)} \le (b-a) + \mathrm{V}(f,[a,b],\Delta)$. As Δ is arbitrary, we know $\mathrm{TV}(f,[a,b]) \le l(f) \le (b-a) + \mathrm{TV}(f,[a,b])$, which implies that $\mathrm{TV}(f) < \infty$ iff $l(f) < \infty$. $\square$

There is a very elegant characterization of functions of bounded variation: they can always be written as the differences of two non-decreasing functions, as shown in the following theorem.

Theorem 5.32 (Jordan)**.** *Suppose* $f : [a,b] \to \mathbb{R}$*. Then* $f \in \mathrm{BV}([a,b])$ *iff there exist two non-decreasing functions* $g, h : [a,b] \to \mathbb{R}$ *such that* $f = g - h$*.*

Proof. First we show the necessity. Suppose $f \in \mathrm{BV}([a,b])$. Denote $T_f(x) = \mathrm{TV}(f,[a,x])$ for any $x \in [a,b]$, which is therefore well defined since $\mathrm{TV}(f) < \infty$. Then define $g(x) = \frac{1}{2}(T_f(x) + f(x))$ and $h(x) = \frac{1}{2}(T_f(x) - f(x)$. We can show that both g and h are non-decreasing: for $x < y$, there is

$$\begin{aligned} g(y) - g(x) &= \frac{1}{2}(T_f(y) + f(y)) - \frac{1}{2}(T_f(x) + f(x)) \\ &= \frac{1}{2}\mathrm{TV}(f,[x,y]) + \frac{1}{2}(f(y) - f(x)) \\ &\geq \frac{1}{2}\mathrm{V}(f,[x,y],\Delta) - \frac{1}{2}|f(y) - f(x)| \geq 0 \end{aligned}$$

where $\Delta : x = x_0 < x_1 < \cdots < x_n = y$ is a partition of $[x,y]$. Similarly h is non-decreasing, and obviously $f = g - h$.

Now we show the sufficiency. If g, h are non-decreasing, then $g, h \in \mathrm{BV}([a,b])$. Hence $f = g - h \in \mathrm{BV}([a,b])$ as $\mathrm{BV}([a,b])$ is a linear space. $\square$

From Theorem 5.23, we know both g and h are differentiable a.e. since they are monotone. Hence $f = g - h$ is differentiable. Therefore f is differentiable a.e. if $f \in \mathrm{BV}([a,b])$.

Lemma 5.33. *Suppose* $f \in L([a,b])$*. Define* $F_h(x) = \frac{1}{h}\int_x^{x+h} f(t)\,\mathrm{d}t$ *(Assume* $f(x) = f(a)$ *if* $x < a$ *and* $f(x) = f(b)$ *if* $x > b$*). Then* $\lim_{h\to 0}\int_a^b |F_h(x) - f(x)| = 0$*.*

Proof. Since $f \in L([a,b])$, we know for any $\epsilon > 0$, there exists $\delta > 0$, such that $\int_a^b |f(x+h) - f(x)| < \epsilon$ for any h with $|h| < \delta$. Note that $F_h(x) - f(x) = \frac{1}{h}\int_x^{x+h}(f(t) - f(x))\,\mathrm{d}t$. Therefore, for any $t < h < \delta$, there is

$$\begin{aligned} \int_a^b |F_h(x) - f(x)|\,\mathrm{d}x &\leq \int_a^b \frac{1}{h}\int_x^{x+h} |f(t) - f(x)|\,\mathrm{d}t\,\mathrm{d}x \\ &= \int_a^b \frac{1}{h}\int_0^h |f(x+t) - f(x)|\,\mathrm{d}t\,\mathrm{d}x \\ &= \int_0^h \frac{1}{h}\int_a^b |f(x+t) - f(x)|\,\mathrm{d}x\,\mathrm{d}t \\ &\leq \int_0^h \frac{1}{h}\cdot\epsilon\,\mathrm{d}t = \epsilon \end{aligned}$$

where we applied Theorem 4.51 (Tonelli) to obtain the second equality. $\square$

Now we have the main theorem of this subsection.

Theorem 5.34. *Let $f \in L([a,b])$. Define $F(x) = \int_a^x f(t)\,\mathrm{d}t$. Then $F'(x) = f(x)$ a.e. $[a,b]$.*

Proof. Note that $F'(x) = \lim_{h\to 0} F_h(x)$ exists a.e. $[a,b]$. Hence

$$\int_a^b |f(x) - F'(x)| = \int_a^b \lim_{h\to 0} |f(x) - F_h(x)|\,\mathrm{d}x \le \liminf_{h\to 0} \int_a^b |f(x) - F_h(x)|\,\mathrm{d}x = 0,$$

which implies that $F'(x) = f(x)$ a.e. $[a,b]$. $\square$

Corollary 5.35. *Suppose $f \in L([a,b])$. Then $\lim_{h\to 0} \frac{1}{h}\int_0^h |f(x+t) - f(x)|\,\mathrm{d}t = 0$ a.e. $[a,b]$.*

Proof. For any $r \in \mathbb{Q}$, we know $|f(x) - r| \in L([a,b])$. Hence, for almost every $x \in [a,b]$, there is

$$\lim_{h\to 0} \frac{1}{h}\int_0^h |f(x+h) - r|\,\mathrm{d}t = |f(x) - r|$$

by Lemma 5.33. Denote $Z_r = \{x : \lim_{h\to 0} \frac{1}{h}\int_0^h |f(x+t) - r|\,\mathrm{d}t \neq |f(x) - r|\}$. Then $\mu(Z_r) = 0$. Let $Z = (\cup_{r\in\mathbb{Q}} Z_r) \cup \{x : f(x) = \pm\infty\}$, there is also $\mu(Z) = 0$.

For any $x \notin Z$ (i.e., $\lim_{h\to 0} \frac{1}{h}\int_0^h |f(x+h) - r|\,\mathrm{d}t = |f(x) - r|$ for all $r \in \mathbb{Q}$ and $|f(x)| < \infty$) and $\epsilon > 0$, there exists $r \in \mathbb{Q}$ and $\delta > 0$, such that $|f(x) - r| < \frac{\epsilon}{3}$ and $|\frac{1}{h}\int_0^h |f(x+t) - r|\,\mathrm{d}t - |f(x) - r|| < \frac{\epsilon}{3}$ for all h with $|h| < \delta$. Hence

$$\begin{aligned}
\frac{1}{h}\int_0^h |f(x+t) - f(x)|\,\mathrm{d}t \le& \frac{1}{h}\int_0^h |f(x+t) - r|\,\mathrm{d}t - |f(x) - r| + 2|f(x) - r| \\
\le& \Big|\frac{1}{h}\int_0^h |f(x+t) - r|\,\mathrm{d}t - |f(x) - r|\Big| + 2|f(x) - r| \\
<& \frac{\epsilon}{3} + 2\cdot\frac{\epsilon}{3} = \epsilon.
\end{aligned}$$

Therefore $\lim_{h\to 0} \frac{1}{h}\int_0^h |f(x+t) - f(x)|\,\mathrm{d}t = 0$ on Z^c. $\square$

Remarks. We call x a *Lebesgue point* if x satisfies $\lim_{h\to 0} \frac{1}{h}\int_0^h |f(x+t) - f(x)|\,\mathrm{d}t = 0$. The corollary above says that f has Lebesgue points a.e. $[a,b]$ if $f \in L([a,b])$. Note that the corollary can also be proved by invoking Lemma 4.15 (Fatou) on $G(h,x) := \frac{1}{h}\int_0^h |f(x+t) - f(x)|\,\mathrm{d}t$.

Example 5.36. Suppose $f \in L(\mathbb{R})$. For $[a,b]$, if $\lim_{h\to 0} \frac{1}{h}\int_a^b |f(x+h) - f(x)|\,\mathrm{d}x = 0$, then there exists constant $c > 0$ such that $f(x) = c$ a.e. $[a,b]$.

Proof. Consider any two Lebesgue points x_1, x_2 on $[a,b]$ where $x_1 < x_2$. Then

$$\begin{aligned}
\Big|\frac{1}{h}\int_{x_1}^{x_2} (f(x+h) - f(x))\,\mathrm{d}x\Big| &\le \frac{1}{h}\int_{x_1}^{x_2} |f(x+h) - f(x)|\,\mathrm{d}x \\
&\le \frac{1}{h}\int_a^b |f(x+h) - f(x)|\,\mathrm{d}x \to 0
\end{aligned}$$

as $h \to 0$. On the other hand,

$$\begin{aligned}\Big|\frac{1}{h}\int_{x_1}^{x_2} f(x+h) - f(x)\,\mathrm{d}t\Big| &= \Big|\frac{1}{h}\int_{x_1+h}^{x_2+h} f(x)\,\mathrm{d}x - \frac{1}{h}\int_{x_1}^{x_2} f(x)\,\mathrm{d}x\Big| \\ &= \Big|\frac{1}{h}\Big(\int_{x_2}^{x_2+h} f(x)\,\mathrm{d}x - \int_{x_1}^{x_1+h} f(x)\,\mathrm{d}x\Big)\Big| \\ &\to |f(x_2) - f(x_1)|\end{aligned}$$

as $h \to 0$. Hence $f(x_2) = f(x_1) = c$. Since $[a,b]$ has Lebesgue point a.e., we know $f(x) = c$ for all Lebesgue point x. $\square$

Example 5.37. Let $f \in L([a,b])$ and $F(x) = \int_a^x f(t)\,\mathrm{d}t$. Then $F \in \mathrm{BV}([a,b])$ and $\mathrm{TV}(F) \leq \int_a^b |f(x)|\,\mathrm{d}x$.

Proof. For any partition $\Delta : a = x_0 < x_1 < \cdots < x_n = b$, there is

$$\sum_{i=1}^n |F(x_i) - F(x_{i-1})| = \sum_{i=1}^n \Big|\int_{x_{i-1}}^{x_i} f(t)\,\mathrm{d}t\Big| \leq \sum_{i=1}^n \int_{x_{i-1}}^{x_i} |f(t)|\,\mathrm{d}t = \int_a^b |f(t)|\,\mathrm{d}t.$$

Therefore $\mathrm{V}(F,[a,b],\Delta) \leq \int_a^b |f(x)|\,\mathrm{d}x$. Hence, $\mathrm{TV}(F) \leq \int_a^b |f|\,\mathrm{d}x$. $\square$

5.5 Absolute continuity

We would like to ask the following questions: suppose $f \in [a,b] \to \mathbb{R}$, then in what case, there exists a function g such that $f(x) - f(a) = \int_a^x g(t)\,\mathrm{d}t$ for a.e. x in $[a,b]$. We have shown before that if such g exists, then f is bounded, has bounded variation, and is continuous. But is the converse true?

