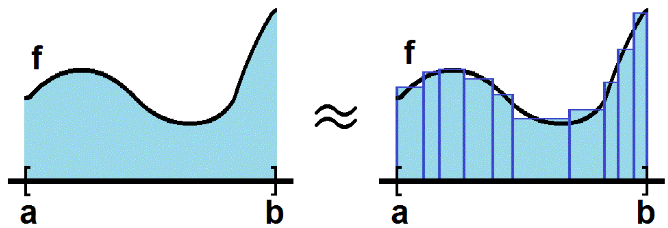
Real Analysis

Dr Nikolai Chernov[1]

0 Introduction

In Calculus, a definite integral $\int_a^b f(x)\,dx$ is computed by dividing the interval $I = [a, b]$ into small subintervals $I_1, \dots, I_n$ and approximating

$$\int_a^b f(x)\,dx \approx \sum_{i=1}^{n} f(x_i)\,|I_i|$$

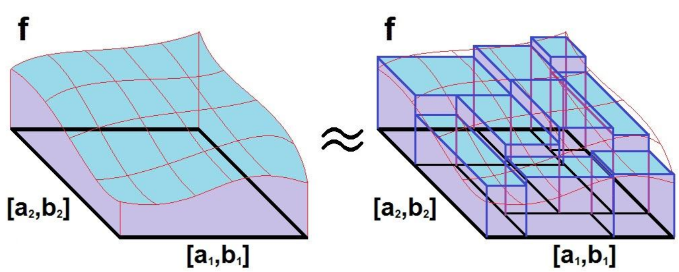


Here $|I_i|$ denotes the length of I_i and x_i is any point in I_i. The limit, as the lengths of each I_i approaches zero, gives the exact value of $\int_a^b f(x)\,dx$.

Examples: In mechanics, if $f(x)$ is the density (mass per unit length) of a rod $[a, b]$ then $\int_a^b f(x)\,dx$ is the total mass of the rod. In probability theory, if $f(x)$ is the density of a 1D probability distribution then $\int_a^b f(x)\,dx$ is the total probability on the interval $[a, b]$.

Similarly, a definite double integral $\iint_R f(x, y)\,dx\,dy$ over a rectangle $R = [a_1, b_1] \times [a_2, b_2]$ is computed by partitioning R into small subrectangles $R_1, \dots, R_n$, choosing any $(x_i, y_i) \in R_i$ for $i = 1, ..., n$, and approximating

$$\iint_R f(x, y)\,dx\,dy \approx \sum_{i=1}^{n} f(x_i, y_i)\,\mathrm{Area}(R_i),$$



Here (x_i, y_i) is any point in R_i. The limit, as the area of each R_i approaches zero, gives the exact value of $\iint_R f(x, y)\,dx\,dy$

Examples: In mechanics, if $f(x, y)$ is the density (mass per unit area) of a rectangular plate R, then $\iint_R f(x, y)\,dx\,dy$ is the total mass of the plate. In probability, if $f(x, y)$ is the density of a 2D probability distribution, then $\iint_R f(x, y)\,dx\,dy$ is the total probability over the rectangle R.

[1] Formatting and illustrations by Michael Pogwizd.

In the studies of probabilities, we often need to replace a 1D interval or 2D rectangle with other objects that can be more complicated. So it becomes necessary to generalize the notion of definite integral.

What is common between the above two examples? A function f has a domain D (an interval or a rectangle) that can be divided into smaller pieces D_i, and each piece is measured by a positive number (its length or area), the smaller the piece the smaller its measure. The measure of the whole domain D is the sum of measures of its pieces:

$$\mu(D) = \sum_{i=1}^{n} \mu(D_i)$$

where μ denotes that measure (the length or area). The integral can then be approximated by

$$\sum_{i=1}^{n} f(z_i)\, \mu(D_i)$$

where z_i is a point in D_i.

The purpose of Real Analysis is to develop a machinery (a theoretical apparatus) for integrating functions on arbitrary domains. We will generalize the "length of a line interval" and the "area of a rectangle" to an abstract notion of a "measure of an arbitrary set". And we will learn how to integrate functions on arbitrary domains.

The following chart summarizes the basic goals and terms of Real Analysis:

1D	2D		General
Interval I	Rectangle R	$\Longrightarrow$	Space X
Length	Area	$\Longrightarrow$	Measure μ
Integral $\int_a^b f\,dx$	Double Integral $\iint_R f(x,y)\,dx\,dy$	$\Longrightarrow$	Lebesgue Integral $\int_X f\,d\mu$

The theory of (abstract) measures and Lebesgue integration provides a solid foundation for modern Probability Theory. A good knowledge of Real Analysis is an absolute must for anyone who plans to do research in Dynamical Systems or Mathematical Physics.

1 Lebesgue measure in $\mathbb{R}$ and $\mathbb{R}^2$: basic constructions

DEFINITION 1.1. An **interval** $I \subset \mathbb{R}$ is a set of the form $[a, b]$ or $[a, b)$ or $(a, b]$ or (a, b), where $a \leq b$ are real numbers. Its length is $|I| = b - a$

- Including or excluding the endpoints of the interval does not affect its length.
- Notation: $A \uplus B$ always means *disjoint union* of two sets A and B i.e., such that $A \cap B = \emptyset$. Furthermore, $\uplus_{n=1}^{\infty} A_n$ always means the union of pairwise disjoint sets A_n, i.e., such that $A_n \cap A_m = \emptyset$ for all $m \neq n$.

Lemma 1.2. *If* $I = \uplus_{n=1}^{N} I_n$, *then* $|I| = \sum_{n=1}^{N} |I_n|$

Proof. If $N = 2$ and c is the common endpoint of I_1 and I_2, then $|I_1| + |I_2| = (b - c) + (c - a) = b - a = |I|$. For $N \geq 2$ use induction. □

DEFINITION 1.3. A (linear) **elementary set** is a finite union of disjoint intervals. i.e. $J = \uplus_{n=1}^{N} I_n$ The total length of J is $|J| = \sum_{n=1}^{N} |I_n|$

- I_n are not necessarily adjacent intervals, there may be gaps in between.

DEFINITION 1.4. A **rectangle** $R \subset \mathbb{R}^2$ is a set of the form $R = I_1 \times I_2$, where I_1, I_2 are intervals. The area of a rectangle is $\text{Area}(R) = |I_1| \times |I_2|$

- We only consider rectangles with horizontal and vertical sides.
- Including or excluding the sides of the rectangles does not affect their areas.

Lemma 1.5. *If* $R = \uplus_{n=1}^{N} R_n$, *then* $\text{Area}(R) = \sum_{n=1}^{N} \text{Area}(R_n)$.

Proof. For $N = 2$ the proof is a trivial calculation. For $N \geq 2$ we first need to extend the sides of R_n's so that they run completely across R (see Figure 1). This causes partitioning of some of the R_n's into smaller rectangles, which can be done one by one and using the already proved version of the lemma for $N = 2$. In the end we get a partition of R by $k_1 \geq 1$ vertical lines and $k_2 \geq 1$ horizontal lines (like a grid). Now the proof is a simple calculation using Lemma 1.2. □

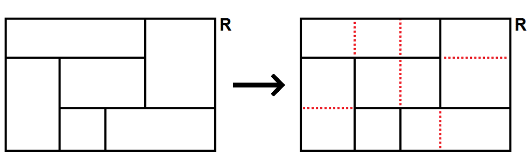


Figure 1: A partition of a rectangle into smaller rectangles. Inner sides are extended so that they run completely across R (top to bottom and left to right).

Definition 1.6. A (planar) **elementary set** is a finite union of disjoint rectangles, i.e. $B = \uplus_{i=1}^n R_i$. The area of an elementary set is $\text{Area}(B) = \sum_{i=1}^n \text{Area}(R_i)$

Theorem 1.7.

(a) *The length of a linear elementary set does not depend on how it is partitioned into disjoint intervals.*

(b) *The area of a planar elementary set does not depend on how it is partitioned into disjoint rectangles.*

Proof. We prove part (b). Let $B = \cup_{i=1}^n R_i$ and $B = \cup_{j=1}^m R'_j$ be two partitions of an elementary set B into rectangles. Note that for each pair i, j the intersection $R_{ij} = R_i \cap R'_j$ is a rectangle (which maybe empty). Now by Lemma 1.5:

$$\sum_{i=1}^n \text{Area}(R_i) \overset{1.5}{=} \sum_{i=1}^n \sum_{j=1}^m \text{Area}(R_{ij}) = \sum_{j=1}^m \sum_{i=1}^n \text{Area}(R_{ij}) \overset{1.5}{=} \sum_{j=1}^m \text{Area}(R'_j).$$

The easier part (a) is left as an exercise. □

Corollary 1.8.

(a) *If $J = \uplus_{i=1}^n J_i$ is a finite union of disjoint linear elementary sets, then*
$$|J| = \sum_{i=1}^n |J_i|$$

(b) *If $B = \uplus_{i=1}^n B_i$ is a finite union of disjoint planar elementary sets, then*
$$\text{Area}(B) = \sum_{i=1}^n \text{Area}(B_i)$$

Theorem 1.9. *Finite unions, intersections and differences of elementary sets are elementary sets.*

Proof. A direct inspection. We omit details. □

- A countable union of elementary sets is not necessarily an elementary set.

Exercise 1. Show that the open disk $x^2 + y^2 < 1$ is a countable union of planar elementary sets. Show that the closed disk $x^2 + y^2 \leq 1$ is a countable intersection of planar elementary sets.

DEFINITION 1.10. A **ring** is a nonempty collection of subsets of a set X closed under finite unions, intersections and differences.

DEFINITION 1.11. An **algebra** is a ring containing X itself.

EXAMPLE 1. Linear elementary sets $J \subset \mathbb{R}$ make a ring. Planar elementary sets $B \subset \mathbb{R}^2$ make a ring.

EXAMPLE 2. Linear elementary sets $J \subset \mathbb{R}$ do not make an algebra. If we consider all finite unions of <u>finite intervals</u> and <u>infinite intervals</u>, then we will get an algebra. (An infinite interval is a set $[a, \infty)$ or (a, ∞) or $(-\infty, a]$ or $(-\infty, a)$, where $a \in \mathbb{R}^1$. The real line $\mathbb{R}^1$ itself is also an infinite interval.)

Theorem 1.12.

(a) *If a linear elementary set J is covered by intervals $I_1, \dots, I_n$ (i.e. $J \subset \cup_{i=1}^n I_i$), then* $|J| \le \sum_{i=1}^n |I_i|$

(b) *If a planar elementary set B is covered by rectangles $R_1, \dots, R_n$ (i.e. $B \subset \cup_{i=1}^n R_i$), then* $\mathrm{Area}(B) \le \sum_{i=1}^n \mathrm{Area}(R_i)$

Proof. We prove (b). By Theorem 1.9, $B_1 = B \cap R_1$ and $B_i = B \cap (R_i \setminus \cup_{j=1}^{i-1} R_j)$ are elementary sets. Obviously, $B_i \subset R_i$, hence $\mathrm{Area}(B_i) \le \mathrm{Area}(R_i)$. Since B_i are disjoint and $B = \uplus_{i=1}^n B_i$, we have

$$\mathrm{Area}(B) \overset{1.8(\mathrm{b})}{=} \sum_{i=1}^n \mathrm{Area}(B_i) \le \sum_{i=1}^n \mathrm{Area}(R_i).$$

The easier part (a) is left as an exercise. □

Next we generalize the concepts of length in $\mathbb{R}^1$ and area in $\mathbb{R}^2$. This can be done in parallel, as above, but for the sake of brevity we only do it for the 2D case (i.e., for the area in $\mathbb{R}^2$). The conversion of all our definitions and theorems to the simpler 1D case is left as an exercise.

For convenience, we want to avoid infinite areas at early stages of our constructions. In this section we just fix a large finite rectangle $X \subset \mathbb{R}^2$ and only consider subsets $A \subset X$. Note that elementary sets $B \subset X$ make an algebra (because X is an elementary set).

Definition 1.13. The **outer measure** of a set $A \subset X$ is

$$\mu^*(A) = \inf\left\{\sum_{i=1}^{\infty} \text{Area}(R_i)\right\}$$

where the infimum is taken over all countable covers of A by rectangles. (That is, such that $A \subset \cup_{i=1}^{\infty} R_i$.)

- $\mu^*(A) \leq \text{Area}(X)$ (we can use countable cover $\{X, \emptyset, \emptyset, \ldots\}$)
- $\mu^*(B) \leq \text{Area}(B)$ (if $B = \uplus_{i=1}^{n} R_i$, then we can use countable cover $\{R_1, \ldots, R_n, \emptyset, \emptyset, \ldots\}$)

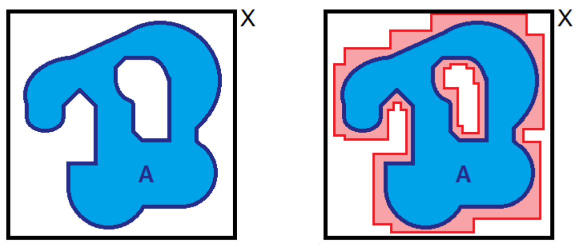


Figure 2: Visualization for calculating the outer measure of a set A in X

The following theorem is the first non-trivial fact of Real Analysis:

Theorem 1.14. *For any elementary set $B \subset X$ we have $\mu^*(B) = \text{Area}(B)$.*

Proof. Suppose, by way of contradiction, that $\mu^*(B) < \text{Area}(B)$
Let $\delta = \text{Area}(B) - \mu^*(B) > 0$. There is a countable cover of B by rectangles $B \subset \cup_{i=1}^{\infty} R_i$ such that $\sum_{i=1}^{\infty} \text{Area}(R_i) < \text{Area}(B) - \delta/2$.

Lemma. *For any rectangle R and any $\varepsilon > 0$ there is an **open** covering rectangle $R' \supset R$ such that $\text{Area}(R' \setminus R) < \varepsilon$ and a **closed** subrectangle $R'' \subset R$ such that $\text{Area}(R \setminus R'') < \varepsilon$. Similar statements hold for elementary sets.*
(Proof is a simple topology exercise and is omitted.)

Thus we can find a closed elementary subset $B' \subset B$ with $\text{Area}(B') > \text{Area}(B) - \delta/4$. For each R_i we can find an open covering rectangle $R_i' \supset R_i$ such that $\text{Area}(R_i') < \text{Area}(R_i) + \delta/2^{i+2}$. Now

$$\sum_{i=1}^{\infty} \text{Area}(R_i') < \sum_{i=1}^{\infty} \text{Area}(R_i) + \delta/4 < \text{Area}(B').$$

Since B' is compact and covered by a countable union of open sets R_i', there exists a finite subcover, i.e., $B' \subset \cup_{i=1}^{n} R_i'$ for some $n < \infty$. According to Theorem 1.12 $\sum_{i=1}^{n} \text{Area}(R_i') \geq \text{Area}(B')$, a contradiction. $\square$

Theorem 1.15. *If $A_1 \subset A_2$, then $\mu^*(A_1) \leq \mu^*(A_2)$*

Proof. Any countable cover of A_1 by rectangles is also a cover for A_2. $\square$

Theorem 1.16. *If $A \subset \bigcup_{i=1}^{n} A_i$, then $\mu^*(A) \leq \sum_{i=1}^{n} \mu^*(A_i)$*
This property is called subadditivity.

Proof. We can combine countable covers of $A_1, \ldots, A_n$ into one countable cover of A. (Recall that a finite union of countable sets is countable.) $\square$

Theorem 1.17. *If $A \subset \cup_{i=1}^{\infty} A_i$, then $\mu^*(A) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$*
This property is called σ-subadditivity.

Proof. We can combine countable covers of $A_1, A_2, \ldots$ into one countable cover of A. (Recall that a countable union of countable sets is countable.) $\square$

Corollary 1.18. *If $B \subset \cup_{i=1}^{\infty} B_i$, where B and all B_i's are elementary sets, then* $\text{Area}(B) \leq \sum_{i=1}^{\infty} \text{Area}(B_i)$.

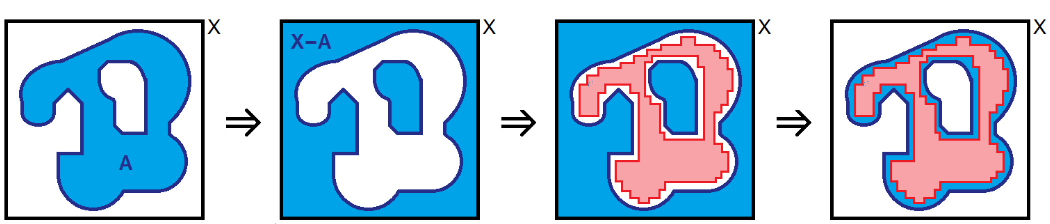
Proof. Combine Theorem 1.17 and Theorem 1.14. $\square$

DEFINITION 1.19. The **inner measure** of a set $A \subset X$ is

$$\mu_*(A) = \text{Area}(X) - \mu^*(A^c)$$

where $A^c = X \setminus A$ is the complement of A.