Example 5.38. The Cantor function ϕ is continuous and satisfies $\phi'(x) = 0$ a.e. but $\phi(0) = 0$ and $\phi(1) = 1$.

So we need stronger condition than continuity. This is called the absolute continuity.

Lemma 5.39. *Suppose $f : [a,b] \to \mathbb{R}$, and $f' = 0$ a.e. If f is not constant, then there exists $\epsilon_0 > 0$ such that for any $\delta > 0$ there exist a finite number of mutually disjoint intervals $(x_1,y_1),\ldots,(x_n,y_n)$, $\sum_{i=0}^p |f(u_i) - f(v_i)| \geq \epsilon_0$.*

Proof. Suppose $c \in (a,b)$ such that $f(c) \neq f(a)$. Then choose $\epsilon_0 \in (0, \frac{|f(c)-f(a)|}{2})$ and $r \in (0, \frac{\epsilon_0}{b-a})$. Define the set $E_c = \{x \in (a,c) : f'(x) = 0\}$ and the collection of closed intervals

$$\mathcal{F} = \{[x, x+h] \subset (a,c) : |f(x+h) - f(x)| < rh \text{ for some } h > 0\}$$

Hence $\mathcal{F}$ is a Vitali cover of E_c. Then for any $\delta > 0$ there exist mutually disjoint intervals $[x_1, x_1+h_1],\ldots,[x_p, x_p+h_p]$, such that $\mu(E_c \setminus \cup_{i=1}^p [x_i, x_i+h_i]) < \delta$. WLOG, assume $a = x_0 < x_1 < x_1+h_1 < \cdots < x_p < x_p+h_p < x_{p+1} = c$. Note that by letting $u_i = x_{i+1}$ and $v_i = x_i + h_i$ ($h_0 = 0$, $u_0 - v_0 = x_1 - x_0$), we have $u_p = x_{p+1}$ and $v_p = x_p + h_p$, and $\sum_{i=1}^{p} |u_i - v_i| < \delta$.

On the other hand, there is

$$2\epsilon_0 < |f(c) - f(a)| \le \sum_{i=0}^{p} |f(u_i) - f(v_i)| + \sum_{i=1}^{p} |f(x_i + h_i) - f(x_i)|$$
$$\le \sum_{i=0}^{p} |f(u_i) - f(v_i)| + r \sum_{i=1}^{p} h_i \le \sum_{i=0}^{p} |f(u_i) - f(v_i)| + r(b - a).$$

Note that $r(b-a) < \epsilon_0$, we know $\sum_{i=0}^{p} |f(u_i) - f(v_i)| \ge \epsilon_0$. □

Definition 5.40. $f : [a, b] \to \mathbb{R}$ is *absolutely continuous* if for any $\epsilon > 0$, there exists $\delta > 0$, such that for any mutually disjoint intervals (x_i, y_i), $i = 1, \ldots, p$, satisfying $\sum_{i=0}^{p} |y_i - x_i| < \delta$, there is $\sum_{i=1}^{p} |f(y_i) - f(x_i)| < \epsilon$. The set of absolutely continuous functions is denoted by $\mathrm{AC}([a, b])$.

Theorem 5.41. *The following statements hold:*

1. *If $f \in \mathrm{AC}([a, b])$ then f is continuous.*
2. *$\mathrm{AC}([a, b])$ is a linear space.*

Example 5.42. If f is Lipschitz continuous then $f \in \mathrm{AC}([a, b])$.

Proof. $\sum_{i=1}^{p} |f(y_i) - f(x_i)| \le M \sum_{i=1}^{p} |y_i - x_i| \le M\delta$. □

Theorem 5.43. *Suppose $f \in L([a, b])$ then $F(x) = \int_a^x f(t)\,\mathrm{d}t \in \mathrm{AC}([a, b])$.*

Proof. Since $f \in L([a, b])$, we know for any $\epsilon > 0$, there exists $\delta > 0$, such that $\int_E |f| < \epsilon$ for any $E \subset [a, b]$ satisfying $\mu(E) < \delta$. For any disjoint intervals $\{(x_i, y_i) : i = 1, \ldots, p\}$, if $\sum_{i=1}^{p} |y_i - x_i| < \delta$, then $\mu(E) < \delta$ where $E = \sum_{i=1}^{p} [x_i, y_i]$. This implies

$$\sum_{i=1}^{p} |F(y_i) - F(x_i)| \le \sum_{i=1}^{p} \int_{x_i}^{y_i} |f(x)|\,\mathrm{d}x = \int_E |f(x)|\,\mathrm{d}x < \epsilon,$$

which completes the proof. □

Theorem 5.44. *If $f \in \mathrm{AC}([a, b])$ then $f \in \mathrm{BV}([a, b])$.*

Proof. Let $\epsilon = 1$, then there exists $\delta > 0$ such that $\sum_{i=1}^{p} |f(y_i) - f(x_i)| < 1$ for any mutually disjoint intervals $\{[x_i, y_i] : 1 \le i \le p\}$ satisfying $\sum_{i=1}^{p} |y_i - x_i| < \delta$. (Clearly it is true for $p = 1$.) Consider the partition $\Delta : a = x_0 < x_1 < \cdots < x_n = b$ where $|x_i - x_{i-1}| < \delta$ for all i, we know that $\mathrm{TV}(f, [x_{i-1}, x_i]) < 1$. Hence $\mathrm{TV}(f, [a, b]) = \sum_{i=1}^{n} \mathrm{TV}(f, [x_{i-1}, x_i]) < n < \infty$. □

Corollary 5.45. *If $f \in \mathrm{AC}([a, b])$, then f is differentiable a.e. $[a, b]$ and $f' \in L([a, b])$.*

Theorem 5.46 (Fundamental theorem of calculus)**.** *If $f \in \mathrm{AC}([a, b])$, then $f(x) - f(a) = \int_a^x f'(t)\,\mathrm{d}t$ for any $x \in [a, b]$.*

Proof. If $f \in \mathrm{AC}([a,b])$, then f' exists a.e. $[a,b]$ and $f' \in L([a,b])$ by Corollary 5.45. Define $g(x) = \int_a^x f'(t)\,\mathrm{d}t$ then $g \in \mathrm{AC}([a,b])$ by Theorem 5.43. Since $f - g \in \mathrm{AC}$, $f' - g' = 0$ a.e., we know $f - g \equiv c$ for some constant c (otherwise $f - g$ is not absolutely continuous due to Lemma 5.39, a contradiction). Hence $c = f(a) - g(a) = f(a)$ as $g(a) = 0$, which implies that $f(x) = f(a) + g(x) = f(a) + \int_a^x f'(t)\,\mathrm{d}t$. □

Remarks. The results above can be summarized as follows: $f \in \mathrm{AC}([a,b])$ iff there exists $g \in L([a,b])$ such that $f(x) = f(a) + \int_a^x g(t)\,\mathrm{d}t$ for all $x \in [a,b]$. In this case, $f' = g$ a.e. $[a,b]$.

Example 5.47. Suppose $g_k \in \mathrm{AC}([a,b])$ for all k. If there exists $c \in [a,b]$ such that $\sum_k g_k(c)$ converges and $\sum_k \int_a^b |g_k'(x)|\,\mathrm{d}x < \infty$, then $\sum_k g_k(x)$ exists for all x. Let $g(x) = \sum_k g_k(x)$, then $g \in \mathrm{AC}([a,b])$ and $g'(x) = \sum_k g_k'(x)$ a.e. $[a,b]$.

Proof. Since $\sum_{k=1}^\infty \int_a^b |g_k'(x)|\,\mathrm{d}x < \infty$, we know by Corollary 4.37 that $h(x) = \sum_{k=1}^\infty g_k'(x)$ exists, $h \in L([a,b])$, and $\sum_{k=1}^\infty \int_c^x g_k'(t)\,\mathrm{d}t = \int_c^x h(t)\,\mathrm{d}t$. Since $g_k \in \mathrm{AC}([a,b])$, we know $g_k(x) = g_k(c) + \int_c^x g_k'(t)\,\mathrm{d}t$ for all x. This implies that

$$\sum_{k=1}^n g_k(x) = \sum_{k=1}^n g_k(c) + \sum_{k=1}^n \int_c^x g_k'(t)\,\mathrm{d}t \to \sum_{k=1}^\infty g_k(c) + \int_c^x h(t)\,\mathrm{d}t$$

as $n \to \infty$ for all x. Therefore $g(x) = \sum_{k=1}^\infty g_k(x) = \sum_{k=1}^\infty g_k(c) + \int_c^x h(t)\,\mathrm{d}t$ exists and $g \in \mathrm{AC}([a,b])$. Moreover $g'(x) = h(x) = \sum_{k=1}^\infty g_k'$. □

Example 5.48. Composition of absolutely continuous functions is not necessarily absolutely continuous. For example, let $f(y) = y^{1/3}$ for $y \in [-1,1]$, and

$$g(x) = \begin{cases} x^3 \cos^3(\frac{\pi}{x}), & \text{if } x \in (0,1], \\ 0, & \text{if } x = 0. \end{cases}$$

Then both f and g are absolutely continuous (they are Lipschitz continuous as $|f'|$ and $|g'|$ are bounded), but $(f \circ g)(x) = x\cos(\frac{\pi}{x})$ is not.

Example 5.49. Absolute continuity is not closed under uniform convergence. Consider the functions

$$f_k(x) = \begin{cases} 0, & \text{if } 0 \le x \le \frac{1}{k}, \\ x\sin(\frac{\pi}{x}), & \text{if } \frac{1}{k} < x \le 1, \end{cases}$$

which are absolutely continuous. Then $f_k \rightrightarrows f := x\sin(\frac{\pi}{x})$. Hence f is uniformly continuous, but not of bounded variation, hence not absolutely continuous.

6 L^p Spaces

6.1 Important inequalities

Definition 6.1. Let $E \in \mathcal{M}$. If $p \in (0, \infty]$, then the L^p norm of f on E is defined by

$$\|f\|_p = \left(\int_E |f|^p \right)^{1/p}.$$

$L^p(E) = \{f : \|f\|_p < \infty\}$ is called the L^p *space*. If $p = \infty$, the L^∞ norm (also called the *essential supremum*) of f on E is defined by

$$\|f\|_\infty = \inf\{M \in \mathbb{R} : |f| \leq M \text{ a.e. } E\}$$

and $L^\infty = \{f : \|f\|_\infty < \infty\}$ is the L^∞ *space*.

Theorem 6.2. *If $\mu(E) < \infty$, then $\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$.*

Proof. Denote $M = \|f\|_\infty$. First we have that

$$\|f\|_p = \left(\int_E |f|^p \right)^{1/p} \leq \left(\int_E |M|^p \right)^{1/p} = M(\mu(E))^{1/p} \to M$$

as $p \to \infty$. Hence $\limsup_{p\to\infty} \|f\|_p \leq M$.

On the other hand, for any $\epsilon > 0$, let $A = \{x \in E : |f(x)| > M - \epsilon\}$. Then $\mu(A) > 0$ (by definition of M). Hence

$$\|f\|_p = \left(\int_E |f|^p \right)^{1/p} \geq \left(\int_A |f|^p \right)^{1/p} \geq (M - \epsilon)(\mu(A))^{1/p} \to M - \epsilon$$

as $p \to \infty$. Hence $\liminf_{p\to\infty} \|f\|_p \geq M - \epsilon$. As ϵ is arbitrary, we have $\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$. $\square$

Theorem 6.3 (L^p space is linear)**.** *Let $p \in (0, \infty]$, and $f, g \in L^p(E)$, then $\alpha f + \beta g \in L^p(E)$ for any $\alpha, \beta \in \mathbb{R}$.*

Proof. Note that for any $u, v \geq 0$, there is

$$(u + v)^p \leq (2 \max(u, v))^p = 2^p \max(u^p, v^p) \leq 2^p(u^p + v^p).$$

If $p \in (0, \infty)$, then there is

$$|\alpha f + \beta g|^p \leq 2^p(|\alpha|^p|f|^p + |\beta|^p|g|^p),$$

integrating on both sides shows $\alpha f + \beta g \in L^p$. If $p = \infty$, then $|\alpha f + \beta g| \leq |\alpha|\|f\|_\infty + |\beta|\|g\|_\infty$ a.e. $\square$

We only consider the case $p \in [1, \infty]$ hereafter unless otherwise noted.

Definition 6.4 (Conjugate)**.** The two numbers $p, q > 1$ are called *conjugate* if $\frac{1}{p} + \frac{1}{q} = 1$.

Note that $q = \frac{p}{p-1}$. If $p = 2$, then $q = 2$. If $p = 1$, then $q = \infty$.

Theorem 6.5 (Young's inequality)**.** *Let p, q be conjugate. For any $u, v \geq 0$, there is $uv \leq \frac{u^p}{p} + \frac{v^q}{q}$.*

Proof. If either u or v is zero, then trivial. Now suppose both are nonzero. Note that e^x is convex, and $\frac{1}{p} + \frac{1}{q} = 1$, therefore

$$uv = e^{\frac{1}{p}\log(u^p)+\frac{1}{q}\log(v^q)} \leq \frac{1}{p}e^{\log(u^p)} + \frac{1}{q}e^{\log(v^q)} = \frac{u^p}{p} + \frac{v^q}{q},$$

which completes the proof. $\square$

Theorem 6.6 (Hölder's inequality)**.** *For any $p \in [1, \infty]$ and q be its conjugate. If $f \in L^p(E)$ and $g \in L^q(E)$, then $\|fg\|_1 \leq \|f\|_p \|g\|_q$.*

Proof. It is trivial if p or q is ∞. Now suppose $p, q \in (1, \infty)$ and $\|f\|_p, \|g\|_q \neq 0$. Then

$$\int_E \frac{|f|}{\|f\|_p} \cdot \frac{|g|}{\|g\|_q} \leq \int_E \left(\frac{1}{p}\frac{|f|^p}{\|f\|_p^p} + \frac{1}{q}\frac{|g|^q}{\|g\|_q^q} \right) = \frac{1}{p} + \frac{1}{q} = 1$$

where we used Hölder's inequality above. Multiplying the constant $\|f\|_p\|g\|_q$ on both sides yields the inequality. $\square$

Corollary 6.7 (Schwarz inequality)**.** *If $f, g \in L^2(E)$, then $\|fg\|_1 \leq \|f\|_2\|g\|_2$.*

Theorem 6.8. *If $\mu(E) < \infty$ and $0 < p_1 < p_2 \leq \infty$, then $L^{p_2}(E) \subset L^{p_1}(E)$ and*

$$\|f\|_{p_1} \leq (\mu(E))^{\frac{1}{p_1} - \frac{1}{p_2}} \|f\|_{p_2}$$

Proof. The proof is trivial if $p_2 = \infty$. Now suppose $0 < p_1 < p_2 < \infty$. Then

$$\|f\|_{p_1} = \left(\int |f|^{p_1} \right)^{1/p_1} \leq \left(\left(\int |f|^{p_1 r} \right)^{1/r} \left(\int 1^s \right)^{1/s} \right)^{1/p_1},$$

where $r, s > 1$ are conjugate. By choosing $r = \frac{p_2}{p_1} > 1$ and its conjugate $s = \frac{r}{r-1} = \frac{p_2}{p_2 - p_1}$, we obtain the claimed inequality. $\square$

Example 6.9. Suppose $f \in L^r \cap L^s$ where $0 < r < p < s < \infty$. Let $\lambda \in (0, 1)$ such that $\frac{1}{p} = \frac{\lambda}{r} + \frac{1-\lambda}{s}$. Then $\|f\|_p \leq \|f\|_r^\lambda \|f\|_s^{1-\lambda}$.

Proof. Note that $\frac{r}{\lambda p}$ and $\frac{s}{(1-\lambda)p}$ are conjugate. Hence

$$\begin{aligned}
\|f\|_p^p = \int |f|^p &= \int |f|^{\lambda p} |f|^{(1-\lambda)p} \\
&\leq \left(\int |f|^{\lambda p \cdot \frac{r}{\lambda p}} \right)^{\frac{\lambda p}{r}} \left(\int |f|^{(1-\lambda)p \cdot \frac{s}{(1-\lambda)p}} \right)^{\frac{(1-\lambda)p}{s}} \\
&= \|f\|_r^{\lambda p} \|f\|_s^{(1-\lambda)p}
\end{aligned}$$

Taking p-th root on both sides completes the proof. $\square$

Example 6.10. Let $0 < r < p < s < \infty$ and $f \in L^p(E)$. Then for any $t > 0$, there exist g, h such that $f = g + h$, and $\|g\|_r^r \le t^{r-p}\|f\|_p^p$ and $\|h\|_s^s \le t^{s-p}\|f\|_p^p$.

Proof. For any x, define $g(x) = f(x)$ if $f(x) > t$ and $g(x) = 0$ otherwise. Let $h = f - g$. Then, by $r - p < 0$, there is

$$\|g\|_r^r = \int_E |g|^r = \int_{\{f>t\}} |g|^{r-p}|g|^p \le t^{r-p}\int_E |f|^p = t^{r-p}\|f\|_p^p.$$

Similarly we can show the inequality for h. □

Example 6.11. Suppose $f_k \in L^2([0,1])$ for all k, $f_k \xrightarrow{\mu} 0$ in $[0,1]$, $\|f_k\|_2 \le 1$. Show that $\lim_k \int_0^1 |f_k| = 0$.

Proof. For any $\epsilon > 0$, let $E_k(\epsilon) = \{x \in [0,1] : |f_k(x)| \ge \epsilon\}$. Then $f_k \xrightarrow{\mu} 0$ implies that $\lim_k E_k(\epsilon) = 0$. Hence

$$\begin{aligned}
0 \le \int_0^1 |f_k| &= \int_{[0,1]\setminus E_k(\epsilon)} |f_k| + \int_{E_k(\epsilon)} |f_k| \\
&\le \int_{[0,1]\setminus E_k(\epsilon)} \epsilon + \int \chi_{E_k(\epsilon)}|f_k| \\
&\le \epsilon + (\mu(E_k(\epsilon)))^{1/2}\|f_k\|_2 \to \epsilon
\end{aligned}$$

as $k \to \infty$, where we used Hölder's inequality to obtain the last inequality above and $\|f_k\|_2 \le 1$ to obtain the limit. Hence $0 \le \limsup_k \int_0^1 |f_k| \le \epsilon$. As ϵ is arbitrary, we know $\lim_k \int_0^1 |f_k| = 0$. □

Theorem 6.12 (Minkowski's inequality)**.** *Let $p \in [1,\infty]$. If $f, g \in L^p$, then $\|f+g\|_p \le \|f\|_p + \|g\|_p$.*

Proof. The proof is trivial if $p = 1$ or $p = \infty$. Suppose $p \in (1,\infty)$ and $q = \frac{p}{p-1}$ is its conjugate. WLOG we assume $\|f+g\|_p > 0$. Then

$$\int |f+g|^p = \int |f+g|^{p-1}|f+g| \le \int |f+g|^{p-1}|f| + \int |f+g|^{p-1}|g|.$$

Now for the first term on the RHS, we have

$$\int |f+g|^{p-1}|f| \le \left(\int |f+g|^{(p-1)q}\right)^{1/q}\left(\int |f|^p\right)^{1/p} = \|f+g\|_p^{p-1}\|f\|_p,$$

where we used Hölder's inequality. Similarly, there is $\int |f+g|^{p-1}|g| \le \|f+g\|_p^{p-1}\|g\|_p$. Therefore

$$\|f+g\|_p^p = \int |f+g|^p \le \|f+g\|_p^{p-1}(\|f\|_p + \|g\|_p).$$

Dividing both sides by $\|f+g\|_p^{p-1}$ yields the Minkowski's inequality. □

6.2 L^p space

We identify two functions $f, g \in L^p(E)$ if $f = g$ a.e. E. Suppose we define $d : L^p(E) \times L^p(E) \to \mathbb{R}$ by $d(f, g) = \|f - g\|_p$ for any $f, g \in L^p(E)$. Then it is easy to verify that d is a metric: (i) $d(f, g) \geq 0$, and $d(f, g) = 0$ iff $f = g$ a.e. E; (ii) $d(f, g) = d(g, f)$; (iii) $d(f, g) \leq d(f, h) + d(h, g)$ for all $f, g, h \in L^p(E)$ by using Theorem 6.12 (Minkowski).

Definition 6.13. Let $p \in [1, \infty]$ and $d(f, g) = \|f - g\|_p$ for any $f, g \in L^p(E)$. Then $(L^p(E), d)$ is a metric space.