Note: The inner measure can be thought as the 'opposite' of the outer measure. The following is a way to visualize it: Cover $X \setminus A$ with a countable number of rectangles (as with the outer measure). The complements of these rectangles are seen to be within A.



Theorem 1.20. *For any set $A \subset X$ we have $0 \le \mu_*(A) \le \mu^*(A) \le \text{Area}(X)$.*

Proof. The only non-trivial claim here is $\mu_*(A) \le \mu^*(A)$ By Definition 1.19, this is equivalent to $\text{Area}(X) - \mu^*(A^c) \le \mu^*(A)$ i.e. $\mu^*(A) + \mu^*(A^c) \ge \text{Area}(X)$. This follows from Theorem 1.16. □

Definition 1.21. A set $A \subset X$ is **measurable** iff $\mu_*(A) = \mu^*(A)$
The **Lebesgue measure** is defined by $\mathbf{m}(A) = \mu_*(A) = \mu^*(A)$.

Lemma 1.22. *A set $A \subset X$ is measurable if and only if*

$$\mu^*(A) + \mu^*(A^c) = \text{Area}(X).$$

Proof. By Definition 1.19, the relation $\mu_*(A) = \mu^*(A)$ is equivalent to $\text{Area}(X) - \mu^*(A^c) = \mu^*(A)$, which is the same as $\mu^*(A) + \mu^*(A^c) = \text{Area}(X)$. □

- The symmetry above implies: A is measurable $\Leftrightarrow$ A^c is measurable

Theorem 1.23. *Every elementary set $B \subset X$ is measurable and its Lebesgue measure is equal to its area, i.e.* $\mathbf{m}(B) = \text{Area}(B)$.

Proof. If B is an elementary set, then $B^c = X \setminus B$ is also an elementary set. Now the result follows from Theorem 1.14, Corollary 1.8, and Lemma 1.22. □

- An empty set is measurable and $\mathbf{m}(\emptyset) = 0$

Exercise 2. Prove that every countable set $A \subset X$ is measurable and $\mathbf{m}(A) = 0$.

Exercise 3. Let $A \subset X$ consist of points (x, y) such that either x or y is a rational number. Is A measurable? What is its Lebesgue measure?

Lemma 1.24. *For any two sets $A_1, A_2 \subset X$ (not necessarily measurable), we have*

$$|\mu^*(A_1) - \mu^*(A_2)| \le \mu^*(A_1 \Delta A_2)$$

Proof. Recall that $A_1 \Delta A_2 = (A_1 \setminus A_2) \cup (A_2 \setminus A_1)$.

By Theorem 1.16 and Theorem 1.15:

$$\mu^*(A_1) \le \mu^*(A_1 \cap A_2) + \mu^*(A_1 \setminus A_2) \le \mu^*(A_2) + \mu^*(A_1 \Delta A_2).$$

Similarly, $\mu^*(A_2) \le \mu^*(A_1) + \mu^*(A_1 \Delta A_2)$. □

Theorem 1.25. (Approximation) *A set $A \subset X$ is measurable if and only if for any $\varepsilon > 0$ there exists an elementary set $B_\varepsilon \subset X$ such that $\mu^*(A \Delta B_\varepsilon) < \varepsilon$*

Proof. (See Figure 3 for visualization of theorem)

Note: In this case we also have $|\mathbf{m}(A) - \text{Area}(B_\varepsilon)| \le \varepsilon$ by Lemma 1.24

$\boxed{\text{"}\Leftarrow\text{"}}$ Suppose $\forall \varepsilon > 0 \quad \exists B_\varepsilon \subset X$ (elementary set) s.t. $\mu^*(A \Delta B_\varepsilon) < \varepsilon$
Since $A^c \Delta B_\varepsilon^c = A \Delta B_\varepsilon$, we have $\mu^*(A^c \Delta B_\varepsilon^c) = \mu^*(A \Delta B_\varepsilon) < \varepsilon$

$$|\mu^*(A) - \text{Area}(B_\varepsilon)| \overset{1.14}{=} |\mu^*(A) - \mu^*(B_\varepsilon)| \overset{1.24}{\le} \mu^*(A \Delta B_\varepsilon) < \varepsilon$$

Thus we have: $|\mu^*(A) - \text{Area}(B_\varepsilon)| < \varepsilon$ and $|\mu^*(A^c) - \text{Area}(B_\varepsilon^c)| < \varepsilon$
Using the triangle inequality, we get

$$|\mu^*(A) + \mu^*(A^c) - \text{Area}(X)| \le |\mu^*(A) - \text{Area}(B_\varepsilon)| + |\mu^*(A^c) - \text{Area}(B_\varepsilon^c)| < 2\varepsilon$$

Since this is true for any $\varepsilon > 0$, we must have $\mu^*(A) + \mu^*(A^c) = \text{Area}(X)$
Lemma 1.22 now implies that A is a measurable set.

$\boxed{\text{"}\Rightarrow\text{"}}$ Let $\varepsilon > 0$

Let $A \subset X$ be a measurable set. By Lemma 1.22, A^c is also measurable.
Choose a disjoint countable cover $\uplus_{i=1}^\infty R_i \supset A$ so that $\sum_i \text{Area}(R_i) < \mu^*(A) + \varepsilon$
Choose a disjoint countable cover $\uplus_{j=1}^\infty R'_j \supset A^c$ so that $\sum_j \text{Area}(R'_j) < \mu^*(A^c) + \varepsilon$
Adding the two covers, we get:

$$\sum_{i=1}^\infty \text{Area}(R_i) + \sum_{j=1}^\infty \text{Area}(R'_j) < \mu^*(A) + \mu^*(A^c) + 2\varepsilon \overset{1.22}{=} \text{Area}(X) + 2\varepsilon. \qquad (1)$$

Now $\exists N > 0$ s.t. $\sum_{i>N} \text{Area}(R_i) < \varepsilon$
Define $B = \cup_{i=1}^N R_i$, $T = \cup_{i>N} R_i$ (the "tail") , $S = \cup_{j=1}^\infty (B \cap R'_j)$

Note that $\mu^*(T) \overset{1.17}{\le} \sum_{i>N} \text{Area}(R_i) < \varepsilon$ and $A \Delta B = (A \setminus B) \cup (B \setminus A)$
It is easy to check that $A \setminus B \subset T$ and $B \setminus A \subset S$

To estimate $\mu^*(S)$ note that

$$X = \left(\cup_{i=1}^\infty R_i\right) \cup \left(\cup_{j=1}^\infty (R'_j \setminus B)\right),$$

Since all the sets in this formula are elementary sets, using σ-subadditivity we get

$$\sum_{i=1}^\infty \text{Area}(R_i) + \sum_{j=1}^\infty \text{Area}(R'_j \setminus B) \ge \text{Area}(X) \qquad (2)$$

Subtracting (2) from (1) yields

$$\sum_{j=1}^{\infty} \text{Area}(R'_j) - \sum_{j=1}^{\infty} \text{Area}(R'_j \setminus B) = \sum_{j=1}^{\infty} \text{Area}(R'_j \cap B) < 2\varepsilon$$

which shows that $\mu^*(S) < 2\varepsilon$. Therefore

$$\mu^*(A\Delta B) \leq \mu^*(T) + \mu^*(S) < \varepsilon + 2\varepsilon = 3\varepsilon$$

Since ε is arbitrary, we conclude: $\exists B \subset X$ (elementary set) s.t. $\mu^*(A\Delta B) \leq \varepsilon$ □

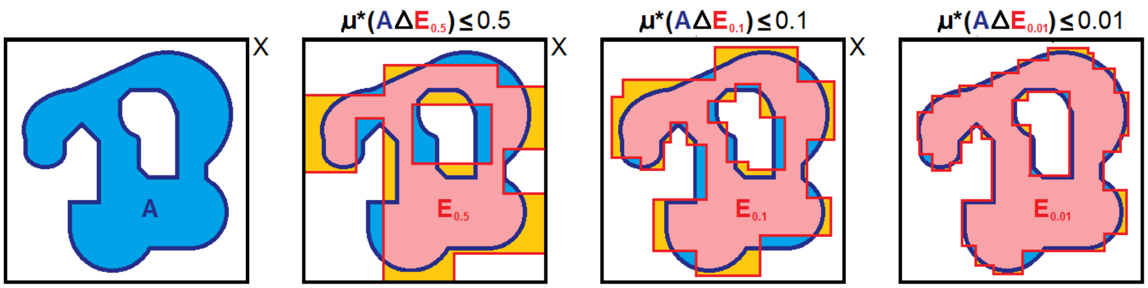


Figure 3: Visualization for Theorem 1.25

Theorem 1.26. *Finite unions, intersections and differences of measurable sets are measurable.*

Proof. Routine verification by approximation (Theorem 1.25). □

Theorem 1.27. (Additivity) *If $A_1, \ldots, A_n$ are disjoint measurable sets, then* $\mathbf{m}\big(\uplus_{i=1}^n A_i\big) = \sum_{i=1}^n \mathbf{m}(A_i)$.

Proof. Again, by approximation (Theorem 1.25). □

Theorem 1.28. *Countable unions and countable intersections of measurable sets are measurable.*

Proof. Let $A = \cup_{i=1}^{\infty} A_i$ be a countable union of measurable sets $A_1, A_2, \ldots$.
Define $A'_i = A_i \setminus \cup_{j=1}^{i-1} A_i$, then $A = \uplus_{i=1}^{\infty} A'_i$ is a disjoint union of sets that are measurable by Theorem 1.26. Moreover,

$$\sum_{i=1}^{n} \mathbf{m}(A'_i) \leq \mu^*(A) < \infty$$

so the series $\sum_{i=1}^{\infty} \mathbf{m}(A_i')$ converges.

Let $\varepsilon > 0$. Since $A_1', A_2', ...$ are disjoint sets whose total measure is finite, (their union is a subset of X, which has finite area), we can choose N so that $\sum_{i>N} \mathbf{m}(A_i') < \varepsilon$. If we denote $T = \cup_{i>N} A_i'$, then by Theorem 1.17

$$\mu^*(T) \leq \sum_{i>N} \mu^*(A_i') = \sum_{i>N} \mathbf{m}(A_i') < \varepsilon.$$

The set $A^\diamond = \cup_{i=1}^{N} A_i$ is measurable (by Theorem 1.26), so it can be approximated by an elementary set B_ε such that $\mu^*(A^\diamond \Delta B_\varepsilon) < \varepsilon$. Now

$$\mu^*(A \Delta B_\varepsilon) \leq \mu^*(A^\diamond \Delta B_\varepsilon) + \mu^*(T) < 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the set A is measurable by Approximation (Theorem 1.25).

Finally, countable intersections are complements of countable unions. □

Exercise 4. Prove that every open set $A \subset X$ is measurable (hint: represent it by a countable union of rectangles). Prove that every closed set is measurable, too.

Theorem 1.29. (σ-additivity) *If $A_1, A_2, \ldots$ are disjoint measurable sets, then* $\mathbf{m}\big(\uplus_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} \mathbf{m}(A_i)$

Proof. Note that $A = \uplus_{i=1}^{\infty} A_i$ is a measurable set by Theorem 1.28.
Thus by Theorem 1.17

$$\mathbf{m}(A) = \mu^*(A) \leq \sum_{i=1}^{\infty} \mu^*(A_i) = \sum_{i=1}^{\infty} \mathbf{m}(A_i).$$

On the other hand: $\forall n \geq 1$ we have $\uplus_{i=1}^{n} A_i \subset A$, therefore

$$\sum_{i=1}^{n} \mathbf{m}(A_i) \overset{1.27}{=} \mathbf{m}\big(\uplus_{i=1}^{n} A_i\big) = \mu^*\big(\uplus_{i=1}^{n} A_i\big) \overset{1.15}{\leq} \mu^*(A) = \mathbf{m}(A).$$

This implies $\mathbf{m}(A) = \sum_{i=1}^{\infty} \mathbf{m}(A_i)$, as claimed. □

Theorem 1.30. (Continuity - I) *Let* $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$ *be a sequence of measurable sets (called* monotonically decreasing sequence*). Then* $\lim_{n\to\infty} \mathbf{m}(A_n) = \mathbf{m}(A)$*, where* $A = \cap_{n=1}^{\infty} A_n$

Proof. Let $A_i' = A_i \setminus A_{i+1}$ for all $i \geq 1$. Obviously, $A_n = A \uplus (\uplus_{i=n}^{\infty} A_i')$

Hence by Theorem 1.29

$$\mathbf{m}(A_n) = \mathbf{m}(A) + \sum_{i=n}^{\infty} \mathbf{m}(A_i')$$

Since the series converges, its tail tends to zero as $n \to \infty$. □

Theorem 1.31. (Continuity - II) *Let* $A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots$ *be a sequence of measurable sets (called* monotonically increasing sequence*). Then* $\lim_{n\to\infty} \mathbf{m}(A_n) = \mathbf{m}(A)$*, where* $A = \cup_{n=1}^{\infty} A_n$

Proof. Let $A_1' = A_1$ and $A_i' = A_i \setminus A_{i-1}$ for all $i \geq 2$. Obviously, $A = \uplus_{i=1}^{\infty} A_i'$

Hence by Theorem 1.29

$$\mathbf{m}(A_n) = \sum_{i=1}^{n} \mathbf{m}(A_i') \quad \text{and} \quad \mathbf{m}(A) = \sum_{i=1}^{\infty} \mathbf{m}(A_i')$$

Now the sum of the series is the limit of its partial sums, thus $\mathbf{m}(A_n) \to \mathbf{m}(A)$. □

2 Lebesgue measure in $\mathbb{R}$ and $\mathbb{R}^2$: some extra facts

First we extend the Lebesgue measure from rectangle X to the entire plane $\mathbb{R}^2$:

DEFINITION 2.1. A set $A \subset \mathbb{R}^2$ is **measurable** if and only if $A \cap X$ is measurable for every rectangle $X \subset \mathbb{R}^2$. Its measure is

$$\mathbf{m}(A) = \lim_{n\to\infty} \mathbf{m}(A \cap X_n),$$

where $X_n = [-n, n] \times [-n, n]$ is a growing sequence of rectangles.

- $\mathbf{m}(A)$ does not depend on the particular growing sequence of rectangles.

Alternative definition of Lebesgue measure in $\mathbb{R}^2$.

We can pave the plane $\mathbb{R}^2$ with rectangles

$$X_{ij} = \{(x, y) \colon i \le x < i + 1, \quad j \le y < j + 1\}$$

and define

$$\mathbf{m}(A) = \sum_{i,j} \mathbf{m}(A \cap X_{ij}).$$

- $\mathbf{m}(A)$ does not depend on a particular pavement of $\mathbb{R}^2$ by rectangles
- Most theorems proved in Section 1 on rectangle X also hold on entire $\mathbb{R}^2$
- $\mathbf{m}(A)$ may be infinite (i.e. $\mathbf{m}(A) = \infty$ for some measurable sets $A \subset \mathbb{R}^2$)

Theorem 2.2. *The Lebesgue measure is translationally invariant. i.e. if $A \subset \mathbb{R}^2$ is measurable and $a \in \mathbb{R}^2$, then the set*

$$A + a = \{x + a \colon x \in A\}$$

is measurable and $\mathbf{m}(A + a) = \mathbf{m}(A)$.

Proof. If R is a rectangle, then $R+a$ is a rectangle, too, and $\text{Area}(R+a) = \text{Area}(R)$, so all our constructions are invariant under translations. $\square$

Lebesgue measure in $\mathbb{R}$.

This can be constructed in $\mathbb{R}$ by using intervals (and their length) instead of rectangles (and area). The Lebesgue measure in $\mathbb{R}$ is translationally invariant, too, i.e. if A is a measurable set and $a \in \mathbb{R}$, then the set $A + a = \{x + a \colon x \in A\}$ is also measurable and $\mathbf{m}(A + a) = \mathbf{m}(A)$.

Next we describe the collection of measurable sets in $\mathbb{R}$ and $\mathbb{R}^2$.

Definition 2.3. Let X be a set. A **σ-algebra** is an algebra of subsets of X closed under countable unions and intersections.

The following lemma presents a minimal set of conditions that we need to verify in order to show that a given collection of sets is a σ-algebra:

Lemma 2.4. *A non-empty collection $\mathfrak{M}$ of subsets of X is a σ-algebra if two conditions hold:*

(i) $A \in \mathfrak{M} \Longrightarrow A^c \in \mathfrak{M}$, *where* $A^c = X \setminus A$;

(ii) $A_1, A_2, \ldots \in \mathfrak{M} \Longrightarrow \cup_{i=1}^{\infty} A_i \in \mathfrak{M}$.