Theorem 6.14. *If $\|f_k - f\|_p \to 0$, then $\|f_k\|_p \to \|f\|_p$.*

Proof. Note that $|\|f_k\|_p - \|f\|_p| \leq \|f_k - f\|_p$ by Minkowski's inequality. □

Theorem 6.15 (L^p space is complete)**.** *If $\{f_k\}$ is Cauchy in L^p, then there exists $f \in L^p$ such that $\|f_k - f\|_p \to 0$.*

Proof. First consider $p \in [1, \infty)$. Since $\|f_k - f_j\|_p \to 0$ as $k, j \to \infty$, we know for any $\epsilon > 0$, denote $E_{k,j}(\epsilon) = \{x \in E : |f_k(x) - f_j(x)| \geq \epsilon\}$, there is

$$\epsilon(\mu(E_{k,j}(\epsilon))^{1/p} \leq \Big(\int_{E_{k,j}(\epsilon)} |f_k - f_j|^p\Big)^{1/p} \leq \Big(\int_E |f_k - f_j|^p\Big)^{1/p} \to 0$$

as $k, j \to \infty$. Hence $\mu(E_{k,j}(\epsilon)) \to 0$ as $k, j \to \infty$. Therefore $\{f_k\}$ is Cauchy in measure, which implies that there exists a subsequence $\{f_{k_j}\}$ and f such that $f_{k_j} \to f$ a.e. E as $j \to \infty$. Therefore

$$\int_E |f_k - f|^p = \int_E \lim_{j\to\infty} |f_k - f_{k_j}|^p \leq \liminf_{j\to\infty} \int_E |f_k - f_{k_j}|^p$$

Taking limit $k \to \infty$ on both sides yields $\|f_k - f\|_p \to 0$. Moreover, $\|f\|_p \leq \|f_k - f\|_p + \|f_k\|_p < \infty$, and hence $f \in L^p(E)$.

Next consider $p = \infty$. Since $\|f_k - f_j\|_\infty \to 0$, there exists $Z \subset E$, such that $\mu(Z) = 0$ and $f_k(x) - f_j(x) \to 0$ as $k, j \to \infty$ on $E \setminus Z$. Let $f(x) = \lim_k f_k(x)$ for $x \in E \setminus Z$ and arbitrary on Z. For any $\epsilon > 0$, there exists K sufficiently large, such that

$$|f_k(x) - f(x)| = \lim_{j\to\infty} |f_k(x) - f_j(x)| \leq \lim_{j\to\infty} \|f_k - f_j\|_\infty \leq \epsilon$$

for all $k \geq K$ and $x \in E \setminus Z$. Hence $\|f_k - f\|_\infty \leq \epsilon$. In addition, $\|f\|_\infty \leq \|f_k - f\|_\infty + \|f_k\|_\infty < \infty$, hence $f \in L^\infty(E)$. □

Definition 6.16. A metric space (X, d) is called *seperable* if X contains a countable dense subset. Namely, X has a countable subset Y, such that for any $x \in X$ and $\epsilon > 0$, there exists $y \in Y$ that satisfies $d(x, y) < \epsilon$.

Lemma 6.17. *Let $p \in [1, \infty)$ and $f \in L^p(E)$, then for any $\epsilon > 0$, the following statements hold:*

1. *There exists $g : E \to \mathbb{R}$ which is continuous and has compact support, such that $\int_E |f-g|^p < \epsilon$.*
2. *There exists a simple function $\phi : E \to \mathbb{R}$ which is of form $\phi(x) = \sum_{i=1}^k c_i \chi_{A_i}$ where every A_i is a finite union of open boxes on regular grids and has compact support, such that $\int_E |f-\phi|^p < \epsilon$.*

Proof. Proof of Item 1 is similar to that of Theorem 4.40. For Item 2, note that the tolerance ϵ allows approximating f by such type of simple function ϕ. □

Theorem 6.18. *Suppose $p \in [1, \infty)$. Then L^p space is separable.*

Proof. (i) Suppose $E = \mathbb{R}^n$. Then for any $f \in L^p(E)$ and $\epsilon > 0$, there exists a simple function $\phi = \sum_{i=1}^k c_i \chi_{A_i}$ such that $\|f - \phi\|_p < \epsilon/2$. Hence there exists $M > 0$, such that $|c_i| \leq M$ and $\mu(A_i) < M^p$ for all $i \leq k$. Note that there exists $r_i \in \mathbb{Q}$ such that $|c_i - r_i| < \epsilon/(2kM)$ for every $i \leq k$. Let $\psi = \sum_{i=1}^k r_i \chi_{A_i}$, then

$$\begin{aligned}
\|\phi - \psi\|_p &= \Big\| \sum_{i=1}^k c_i \chi_{A_i} - \sum_{i=1}^k r_i \chi_{A_i} \Big\|_p \leq \sum_{i=1}^k |c_i - r_i| \|\chi_{A_i}\|_p \\
&= \sum_{i=1}^k |c_i - r_i| \mu(A_i)^{1/p} \leq k \cdot \frac{\epsilon}{2kM} \cdot M = \frac{\epsilon}{2}.
\end{aligned}$$

Hence $\|f - \psi\|_p \leq \|f - \phi\|_p + \|\phi - \psi\|_p < \epsilon$. Note that the set $\Gamma = \{\psi = \sum_{i=1}^k r_i \chi_{A_i} : r \in \mathbb{Q}\}$ is a countable, hence Γ is a countable dense set of $L^p(\mathbb{R}^n)$.

(ii) For general $E \subset \mathbb{R}^n$, consider $g(x) = f(x)$ if $x \in E$ and $g(x) = 0$ otherwise. Then $g : \mathbb{R}^n \to \mathbb{R}$. By (i), there exists a simple function $\psi \in \Gamma$ such that $\int_{\mathbb{R}^n} |g - \psi|^p < \epsilon$. Hence

$$\int_E |f - \psi|^p = \int_E |g - \psi|^p \leq \int_{\mathbb{R}^n} |g - \psi|^p < \epsilon$$

which also implies that Γ is dense in $L^p(E)$. □

Example 6.19. Let $p \in [1, \infty)$ and $f \in L^p(\mathbb{R}^n)$. Show that $\lim_{|t|\to\infty} \int_{\mathbb{R}^n} |f(x) + f(x-t)|^p \,\mathrm{d}x = 2 \int_{\mathbb{R}^n} |f(x)|^p \,\mathrm{d}x$.

Proof. For any $\epsilon > 0$, consider the decomposition $f = g + h$ where g is a continuous function with compact support and $h = f - g$ such that $\|h\|_p < \epsilon/4$. For notation simplicity, we denote $f_t(x) = f(x-t)$, $g_t(x) = g(x-t)$, and $h_t(x) = h(x-t)$ for any fixed t. Since g has compact support, we know that the supports of $g(x)$ and $g_t(x)$ do not overlap if $|t|$ is sufficiently large, which implies that

$$\|g + g_t\|_p^p = \int_{\mathbb{R}^n} |g + g_t|^p = \int_{\mathbb{R}^n} (|g|^p + |g_t|^p) = 2 \int_{\mathbb{R}^n} |g|^p = 2\|g\|_p^p$$

Hence we have

$$\begin{aligned}\Big|\|f+f_t\|_p - 2^{1/p}\|f\|_p\Big| &\le \Big|\|f+f_t\|_p - 2^{1/p}\|g\|_p\Big| + 2^{1/p}\Big|\|g\|_p - \|f\|_p\Big| \\ &= \Big|\|f+f_t\|_p - \|g+g_t\|_p\Big| + 2^{1/p}\Big|\|g\|_p - \|f\|_p\Big| \\ &\le \|h+h_t\|_p + 2^{1/p}\|h\|_p < \frac{\epsilon}{4} + \frac{\epsilon}{4} + 2^{1/p}\cdot\frac{\epsilon}{4} < \epsilon,\end{aligned}$$

which completes the proof. □

6.3 L^2 space and inner product

Definition 6.20 (Inner product)**.** Let $f, g \in L^2(E)$, then the *inner product of f and g* is defined by

$$\langle f, g\rangle = \int_E fg.$$

Note that $|\langle f, g\rangle| \le \int_E |fg| \le \|f\|_2\|g\|_2 < \infty$.

It is easy to verify that the following identities hold:

- $\langle f, g\rangle = \langle g, f\rangle$.
- $\langle f_1 + f_2, g\rangle = \langle f_1, g\rangle + \langle f_2, g\rangle$.
- $\langle \alpha f, g\rangle = \alpha\langle f, g\rangle = \langle f, \alpha g\rangle$ for all $\alpha \in \mathbb{R}$.

Example 6.21. Suppose $f, g \in L^2$ then $2\|fg\|_1 \le t\|f\|_2^2 + \frac{1}{t}\|g\|_2^2$ for all $t > 0$.

Proof. Note that $|fg| = \sqrt{t}|f| \cdot \frac{1}{\sqrt{t}}|g| \le \frac{t|f|^2}{2} + \frac{|g|^2}{2t}$ by Young's inequality. □

Example 6.22. Suppose $f : [0, \infty) \to \mathbb{R}_+$ is integrable, then

$$\left(\int_0^\infty f\,\mathrm{d}x\right)^4 = \pi^2\left(\int_0^\infty f^2\,\mathrm{d}x\right)\left(\int_0^\infty x^2 f^2\,\mathrm{d}x\right).$$

Proof. Recall that for any $\alpha, \beta > 0$, there is $\int_0^\infty \frac{1}{\alpha+\beta x^2}\,\mathrm{d}x = \frac{1}{\sqrt{\alpha\beta}}\int_0^\infty \frac{1}{1+y^2}\,\mathrm{d}y = \frac{1}{\sqrt{\alpha\beta}}\frac{\pi}{2}$ using the change of variable $y = \sqrt{\frac{\beta}{\alpha}}x$. Therefore

$$\begin{aligned}\left(\int_0^\infty f\,\mathrm{d}x\right)^2 &= \left(\int_0^\infty \frac{1}{\sqrt{\alpha+\beta x^2}}\cdot\sqrt{\alpha+\beta x^2}f\,\mathrm{d}x\right)^2 \\ &\le \int_0^\infty \frac{1}{\alpha+\beta x^2}\,\mathrm{d}x \cdot \int_0^\infty (\alpha+\beta x^2)f^2\,\mathrm{d}x \\ &= \frac{\pi}{2}\frac{1}{\sqrt{\alpha\beta}}\left(\alpha\int_0^\infty f^2\,\mathrm{d}x + \beta\int_0^\infty x^2 f^2\,\mathrm{d}x\right)\end{aligned}$$

Letting $\alpha = \int_0^\infty x^2 f^2\,\mathrm{d}x$ and $\beta = \int_0^\infty f^2\,\mathrm{d}x$ yields the claimed inequality. □

Theorem 6.23. *If $\|f_k - f\|_2 \to 0$, then $\langle f_k, g\rangle \to \langle f, g\rangle$ for all $g \in L^2$.*

Proof. Note $|\langle f_k, g\rangle - \langle f, g\rangle| = |\langle f_k - f, g\rangle| \le \|f_k - f\|_2\|g\|_2 \to 0$. □

Definition 6.24. $f, g \in L^2$ is called *orthogonal* if $\langle f, g\rangle = 0$. $\{\phi_\alpha : \alpha \in \mathcal{A}\}$ is called an *orthogonal set* if $\langle \phi_\alpha, \phi_\beta\rangle = 0$ for all distinct $\alpha, \beta \in \mathcal{A}$. If in addition $\|\phi_\alpha\|_2 = 1$, then $\{\phi_\alpha\}$ is called an *orthonormal set.*

Example 6.25. $\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos(kx), \frac{1}{\sqrt{\pi}}\sin(kx) : k \in \mathbb{N}\}$ is an orthonormal set of $L^2([-\pi, \pi])$.