In other words, a σ-algebra is any non-empty collection of subsets of X closed under complementation and countable unions.

Proof. A straightforward verification. $\square$

Definition 2.5.

Measurable sets in $\mathbb{R}^1$ make a σ-algebra called **Lebesgue σ-algebra** (over a line)

Measurable sets in $\mathbb{R}^2$ make a σ-algebra called **Lebesgue σ-algebra** (over a plane)

Some examples of measurable sets.

Recall open sets in $\mathbb{R}$ and $\mathbb{R}^2$ are measurable. Thus closed sets are also measurable.

Any countable intersection of open sets (called G_δ in topology) is measurable.

Any countable union of closed sets (called F_σ in topology) is measurable.

In particular, the set of rational numbers in $\mathbb{R}$ is measurable. So is the set of irrational numbers.

The above list of measurable sets is far from complete. The collection of measurable sets is very rich (as we will see shortly).

The following example (The Cantor set), is an example of a measurable set with many interesting properties:

Cantor set.

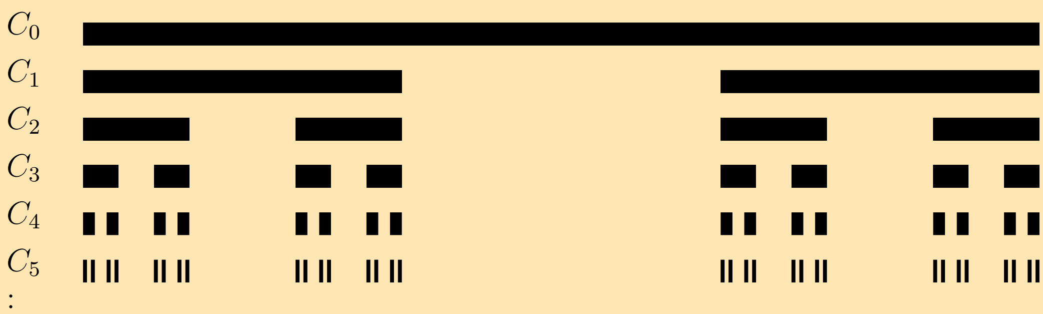
An interesting example of a measurable set is the middle-third Cantor set C. It is a subset of the unit interval $[0,1] \subset \mathbb{R}$ that has many useful properties. It is constructed beginning with closed unit interval ($C_0 = [0,1]$), removing the middle third open interval

$$C_1 = C_0 \setminus (\tfrac{1}{3}, \tfrac{2}{3}) = [0, \tfrac{1}{3}] \cup [\tfrac{2}{3}, 1],$$

then recursively removing the middle third intervals from each remaining closed interval:

$$C_2 = C_1 \setminus \left[(\tfrac{1}{9}, \tfrac{2}{9}) \cup (\tfrac{7}{9}, \tfrac{8}{9})\right], \quad \text{etc.}$$

This results in a decreasing sequence of sets $C_0 \supset C_1 \supset C_2 \supset \cdots$ shown here:



A general formula for the set C_n is

$$C_n = \frac{C_{n-1}}{3} \biguplus \left(\frac{2}{3} + \frac{C_{n-1}}{3}\right),$$

so that C_n is a disjoint union of two copies of C_{n-1} compressed by factor $\frac{1}{3}$.

The Cantor set is the limit of this procedure, it is defined by

$$C = \lim_{n\to\infty} C_n = \cap_{n=0}^{\infty} C_n$$

We record the following properties of C_n's and C:

- At each step we remove open intervals, so each C_n is a closed set
- The Cantor set C is a closed set (being an intersection of closed sets)
- Each C_n is a linear elementary set (a finite union of intervals)
- The length of C_n is decreased by factor $\frac{2}{3}$ at each step: $|C_n| = \frac{2}{3}|C_{n-1}|$
- The length of C_n is, by induction, $|C_n| = (\frac{2}{3})^n$
- By Continuity-I (Theorem 1.30), $\mathbf{m}(C) = \lim_{n\to\infty} \mathbf{m}(C_n) = \lim_{n\to\infty} (\frac{2}{3})^n = 0$

Theorem 2.6. *If $A \subset X$ is measurable and $\mathbf{m}(A) = 0$, then every subset $A' \subset A$ is measurable and $\mathbf{m}(A') = 0$.*

Proof. By Theorem 1.15, $\mu^*(A') \le \mu^*(A) = \mathbf{m}(A) = 0$, hence $\mu^*(A') = 0$. Now by Theorem 1.20, $0 \le \mu_*(A') \le \mu^*(A')$=0, hence $\mu_*(A') = 0$. □

Ternary description of the Cantor set.

The ternary number system has base 3 and uses three digits: 0, 1, and 2. Points $x \in (0,1)$ are described in the ternary system by $x = .d_1d_2d_3\cdots$ where d_i's are ternary digits (0, 1, 2). For example,

$$\tfrac{1}{3} = (.1000000\cdots)_{\text{ternary}} \qquad \tfrac{1}{2} = (.1111111\cdots)_{\text{ternary}}$$

If $x \in (0, \frac{1}{3})$, then $d_1 = 0$; if $x \in (\frac{1}{3}, \frac{2}{3})$, then $d_1 = 1$; if $x \in (\frac{2}{3}, 1)$, then $d_1 = 2$.

To use this in construction of the Cantor set, Step 1 would be to remove all numbers with $d_1 = 1$. Similarly, Step 2 would remove all the numbers with $d_2 = 1$, etc. Repeating these steps this leads to the removal of <u>all</u> numbers $x \in (0,1)$ whose ternary representation $x = .d_1d_2d_3\cdots$ contains at least one 1.

Thus, the Cantor set C consists exactly of the numbers $x \in (0,1)$ whose ternary representation $x = .d_1d_2d_3\cdots$ do not include 1's, i.e., consist of 0's and 2's only.

Note: the number $x = \frac{1}{3}$ does belong to the Cantor set, even though its ternary representation $(.1000000\cdots)_{\text{ternary}}$ given above contains a 1. Why? Because this number has <u>two</u> ternary representations; the other (which has no 1's) is

$$\tfrac{1}{3} = (.0222222222\cdots)_{\text{ternary}}.$$

From this, the description of the Cantor set should be as follows: $x \in C$ if and only if x can be represented in the ternary system by a sequence of zeros and twos.

Which numbers $x \in (0,1)$ have multiple ternary representations? Only rational numbers $x = \frac{m}{3^n}$ where $n \ge 1$ and $1 \le m \le 3^n - 1$. These number have exactly two ternary representations, one ends with zeros and the other ends with twos.

A question arises: how many numbers are in the Cantor set C? Obviously, the endpoints of the intervals making C_n will remain in C. Those are rational numbers $x = \frac{m}{3^n}$ where $n \ge 1$ and $1 \le m \le 3^n - 1$. But there are many more numbers in C, as we will see later.

Recall some basic definitions and facts from set theory:

Cardinality.

Two sets A and B have the same cardinality if there is a bijection $\varphi\colon A \leftrightarrow B$
We write $\mathrm{card}(A) = \mathrm{card}(B)$.

A set A has cardinality smaller than that of B, $\mathrm{card}(A) < \mathrm{card}(B)$, if there is an injection $\varphi\colon A \to B$ but not vice versa.

- For finite sets A and B, the relation $\mathrm{card}(A) = \mathrm{card}(B)$ holds if and only if they have the same number of elements.
- 'Same cardinality' is an equivalence relations. Thus all sets of the same cardinality make an (equivalence) class.

The following two theorems are cited without proof:

Theorem 2.7. *For any two sets A and B there is either an injection $A \to B$, or an injection $B \to A$, or both.*

Theorem 2.8. (Cantor-Bernstein-Schroeder)
If there is an injection $A \to B$ and *an injection $B \to A$, then there is a bijection $A \leftrightarrow B$, i.e.,* $\mathrm{card}(A) = \mathrm{card}(B)$.

- Thus, for any two sets A and B there are three possibilities: .
  $\mathrm{card}(A) = \mathrm{card}(B)$ or $\mathrm{card}(A) > \mathrm{card}(B)$ or $\mathrm{card}(A) < \mathrm{card}(B)$
- Recall: $\mathrm{card}(\mathbb{N}) = \mathrm{card}(\mathbb{Z}) = \mathrm{card}(\mathbb{Q})$ and $\mathrm{card}(\mathbb{N}) < \mathrm{card}(\mathbb{R})$

Finite and countable cardinalities.

For a finite set A of n elements, we simply put $\mathrm{card}(A) = n$.
For the set of integers $\mathbb{N}$, we put $\mathrm{card}(\mathbb{N}) = \aleph_0$ ('aleph-null').
All countable sets have cardinality $\aleph_0$. In particular: $\mathrm{card}(\mathbb{Z}) = \mathrm{card}(\mathbb{Q}) = \aleph_0$.

- Any countable union of countable sets is countable.

Continuum.

The cardinality of $\mathbb{R}$ is denoted by $\mathfrak{C}$ and is called *continuum.* Thus,
 $\mathrm{card}(I) = \mathfrak{C}$ for any interval $I \subset \mathbb{R}$ of positive length.
 $\mathrm{card}(\mathbb{R}^n) = \mathfrak{C}$ for every $n \geq 1$
Note that $\aleph_0 < \mathfrak{C}$

Continuum hypothesis.

States that there is no set A such that $\aleph_0 < \operatorname{card}(A) < \mathfrak{C}$. In particular, it implies that every subset of $\mathbb{R}$ is either finite or countable or has the same cardinality as $\mathbb{R}$ itself. The continuum hypothesis cannot be proved or disproved if one uses standard mathematical axioms. Thus it may (or may not) be adopted as an independent axiom. We do not assume it for the purposes of our course.

The Power set.

For any set A, the *power set* of A, denoted by 2^A, is the set of all subsets of A.

Note: if $\operatorname{card}(A) = n < \infty$, then $\operatorname{card}(2^A) = 2^n$.

Theorem 2.9. (Cantor)
For any set A we have $\operatorname{card}(A) < \operatorname{card}(2^A)$.

Proof. It is enough to show that no function $f\colon A \to 2^A$ can be surjective. That is, given a function $f\colon A \to 2^A$, we need to prove the existence of at least one subset $B \subset A$ that $B \neq f(a)$ for any $a \in A$. Such a subset is given by the following construction:

$$B = \{b \in A\colon b \notin f(b)\}.$$

If $B = f(a)$ for some $a \in A$, then both relations $a \in B$ and $a \notin B$ would contradict the definition of B. $\square$

Generalized continuum hypothesis.

States that for any set A there is no set B such that $\operatorname{card}(A) < \operatorname{card}(B) < \operatorname{card}(2^A)$. This implies that cardinalities make a simple sequence:

$\aleph_0$ (the set $\mathbb{N}$ and other countable sets),

$\aleph_1$ (the set $2^{\mathbb{N}}$),

$\aleph_2$ (the set $2^{2^{\mathbb{N}}}$), etc.

This hypothesis cannot be proved or disproved either. We do not assume it for the purposes of our course.

Equivalence of $\mathbb{R}$ and $2^{\mathbb{N}}$.

The set $2^{\mathbb{N}}$ can be identified with the set of all infinite sequences of zeros and ones. Indeed, for any subset $A \subset \mathbb{N}$ we construct a unique sequence $a_1, a_2, \ldots$ such that $a_n = 1$ if $n \in A$ and $a_n = 0$ if $n \notin A$. For example, the set of all even numbers is represented by the sequence

$$\{0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ \cdots\}.$$

The set of all prime numbers is represented by the sequence

$$\{1\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ \cdots\}.$$

On the other hand, the sequences of zeros and ones code real numbers in the interval $[0, 1]$, because every $x \in [0, 1]$ has a binary representation $x = .a_1 a_2 \ldots$, where a_n's are zeros and ones. For example,

$$\tfrac{3}{4} = (.11000000\cdots)_{\text{binary}}, \qquad \tfrac{1}{3} = (.01010101010\cdots)_{\text{binary}}.$$

Thus the set $2^{\mathbb{N}}$ can be bijectively mapped onto $[0, 1]$, hence it its cardinality is $\mathfrak{C}$.

EXERCISE 5. Let $\mathfrak{U}$ denote the set of all open subsets $U \subset \mathbb{R}$. Prove that $\mathrm{card}(\mathfrak{U}) = \mathfrak{C}$. Do the same for open sets in $\mathbb{R}^2$. Hint: use a countable basis for the respective topology.

Cardinality of the Cantor set.

Recall that the Cantor set consists of numbers $x \in [0, 1]$ whose ternary representation $x = 0.b_1 b_2 \ldots$ does not contain ones, i.e., it consists of zeros and twos only. Thus the Cantor set can be bijectively mapped onto $2^{\mathbb{N}}$, hence its cardinality is $\mathfrak{C}$.

Recall that the Cantor set has Lebesgue measure zero, $\mathbf{m}(C) = 0$. This suggests that the Cantor set is "tiny" or "slim". But it has the same cardinality as $\mathbb{R}$.

EXERCISE 6. Let $\mathfrak{L}$ denote the set of all Lebesgue measurable sets (in $\mathbb{R}$). Prove that $\mathrm{card}(\mathfrak{L}) > \mathfrak{C}$, in fact $\mathrm{card}(\mathfrak{L}) = \mathrm{card}(2^{\mathbb{R}})$. (Hint: use the Cantor set and the Cantor theorem).

The collection of all measurable sets has the same cardinality as the collection of all subsets of $\mathbb{R}$. It is not clear yet if there are any non-measurable sets at all.

The following theorem states there exist such sets:

Theorem 2.10. *There exists a non-measurable set $V \subset \mathbb{R}$.*

Proof. The set V is called Vitali set. Its construction is not so simple.

We say that two real numbers $x, y \in \mathbb{R}$ are rationally equivalent if $x - y \in \mathbb{Q}$, i.e., they differ by a rational number. One can check directly that this is an equivalence relation. Each equivalence class is countable and can be thought of as a "shifted copy" of the set of rational numbers. The collection of the classes is uncountable, its cardinality is $\mathfrak{C}$.

Each equivalence class is dense, so they all intersect the unit interval $[0, 1]$. We want to choose one (arbitrary) representative from each equivalence class in $[0, 1]$. The chosen numbers make a set (a subset of $[0, 1]$). This is **Vitali set**, denoted by V. Note that $V \subset [0, 1]$, and for each equivalence class $C \subset \mathbb{R}$ the intersection $V \cap C$ consists of a single point. Note that for any $x, y \in V$ the difference $x - y$ is irrational (otherwise these two points would belong to the same class).

Next we argue that the Vitali set cannot be measurable. Let $r_1, r_2, \ldots$ denote all rational number in $[-1, 1]$. From the construction of V it follows that the translated sets $V_i = V + r_i$ are pairwise disjoint. Further note that

$$[0, 1] \subset \uplus_{i=1}^{\infty} V_i \subset [-1, 2].$$

(To see the first inclusion, consider any real number $x \in [0, 1]$ and let v be the representative in V for the equivalence class containing x; then $x - v = r$ for some rational number $r \in [-1, 1]$.)

Now suppose V is measurable. Its measure $\mathbf{m}(V)$ is a nonnegative real number. All the "shifted" copies of V, i.e., the sets $V_i = V + r_i$, are measurable, too, and their measure is the same as that of V, i.e., $\mathbf{m}(V_i) = \mathbf{m}(V)$, due to the translation invariance. This implies

$$\mathbf{m}([0, 1]) = 1 \leq \sum_{i=1}^{\infty} \mathbf{m}(V_i) \leq 3 = \mathbf{m}([-1, 2]).$$

But the infinite sum in the middle can only be zero (if $\mathbf{m}(V) = 0$) or infinity (if $\mathbf{m}(V) > 0$), a contradiction. $\square$

The construction of the Vitali set V looks simple enough, but it touches upon a subtle and controversial issue in mathematical logic – Axiom of Choice. Indeed, how can we choose exactly one representative from each equivalence class? Is there a rule? An algorithm? There are uncountably many classes out there, so no formal procedure can handle all of them.

Axiom of Choice.

Asserts that such a selection is always possible. This principle does not follow from other, standard logical axioms, so it has to be adopted as a separate one. Axiom of Choice was formally introduced by Zermelo in 1904. Although originally controversial, it is now used by most mathematicians without reservation, and it is included in the standard form of axiomatic set theory. However, there are branches of mathematics where the Axiom of Choice is avoided.

If we do not adopt the Axiom of Choice, there would be no way to construct non-measurable sets or even prove their existence. Then we could just suppose that all sets $A \subset \mathbb{R}$ are measurable...

For the purposes of this course we adopt the Axiom of Choice and hence admit the existence of non-measurable sets.

Lebesgue measure in $\mathbb{R}^k$, $k \geq 3$.

We have constructed the Lebesgue measure in $\mathbb{R}$ and $\mathbb{R}^2$. Our construction easily extends to $\mathbb{R}^k$ for $k \geq 3$. In $\mathbb{R}^3$, for example, the Lebesgue measure generalizes **volume**, and instead of intervals or rectangles, we have to use rectangular boxes. All the basic facts for the Lebesgue measure $\mathbf{m}$ remain valid in spaces $\mathbb{R}^k$, $k \geq 3$.

Interestingly, though, the are new examples of non-measurable sets in $\mathbb{R}^3$ due to the striking fact known as **Banach–Tarski paradox**:

Theorem 2.11. (Banach–Tarski)
Given a solid ball in $\mathbb{R}^3$, there exists a decomposition of the ball into a finite number of non-overlapping pieces (i.e., subsets), which can then be put back together in a different way to yield two identical copies of the original ball. The reassembly process involves only moving the pieces around and rotating them, without changing their shape or size.

Figure 4: Banach–Tarski paradox illustrated

The Banach–Tarski is often stated, colloquially, as "a pea can be chopped up and reassembled into the Sun".

Now suppose the Lebesgue measure $\mathbf{m}$ in $\mathbb{R}^3$ is not only translation invariant but also rotation invariant (which is a reasonable assumption). Then if the above pieces of the ball were all measurable, we would immediately arrive at a contradiction with the additivity of the measure, because the volume of the original ball *doubles* after the reassembly.

To resolve the contradiction, we either have to assume that the Lebesgue measure in $\mathbb{R}^3$ is *not* rotation invariant or admit that some of the above pieces of the ball are not measurable. Thus the existence of non-measurable sets becomes a more pressing issue in $\mathbb{R}^3$ than it was in $\mathbb{R}$.

For the purposes of this course we assume that the Lebesgue measure $\mathbf{m}$ is rotation invariant in both $\mathbb{R}^2$ and $\mathbb{R}^3$, though we will hardly need this.

3 General measures

Definition 3.1. A set X with a σ-algebra $\mathfrak{M}$ of its subsets is called a **measurable space**. Sets $A \in \mathfrak{M}$ are said to be **measurable**.

Example 3. For any set X, there are two trivial σ-algebras. One is *minimal*, it consists of the sets X and $\emptyset$ only. The other is *maximal*, it contains all the subsets of X. The latter is denoted by 2^X.

Definition 3.2. Let $(X, \mathfrak{M})$ be a measurable space. A **measure** is a function μ, defined on $\mathfrak{M}$, whose range is $[0, \infty]$ and which is σ-additive. The latter means that for any sequence of disjoint measurable sets $A_1, A_2, \ldots \in \mathfrak{M}$ we have

$$\mu\big(\uplus_{n=1}^{\infty} A_n\big) = \sum_{n=1}^{\infty} \mu(A_n).$$

Note that $\uplus_{n=1}^{\infty} A_n \in \mathfrak{M}$ because $\mathfrak{M}$ is a σ-algebra.
To avoid trivialities, we shall also assume that $\mu(A) < \infty$ for at least one $A \in \mathfrak{M}$.
A measurable space with a measure is called a **measure space**.

Proposition 3.3. *Let $(X, \mathfrak{M}, \mu)$ be a measure space. Then*

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

(b) *For any finite collection of disjoint measurable sets $A_1, \ldots, A_n \in \mathfrak{M}$ and $A = \uplus_{i=1}^{n} A_i$ we have $\mu(A) = \sum_{i=1}^{n} \mu(A_i)$.*

(c) *For every measurable sets $A \subset B$ we have $\mu(A) \leq \mu(B)$.*

Proof.
(a): Let $A \in \mathfrak{M}$ be such that $\mu(A) < \infty$. Then the sets $A_1 = A$, $A_2 = \emptyset$, $A_3 = \emptyset, \ldots$ are disjoint and $\uplus_{n=1}^{\infty} A_n = A$. By the σ-additivity

$$\mu(A) = \mu(A) + \mu(\emptyset) + \mu(\emptyset) + \cdots$$

which implies $\mu(\emptyset) = 0$. Note: without assuming that $\exists A \in \mathfrak{M}\colon \mu(A) < \infty$ the claim (a) is false: there exists a (trivial) measure μ such that $\forall A \in \mathfrak{M}\colon \mu(A) = \infty$.

(b): Given $A_1, \ldots, A_n \in \mathfrak{M}$, define $A_{n+1} = A_{n+2} = \cdots = \emptyset$, then by the σ-additivity we have

$$\mu\big(\uplus_{i=1}^{n} A_i\big) = \mu\big(\uplus_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} \mu(A_i) = \sum_{i=1}^{n} \mu(A_i)$$

the last identity is based on the fact $\mu(\emptyset) = 0$ proven in (a).

(c): Due to (b) we have $\mu(B) = \mu(A) + \mu(B \setminus A)$ and $\mu(B \setminus A) \geq 0$. $\square$

Exercise 7. Let $X = \{1, 2, 3\}$. Construct all σ-algebras of X.

Finite and countable spaces.

If X is finite or countable, we will always use the maximal σ-algebra 2^X. In that case for any measure μ on $(X, 2^X)$

$$A = \{x_1, x_2, \dots\} \quad \Longrightarrow \quad \mu(A) = \mu(\{x_1\}) + \mu(\{x_2\}) + \cdots$$

Thus any measure μ on $(X, 2^X)$ is determined by its values $\mu(\{x\})$ on one-point sets $\{x\}, x \in X$. One-point sets are also called **singletons**.

Restriction of a measure.

Let $E \in \mathfrak{M}$ be a measurable set. One can check, by direct inspection, that the collection

$$\mathfrak{M}_E = \{A \cap E \colon A \in \mathfrak{M}\}$$

is a σ-algebra over E. Note that $\mathfrak{M}_E \subset \mathfrak{M}$, thus $\mu(B)$ is defined for every $B \in \mathfrak{M}_E$. It is now easy to see that the restriction of μ to $\mathfrak{M}_E$ is a measure.

Theorem 3.4. *Let $\{\mathfrak{M}_\alpha\}$ be an arbitrary collection of σ-algebras of a set X. Then their intersection $\cap_\alpha \mathfrak{M}_\alpha$ is a σ-algebra of X as well.*

Proof. Direct inspection. Note that the collection of σ-algebras here may be finite, countable, or uncountable; its cardinality is not essential. □

Theorem 3.5. *Let $\mathfrak{G}$ be any collection of subsets of X. Then there exists a unique σ-algebra $\mathfrak{M}^* \supset \mathfrak{G}$ such that for any other σ-algebra $\mathfrak{M} \supset \mathfrak{G}$ we have $\mathfrak{M}^* \subset \mathfrak{M}$. (In other words, $\mathfrak{M}^*$ is the minimal σ-algebra containing $\mathfrak{G}$.)*

Proof. The σ-algebra $\mathfrak{M}^*$ is the intersection of all σ-algebras containing $\mathfrak{G}$. □

Definition 3.6. We say that the minimal σ-algebra $\mathfrak{M}^*$ containing the given collection $\mathfrak{G}$ is **generated** by $\mathfrak{G}$. We also denote it by $\mathfrak{M}(\mathfrak{G})$.

Exercise 8. Let $X = [0, 1]$ and $\mathfrak{G}$ consist of all one-point sets, i.e. $\mathfrak{G} = \big\{\{x\}, x \in X\big\}$. Describe the σ-algebra $\mathfrak{M}(\mathfrak{G})$.

It is interesting to compare the notion of σ-algebra with topology. The standard topology in $\mathbb{R}$ is not a σ-algebra. On the other hand, the σ-algebra in the previous exercise is not a topology.

DEFINITION 3.7. Let X be a topological space. The σ-algebra $\mathfrak{M}$ generated by the collection of all open subsets $U \subset X$ is called the **Borel** σ**-algebra**. Its members are called **Borel sets**.

- We can take complements, so all closed sets are Borel sets.
- We can take countable intersections, so all G_δ sets are Borel sets.
- We can take countable unions, so all F_σ sets are Borel sets.

EXAMPLE 4. In $\mathbb{R}$, every countable set is Borel. The Cantor set is Borel. In fact, virtually any set that can be precisely described is Borel. It is hard to find any specific non-Borel set.

EXERCISE 9. Show that the Borel σ-algebra in $\mathbb{R}$ is generated by the collection of all intervals (r_1, r_2) with rational endpoints $r_1, r_2 \in \mathbb{Q}$.

Borel σ-algebra in $\mathbb{R}$.

It contains many more sets than just open, closed, G_δ, and F_σ sets. If we take countable unions of G_δ sets, we will get new sets that are also Borel. If we take countable intersections of F_σ sets, we will get new sets that are also Borel. Then we can take countable unions and countable intersections of those new sets, and get some more new sets, all of which will be Borel, too. This process will never stop. Its full description requires the so called *transfinite recursion* which is beyond the scope of this course.

Recall that the collection of open sets in $\mathbb{R}$ has cardinality $\mathfrak{C}$. So does the collection of closed sets. Constructing new sets by countable unions and intersections will not increase the cardinality of the collection, so the collection of all G_δ and F_σ sets still has cardinality $\mathfrak{C}$. Further steps in the above process will not increase the cardinality of the collection either. This leads to the following theorem (its formal proof is omitted):

Theorem 3.8. *The cardinality of the Borel σ-algebra of $\mathbb{R}$ is $\mathfrak{C}$ (i.e. continuum).*

EXERCISE 10. Show that every Borel set in $\mathbb{R}$ is Lebesgue measurable, but not vice versa.

EXERCISE 11. [Bonus] Does there exist an infinite σ-algebra which has only countably many members?

DEFINITION 3.9. Let $(X, \mathfrak{M})$ be a measurable space. For any $X \in \mathfrak{M}$, define $\mu(A) = \infty$ if A is an infinite set, and $\mu(A) = \text{card}(A)$ if A is finite. Then μ is called **counting measure**.

DEFINITION 3.10. Let $(X, \mathfrak{M})$ be a measurable space and $x \in X$. The measure δ_x defined by

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{else} \end{cases}$$

is called the **delta-measure** or the **Dirac measure** (concentrated at x).

Theorem 3.11. (σ-subadditivity) *Let $A_1, A_2, \ldots \in \mathfrak{M}$ be measurable sets. Then*

$$\mu\left(\cup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i)$$

Proof. Define $A_1' = A_1$ and $A_i' = A_i \setminus (A_1 \cup \cdots \cup A_{i-1})$ for $i \geq 2$. Note the following:

- A_i' are disjoint sets;
- $A_i' \subset A_i$ for each $i \geq 1$; therefore $\mu(A_i') \leq \mu(A_i)$ by Proposition 3.3(c);
- $\cup_{i=1}^{\infty} A_i = \cup_{i=1}^{\infty} A_i'$.

Now

$$\mu\left(\cup_{i=1}^{\infty} A_i\right) = \mu\left(\cup_{i=1}^{\infty} A_i'\right) = \sum_{i=1}^{\infty} \mu(A_i') \leq \sum_{i=1}^{\infty} \mu(A_i)$$

□

Theorem 3.12. (Continuity - I) *Let $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$ be a sequence of measurable sets, and $\mu(A_1) < \infty$. Then $\lim_{n\to\infty} \mu(A_n) = \mu(A)$, where $A = \cap_{n=1}^{\infty} A_n$.*

Proof. Similar to Theorem 1.30. □

Theorem 3.13. (Continuity - II) *Let $A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots$ be a sequence of measurable sets. Then $\lim_{n\to\infty} \mu(A_n) = \mu(A)$, where $A = \cup_{n=1}^{\infty} A_n$.*

Proof. Similar to Theorem 1.31. □

EXERCISE 12. Show that the assumption $\mu(A_1) < \infty$ in Theorem 3.12 is indispensable. Hint: consider the counting measure on $\mathbb{N}$ and take sets $A_n = \{n, n+1, \ldots\}$.

In many cases σ-algebras tend to be rather large, an explicit definition (or description) of $\mu(A)$ for all $A \in \mathfrak{M}$ is often an impossible task. It is common to define $\mu(A)$ on a smaller collection of sets, $\mathfrak{E}$, and extend it to $\mathfrak{M}(\mathfrak{E})$ automatically by referring to general theorems.

Definition 3.14. A **semi-algebra** is a nonempty collection $\mathfrak{E}$ of subsets of X with two properties: (i) it is closed under intersections; i.e. if $A, B \in \mathfrak{E}$, then $A \cap B \in \mathfrak{E}$; and (ii) if $A \in \mathfrak{E}$, then $A^c = \cup_{i=1}^n A_i$, where each $A_i \in \mathfrak{E}$ and $A_1, \ldots, A_n$ are pairwise disjoint subsets of X.

Example 5. The collection of all finite and infinite intervals in $\mathbb{R}$ (cf. Example 2) make a semi-algebra.

Exercise 13. Let $X \subset \mathbb{R}^2$ be a rectangle. Verify that the collection of all subrectangles $R \subset X$ is a semi-algebra.

Theorem 3.15. (Extension) *Let $\mathfrak{E}$ be a semi-algebra of X. Let ν be a function on $\mathfrak{E}$, whose range is $[0, \infty)$ and which is σ-additive, i.e. for any $A \in \mathfrak{E}$ such that $A = \uplus_{i=1}^\infty A_i$ for some $A_i \in \mathfrak{E}$, we have $\mu(A) = \sum_{i=1}^\infty \mu(A_i)$. Then there is a unique measure μ on the σ-algebra $\mathfrak{M}(\mathfrak{E})$ that agrees with ν on $\mathfrak{E}$, i.e. $\mu(A) = \nu(A)$ for all $A \in \mathfrak{E}$.*

We accept this theorem without proof. Its proof is basically the repetition of our construction of the Lebesgue measure in $\mathbb{R}^2$.

Corollary 3.16. *Let $X \subset \mathbb{R}^k$ be a rectangular box. There is a unique measure μ on the Borel σ-algebra over X such that for every rectangular box $R \subset X$ we have $\mu(R) = \text{Volume}(R)$.*

The following useful theorem is also given without proof:

Theorem 3.17. *Let $(X, \mathfrak{M})$ be a measurable space and $\mathfrak{E}$ a collection of subsets of X that generates $\mathfrak{M}$, i.e. such that $\mathfrak{M}(\mathfrak{E}) = \mathfrak{M}$. Suppose two measures, μ_1 and μ_2, agree on $\mathfrak{E}$, i.e. $\mu_1(A) = \mu_2(A)$ for all $A \in \mathfrak{E}$, and $\mu_1(X) = \mu_2(X)$. Then $\mu_1 = \mu_2$.*

- For any measure space $(X, \mathfrak{M}, \mu)$ and a real number $c > 0$ we can define a new measure ν on $\mathfrak{M}$ by $\nu(A) = c\mu(A)$ for all $A \in \mathfrak{M}$. (The verification is straightforward.)

Theorem 3.18. *Let μ be a translation invariant measure defined on the Borel σ-algebra over $\mathbb{R}$. Assume that $\mu(I) < \infty$ for at least one interval $I \neq \emptyset$. Then there exists a constant $c \geq 0$ such that $\mu(E) = c\mathbf{m}(E)$ for every Borel set E; here $\mathbf{m}$ is the Lebesgue measure.*

Proof. Due to the translation invariance, all singletons have the same measure, i.e., $\mu(\{x\}) = \mu(\{y\})$ for all $x, y \in \mathbb{R}$. If, $\mu(\{x\}) > 0$ for any (and then for all) $x \in \mathbb{R}$, then the measure of every infinite set would be infinite, which contradicts the assumption $\mu(I) < \infty$. Thus all singletons have measure zero, hence for every $a < b$ we have $\mu(a, b) = \mu([a, b)) = \mu((a, b]) = \mu([a, b])$.

Now let $\mu(I) < \infty$ for a nonempty interval (a, b). Note that $\mathbb{R} = \biguplus_{n=-\infty}^{\infty}(a + n|I|, b + n|I|]$. If $\mu(I) = 0$, then $\mu(\mathbb{R}) = 0$, hence $\mu(A) = 0 \cdot \mathbf{m}(A)$ for any Borel set, and the theorem follows with $c = 0$.