Theorem 6.26. *An orthonormal set of $L^2(E)$ is at most countable.*

Proof. Suppose $\{\phi_\alpha \in L^2(E) : \alpha \in \mathcal{A}\}$ is an orthonormal set. Then for any distinct $\alpha, \beta \in \mathcal{A}$, there is

$$\|\phi_\alpha - \phi_\beta\|_2^2 = \|\phi_\alpha\|_2^2 + \|\phi_\beta\|_2^2 = 2.$$

Since $L^2(E)$ is separable, there exists a countable dense set $\Gamma \subset L^2(E)$ such that for any $\alpha \in \mathcal{A}$, there exists $x_\alpha \in \Gamma$ satisfying $\|x_\alpha - \phi_\alpha\|_2 < \sqrt{2}/2$. Hence $|\mathcal{A}| \le |\Gamma| = \aleph_0$. □

Example 6.27 (Parallelogram law)**.** Suppose $f, g \in L^2$, then $\|f + g\|_2^2 + \|f - g\|_2^2 = 2(\|f\|_2^2 + \|g\|_2^2)$.

Definition 6.28 (Generalized Fourier series)**.** Suppose $\{\phi_k\}$ is an orthonormal set of L^2. For any $f \in L^2$, let $c_k = \langle f, \phi_k\rangle$ for any $k \in \mathbb{N}$. Then $\{c_k\}$ are called the *generalized Fourier coefficients* of f under $\{\phi_k\}$ and $\sum_{k=1}^\infty c_k\phi_k$ is *generalized Fourier series* of f.

Theorem 6.29. *For any fixed k, let $\mathcal{F}_k = \{\sum_{i=1}^k a_i\phi_i : a_i \in \mathbb{R}\}$. Then $f_k = \sum_{i=1}^k c_i\phi_i$, where $c_i = \langle f, \phi_i\rangle$ for every i, uniquely minimizes $\|f - g\|_2$ among all $g \in \mathcal{F}_k$.*

Proof. For any $f_k = \sum_{i=1}^k a_i\phi_i \in \mathcal{F}_k$, there is

$$\begin{aligned}
\|f - f_k\|_2^2 &= \|f\|_2^2 - 2\langle f, f_k\rangle + \|f_k\|_2^2 \\
&= \|f\|_2^2 - 2\sum_{i=1}^k a_i\langle f, \phi_i\rangle + \sum_{i=1}^k a_i^2 \\
&= \|f\|_2^2 - 2\sum_{i=1}^k a_ic_i + \sum_{i=1}^k a_i^2 \\
&= \|f\|_2^2 + 2\sum_{i=1}^k |a_i - c_i|^2 - \sum_{i=1}^k c_i^2
\end{aligned}$$

which is minimized only if $a_i = c_i$ for all i. □

Theorem 6.30 (Bessel inequality)**.** *Let $\{\phi_k\}$ be an orthonormal set in L^2. Suppose $f \in L^2$ and $\{c_k\}$ is the generalized Fourier coefficients. Then $\sum_{k=1}^\infty c_k^2 \le \|f\|_2^2$.*

Proof. For any $f_k = \sum_{i=1}^k c_i\phi_i$, $0 \le \|f - f_k\|^2 = \|f\|_2^2 - \sum_{i=1}^k c_i^2$. Hence $\sum_{i=1}^k c_i^2 \le \|f\|_2^2$ for all k, which implies that $\sum_{k=1}^\infty c_k^2 \le \|f\|_2^2$. □

Lemma 6.31. *Suppose $\{\phi_k\}$ is an orthonormal set of L^2 and $f \in L^2$. If $f_k = \sum_{i=1}^k c_i\phi_i$ where $c_i = \langle f, \phi_i\rangle$, then $\langle f - f_k, f_k\rangle = 0$.*

Proof. Note that $\langle f, f_k\rangle = \sum_{i=1}^k c_i\langle f, \phi_i\rangle = \sum_{i=1}^k c_i^2 = \langle f_k, f_k\rangle$. □

Theorem 6.32 (Riesz-Fischer)**.** *Suppose $\{\phi_k\}$ is an orthonormal set of L^2. If $\sum_{k=1}^\infty c_k^2 < \infty$. Then there exists $g \in L^2$, such that $\langle g, \phi_k\rangle = c_k$ for all k.*

Proof. Define $s_k = \sum_{i=1}^k c_i\phi_i$, then $s_k \in L^2$. Note that, for any $l \in \mathbb{N}$,

$$\|s_{k+l} - s_k\|_2^2 = \|\sum_{i=k+1}^{k+l} c_i\phi_i\|_2^2 = \sum_{i=k+1}^{k+l} c_i^2 \to 0$$

as $k \to \infty$. Hence $\{s_k\}$ is Cauchy in L^2, and there exists $g \in L^2$ such that $\|s_k - g\|_2 \to 0$ as $k \to \infty$. Let $a_i = \langle g, \phi_i\rangle$ for all $i \in \mathbb{N}$, $g_k = \sum_{i=1}^k a_i\phi_i$, and $h_k = g - g_k$, then

$$\sum_{i=1}^k |c_i - a_i|^2 = \|s_k - g_k\|_2^2 \le \|s_k - g_k\|_2^2 + \|h_k\|_2^2 = \|s_k - g_k - h_k\|_2^2 = \|s_k - g\|_2^2 \to 0$$

where we used the fact that $\langle g_k, h_k\rangle = 0$ from Lemma 6.31 to show that $\langle s_k - g_k, h_k\rangle = 0$ and obtained the second equality. Hence $a_i = c_i$ for all i. □

Definition 6.33 (Complete orthonormal basis)**.** We call $\{\phi_k\}$ a *complete orthonormal basis* if $\{\phi_k\}$ is an orthonormal set, and $\langle f, \phi_k\rangle = 0$ for all k implies $f = 0$ a.e.

Theorem 6.34. *Suppose $\{\phi_k\}$ is a complete orthonormal basis in L^2. Let $f \in L^2$, $c_k = \langle f, \phi_k\rangle$ for all k, then $\lim_k \|\sum_{i=1}^k c_i\phi_i - f\|_2 = 0$.*

Proof. By Theorem 6.30 (Bessel's inequality), we know $\sum_{k=1}^\infty c_k^2 \le \|f\|_2^2 < \infty$. By Theorem 6.32 (Riesz-Fischer), there exists $g \in L^2$ such that $g = \sum_{k=1}^\infty c_k\phi_k$, and $\|\sum_{i=1}^k c_i\phi_i - g\|_2 \to 0$. Note that $\langle f - g, \phi_k\rangle = \langle f, \phi_k\rangle - \langle g, \phi_k\rangle = 0$ for all k, we know $f = g$ a.e. Hence $\|\sum_{i=1}^k c_i\phi_i - f\|_2 \to 0$. □

Definition 6.35 (Linear independency)**.** $\{\phi_i : 1 \le i \le k\}$ is called *linearly independent* if $\sum_{i=1}^k c_i\phi_i = 0$ implies $c_i = 0$ for all $i \le k$. $\{\phi_k : k \in \mathbb{N}\}$ is called *linearly independent* if any finite subset is linearly independent. It is obvious that 0 cannot be in a linearly independent set.

Example 6.36. If $\{\phi_k\}$ is an orthonormal set in L^2, then it is linearly independent.

Proof. Suppose $\sum_{i=1}^k c_i\phi_i = 0$, then multiplying both sides by ϕ_i yields $c_i\|\phi_i\|_2^2 = 0$, which implies that $c_i = 0$, for every i. □

Example 6.37 (Gram-Schmidt)**.** If $\{\psi_k : k \in \mathbb{N}\}$ is a linearly independent set, then we can construct an orthonormal set $\{\phi_k : k \in \mathbb{N}\}$: define $\phi_1 = \psi_1/\|\psi_1\|_2$; suppose we already have constructed $\phi_1, \dots, \phi_{k-1}$, then define $\phi_k = (\psi_k - \sum_{i=1}^{k-1}\langle\psi_k, \phi_i\rangle\phi_i)/\|\psi_k - \sum_{i=1}^{k-1}\langle\psi_k, \phi_i\rangle\phi_i\|_2$. It is easy to verify that $\{\phi_k\}$ is an orthonormal set.

Theorem 6.38. *Suppose $\{\phi_i : i \in \mathbb{N}\}$ is an orthonormal set in L^2. If for any $f \in L^2$ and $\epsilon > 0$, there exists a finite subset $\{\phi_{i_j} : 1 \le j \le k\}$ such that $\|f - \sum_{j=1}^k c_j\phi_{i_j}\|_2 < \epsilon$, then $\{\phi_i\}$ is complete.*

Proof. Assume $\{\phi_i\}$ is not complete. Then there exists nonzero $f \in L^2$ such that $\langle f, \phi_i\rangle = 0$ for all i. On the one hand, there exist a finite subset $\{\phi_{i_j} : 1 \le j \le k\}$ such that $\|f - \sum_{j=1}^k c_j\phi_{i_j}\|_2 < \|f\|_2/2$. Moreover,

$$|\langle f, f - \sum_{j=1}^k c_j\phi_{i_j}\rangle| \le \|f\|_2 \cdot \|f - \sum_{j=1}^k c_j\phi_{i_j}\|_2 \le \frac{\|f\|_2^2}{2}.$$

On the other hand, there is

$$\left|\langle f, f - \sum_{j=1}^k c_j\phi_{i_j}\rangle\right| = \left|\|f\|_2^2 - \langle f, \sum_{j=1}^k c_j\phi_{i_j}\rangle\right| = \|f\|_2^2,$$

which is a contradiction. □

6.4 Dual space of L^p

Theorem 6.39. *Let $p \in [1,\infty)$ and $f \in L^p(E)$. Then there exists $g \in L^q(E)$, $\|g\|_q = 1$, and $\|f\|_p = \int_E fg$.*

Proof. (i) First consider $p = 1$. Then letting $g = \mathrm{sign}(f)$ proves the claim.