If $\mu(I) > 0$, then $\mu(\mathbb{R}) = \infty$. In this case we put $c = \mu(I)/|I|$. Divide I into $k \geq 2$ intervals $\{I_i\}$ of equal length. They also have equal μ measure due to the translation invariance, hence

$$\mu(I_i) = \mu(I)/k = c|I|/k = c|I_i|.$$

Next for any interval $J \subset \mathbb{R}$ of length $|J| = \frac{m}{k}|I|$, we can represent $J = \biguplus_{j=1}^{m} J_j$ where $|J_j| = |I|/k$ and obtain

$$\mu(J) = m\mu(J_1) = mc|J_1| = c|J|.$$

Thus the measure $c^{-1}\mu$ agrees with the Lebesgue measure on all intervals with rational lengths. Lastly we use the result of Exercise 9 and Theorem 3.17. $\square$

Exercise 14. Extend this theorem to $\mathbb{R}^2$: show that if μ is a translation invariant measure defined on the Borel σ-algebra over $\mathbb{R}^2$ such that $\mu(R) < \infty$ for at least one rectangle $R \neq \emptyset$, then there exists a constant $c \geq 0$ such that $\mu(E) = c\mathbf{m}(E)$ for every Borel set $E \subset \mathbb{R}^2$.

- The above fact remains valid in $\mathbb{R}^k$:

Corollary 3.19. *If μ is a translation invariant measure defined on the Borel σ-algebra over $\mathbb{R}^k$ such that $\mu(R) < \infty$ for at least one rectangular box $R \neq \emptyset$, then there exists a constant $c \geq 0$ such that $\mu(E) = c\mathbf{m}(E)$ for every Borel set $E \subset \mathbb{R}^k$.*

Lastly, given a measure μ on a σ-algebra $\mathfrak{M}$, it is often convenient to complete it in the way we constructed the Lebesgue measure.

DEFINITION 3.20. A measure μ on a measurable space $(X, \mathfrak{M})$ is said to be **complete** if every subset of any set of measure zero is measurable, i.e., if for any set $A \in \mathfrak{M}$ such that $\mu(A) = 0$ and any subset $B \subset A$ we have $B \in \mathfrak{M}$ (in this case obviously $\mu(B) = 0$).

DEFINITION 3.21. Sets of measure zero are called **null sets**. Their complements are called **full measure sets**.

- A set A is of full measure iff $\mu(A^c) = 0$.
- It is incorrect to say that A is of full measure iff $\mu(A) = \mu(X)$. (This statement is true only if $\mu(X) < \infty$.)

Theorem 3.22. (Completion) *Let $(X, \mathfrak{M}, \mu)$ be a measure space. Then*

$$\bar{\mathfrak{M}} = \{A \cup E \colon A \in \mathfrak{M}, \quad E \subset N \text{ for some null set } N\}$$

is a σ-algebra. For every $B = A \cup E$ as above define $\bar{\mu}(B) = \mu(A)$. Then $\bar{\mu}$ is a complete measure on $(X, \bar{\mathfrak{M}})$. Moreover, if μ^* is another complete measure that agrees with μ on $\mathfrak{M}$, then μ^* coincides with $\bar{\mu}$ on $\bar{\mathfrak{M}}$.

Proof. Basically, the proof goes by direct inspection. □

DEFINITION 3.23. $\bar{\mu}$ is called the **completion** of μ.

- The completion of the measure constructed in Corollary 3.16 is exactly the Lebesgue measure on $X \subset \mathbb{R}^k$.

Corollary 3.24. *For every Lebesgue measurable set $A \subset \mathbb{R}^k$ there exists a Borel measurable set $B \subset \mathbb{R}^k$ and a Lebesgue null set $N \subset \mathbb{R}^k$ such that $A = B \cup N$.*

- The union $B \cup N$ can be easily made disjoint. Indeed, by Theorem 3.22 there is a Borel null set $N_0 \supset N$. Then we define $B_1 = B \setminus N_0$ and $N_1 = (N_0 \cap B) \cup N$. Now we have $A = B \cup N = B_1 \uplus N_1$, where B_1 is a Borel set and N_1 is a Lebesgue null set.

4 Measurable functions

Recall: given two topological spaces X and Y, a function $f\colon X \to Y$ is said to be continuous iff for any open set $V \subset Y$ its preimage $f^{-1}(V) \subset X$ is open, too.

• The **preimage** (also called **inverse image**) is defined by

$$f^{-1}(V) = \{x \in X\colon f(x) \in V\}.$$

Definition 4.1. Let $(X, \mathfrak{M})$ be a measurable space and Y a topological space. A function $f\colon X \to Y$ is said to be a **measurable function** iff for any open set $V \subset Y$ its preimage is measurable, i.e. $f^{-1}(V) \in \mathfrak{M}$.

• Most interesting functions for us are real-valued ($\mathbb{R}$) and complex-valued ($\mathbb{C}$).

Theorem 4.2. *Let $(X, \mathfrak{M})$ be a measurable space and Y, Z topological spaces. If $f\colon X \to Y$ is measurable and $g\colon Y \to Z$ is continuous, then their composition $g \circ f\colon X \to Z$ is measurable.*

Proof. For any open set $V \subset Z$ the set $g^{-1}(V) \subset Y$ is open, hence the set $f^{-1}(g^{-1}(V)) = (g \circ f)^{-1}(V) \subset X$ is measurable (i.e., belongs in $\mathfrak{M}$). □

Recall that in any topological space, the Borel σ-algebra is generated by open sets.

Definition 4.3. Let X, Y be topological spaces. A function $f\colon X \to Y$ is said to be a **Borel function** iff for any open set $V \subset Y$ its preimage $f^{-1}(V) \subset X$ is a Borel set.

Proposition 4.4. *Continuous functions are Borel functions.*

• There are many Borel functions that are not continuous.

Definition 4.5. Given a subset $A \subset X$ the function

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{else} \end{cases}$$

is called the **characteristic function** of A (or the **indicator** of A).

Exercise 15. Prove that $A \in \mathfrak{M}$ if and only if χ_A is measurable.

Example 6. Let $X = \mathbb{R}$ with $\mathfrak{M}$ being Borel σ-algebra. The function $f = \chi_{\mathbb{Q}}$ (the indicator of the set of rational numbers) is known as **Dirichlet function**. It is discontinuous ar every point $x \in \mathbb{R}$. At the same time it is a Borel function.

In fact, virtually every function that can be precisely described is Borel. It is hard to find any specific non-Borel function.

Theorem 4.6. *Let $(X, \mathfrak{M})$ be a measurable space and $f\colon X \to Y$ a function. Then*

(a) $\mathfrak{G} = \{E \subset Y\colon f^{-1}(E) \in \mathfrak{M}\}$ *is a σ-algebra in Y;*

(b) *if Y is a topological space and f measurable, then $f^{-1}(E) \in \mathfrak{M}$ for any Borel set $E \subset Y$;*

(c) *If Y and Z are topological spaces, $f\colon X \to Y$ is measurable and $g\colon Y \to Z$ is a Borel function, then $g \circ f\colon X \to Z$ is measurable.*

Proof.
First, note that set-theoretic operations are preserved under inverse functions. That is, if $f\colon X \to Y$ is a function, then for any sets $A, B \in Y$ we have

$$\begin{aligned} f^{-1}(A \cup B) &= f^{-1}(A) \cup f^{-1}(B) \\ f^{-1}(A \cap B) &= f^{-1}(A) \cap f^{-1}(B) \\ f^{-1}(A \setminus B) &= f^{-1}(A) \setminus f^{-1}(B) \\ f^{-1}(A^c) &= (f^{-1}(A))^c. \end{aligned}$$

Similar identities hold for countable unions and intersections:

$$\begin{aligned} f^{-1}\left(\cup_{n=1}^{\infty} A_n\right) &= \cup_{n=1}^{\infty} f^{-1}(A_n) \\ f^{-1}\left(\cap_{n=1}^{\infty} A_n\right) &= \cap_{n=1}^{\infty} f^{-1}(A_n). \end{aligned}$$

• Proof of (a): By direct inspection

$$\begin{aligned} A_1, A_2, \ldots \in \mathfrak{G} \quad &\Longrightarrow \quad f^{-1}(A_1), f^{-1}(A_2), \ldots \in \mathfrak{M} \\ &\Longrightarrow \quad f^{-1}(\cup_n A_n) = \cup_n f^{-1}(A_n) \in \mathfrak{M} \\ &\Longrightarrow \quad \cup_n A_n \in \mathfrak{G}. \end{aligned}$$

Similarly,

$$A \in \mathfrak{G} \quad \Longrightarrow \quad f^{-1}(A) \in \mathfrak{M} \quad \Longrightarrow \quad f^{-1}(A^c) = (f^{-1}(A))^c \in \mathfrak{M} \quad \Longrightarrow \quad A^c \in \mathfrak{G}.$$

Lastly we apply Lemma 2.4.

• Proof of (b): the collection $\mathfrak{G} = \{E \subset Y\colon f^{-1}(E) \in \mathfrak{M}\}$ is a σ-algebra (by (a)) and it contains all open sets. Hence it contains the Borel σ-algebra.

• Proof of (c): for any open set $V \subset Z$ the set $g^{-1}(V) \subset Y$ is Borel, hence (by (b)) the set $f^{-1}(g^{-1}(V)) = (g \circ f)^{-1}(V) \subset X$ is measurable, i.e., belongs in $\mathfrak{M}$. $\square$

DEFINITION 4.7. **Extended real line**, denoted by $[-\infty, \infty]$, is $\mathbb{R} \cup \{\infty\} \cup \{-\infty\}$. The topology in $[-\infty, \infty]$ is generated by standard open intervals $(a, b) \subset \mathbb{R}$ and the following "infinite intervals": $(a, \infty]$ and $[-\infty, b)$ for all $a, b \in \mathbb{R}$. Any function $f\colon X \to [-\infty, \infty]$ is called an **extended real-valued function**.

Theorem 4.8. *Let $(X, \mathfrak{M})$ be a measurable space and $f\colon X \to [-\infty, \infty]$. Then f is measurable iff $f^{-1}\big([-\infty, x)\big)$ is a measurable set for every $x \in \mathbb{R}$.*

Proof. By Theorem 4.6(a) the collection $\mathfrak{G} = \{E \subset Y : f^{-1}(E) \in \mathfrak{M}\}$ is a σ-algebra containing all open infinite intervals $[-\infty, x)$. Now we just need to check that these intervals generate the Borel σ-algebra in $[-\infty, \infty]$, i.e., applying countable unions/intersections and complements produces all other open intervals in $[-\infty, x)$. This is a routine exercise in topology. $\square$

EXERCISE 16. In the context of the previous theorem, prove that f is measurable iff $f^{-1}\big([-\infty, x]\big)$ is a measurable set for every $x \in \mathbb{R}$.

EXERCISE 17. In the context of the previous theorem, prove that f is measurable iff $f^{-1}\big([-\infty, x)\big)$ is a measurable set for every rational $x \in \mathbb{Q}$.

EXERCISE 18. Let $(X, \mathfrak{M})$ be a measurable space and $f\colon X \to [-\infty, \infty]$ and $g\colon X \to [-\infty, \infty]$ two measurable functions. Prove the following sets are measurable:

$$\{x : f(x) < g(x)\} \qquad \text{and} \qquad \{x : f(x) = g(x)\}$$

EXERCISE 19. Show the following is a Borel function: $f\colon \mathbb{R} \to \mathbb{R}$ s.t.

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$$

EXERCISE 20. Show the following is a Borel function: $f\colon \mathbb{R} \to \mathbb{R}$ s.t.

$$f(x) = \begin{cases} \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

EXERCISE 21. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a monotonically increasing function, i.e. $f(x_1) \leq f(x_2)$ for $x_1 \leq x_2$. Show that f is a Borel function.

Theorem 4.9. *Let $(X, \mathfrak{M})$ be a measurable space and $u\colon X \to \mathbb{R}$ and $v\colon X \to \mathbb{R}$ two functions. Then $f = (u, v)\colon X \to \mathbb{R}^2$ is measurable if and only if both u and v are measurable.*

Proof.
$\boxed{\Rightarrow}$ If f is measurable, then for any open set $V \subset \mathbb{R}$ we have $u^{-1}(V) = f^{-1}(V \times \mathbb{R}) \in \mathfrak{M}$ because $V \times \mathbb{R}$ is open in $\mathbb{R}^2$. Similarly $v^{-1}(V) = f^{-1}(\mathbb{R} \times V) \in \mathfrak{M}$

$\boxed{\Leftarrow}$ If u and v are measurable, then for any open intervals $I, J \subset \mathbb{R}$ we have $f^{-1}(I \times J) = u^{-1}(I) \cap v^{-1}(J) \in \mathfrak{M}$ (as the intersection of two measurable sets). The sets $I \times J$ generate the Borel σ-algebra in $\mathbb{R}^2$, thus our result follows from Theorem 4.6(a). □

Corollary 4.10. *Let $(X, \mathfrak{M})$ be a measurable space, $u\colon X \to \mathbb{R}$ and $v\colon X \to \mathbb{R}$ two measurable functions, and $\Phi\colon \mathbb{R}^2 \to \mathbb{R}$ a continuous function. Then the composition $\Phi(u(x), v(x))\colon X \to \mathbb{R}$ is a measurable function. In particular, $u + v$, $u - v$, and uv are measurable functions.*

Corollary 4.11. *Let $(X, \mathfrak{M})$ be a measurable space and $u\colon X \to \mathbb{R}$ and $v\colon X \to \mathbb{R}$ two functions. Then $f(x) = u(x) + \mathbf{i}v(x)$ is a measurable function from X to $\mathbb{C}$ if and only if both $u(x)$ and $v(x)$ are measurable functions.*

Corollary 4.12. *Let $(X, \mathfrak{M})$ be a measurable space and $f\colon X \to \mathbb{C}$ and $g\colon X \to \mathbb{C}$ two measurable functions. Then $|f|$, $f+g$, $f-g$, and fg are measurable functions.*

Theorem 4.13. (Polar factorization) *Let $(X, \mathfrak{M})$ be a measurable space and $f\colon X \to \mathbb{C}$ a measurable function. Then $f = g|f|$, where $g\colon X \to \mathbb{C}$ is a measurable function such that $|g| = 1$.*

Proof. The set $E = f^{-1}(\{0\})$ is measurable. Note that $h(z) = z/|z|$ is a continuous function from $\mathbb{C} \setminus \{0\}$ to $\mathbb{C}$, thus the function $h \circ f$ is measurable on $X \setminus E$. Define g as follows: $g = h \circ f$ on $X \setminus E$ and $g = 1$ on E. □

Arithmetic in the Extended Real Line $[-\infty, \infty]$.

Arithmetic operations in $[-\infty, \infty]$ are defined in an obvious way:

- $a + \infty = \infty$ for any $a \in (-\infty, \infty]$,
- $a + (-\infty) = -\infty$ for any $a \in [-\infty, \infty)$,
- $a \cdot \infty = \infty$ for every $a \in (0, \infty]$,
- $a \cdot \infty = -\infty$ for every $a \in [-\infty, 0)$, etc.

The sum $\infty + (-\infty)$ is not defined. Most importantly, we put

$$0 \cdot \infty = 0$$

Limits in the Extended Real Line $[-\infty, \infty]$.

The set $[-\infty, \infty]$ is naturally ordered, any subset $A \subset [-\infty, \infty]$ obviously has $\inf A$ a $\sup A$. The convergence of sequences in $[-\infty, \infty]$ is defined by using its topology. For any sequence $\{a_n\}$ in $[-\infty, \infty]$ we naturally define $\liminf a_n$ and $\limsup a_n$:

$$\limsup_{n\to\infty} a_n = \inf\{b_1, b_2, \ldots\}, \qquad b_n = \sup\{a_n, a_{n+1}, \ldots\}. \tag{4.1}$$

(Here inf can be replaced with lim, because $b_1 \geq b_2 \geq \cdots$ is a monotonically decreasing sequence, hence it has a limit.)

Proposition 4.14. *For any sequence of numbers $a_1, a_2, \ldots \in [-\infty, \infty]$ we have*

(a) $\limsup_{n\to\infty}(-a_n) = -\liminf_{n\to\infty} a_n$.

(b) $\limsup_{n\to\infty}(a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n$.

(c) $\liminf_{n\to\infty}(a_n + b_n) \geq \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n$.

(d) $\lim_{n\to\infty} a_n$ *exists if and only if* $\limsup_{n\to\infty} a_n = \liminf_{n\to\infty} a_n$, *in which case* $\lim_{n\to\infty} a_n$ *equals to both* $\limsup_{n\to\infty} a_n$ *and* $\liminf_{n\to\infty} a_n$.

Theorem 4.15. *Let* $(X, \mathfrak{M})$ *be a measurable space and* $f_n\colon X \to [-\infty, \infty]$ *measurable functions. Then*

$$g = \sup_{n\geq 1} f_n \qquad \text{and} \qquad h = \limsup_{n\to\infty} f_n$$

are measurable functions. (Similarly for $\inf f_n$ *and* $\liminf f_n$*.)*

Proof. For any $c \in \mathbb{R}$ we have

$$\{x\colon g(x) > c\} = \cup_{n=1}^{\infty}\{x\colon f_n(x) > c\},$$

hence g is measurable (cf. Theorem 4.8 and Exercise 9). Similarly, $\inf f_n$ is a measurable function. The function $\limsup f_n$ can be expressed as a combination of sup's and inf's due to (4.1). □

Corollary 4.16. *Let* $(X, \mathfrak{M})$ *be a measurable space and* $f_n\colon X \to [-\infty, \infty]$ *measurable functions. If the limit*

$$g(x) = \lim_{n\to\infty} f_n(x)$$

exists for every $x \in X$*, then* g *is a measurable function.*

Corollary 4.17. *Let* $(X, \mathfrak{M})$ *be a measurable space and* $f, g\colon X \to [-\infty, \infty]$ *are measurable functions. Then*

$$\max\{f(x), g(x)\} \qquad \text{and} \qquad \min\{f(x), g(x)\}$$

are measurable functions. Also,

$$f^+(x) = \max\{f(x), 0\} \qquad \text{and} \qquad f^-(x) = -\min\{f(x), 0\}$$

are measurable functions.

Definition 4.18. The functions

$$f^+(x) = \max\{f(x), 0\} \qquad \text{and} \qquad f^-(x) = -\min\{f(x), 0\}$$

are called **positive and negative parts** of f, respectively.

- For any function $f\colon X \to [-\infty, \infty]$

$$f = f^+ - f^- \qquad \text{and} \qquad |f| = f^+ + f^-. \tag{4.2}$$

Theorem 4.19. *If $f = g - h$ and $g \geq 0$, $h \geq 0$, then $f^+ \leq g$ and $f^- \leq h$.*

Proof. If $f(x) \geq 0$ then

$$f^-(x) = 0 \leq h(x) \quad \text{and} \quad f^+(x) = f(x) = g(x) - h(x) \leq g(x)$$

If $f(x) \leq 0$ then

$$f^+(x) = 0 \leq g(x) \quad \text{and} \quad f^-(x) = -f(x) = h(x) - g(x) \leq h(x)$$

□

- In other words, f^+ and f^- are the "most economical" nonnegative functions whose difference is the given function f.

Definition 4.20. A function $f\colon X \to Y$ is a **simple function** iff its range $f(X)$ is finite.

Proposition 4.21. *If $s\colon X \to \mathbb{R}$ is a measurable simple function whose (distinct) values are $\alpha_1, \ldots, \alpha_n$, then it can be represented by*

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

where $A_i = s^{-1}(\{\alpha_i\})$ are disjoint measurable sets such that $X = \biguplus_{i=1}^{n} A_i$.

Theorem 4.22. *Let $(X, \mathfrak{M})$ be a measurable space and $f\colon X \to [0, \infty]$ a measurable function. Then there exist simple functions $s_n\colon X \to \mathbb{R}$ such that*

$$0 \le s_1 \le s_2 \le \cdots \le f$$

and $s_n(x) \to f(x)$ as $n \to \infty$ for every $x \in X$.

Proof.

$\boxed{\textbf{(i)}}$ Special case: $X = [0, \infty]$ and $f(x) = x$. Define simple functions by

$$\varphi_n = \sum_{j=1}^{n2^n} \frac{j-1}{2^n} \chi_{E_j} + n\chi_{[n,\infty]}$$

where

$$E_j = \left[\frac{j-1}{2^n}, \frac{j}{2^n}\right)$$

Intuitively, we divide the interval $[0, n)$ into small pieces so that each subinterval $[k, k+1)$ of length one is divided into 2^n pieces, and treat the remaining infinite interval $[n, \infty]$ as one piece. The total number of pieces is $n2^n + 1$. The function φ_n collapses each piece into its left endpoint. It is easy to check that $\forall x \in [0, \infty]$

$$0 \le \varphi_1(x) \le \varphi_2(x) \le \cdots \le x, \qquad \lim_{n\to\infty} \varphi_n(x) = x. \tag{4.3}$$

$\boxed{\textbf{(ii)}}$ General case: we define simple functions by $s_n = \varphi_n \circ f$. They are measurable due to Theorem 4.6(c). They are simple because $s_n(X) \subset \varphi_n(X)$, which is a finite set. For any $x \in X$ we rewrite (4.3) as follows:

$$0 \le \varphi_1(f(x)) \le \varphi_2(f(x)) \le \cdots \le f(x), \qquad \lim_{n\to\infty} \varphi_n(f(x)) = f(x).$$

Now replacing $\varphi_n(f(x))$ with $s_n(x)$ we get the desired:

$$0 \le s_1(x) \le s_2(x) \le \cdots \le f(x), \qquad \lim_{n\to\infty} s_n(x) = f(x),$$

$\square$

5 Lebesgue integration of nonnegative functions

Definition 5.1. Let $(X, \mathfrak{M}, \mu)$ be a measure space and $s\colon X \to [0,\infty)$ a nonnegative measurable simple function represented by

$$s = \sum_{i=1}^{n} \alpha_i \, \chi_{A_i}. \tag{5.1}$$

Then we define the **Lebesgue integral** of s over X by

$$\int_X s \, d\mu = \sum_{i=1}^{n} \alpha_i \, \mu(A_i). \tag{5.2}$$

- It is possible that for some i we have $\alpha_i = 0$ and $\mu(A_i) = \infty$. In this case we use the rule $0 \cdot \infty = 0$.

Proposition 5.2. *The formula (5.2) in Definition 5.1 holds even if the values $\alpha_1, \ldots, \alpha_n$ are not distinct.*

Proof. In (5.1) it is usually assumed that the values $\alpha_1, \ldots, \alpha_n$ are distinct. However, if they are not, we can simply "lump together" the subsets A_i's on which s takes the same values. For example, if $\alpha_i = \alpha_j$ for some $i \neq j$, then

$$\alpha_i \mu(A_i) + \alpha_j \mu(A_j) = \alpha_i(\mu(A_i) + \mu(A_j)) = \alpha_i \mu(A_i \cup A_j),$$

and note that $s(x) = \alpha_i$ for all $x \in A_i \cup A_j$. So we can replace the two terms $\alpha_i \mu(A_i)$ and $\alpha_j \mu(A_j)$ in (5.2) with one, $\alpha_i \mu(A_i \cup A_j)$, and the sum (5.2) will have $n-1$ terms total. Repeating this lumping process, in less than n steps we will get a sum where all the α_i's are distinct. □

Definition 5.3. In the context of Definition 5.1, for any $E \in \mathfrak{M}$ we define the **Lebesgue integral** of s over E by

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

Alternatively, we can restrict the function s and measure μ to E (as described in Section 3) and define $\int_E s \, d\mu$ as in Definition 5.1. These two definitions of $\int_X s \, d\mu$ agree (one can check via direct inspection).

Proposition 5.4.

(a) $\int_E s\,d\mu \in [0,\infty]$ *for any such* $s\colon X \to [0,\infty]$.

(b) *If* $\mu(E) = 0$ *then* $\int_E s\,d\mu = 0$.

(c) *If* $s \equiv c \geq 0$ *is a constant function, then* $\int_E c\,d\mu = c\,\mu(E)$.

(d) $\int_E 0\,d\mu = 0$ *even if* $\mu(E) = \infty$.

(e) *For any set* $A \in \mathfrak{M}$ *we have* $\int_X \chi_A\,d\mu = \mu(A)$.

Proof. Direct inspection. □

Lemma 5.5. *Let* $s, t\colon X \to [0,\infty)$ *be two simple functions. Then*

$$\int_E (s+t)\,d\mu = \int_E s\,d\mu + \int_E t\,d\mu$$

Proof. Let $s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ $\quad t = \sum_{j=1}^{m} \beta_j \chi_{B_j}$ $\quad$ s.t. $\biguplus_{i=1}^{n} A_i = \biguplus_{j=1}^{m} B_i = X$

Now denote $E_{i,j} = A_i \cap B_j \cap E$ (so $E = \uplus_{i,j} E_{i,j}$). Note that

$$A_i \cap E = \uplus_j E_{i,j} \qquad B_j \cap E = \uplus_i E_{i,j}$$

On each $E_{i,j}$ we have $s = \alpha_i$ and $t = \beta_j$ hence $s + t = \alpha_i + \beta_j$

Thus $s + t$ is a simple function given by

$$s + t = \sum_{i,j} (\alpha_i + \beta_j)\chi_{E_{i,j}}$$

Therefore

$$\begin{aligned}
\int_E (s+t)\,d\mu &= \sum_{i,j} (\alpha_i + \beta_j)\mu(E_{i,j}) \\
&= \sum_{i,j} \alpha_i \mu(E_{i,j}) + \sum_{i,j} \beta_j \mu(E_{i,j}) \\
&= \sum_i \alpha_i \sum_j \mu(E_{i,j}) + \sum_j \beta_j \sum_i \mu(E_{i,j}) \\
&= \sum_i \alpha_i \mu(A_i \cap E) + \sum_j \beta_j \mu(B_j \cap E) \\
&= \int_E s\,d\mu + \int_E t\,d\mu
\end{aligned}$$

□

Corollary 5.6. *If $0 \le s \le t$ are two simple functions, then $\int_E s\,d\mu \le \int_E t\,d\mu$.*

Proof. The difference $u = t - s$ is also a nonnegative simple function, hence

$$\int_E t\,d\mu = \int_E (s+u)\,d\mu \overset{5.5}{=} \int_E s\,d\mu + \int_E u\,d\mu \ge \int_E s\,d\mu$$

This follows from Proposition 5.4(a) which states $\int_E u\,d\mu \ge 0$ □

Definition 5.7. Let $(X, \mathfrak{M}, \mu)$ be a measure space and $f\colon X \to [0,\infty]$ a nonnegative measurable function (possibly taking value $+\infty$). Then for any $E \in \mathfrak{M}$ we define the **Lebesgue integral**

$$\int_E f\,d\mu = \sup_{s \in L_f} \int_E s\,d\mu,$$

where $\quad L_f = \big\{ s\colon X \to [0,\infty) \text{ measurable simple}, \quad 0 \le s \le f \big\}$

Note that $L_f \ne \emptyset$ as it always contains the function $s \equiv 0$.

Exercise 22. Verify that for simple functions, Definition 5.3 and Definition 5.7 agree.

Theorem 5.8. **Basic properties of the Lebesgue integral**

Let $f, g\colon X \to [0,\infty]$ be measurable functions and $A, B, E \in \mathfrak{M}$ measurable sets.

(a) *if $f \le g$, then $\int_E f\,d\mu \le \int_E g\,d\mu$;*

(b) *if $A \subset B$, then $\int_A f\,d\mu \le \int_B f\,d\mu$;*

(c) *for any constant $c \ge 0$ we have $\int_A cf\,d\mu = c\int_A f\,d\mu$;*

(d) *if $\mu(E) = 0$, then $\int_E f\,d\mu = 0$, even if $f \equiv \infty$;*

(e) *$\int_E f\,d\mu = \int_X \chi_E f\,d\mu$.*

(f) *if $\int_X f\,d\mu < \infty$, then $\mu\{x\colon g(x) = \infty\} = 0$.*

Proof. <u>Basic properties of the Lebesgue integral</u> (Theorem 5.8)

(a) For any simple function s we have $s \in L_f \implies s \le f \le g \implies s \in L_g$ (i.e. $L_f \subset L_g$), this implies $\sup_{s\in L_f}\left\{\int_E s\,d\mu\right\} \le \sup_{t\in L_g}\left\{\int_E t\,d\mu\right\}$
Which, by Definition 5.7, is equivalent to $\int_E f\,d\mu \le \int_E g\,d\mu$

(b) For any $\{A_k\}_{k=1}^n$: $A_k \cap A \subset A_k \cap B \implies \mu(A_k\cap A) \le \mu(A_k \cap B)$ $(\forall k)$
For any $s \in L_f$: $\int_A s\,d\mu = \sum_{k=1}^n \alpha_k \mu(A_k\cap A) \le \sum_{k=1}^n \alpha_k\mu(A_k\cap B) = \int_B s\,d\mu$
Since $\int_A s\,d\mu \le \int_B s\,d\mu$ $(\forall s \in L_f)$ we conclude $\int_A f\,d\mu \le \int_B f\,d\mu$

(c)
- If $c = 0$: $\int_A 0\cdot f\,d\mu = \int_A 0\,d\mu = 0 = 0 \cdot \int_A f\,d\mu$
- If $f = s = \sum_{k=1}^n \alpha_k \chi_{A_k}$ (simple):
  $\int_A cf\,d\mu = \sum_{k=1}^n c\alpha_k\,\mu(A_k\cap A) = c\sum_{k=1}^n \alpha_k\,\mu(A_k \cap A) = c\int_A f\,d\mu$
- If $c \ne 0$ and f is not simple: $s \in L_{cf} \iff \frac{1}{c}s \in L_f$
  Thus, $\int_A cf\,d\mu = \sup_{s\in L_{cf}}\left\{\int_A s\,d\mu\right\} = \sup_{\frac{1}{c}s\in L_f}\left\{\int_A s\,d\mu\right\}$
  $= \sup_{\frac{1}{c}s\in L_f}\left\{c\int_A \frac{1}{c}s\,d\mu\right\} = c\ \sup_{\frac{1}{c}s\in L_f}\left\{\int_A \frac{1}{c}s\,d\mu\right\}$
  $= c\ \sup_{t\in L_f}\left\{\int_A t\,d\mu\right\} = c\int_A f\,d\mu$

(d) For any $s \in L_f$: $A_k \cap E \subset E \implies \mu(A_k \cap E) = 0$ which gives us

$$\int_E s\,d\mu = \sum_{k=1}^n \alpha_k\mu(A_k\cap E) = \sum_{k=1}^n \alpha_k \cdot 0 = 0$$

Thus, $\int_E f\,d\mu = \sup_{s\in L_f}\left\{\int_E s\,d\mu\right\} = \sup_{s\in L_f}\{0\} = 0$

(e) **(i)** For s (simple): $\chi_E s = \sum_{k=1}^n \alpha_k\chi_{A_k}\chi_E = \sum_{k=1}^n \alpha_k \chi_{A_k\cap E}$
thus, $\int_X \chi_E s\,d\mu = \sum_{k=1}^n \alpha_k\mu(A_k\cap E) = \int_E s\,d\mu$

(ii) Note that $s \in L_{\chi_E f} \iff s = \chi_E \tilde{s}$ (for some $\tilde{s} \in L_f$)

$$\int_X \chi_E f\,d\mu = \sup_{s\in L_{\chi_E f}}\left\{\int_X s\,d\mu\right\} \overset{\text{(ii)}}{=} \sup_{\tilde{s}\in L_f}\left\{\int_X \chi_E\tilde{s}\,d\mu\right\}$$
$$\overset{\text{(i)}}{=} \sup_{\tilde{s}\in L_f}\left\{\int_E \tilde{s}\,d\mu\right\} = \int_E f\,d\mu$$

(f) First, note the set $E = \{x\colon f(x) = \infty\} = f^{-1}(\{\infty\})$ is measurable.
For any $N > 0$, we have $f \ge N\chi_E$, thus by Parts (a) and (c)

$$\int_X f\,d\mu \ge \int_X N\chi_E\,d\mu = N\mu(E).$$

If $\mu(E) > 0$, then (because our N is arbitrary) we would have $\int_X f\,d\mu = \infty$.
This contradiction proves (f).

□

EXERCISE 23. Let $x_0 \in X$ and $\mu = \delta_{x_0}$ the δ-measure. Assume that $\{x_0\} \in \mathfrak{M}$. Show that for every measurable function $f\colon X \to [0,\infty]$ we have

$$\int_X f\,d\mu = f(x_0).$$

EXERCISE 24. Let $X = \mathbb{N}$ and μ the counting measure on the σ-algebra $\mathfrak{M} = 2^{\mathbb{N}}$. Show that for every function $f\colon X \to [0,\infty]$ we have

$$\int_X f\,d\mu = \sum_{n=1}^{\infty} f(n).$$

Theorem 5.9. *Let $(X, \mathfrak{M}, \mu)$ be a measure space and $s\colon X \to [0,\infty)$ a nonnegative measurable simple function. Then*

$$\varphi(E) = \int_E s\,d\mu$$

is a measure on $\mathfrak{M}$.

Proof. Let $s = \sum_{k=1}^{n} \alpha_k \chi_{A_k}$ for any $E \in \mathfrak{M}$: $\varphi(E) = \sum_{k=1}^{n} \alpha_k \mu(A_k \cap E)$.