(ii) Next consider $p \in (1,\infty)$. Let $q = \frac{p}{p-1}$ be the conjugate, and define $g = \mathrm{sign}(f) \cdot \frac{|f|^{p-1}}{\|f\|_p^{p-1}}$. Then $(p-1)q = p$ and

$$\int_E |g|^q = \frac{1}{\|f\|_p^p}\int_E |f|^p = 1,$$
$$\int_E fg = \int_E \frac{|f|^p}{\|f\|_p^{p-1}} = \frac{\|f\|_p^p}{\|f\|_p^{p-1}} = \|f\|_p,$$

which prove the claim. □

Remarks. Note that Hölder's inequality implies that $\|fg\|_1 \le \|f\|_p\|g\|_q$, and hence $\|f\|_p \ge \sup_{g\in L^q}\frac{\|fg\|_1}{\|g\|_q} = \sup_{\|g\|_q=1}\|fg\|_1$. The theorem above implies that the supremum can be replaced by maximum, and shows the maximizer for $p \in [1,\infty)$.

Theorem 6.40. *Suppose $f \in L^\infty(E)$, then $\|f\|_\infty = \sup_{\|g\|_1=1}|\int_E fg|$.*

Proof. Note that, if $\|g\|_1 = 1$, then $|\int_E fg| \le \int_E |fg| \le \|f\|_\infty \|g\|_1 = \|f\|_\infty$. Hence $\|f\|_\infty \ge \sup_{\|g\|_1=1} |\int_E fg|$.

On the other hand, let $M = \|f\|_\infty$. Then for any $\epsilon > 0$, there exists $A \subset E$, such that $\mu(A) = a > 0$ and $|f| > M - \epsilon$ on A. Define $g = \operatorname{sign}(f)\chi_A/a$, then

$$\Big|\int_E fg\Big| = \frac{1}{a}\int_A |f| \ge (M-\epsilon)\cdot a \cdot \frac{1}{a} = M - \epsilon,$$

where $\|g\|_1 = 1$. Hence $M - \epsilon \le \sup_{\|g\|_1=1} |\int_E fg| \le M$. As ϵ is arbitrary, we know $\|f\|_\infty = \sup_{\|g\|_1=1} |\int_E fg|$. $\square$

Example 6.41. We cannot replace the supremum by maximum in the theorem above: consider $E = [0,1]$ and $f(x) = x$, then $\|f\|_\infty = 1$ but for any $g \in L^1$ and $\|g\|_1 = 1$, there is $|\int_0^1 fg\,\mathrm{d}x| \le \int_0^1 x|g(x)|\,\mathrm{d}x < 1$ (otherwise $\int_0^1 (1-x)|g(x)|\,\mathrm{d}x = 0$ implies $g = 0$ a.e., a contradiction).

Definition 6.42 (Dual space of L^p)**.** We call L^q the *dual space* of L^p if q is the conjugate of p.

Theorem 6.43. *Suppose $g : E \to \mathbb{R}$ is a measurable function. Let $p \in [1,\infty]$ and q be its conjugate. If there exists $M > 0$ such that for any simple function ϕ there is $|\int_E g\phi| \le M\|\phi\|_p$, then $g \in L^q$ and $\|g\|_q \le M$.*

Proof. (i) First consider $p \in (1,\infty)$. Let $\{\psi_k\}$ be a sequence of simple functions such that $\psi_k \uparrow |g|^{\frac{1}{p-1}}$ and $\phi_k = \operatorname{sign}(g)\psi_k$. Then

$$\int_E g\phi_k \to \int_E |g|^{\frac{p}{p-1}} = \int_E |g|^q \quad \text{and} \quad \int_E g\phi_k \le M\|\phi_k\|_p = M\|g\|_q^{q/p}$$

from which we can obtain $\|g\|_q \le M$.

(ii) Now consider $p = \infty$. Let $\phi = \operatorname{sign}(g)$ then $\int_E g\phi = \int_E |g| \le M\|\phi\|_\infty = M$, i.e., $\|g\|_1 \le M$.

(iii) Next consider $p = 1$. WLOG assume $g \ge 0$, a.e. If $g \notin L^\infty(E)$, then for any $k \in \mathbb{N}$ let $A_k = \{x \in E : g(x) \ge k\}$. Then A_k is non-increasing, and $\mu(A_k) > 0$ for all k. Let $\phi_k = \chi_{A_k}$, then

$$k\mu(A_k) \le \int_{A_k} g = \int_E g\phi_k \le M\|\phi_k\|_1 = M\mu(A_k)$$

which implies that $M \ge k$ for all $k \in \mathbb{N}$, contradiction. Therefore $g \in L^\infty(E)$.

Now we need to show $\|g\|_\infty \le M$. If not, then $\|g\|_\infty = M' > M$. Let $\epsilon = (M' - M)/2$, then there exists $A \subset E$ such that $\mu(A) = a > 0$ and $|g(x)| \ge M + \epsilon$ for all $x \in A$ since $\|g\|_\infty = M' > M + \epsilon$. Let $\phi(x) = \operatorname{sign}(g)\chi_A/a$, then $\|\phi\|_1 = 1$ and

$$\int_E g\phi = \frac{1}{a}\int_A |g| \ge M + \epsilon = (M+\epsilon)\|\phi\|_1,$$

which is a contradiction. Hence $\|g\|_\infty \le M$. $\square$

Theorem 6.44 (Generalized Minkowski's inequality)**.** *Suppose $p \in [1, \infty)$ and $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. If for almost every $y \in \mathbb{R}^n$, $f(x, y) \in L^p(\mathbb{R}^n)$, and $M = \int_{\mathbb{R}^n} (\int_{\mathbb{R}^n} |f(x, y)|^p \,\mathrm{d}x)^{1/p} \,\mathrm{d}y < \infty$, then*

$$\left(\int_{\mathbb{R}^n} \left|\int_{\mathbb{R}^n} f(x, y) \,\mathrm{d}y\right|^p \mathrm{d}x\right)^{1/p} \le M = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |f(x, y)|^p \,\mathrm{d}x\right)^{1/p} \mathrm{d}y.$$

Proof. Proof is trivial for $p = 1$. For $p \in (1, \infty)$, let $F(x) = \int_{\mathbb{R}^n} f(x, y) \,\mathrm{d}y$. Then for any simple function ϕ, there is

$$\begin{aligned}
\left|\int F(x)\phi(x) \,\mathrm{d}x\right| &\le \int |F(x)||\phi(x)| \,\mathrm{d}x \le \int \left(\int |f(x, y)| \,\mathrm{d}y\right) |\phi(x)| \,\mathrm{d}x \\
&= \int \left(\int |f(x, y)||\phi(x)| \,\mathrm{d}x\right) \mathrm{d}y \\
&\le \int \left(\int |f(x, y)|^p \,\mathrm{d}x\right)^{1/p} \mathrm{d}y \|\phi\|_q = M\|\phi\|_q,
\end{aligned}$$

where we applied Theorem 4.51 (Tonelli) to obtain the first equality and Hölder's inequality to obtain the last inequality. Hence $F(x) \in L^p$ and $\|F\|_p \le M$. Applying Theorem 6.43 yields the claimed inequality. □

Example 6.45 (Reduction to Minkowski's inequality)**.** Suppose $f, g \in L^p(\mathbb{R})$. Define the function $h : \mathbb{R} \times [0, 2] \to \mathbb{R}$ by

$$h(x, y) = \begin{cases} f(x) & \text{if } 0 \le y \le 1, \\ g(x) & \text{if } 1 < y \le 2. \end{cases}$$

Then $\int_0^2 h(x, y) \,\mathrm{d}y = f(x) + g(x)$ and

$$\left(\int |h(x, y)|^p \,\mathrm{d}x\right)^{1/p} = \begin{cases} \|f\|_p & \text{if } 0 \le y \le 1, \\ \|g\|_p & \text{if } 1 < y \le 2. \end{cases}$$

Hence the generalized Minkowski's inequality implies $\|f + g\|_p \le \|f\|_p + \|g\|_p$.

Example 6.46. Define the function $f : (-\infty, \infty) \times [0, 2] \to \mathbb{R}$ by

$$f(x, y) = \begin{cases} a_k, & \text{if } k \le x < k + 1,\ 0 \le y \le 1, \\ b_k, & \text{if } k \le x < k + 1,\ 1 < y \le 2, \end{cases}$$

where $a_k, b_k \ge 0$ for all k. Then $\int_0^2 f(x, y) \,\mathrm{d}y = a_k + b_k$ and

$$\left(\int_0^\infty |f(x, y)|^p \,\mathrm{d}x\right)^{1/p} = \begin{cases} (\sum_{k=1}^\infty |a_k|^p)^{1/p}, & \text{if } k \le x < k + 1,\ 0 \le y \le 1, \\ (\sum_{k=1}^\infty |b_k|^p)^{1/p}, & \text{if } k \le x < k + 1,\ 1 < y \le 2. \end{cases}$$

Hence the generalized Minkowski's inequality implies $(\sum_{k=1}^\infty |a_k + b_k|^p)^{1/p} \le (\sum_{k=1}^\infty |a_k|^p)^{1/p} + (\sum_{k=1}^\infty |b_k|^p)^{1/p}$.

7 Probability Theory

7.1 Basic concepts

Example 7.1 (Terms in measure theory vs probability theory)**.** See Table 1.

Table 1: Terminology correspondences in measure theory and probability theory

Measure theory	Probability theory
Measure space $(X, \mathcal{M}, \mu)$	Probability space $(\Omega, \mathcal{F}, \mathrm{P})$ $(\mathrm{P}(\Omega) = 1)$
σ-algebra $\mathcal{M}$	σ-field $\mathcal{F}$
Measurable set $E \in \mathcal{M}$	Event $E \in \mathcal{F}$
Measurable real-valued function f	Random variable X
Measure on $\mathbb{R}$ induced by f	Probability distribution P_X
Integral $\int_X f(x)\,\mathrm{d}\mu(x)$	Expectation $\mathrm{E}(X) = \int_\Omega X(\omega)\,\mathrm{dP}(\omega)$
$f \in L^p$	X has finite pth moment
Convergence in measure	Convergence in probability
Almost everywhere (a.e.)	Almost surely (a.s.)
Borel probability measure	Distribution
Fourier transform of a measure	Characteristic function of P_X
Laplace transform of a measure	Moment generating function of P_X

Definition 7.2 (Expectation and variance)**.** Suppose X is a random variable on $(\Omega, \mathcal{F}, \mathrm{P})$, then the *expectation of* X is defined by $\mathrm{E}(X) = \int_\Omega X(\omega)\,\mathrm{dP}(\omega)$, and the *variance of* X is defined by $\mathrm{V}(X) = \mathrm{E}[(X - \mathrm{E}(X))^2]$. Note that is is easy to verify that $\mathrm{V}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$.