(i) $\varphi(\emptyset) = \sum_{k=1}^{n} \alpha_k \mu(A_k \cap \emptyset) = \sum_{k=1}^{n} \alpha_k \mu(\emptyset) = \sum\limits_{k=1}^{n} \alpha_k \cdot 0 = 0$

(ii) let $E = \biguplus_{n=1}^{\infty} E_n$ where $E, E_n \in \mathfrak{M}$ $(\forall n \in \mathbb{N})$. Then

$$\begin{aligned}
\varphi(E) &= \varphi\big(\textstyle\biguplus_{m=1}^{\infty} E_m\big) = \sum_{k=1}^{n} \alpha_k \mu\big(A_k \cap \biguplus_{m=1}^{\infty} E_m\big) \\
&= \textstyle\sum_{k=1}^{n} \alpha_k \mu\big(\biguplus_{m=1}^{\infty} (A_k \cap E_m)\big) = \sum_{k=1}^{n} \big(\alpha_k \sum_{m=1}^{\infty} \mu(A_k \cap E_m)\big) \\
&= \textstyle\sum_{m=1}^{\infty} \big(\sum_{k=1}^{n} \alpha_k \mu(A_k \cap E_m)\big) = \sum_{m=1}^{\infty} \big(\int_{E_m} s\,d\mu\big) \\
&= \textstyle\sum_{m=1}^{\infty} \varphi(E_m) \qquad \text{(i.e. } \varphi \text{ is } \sigma\text{-additive, and thus a measure)}
\end{aligned}$$

$\square$

Theorem 5.10. **Lebesgue's Monotone Convergence**

Let $(X, \mathfrak{M}, \mu)$ be a measure space and $\{f_n\}$ a sequence of meas. functions on X.

Suppose (a) $0 \le f_1(x) \le f_2(x) \le \cdots \le \infty$ *for every* $x \in X$

(b) $f_n(x) \to f(x)$ $(n \to \infty)$ *for every* $x \in X$

Then f is measurable and

$$\int_X f_n\,d\mu \to \int_X f\,d\mu \qquad (n \to \infty)$$

- This can be written as $\lim\limits_{n\to\infty} \int_X f_n\,d\mu = \int_X \lim\limits_{n\to\infty} f_n\,d\mu$

Proof. Lebesgue's Monotone Convergence (Theorem 5.10)

By Corollary 4.16, the limit function f is measurable, thus $\int_X f\,d\mu$ exists.

By Theorem 5.8(a): $\int_X f_n\,d\mu \leq \int_X f_{n+1}\,d\mu$ (since $f_n \leq f_{n+1}$)
This means the integrals $\int_X f_n\,d\mu$ make a monotonically increasing sequence. Thus

$$\exists \alpha \in [0,\infty]: \quad \int_X f_n\,d\mu \to \alpha \qquad \text{as} \quad n \to \infty$$

Next, we need to show: **(i)** $\alpha \leq \int_X f\,d\mu$ **(ii)** $\alpha \geq \int_X f\,d\mu$

(i) By Theorem 5.8(b): $f_n \leq f \implies \int_X f_n\,d\mu \leq \int_X f\,d\mu$ $(\forall n > 0)$
$\implies \alpha \leq \int_X f\,d\mu$ (i)

(ii) Let $s \in L_f$ and $0 < c < 1$.
Consider the sets $E_n = \{x : f_n \geq cs(x)\}$ $(n = 1, 2, \ldots)$

(a) each E_n is measurable (by Exercise 18)

(b) For $x \in E_n$: $cs(x) \leq f_n(x) \leq f_{n+1}(x) \Rightarrow x \in E_{n+1}$ $(\forall n > 0)$
$\implies E_n \subset E_{n+1}$ $(\forall n > 0)$
$\implies E_1 \subset E_2 \subset \ldots$ (increasing sequence)

(c) We claim that $X = \cup_{n=1}^{\infty} E_n$. Let $x \in X$.
- $f(x) = 0 \Rightarrow s(x) = 0,\ f_n(x) = 0$ $(\forall n > 0)$ $\Rightarrow x \in E_n$ $(\forall n > 0)$
- $f(x) > 0 \Rightarrow cs(x) < f(x)$ (since $c < 1$) $\Rightarrow cs(x) < f_n(x)$ (for some n)
  $\Rightarrow x \in E_n$ (for some n) $\Rightarrow x \in \cup_{n=1}^{\infty} E_n$

(d) For each $n \geq 1$: $\alpha \geq \int_X f_n\,d\mu \overset{5.8(b)}{\geq} \int_{E_n} f_n\,d\mu \overset{5.8(e)}{=} \int_X f_n \chi_{E_n}\,d\mu$
$\overset{5.8(a)}{\geq} \int_X cs\chi_{E_n}\,d\mu \overset{5.8(e)}{\geq} \int_{E_n} cs\,d\mu \overset{5.8(c)}{=} c\int_{E_n} s\,d\mu$
$= c\,\varphi(E_n)$ (where we denote $\varphi(E_n) = \int_{E_n} s\,d\mu$)

By (a), (b), (c), and Continuity-II (3.13): $\varphi(E_n) \to \varphi(X) = \int_{E_n} s\,d\mu$ $(n \to \infty)$
multiplying both sides by α, we obtain $\alpha \geq c\varphi(E_n) \to c\int_X s\,d\mu$ $(n \to \infty)$

Thus, $\alpha \geq c\int_X s\,d\mu$ $(\forall s \in L_f,\ c < 1)$ $\overset{c \to 1}{\implies} \alpha \geq \int_X s\,d\mu$ $(\forall s \in L_f)$
$\implies \alpha \geq \int_X f\,d\mu$ (ii)

By (i) and (ii), $\int_X f_n\,d\mu \to \alpha = \int_X f\,d\mu$ $(n \to \infty)$ $\square$

Theorem 5.11. **Additivity**

Let $(X, \mathfrak{M}, \mu)$ be a measure space and $f, g\colon X \to [0, \infty]$ two nonnegative measurable functions. Then

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

Proof. Recall that for simple functions the additivity was proved in Lemma 5.5. General functions f and g can be approximated by simple functions due to Theorem 4.22:

$$0 \le s_1 \le s_2 \le \cdots \le f, \qquad s_n \to f$$

$$0 \le t_1 \le t_2 \le \cdots \le g, \qquad t_n \to g$$

Note that $s_n + t_n$ is a simple function for each $n \ge 1$, and

$$0 \le (s_1+t_1) \le (s_2+t_2) \le \cdots \le (f+g), \qquad (s_n+t_n) \to (f+g)$$

Now we have by Lebesgue's Monotone Convergence Theorem

$$\begin{array}{ccccc} \displaystyle\int_X (s_n+t_n)\,d\mu & \overset{5.5}{=} & \displaystyle\int_X s_n\,d\mu & + & \displaystyle\int_X t_n\,d\mu \\ \downarrow & & \downarrow & & \downarrow \\ \displaystyle\int_X (f+g)\,d\mu & = & \displaystyle\int_X f\,d\mu & + & \displaystyle\int_X g\,d\mu \end{array}$$

The additivity theorem is proved. □

Corollary 5.12. **Linearity**

Let $(X, \mathfrak{M}, \mu)$ be a measure space; $f, g\colon X \to [0, \infty]$ two nonnegative measurable functions; and $c, d \ge 0$ two nonnegative constants. Then

$$\int_X (cf+dg)\,d\mu = c\int_X f\,d\mu + d\int_X g\,d\mu.$$

Proof. Combine Theorem 5.11 with Theorem 5.8(c) □

Theorem 5.13. *Let $(X, \mathfrak{M}, \mu)$ be a measure space and $f_n \colon X \to [0, \infty]$ a sequence of nonnegative measurable functions on X and*

$$f(x) = \sum_{n=1}^{\infty} f_n(x) \quad \text{for every} \quad x \in X.$$

Then f is measurable and

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

(That is, summation and integration “commute” for nonnegative functions.)

Proof. Denote $g_N = \sum_{n=1}^{N} f_n$. By induction, Theorem 5.11 extends to finite sums,

$$\int_X g_N \, d\mu = \sum_{n=1}^{N} \int_X f_n \, d\mu$$

Note that $g_1 \leq g_2 \leq \cdots$ and $g_N \to f$ as $N \to \infty$. Now

$$\sum_{n=1}^{\infty} \int_X f_n \, d\mu = \lim_{N\to\infty} \sum_{n=1}^{N} \int_X f_n \, d\mu = \lim_{N\to\infty} \int_X g_N \, d\mu = \int_X f \, d\mu$$

(the last identity follows from Lebesgue’s Monotone Convergence Theorem 5.10) □

Corollary 5.14. *If $a_{ij} \geq 0$ for all $i, j = 1, 2, \ldots,$ then*

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}.$$

Proof. Let $(X, \mathfrak{M}, \mu) = (\mathbb{N}, 2^{\mathbb{N}}, \mu)$, where μ is the counting measure.
For each $i \geq 1$ define a function $f_i \colon \mathbb{N} \to [0, \infty)$ by $f_i(j) = a_{ij}$ $(\forall j \in \mathbb{N})$.
Then by Exercise 24 $\int_{\mathbb{N}} f_i \, d\mu = \sum_{j=1}^{\infty} a_{ij}$ Therefore

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \sum_{i=1}^{\infty} \int_{\mathbb{N}} f_i \, d\mu \overset{5.13}{=} \int_{\mathbb{N}} \Big(\sum_{i=1}^{\infty} f_i \Big) \, d\mu = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}.$$

in the end we used the result of Exercise 24 once again. □

• Without assumption $a_{ij} \geq 0$ Corollary 5.14 fails. As we remember from calculus, the sum of an infinite series may depend on the order in which its terms are added.

Motivating example.Let

- $(X, \mathfrak{M}, \mu)$ a measure space
- $f_n \colon X \to [0, \infty]$ nonnegative measurable functions
- $f_n(x) \to f(x)$, as $n \to \infty$, for every $x \in X$ (i.e., f_n converge to f *pointwise*)

Here is a big question:

$$\boxed{\text{Is it true that } \int f_n\, d\mu \to \int f\, d\mu?}$$

Suppose for example that $X = [0,1]$, $\mu = \mathbf{m}$ is the Lebesgue measure, and $f(x) \equiv 1$, so that $\int_{[0,1]} f\, d\mathbf{m} = 1$. Let $\varepsilon > 0$. The pointwise convergence means that for every $x \in [0,1]$ there is N_x such that $|f_n(x) - f(x)| < \varepsilon$ for all $n \geq N_x$. So if we wait long enough, then the current value $f_n(x)$ will be almost equal to the limit value $f(x) = 1$. Of course the waiting period N_x depends on $x \in [0,1]$, but the longer we wait the more and more points $x \in [0,1]$ will have the above property $|f_n(x) - f(x)| < \varepsilon$. If we wait long enough, then those points will make an overwhelming majority in $[0,1]$, i.e., the set of points $A = \{x \in [0,1] \colon |f_n(x) - f(x)| < \varepsilon\}$ will have measure $\mathbf{m}(A) > 1 - \varepsilon$ for sufficiently large n. And then

$$\int_{[0,1]} f_n\, d\mathbf{m} \geq \int_A f_n\, d\mathbf{m} \geq \int_A (1-\varepsilon)\, d\mathbf{m} = (1-\varepsilon)\mathbf{m}(A) > (1-\varepsilon)^2.$$

Thus the limit value of the integral <u>cannot be smaller</u> than one:

$$\liminf_{n\to\infty} \int_{[0,1]} f_n\, d\mathbf{m} \geq 1 = \int_{[0,1]} f\, d\mathbf{m}.$$

But <u>can it be larger</u> than one? The answer is Yes!

Suppose, for example, $f_n(x) = 1 + n\chi_{(0,\frac{1}{n})}$. This function takes a constant value, 1, everywhere except a small open interval $(0, \frac{1}{n})$, on which its value is big (equal to $1+n$). Then $\int_{[0,1]} f_n\, d\mathbf{m} = 2$ for every $n \geq 1$, but the limit function is still $f(x) \equiv 1$ and its integral still equals 1. We can say that the "mass" of each function f_n equals 2, but only half of that "mass" reaches the limit function f; the other half "escapes through holes" or "falls through cracks". The role of "holes" or "cracks" is played by the vanishing intervals $(0, \frac{1}{n})$.

Theorem 5.15. **Fatou's Lemma**

Let $(X, \mathfrak{M}, \mu)$ be a measure space and $f_n \colon X \to [0, \infty]$ a sequence of nonnegative measurable functions on X. Then

$$\int_X \left(\liminf_{n\to\infty} f_n \right) d\mu \le \liminf_{n\to\infty} \int_X f_n \, d\mu.$$

Proof. Recall the definition of lim inf (Equation 4.1)

$$\liminf_{n\to\infty} f_n = \lim_{n\to\infty} g_n, \qquad g_n = \inf\{f_n, f_{n+1}, ...\}.$$

Note that $g_1 \le g_2 \le \ldots$ is a monotonically increasing sequence, hence by Lebesgue Monotone Convergence (Theorem 5.10):

$$\lim_{n\to\infty} \int_X g_n \, d\mu = \int_X (\lim_{n\to\infty} g_n) \, d\mu = \int_X (\liminf_{n\to\infty} f_n) \, d\mu.$$

Also note that $g_n \le f_n$ for each $n = 1, 2, \ldots$, hence

$$\begin{aligned}
&\overset{5.8(a)}{\Longrightarrow} \quad \int_X g_n \, d\mu \le \int_X f_n \, d\mu \\
&\Longrightarrow \quad \liminf_{n\to\infty} \int_X g_n \, d\mu \le \liminf_{n\to\infty} \int_X f_n \, d\mu \\
&\overset{4.14(d)}{\Longrightarrow} \quad \lim_{n\to\infty} \int_X g_n \, d\mu \le \liminf_{n\to\infty} \int_X f_n \, d\mu \\
&\Longrightarrow \quad \int_X (\liminf_{n\to\infty} f_n) \, d\mu \le \liminf_{n\to\infty} \int_X f_n \, d\mu
\end{aligned}$$

□

EXERCISE 25. Let $E \subset X$ be such that $\mu(E) > 0$ and $\mu(E^c) > 0$. Put $f_n = \chi_E$ if n is odd and $f_n = 1 - \chi_E$ if n is even. What is the relevance of this example to Fatou's lemma?

EXERCISE 26. Construct an example of a sequence of nonnegative measurable functions $f_n \colon X \to [0, \infty)$ such that $f(x) = \lim_{n\to\infty} f_n(x)$ exists pointwise, but

$$\int_X f \, d\mu < \liminf_{n\to\infty} \int_X f_n \, d\mu.$$

Theorem 5.16. *Let $(X, \mathfrak{M}, \mu)$ be a measure space and $f\colon X \to [0,\infty]$ a nonnegative measurable function on X . Then*

$$\varphi(E) = \int_E f\, d\mu$$

is a measure on $\mathfrak{M}$.

Furthermore, for every nonnegative measurable function g on X, we have

$$\int_X g\, d\varphi = \int_X gf\, d\mu$$

- The last identity allows us to write $d\varphi = f\, d\mu$, rather informally.

Proof.

First we prove that φ is a measure:

(i) $\varphi(\emptyset) = \int_\emptyset f\, d\mu \overset{5.8(d)}{=} 0$

(ii) Let $E = \biguplus_{n=1}^\infty E_n$ for some $E, E_n \in \mathfrak{M}$. Then

$$\begin{aligned}
\varphi(E) &= \int_E f\, d\mu \overset{5.8(e)}{=} \int_X \chi_E f\, d\mu = \int_X \chi_{(\uplus_n E_n)} f\, d\mu \\
&= \int_X \Big(\sum_{n=1}^\infty \chi_{E_n}\Big) f\, d\mu = \int_X \Big(\sum_{n=1}^\infty \chi_{E_n} f\Big)\, d\mu \\
&\overset{5.13}{=} \sum_{n=1}^\infty \int_X \chi_{E_n} f\, d\mu \overset{5.8(e)}{=} \sum_{n=1}^\infty \int_{E_n} f\, d\mu = \sum_{n=1}^\infty \varphi(E_n)
\end{aligned}$$

Next we prove that $\int_X g\, d\varphi = \int_X gf\, d\mu$

First suppose $g = s = \sum_{i=1}^m \alpha_i \chi_{A_i}$ is a simple, then

$$\begin{aligned}
\int_X s\, d\varphi &\overset{5.12}{=} \sum_{i=1}^m \alpha_i \varphi(A_i) = \sum_{i=1}^m \alpha_i \int_{A_i} f\, d\mu \\
&\overset{5.8(c)}{=} \sum_{i=1}^m \int_{A_i} \alpha_i f\, d\mu \overset{5.8(e)}{=} \sum_{i=1}^m \int_X \alpha_i \chi_{A_i} f\, d\mu \\
&\overset{5.11}{=} \int_X \left(\sum_{i=1}^m \alpha_i \chi_{A_i} f\right) d\mu = \int_X sf\, d\mu
\end{aligned}$$

Now let $g \geq 0$ be arbitrary measurable function. By Theorem 4.22, g can be approximated by simple functions

$$0 \leq s_1 \leq s_2 \leq \cdots \leq g, \qquad s_n \to g,$$

hence by Lebesgue's Monotone Convergence Theorem

$$\int_X s_n f\, d\mu = \int_X s_n\, d\varphi \to \int_X g\, d\varphi.$$

On the other hand, since $f \geq 0$, we have

$$0 \leq s_1 f \leq s_2 f \leq \cdots \leq gf, \qquad s_n f \to gf,$$

hence again by Lebesgue's Monotone Convergence (Theorem 5.10)

$$\int_X s_n f \, d\mu \to \int_X gf \, d\mu.$$

Thus we conclude

$$\int_X g \, d\varphi = \int_X gf \, d\mu$$

proving the theorem. $\square$

Corollary 5.17. *Let $(X, \mathfrak{M}, \mu)$ be a measure space and $f\colon X \to [0, \infty]$ a nonnegative measurable function on X . Then for every $A, B \in \mathfrak{M}$, $A \cap B = \emptyset$*

$$\int_{A \uplus B} f \, d\mu = \int_A f \, d\mu + \int_B f \, d\mu.$$

Proof. Using the measure $\varphi(E) = \int_E f \, d\mu$ defined in Theorem 5.16

$$\begin{aligned} \textstyle\int_{A \uplus B} f \, d\mu &= \varphi(A \uplus B) \\ &= \varphi(A) + \varphi(B) && \text{Since } \varphi \text{ is a measure on } (X, \mathfrak{M}) \\ &= \textstyle\int_A f \, d\mu + \int_B f \, d\mu \end{aligned}$$

$\square$

So far we have only used Lebesgue integral for nonnegative functions. In the next section we define Lebesgue integral for general real-valued and complex-valued functions.