Definition 7.3 (Image measure)**.** Suppose $(\Omega, \mathcal{F}, \mathrm{P})$ is a probability space, and $(\Omega', \mathcal{F}')$ is a measure space. Let $\phi : \Omega \to \Omega'$ be a measurable function, i.e., $\phi^{-1}(E) \in \mathcal{F}$ if $E \in \mathcal{F}'$. Then ϕ induces a probability measure on $(\Omega', \mathcal{F}')$, called *image measure*, defined by $\mathrm{P}_\phi(E) = \mathrm{P}(\phi^{-1}(E))$ for all $E \in \mathcal{F}'$.

Theorem 7.4. *Suppose $f : \mathcal{F}' \to \mathbb{R}$ is a measurable function, then $\int_{\Omega'} f'\,\mathrm{dP}_\phi = \int_\Omega f \circ \phi\,\mathrm{d}P$.*

Proof. Let $E \in \mathcal{F}'$ and $f = \chi_E : \mathcal{F}' \to \mathbb{R}$. Note that for any $\omega \in \Omega$ there is $\chi_E(\phi(\omega)) = 1$ iff $\phi(\omega) \in E$ iff $\omega \in \phi^{-1}(E)$, i.e., $\chi_E \circ \phi = \chi_{\phi^{-1}(E)} : \Omega \to \mathbb{R}$. Hence

$$\begin{aligned}\int_{\Omega'} f\,\mathrm{dP}_\phi &= \int_{\Omega'} \chi_E\,\mathrm{dP}_\phi = \int_E \mathrm{dP}_\phi = \mathrm{P}_\phi(E) = \mathrm{P}(\phi^{-1}(E)) \\ &= \int_{\phi^{-1}(E)} \mathrm{dP} = \int_\Omega \chi_{\phi^{-1}(E)}\,\mathrm{dP} = \int_\Omega \chi_E \circ \phi\,\mathrm{dP}.\end{aligned}$$

Therefore the identity holds for $f = \chi_E$. It is straightforward to show that it holds for simple functions by linearity. Taking limit of a sequence of simple functions and applying Theorem 4.28 (DCT) prove the claim for general measurable functions. $\square$

Definition 7.5 (Distribution)**.** Suppose X is a random variable on $(\Omega, \mathcal{F}, \mathrm{P})$. Let P_X be the image measure of X on $\mathbb{R}$, called the *distribution of* X. The function $F(t) = \mathrm{P}_X((-\infty, t)) = \mathrm{P}(X < t)$ is called the *distribution function of* X. A family of random variables $\{X_\alpha : \alpha \in A\}$ is called *identically distributed* if their image measures $\{\mathrm{P}_{X_\alpha} : \alpha \in A\}$ are identical.

Definition 7.6 (Joint distribution)**.** Suppose $\{X_k : 1 \le k \le n\}$ are random variables on $(\Omega, \mathcal{F}, \mathrm{P})$. Then $(X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$, and the image measure $\mathrm{P}_{X_1,\ldots,X_n}$ is called the *joint distribution of* $X_1, \ldots, X_n$.

Remarks. The behaviors of random variables are completely determined by their (joint) distributions. Therefore, we often use

$$\mathrm{E}(X) = \int_\Omega X(\omega)\,\mathrm{dP}(\omega) = \int_\mathbb{R} t\,\mathrm{dP}_X(t), \quad \mathrm{V}(X) = \int_\mathbb{R} (t - \mathrm{E}(X))^2\,\mathrm{dP}_X(t).$$

We also use

$$\mathrm{E}(X+Y) = \int_{\mathbb{R}^2} (t+s)\,\mathrm{dP}_{X,Y}(t,s).$$

Definition 7.7 (Independency)**.** Suppose $(\Omega, \mathcal{F}, \mathrm{P})$ is a probability space. A set of events $\{E_\alpha : \alpha \in A\}$ are called *independent* if for any finite subset of distinct events $\{E_{\alpha_k} : 1 \le k \le n,\ \alpha_k \in A\}$ there is

$$\mathrm{P}(E_{\alpha_1} \cap \cdots \cap E_{\alpha_n}) = \prod_{k=1}^n \mathrm{P}(E_{\alpha_k}).$$

A set of random variables $\{X_\alpha : \alpha \in A\}$ are called *independent* if the events $\{X_\alpha^{-1}(B_\alpha) : \alpha \in A\}$ are independent. Note that this is different from and stronger than pairwise independency. An alternative definition of independent random variables is that for any finite subset of these random variables, say $X_1, \ldots, X_n$, which are distinct, there is

$$\mathrm{P}_{X_1,\ldots,X_n}(B_1 \times \cdots \times B_n) = \prod_{k=1}^n \mathrm{P}_{X_k}(B_k).$$

We can see this because on the one hand we have

$$\begin{aligned}\mathrm{P}(X_1^{-1}(B_1) \cap \cdots \cap X_n^{-1}(B_n)) &= \mathrm{P}((X_1, \ldots, X_n)^{-1}(B_1 \times \cdots \times B_n)) \\ &= \mathrm{P}_{X_1,\ldots,X_n}(B_1 \times \cdots \times B_n)\end{aligned}$$

and on the other hand we have

$$\prod_{k=1}^n \mathrm{P}(X_k^{-1}(B_k)) = \prod_{k=1}^n \mathrm{P}_{X_k}(B_k) = (\prod_{k=1}^n \mathrm{P}_{X_k})(B_1 \times \cdots \times B_n).$$

Hence $X_1, \ldots, X_n$ are independent iff the two quantities above are identical, i.e., $\mathrm{P}_{X_1,\ldots,X_n} = \prod_{k=1}^n \mathrm{P}_{X_k}$.

Theorem 7.8. *Suppose $X_1, \dots, X_n$ are independent random variables, and $f_k : \mathbb{R} \to \mathbb{R}$ are measurable, then $f_1(X_1), \dots, f_n(X_n)$ are also independent.*

Proof. Let $Y_k = f_k(X_k)$. For any $B_k \in \mathcal{F}$, there is

$$\begin{aligned}\mathrm{P}_{Y_1,\dots,Y_n}(B_1 \times \cdots \times B_n) &= \mathrm{P}_{X_1,\dots,X_n}(f_1^{-1}(B_1) \times \cdots \times f_n^{-1}(B_n)) \\ &= \prod_{k=1}^{n} \mathrm{P}_{X_k}(f_k^{-1}(B_k)) = \prod_{k=1}^{n} \mathrm{P}_{Y_k}(B_k),\end{aligned}$$

which completes the proof. □

Theorem 7.9. *Suppose $\{X_k : 1 \le k \le n\}$ are independent and $X_k \in L^1$, then $\prod_{k=1}^n X_k \in L^1$ and $\mathrm{E}(\prod_{k=1}^n X_k) = \prod_{k=1}^n \mathrm{E}(X_k)$.*

Proof. Note that

$$\begin{aligned}\mathrm{E}\Big(\prod_{k=1}^{n} |X_k|\Big) &= \int \prod_{k=1}^{n} |X_k| \,\mathrm{d}\mathrm{P}_{X_1,\dots,X_n} = \int \prod_{k=1}^{n} |X_k| \,\mathrm{d}P_{X_1} \dots \mathrm{d}P_{X_n} \\ &= \prod_{k=1}^{n} \int |X_k| \,\mathrm{d}P_{X_k} = \prod_{k=1}^{n} \mathrm{E}(|X_k|) < \infty,\end{aligned}$$

which implies that $\prod_{k=1}^n X_k \in L^1$. Remove the absolute values and redo this to show $\mathrm{E}(\prod_{k=1}^n X_k) = \prod_{k=1}^n \mathrm{E}(X_k)$. □

Theorem 7.10. *Suppose $\{X_k : 1 \le k \le n\}$ are independent and $X_k \in L^2$, then $\mathrm{V}(\sum_{k=1}^n X_k) = \sum_{k=1}^n \mathrm{V}(X_k)$.*

Proof. Let $Y_k = X_k - \mathrm{E}(X_k)$. Then $Y_1, \dots, Y_n$ are independent, $\mathrm{E}(Y_k) = 0$, and $\mathrm{E}(Y_k^2) = \mathrm{V}(X_k$. Moreover $\mathrm{E}(Y_k Y_j) = \mathrm{E}(Y_k)\mathrm{E}(Y_j) = 0$ whenever $k \ne j$ since they are independent. Therefore

$$\begin{aligned}\mathrm{V}(X_1 + \cdots + X_n) &= \mathrm{E}\left[\Big(\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mathrm{E}(X_k)\Big)^2\right] = \mathrm{E}\left[\Big(\sum_{k=1}^{n} Y_k\Big)^2\right] \\ &= \sum_{k,j=1}^{n} \mathrm{E}(Y_k Y_j) = \sum_{k=1}^{n} \mathrm{E}(Y_k^2) = \sum_{k=1}^{n} \mathrm{V}(X_k),\end{aligned}$$

which completes the proof. □

7.2 The law of large numbers

Theorem 7.11 (Chebyshev's inequality)**.** *Suppose X is a random variable with mean $\mathrm{E}(X)$ and variance $\mathrm{V}(X)$. Then for any $\epsilon > 0$, there is*

$$\mathrm{P}(|X - \mathrm{E}(X)| \ge \epsilon) < \frac{\mathrm{V}(X)}{\epsilon^2}.$$

Proof. Note that

$$\mathrm{P}(|X - \mathrm{E}(X)| \geq \epsilon) = \int_{|X-\mathrm{E}(X)|\geq\epsilon} \mathrm{dP} \leq \int_{\mathbb{R}} \Big(\frac{t - \mathrm{E}(X)}{\epsilon}\Big)^2 \mathrm{dP}_X(t) = \frac{\mathrm{V}(X)}{\epsilon^2},$$

which proves the claim. □

Theorem 7.12 (Weak law of large numbers)**.** *Suppose* $\{X_k : k \in \mathbb{N}\}$ *are independent random variables with means* μ_k *and variances* σ_k^2*. If* $\frac{1}{n}\sum_{k=1}^n \sigma_k^2 \to 0$ *as* $n \to \infty$*, then for any* $\epsilon > 0$*,*

$$\mathrm{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^n (X_k - \mu_k)\Big| \geq \epsilon\Big) \leq \frac{1}{\epsilon^2}\frac{1}{n}\sum_{k=1}^n \sigma_k^2 \to 0.$$

Proof. Applying Theorem 7.11 (Chebyshev) to $\frac{1}{n}\sum_{k=1}^n (X_k - \mu_k)$, which has mean 0 and variance $\frac{1}{n}\sum_{k=1}^n \sigma_k^2$, to obtain the claimed inequality. □

Theorem 7.13 (Borel-Cantelli)**.** *Suppose* $(\Omega, \mathcal{F}, \mathrm{P})$ *is a probability space, and* $\{E_k : k \in \mathbb{N}\}$ *are events. Then*

1. *If* $\sum_{k=1}^\infty \mathrm{P}(E_k) < \infty$*, then* $\mathrm{P}(\limsup_k E_k) = 0$*.*
2. *If* $\{E_k\}$ *are independent and* $\sum_{k=1}^\infty \mathrm{P}(E_k) = \infty$*, then* $\mathrm{P}(\limsup_k E_k) = 1$*.*

Proof. Item 1 can be easily verified: as $k \to \infty$, there is

$$\mathrm{P}(\limsup_{k\to\infty} E_k) = \mathrm{P}(\bigcap_{k=1}^\infty \bigcup_{j=k}^\infty E_j) \leq \mathrm{P}(\bigcup_{j=k}^\infty E_j) \leq \sum_{j=k}^\infty \mathrm{P}(E_j) \to 0.$$

For Item 2, we know $\{E_k^c\}$ are independent since $\{E_k\}$ are so. Hence

$$\mathrm{P}(\bigcup_{j=k}^n E_j^c) = \prod_{j=k}^n \mathrm{P}(E_j^c) = \prod_{j=k}^n (1 - \mathrm{P}(E_j)) \leq \prod_{j=k}^n e^{-\mathrm{P}(E_j)} = e^{-\sum_{j=k}^n \mathrm{P}(E_j)} \to 0$$

as $n \to \infty$. Hence $\mathrm{P}(\liminf_k E_k^c) = \mathrm{P}(\cup_{k=1}^\infty \cap_{j=k}^\infty E_j^c) \leq \sum_{k=1}^\infty \mathrm{P}(\cap_{j=k}^\infty E_j^c) = 0$, which implies that $\mathrm{P}(\limsup_k E_k) = \mathrm{P}((\liminf_k E_k^c)^c) = 1$. □

Theorem 7.14 (Kolmogorov's inequality)**.** *Suppose* $\{X_k : 1 \leq k \leq n\}$ *are independent random variables with mean* 0 *and variances* σ_k^2 *for all* k*. Let* $S_k = \sum_{j=1}^k X_j$ *for* $k = 1, \dots, n$*. Then for any* $\epsilon > 0$*, there is*

$$\mathrm{P}\Big(\max_{1\leq k\leq n} |S_k| \geq \epsilon\Big) \leq \epsilon^{-2}\sum_{k=1}^n \sigma_k^2.$$

Proof. Note that $\mathrm{E}(X_k) = 0$ and $\mathrm{V}(X_k) = \sigma_k^2$, hence $\mathrm{E}(S_k) = 0$ and $\mathrm{V}(S_k) = \mathrm{E}(S_k^2) = \sum_{j=1}^k \sigma_j^2$. Moreover, S_k and $S_n - S_k$ are independent. Now let $A_k =$

$\{|S_k| \geq \epsilon\} \cap \{|S_j| < \epsilon : 1 \leq j < k\}$, then $A_k \cap A_j = \emptyset$ whenever $k \neq j$. Thus $\{\max_{1\leq k\leq n} |S_k| \geq \epsilon\} = \cup_{k=1}^n A_k$ is a disjoint union. Therefore

$$\mathrm{P}\left(\max_{1\leq k\leq n} |S_k| \geq \epsilon\right) = \mathrm{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mathrm{P}(A_k) \leq \frac{1}{\epsilon^2}\sum_{k=1}^{n} \mathrm{E}(\chi_{A_k} S_k^2),$$

where we used $P(A_k) = \int_{A_k} \mathrm{dP} \leq \int_{A_k} \frac{|S_k|^2}{\epsilon^2}\,\mathrm{dP} = \frac{1}{\epsilon^2}\mathrm{E}(\chi_{A_k} S_k^2)$ to obtain the inequality. On the other hand, we have

$$\begin{aligned}\mathrm{E}(S_n^2) &\geq \mathrm{E}\left[\Big(\sum_{k=1}^{n} \chi_{A_k}\Big)S_n^2\right] = \mathrm{E}\left[\sum_{k=1}^{n} \chi_{A_k}\Big(S_n^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2\Big)\right] \\ &\geq \sum_{k=1}^{n} \mathrm{E}(\chi_{A_k} S_k^2) + 2\sum_{k=1}^{n} \mathrm{E}(\chi_{A_k} S_k(S_n - S_k)) = \sum_{k=1}^{n} \mathrm{E}(\chi_{A_k} S_k^2),\end{aligned}$$

where we used the fact $\mathrm{E}(\chi_{A_k} S_k(S_n - S_k)) = \mathrm{E}(\chi_{A_k} S_k)\mathrm{E}(S_n - S_k) = 0$ due to the independency between $\chi_{A_k} S_k = S_n - S_k$. Combining the two inequalities above and recalling $\mathrm{E}(S_n^2) = \sum_{k=1}^n \sigma_k^2$ completes the proof. □

Theorem 7.15 (Strong law of large numbers)**.** *If $\{X_n : n \in \mathbb{N}\}$ is a sequence of independent L^2 random variables with mean μ_n and variances σ_n^2 such that $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$, then $\frac{1}{n}\sum_{k=1}^n (X_k - \mu_k) \to 0$ a.s. as $n \to \infty$.*

Proof. Denote $S_n = \sum_{k=1}^n (X_k - \mu_k)$. It suffices to show that, for any $\epsilon > 0$, $\mathrm{P}(\limsup_n \{\frac{|S_n|}{n} \geq \epsilon\}) = 0$. Now we define

$$A_k = \left\{\max_{2^{k-1}\leq n<2^k} \frac{|S_n|}{n} \geq \epsilon\right\} \subset \left\{\max_{1\leq n<2^k} |S_n| \geq 2^{k-1}\epsilon\right\}.$$

Then it is clear that

$$\limsup_{n\to\infty}\left\{\frac{|S_n|}{n} \geq \epsilon\right\} = \bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\left\{\frac{|S_n|}{n} \geq \epsilon\right\} = \bigcap_{m=1}^{\infty}\bigcup_{k=m}^{\infty} A_k = \limsup_{k\to\infty} A_k.$$

On the other hand, by Theorem 7.14 (Kolmogorov's inequality), we know $\mathrm{P}(A_k) \leq (2^{k-1}\epsilon)^{-2}\sum_{n=1}^{2^k} \sigma_n^2$, summing of which over k yields

$$\sum_{k=1}^{\infty} \mathrm{P}(A_k) \leq \sum_{k=1}^{\infty}\sum_{n=1}^{2^k} \frac{\sigma_n^2}{2^{2(k-1)}\epsilon^2} = \frac{4}{\epsilon^2}\sum_{n=1}^{\infty}\Big(\sum_{k\geq\log_2 n} \frac{1}{2^{2k}}\Big)\sigma_n^2 \leq \frac{16}{\epsilon^2}\sum_{n=1}^{\infty}\frac{\sigma_n^2}{n^2} < \infty.$$

Hence, by Theorem 7.13 (Borel-Cantelli), we know $\mathrm{P}(\limsup_k A_k) = 0$. □

7.3 Central limit theorem

Definition 7.16 (Moment generating function)**.** The *moment generating function* of a random variable X with distribution function F is defined by $\mathrm{E}[e^{tX}] = \int_{\infty}^{\infty} e^{tx}\,\mathrm{d}F(x)$ for every $t \in \mathbb{R}$.

Remarks. The name of moment generating function is due to the fact that $\mathrm{E}[X^k] = M_X^{(k)}(0)$, the kth derivative of M_X at $t = 0$ for all $k = 0, 1, \ldots$. This also implies that $M_X(t) = \sum_{k=0}^{\infty} \frac{\mathrm{E}[X^k]}{k!} t^k$. Moment generating function M_X is essentially the Laplace transform of the distribution function F. Thus, two random variables are identical iff their moment generating functions are identical.

Remarks. It is straightforward to verify that $M_{aX+b}(t) = e^{bt} M_X(at)$ for any $a, b \in \mathbb{R}$ and $M_{X+Y}(t) = M_X(t) M_Y(t)$ for any independent random variables X and Y.

Theorem 7.17 (Central limit theorem)**.** *Let $\{X_k\}$ be a sequence of independent and identically distributed L^2 random variables with mean μ and variance σ^2, then $Y_n := (\sigma\sqrt{n})^{-1} \sum_{k=1}^{n} (X_k - \mu)$ has mean 0 and variance 1. Moreover, for any $a \in \mathbb{R}$, there is*

$$\lim_{n\to\infty} \mathrm{P}(Y_n \le a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-t^2/2} \,\mathrm{d}t.$$

That is, $\lim_n \mathrm{P}(Y_n \le a) = \mathrm{P}(Z \le a)$ where $Z \sim N(0, 1)$ is the standard normal random variable.

Proof. We assume $\mu = 0$ and $\sigma^2 = 1$ since it is straightforward to extend to the general case by changing variable X_k with $(X_k - \mu)/\sigma$. Let F be the distribution function of Z and F_n the distribution function of Y_n, then we need to show that $F_n \to F$ pointwisely. To this end, we consider their moment generating functions M_Z and M_{Y_n}. We know $M_{Y_n}(t) = M_X(t/\sqrt{n})^n$. Noting that $M_X(t) = 1 + t^2/2 + o(t^2)$, we have $M_{Y_n}(t) = (1 + t^2/(2n) + o(t^2))^n$. On the other hand, $M_Z(t) = 1 + t^2/2 + o(t^2)$. Hence $M_{Y_n}(t) \to M_Z(t)$ as $n \to \infty$ for all $t \in \mathbb{R}$ sufficiently close to 0, and applying inverse Laplacian transform to the moment generating functions yields the claim. $\square$