6 Lebesgue integration of real/complex valued functions

DEFINITION 6.1. We say that a measurable function $f\colon X \to \mathbb{R}$ or $f\colon X \to \mathbb{C}$ is **Lebesgue integrable** if

$$\int_X |f|\, d\mu < \infty.$$

The set of all integrable functions is denoted by $L^1_\mu(X)$.

- The function $|f|$ is measurable due to Corollary 4.12.
- The above integral is defined since $|f| \geq 0$ its value is either a nonnegative finite number or infinity.

Recall that for a real-valued function $f\colon X \to \mathbb{R}$ we have $f = f^+ - f^-$ and $|f| = f^+ + f^-$ where $f^+ \geq 0$ and $f^- \geq 0$. This leads to the following:

Lemma 6.2. *For any real-valued measurable function $f\colon X \to \mathbb{R}$*

$$\int_X |f|\, d\mu = \int_X f^+\, d\mu + \int_X f^-\, d\mu.$$

The left integral is finite if and only if both right integrals are finite:

$$\int_X |f|\, d\mu < \infty \quad \Longleftrightarrow \quad \int_X f^+\, d\mu < \infty \;\; \text{and} \;\; \int_X f^-\, d\mu < \infty.$$

DEFINITION 6.3. Let $f\colon X \to \mathbb{R}$ be an integrable real-valued function. Then its **Lebesgue integral** over any measurable set $E \in \mathfrak{M}$ is defined by

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

Lemma 6.4. *For any complex-valued measurable function $f\colon X \to \mathbb{C}$, denote by $u = \mathtt{Re}(f)$ and $v = \mathtt{Im}(f)$ (such that $f = u + \mathbf{i}v$). Then*

$$\max\{|u|, |v|\} \;\leq\; |f| \;\leq\; |u| + |v|$$

Therefore, $|f|$ is integrable if and only if both u and v are integrable:

$$\int_X |f|\, d\mu < \infty \quad \Longleftrightarrow \quad \int_X |u|\, d\mu < \infty \quad \text{and} \quad \int_X |v|\, d\mu < \infty.$$

DEFINITION 6.5. Let $f\colon X \to \mathbb{C}$ be an integrable complex-valued function. Let $u = \texttt{Re}\, f$ and $v = \texttt{Im}\, f$ be its real and imaginary parts ($f = u + \mathbf{i}v$). Then the **Lebesgue integral** of f over any measurable set $E \in \mathfrak{M}$ is defined by

$$\int_E f\, d\mu = \int_E u\, d\mu + \mathbf{i} \int_E v\, d\mu$$

- In other words, we integrate the real part and the imaginary part separately:

$$\texttt{Re} \int_E f\, d\mu = \int_E \texttt{Re}\, f\, d\mu \quad \text{and} \quad \texttt{Im} \int_E f\, d\mu = \int_E \texttt{Im}\, f\, d\mu. \tag{6.1}$$

- The integral of a real-valued function is a (finite) real number.
- The integral of a complex-valued function is a (finite) complex number.

Two types of Lebesgue integrals?.

For nonnegative measurable functions $f\colon X \to [0, \infty]$ we have defined the Lebesgue integral $\int_X f\, d\mu$ twice: in Definition 5.3 and here in Definition 6.3. Do these definitions agree? Not quite.

By Definition 5.3, the integral $\int_X f\, d\mu$ is always defined, though its value may be finite or infinite. When its value is finite, then by Definition 6.3 (note that $f = f^+$ in this case) f is integrable and its integral is the same as the one due to Definition 5.3. In this sense our definitions agree.

However, if the value of $\int_X f\, d\mu$ by Definition 5.3 is infinite, then by Definition 6.3 the function f is **not integrable**, so the integral $\int_X f\, d\mu$ is **not defined**. Thus the new Definition 6.3 is just a restriction of the old Definition 5.3 to the category of functions whose integral is finite.

This little disagreement will not cause us trouble. In fact, with the following extension, Definition 6.3 will fully agree with Definition 5.3.

Extension of Definition 6.3.

If one of the integrals $\int_E f^+ \, d\mu$ and $\int_E f^- \, d\mu$ is infinite, but the other finite, the formula in Definition 6.3 still can be applied. Precisely,

$$\int_E f^+ \, d\mu = \infty, \quad \int_E f^- \, d\mu < \infty \quad \Longrightarrow \quad \int_E f \, d\mu = \infty$$

and

$$\int_E f^+ \, d\mu < \infty, \quad \int_E f^- \, d\mu = \infty \quad \Longrightarrow \quad \int_E f \, d\mu = -\infty$$

Occasionally this extension is used, and we will need it in Section 11.

Note though that if both integrals $\int_E f^+ \, d\mu$ and $\int_E f^- \, d\mu$ are infinite, then the formula in Definition 6.3 makes no sense, because $\infty - \infty$ is not defined.

Lemma 6.6. *Let $f \in L^1_\mu(X)$. If $A \cap B = \emptyset$ are disjoint measurable sets, then*

$$\int_{A \uplus B} f \, d\mu = \int_A f \, d\mu + \int_B f \, d\mu.$$

Proof. We just need to break down all the three integral into $\texttt{Re}\, f$ and $\texttt{Im}\, f$, and then integrate the positive and negative parts separately applying Corollary 5.17. □

Theorem 6.7. **Linearity**

Let $f, g \in L^1_\mu(X)$ and $\alpha, \beta \in \mathbb{C}$. Then

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

In particular, $\alpha f + \beta g \in L^1_\mu(X)$.

Proof. Note: the proof is tricky!

First we show that $\alpha f + \beta g$ is integrable. By standard triangle inequality

$$|\alpha f + \beta g| \leq |\alpha| \, |f| + |\beta| \, |g|$$

therefore due to Theorem 5.8(a,c) and Theorem 5.11

$$\int_X |\alpha f + \beta g| \, d\mu \leq |\alpha| \int_X |f| \, d\mu + |\beta| \int_X |g| \, d\mu < \infty.$$

Next it is enough to prove two facts:

$$\int_X (f+g)\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu \tag{6.2}$$

$$\int_X (\alpha f)\,d\mu = \alpha \int_X f\,d\mu \tag{6.3}$$

the combination of which readily implies our theorem.

First we prove (6.2). Since we integrate the real part and imaginary part separately, and since $\texttt{Re}\,(f+g) = \texttt{Re}\,f + \texttt{Re}\,g$ and $\texttt{Im}\,(f+g) = \texttt{Im}\,f + \texttt{Im}\,g$, it will be enough to treat real valued functions $f, g\colon X \to \mathbb{R}$ only.

Next, it is tempting to break down each real-valued function into its positive and negative parts. However, this will not work, because $(f+g)^+ \neq f^+ + g^+$ and $(f+g)^- \neq f^- + g^-$, generally speaking.

We use a different trick. Denote $h = f + g$, then

$$h^+ - h^- = (f^+ - f^-) + (g^+ - g^-) \quad \Longrightarrow \quad h^+ + f^- + g^- = h^- + f^+ + g^+.$$

Now on each side of the last equation we have a sum of nonnegative functions, thus we can use Theorem 5.11 to get

$$\int_X h^+\,d\mu + \int_X f^-\,d\mu + \int_X g^-\,d\mu = \int_X h^-\,d\mu + \int_X f^+\,d\mu + \int_X g^+\,d\mu.$$

Rearranging these integrals gives

$$\int_X h^+\,d\mu - \int_X h^-\,d\mu = \int_X f^+\,d\mu - \int_X f^-\,d\mu + \int_X g^+\,d\mu - \int_X g^-\,d\mu.$$

Using Definition 6.3 gives

$$\int_X h\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu,$$

which proves (6.2).

Now we prove (6.3). Suppose first that $\alpha \geq 0$ and $f = f^+ - f^-$ is a real-valued function. Then $(\alpha f)^+ = \alpha f^+$ and $(\alpha f)^- = \alpha f^-$, hence

$$\int_X (\alpha f)\,d\mu = \int_X (\alpha f)^+\,d\mu - \int_X (\alpha f)^-\,d\mu \overset{5.8(c)}{=} \alpha \int_X f^+\,d\mu - \alpha \int_X f^-\,d\mu = \alpha \int_X f\,d\mu$$

If $\alpha < 0$, then $(\alpha f)^+ = -\alpha f^-$ and $(\alpha f)^- = -\alpha f^+$, hence

$$\int_X (\alpha f)\,d\mu = \int_X (\alpha f)^+\,d\mu - \int_X (\alpha f)^-\,d\mu \overset{5.8(c)}{=} -\alpha \int_X f^-\,d\mu + \alpha \int_X f^+\,d\mu = \alpha \int_X f\,d\mu$$

Thus we proved (6.3) for any real α and any real-valued function f. To handle complex constants $\alpha = a + \mathbf{i}b$ and complex-valued functions $f = u + \mathbf{i}v$, note that

$$\alpha f = (au - bv) + \mathbf{i}(av + bu),$$

hence by Definition 6.5

$$\int_X (\alpha f)\,d\mu = \int_X (au - bv)\,d\mu + \mathbf{i} \int_X (av + bu)\,d\mu,$$

and now we can use the already proved linearity for real-valued functions. □

EXERCISE 27. Let $f, g \in L^1_\mu(X)$ be real-valued functions and $f \le g$. Show that

$$\int_X f\, d\mu \le \int_X g\, d\mu.$$

EXERCISE 28. Let $f_n \colon X \to [0,\infty]$ be a sequence of measurable functions such that $f_1 \ge f_2 \ge \cdots \ge 0$ and $\lim_{n\to\infty} f_n(x) = f(x)$ for every $x \in X$. Suppose $f_1 \in L^1_\mu(X)$. Show that

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

Theorem 6.8. **Integral triangle inequality**

If $f \in L^1_\mu(X)$ is a real-valued or complex-valued function, then

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

Proof.

Note: The analogy with the triangle inequality is clear if we replace integration with summation.

$\boxed{\textbf{Case 1}}$ $f : X \to [-\infty, \infty]$ is a real-valued function with possible infinite values.

Then
$$\begin{aligned}
\left|\int_X f\, d\mu\right| &= \left|\int_X f^+\, d\mu - \int_X f^-\, d\mu\right| \\
&\le \left|\int_X f^+\, d\mu\right| + \left|\int_X f^-\, d\mu\right| && \text{(triangle inequality in } \mathbb{R}) \\
&\le \int_X f^+\, d\mu + \int_X f^-\, d\mu && (f^+ \ge 0 \text{ and } f^- \ge 0) \\
&= \int_X (f^+ + f^-)\, d\mu && \text{(Additivity: Theorem 5.11)} \\
&= \int_X |f|\, d\mu && \text{(Eq. (4.2))}
\end{aligned}$$

$\boxed{\textbf{Case 2}}$ $f : X \to \mathbb{C}$ is a complex-valued function with finite values.

Then $\int_X f\, d\mu \in \mathbb{C} \implies \exists \theta \in [0, 2\pi)$ s.t. $\int_X f\, d\mu = e^{\mathbf{i}\theta} \left|\int_X f\, d\mu\right|$

Therefore
$$\begin{aligned}
\left|\int_X f\, d\mu\right| &= e^{-\mathbf{i}\theta} \int_X f\, d\mu \\
&= \int_X e^{-\mathbf{i}\theta} f\, d\mu && \text{(By Linearity, Theorem 6.7)} \\
&= \mathtt{Re} \int_X e^{-\mathbf{i}\theta} f\, d\mu && \text{(Because it is a real number)} \\
&= \int_X \mathtt{Re}\left(e^{-\mathbf{i}\theta} f\right) d\mu && \text{(By Eq. (6.1))} \\
&\le \int_X |f|\, d\mu && \text{(Because } \mathtt{Re}\,(z) \le |z| \text{ for any } z \in \mathbb{C})
\end{aligned}$$

Note: in the last step we also used the result of Exercise 27. $\square$

Theorem 6.9. **Lebesgue's Dominated Convergence**

Let $f_n \colon X \to \mathbb{C}$ be a sequence of measurable functions such that

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

Suppose there exists $g \in L^1_\mu(X)$ such that

$$|f_n(x)| \le g(x) \qquad \forall x \in X, \quad \forall n = 1, 2, \dots$$

i.e., the functions $f_n(x)$ are dominated by $g(x)$. Then

(a) $f \in L^1_\mu(X)$

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

(c) $\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu$

Proof.

(a) The limit $f(x)$ of measurable functions $f_n(x)$ is measurable by Corollary 4.16. By standard calculus rules we also have $|f(x)| \le g(x)$, hence

$$\int_X |f(x)| \, d\mu \overset{5.8(a)}{\le} \int_X g(x) \, d\mu < \infty$$

therefore $f \in L^1_\mu(X)$

(b) Similarly to Part (a), for every $n \ge 0$,

$$|f_n - f| \le |f_n| + |f| \le 2g \qquad \Longrightarrow \qquad |f_n - f| \in L^1_\mu(X).$$

Note also that

$$2g - |f_n - f| \ge 0,$$

thus Fatou's lemma will apply to these functions. We also have

$$\lim_{n\to\infty} (2g - |f_n - f|) = 2g - \lim_{n\to\infty} |f_n - f| = 2g,$$

therefore

$$\begin{aligned}
\textstyle\int_X 2g \, d\mu &= \textstyle\int_X \lim_{n\to\infty} (2g - |f_n - f|) \, d\mu \\
&= \textstyle\int_X \liminf_{n\to\infty} (2g - |f_n - f|) \, d\mu && \text{By Proposition 4.14(d)} \\
&\le \textstyle\liminf_{n\to\infty} \int_X (2g - |f_n - f|) \, d\mu && \text{By Fatou's Lemma} \\
&= \textstyle\liminf_{n\to\infty} \left(\int_X 2g \, d\mu - \int_X |f_n - f| \, d\mu \right) && \text{By Linearity, 6.7} \\
&= \textstyle\int_X 2g \, d\mu - \limsup_{n\to\infty} \int_X |f_n - f| \, d\mu && \text{By Proposition 4.14(a).}
\end{aligned}$$

Canceling $\int_X 2g\, d\mu$ gives

$$\limsup_{n\to\infty} \int_X |f_n - f|\, d\mu \le 0.$$

Since this is a sequence of nonnegative numbers, this relation is only possible if

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

(c) Using the linearity and the integral triangle inequality

$$0 \le \left| \int_X f_n\, d\mu - \int_X f\, d\mu \right| \overset{6.7}{=} \left| \int_X (f_n - f)\, d\mu \right| \overset{6.8}{\le} \int_X |f_n - f|\, d\mu \to 0$$

where in the end we used Part (b). □

Corollary 6.10. *Let $f \in L^1_\mu(X)$ and let*

$$E_1 \supset E_2 \supset \cdots$$

be a sequence of measurable sets such that $\mu(E_n) \to 0$ as $n \to \infty$. Then

$$\int_{E_n} f\, d\mu \to 0 \qquad \textit{as} \quad n \to \infty$$

Proof. Denote $E = \cap_{n=1}^\infty E_n$. By Continuity-I (Theorem 3.12) we have

$$\mu(E) = \lim_{n\to\infty} \mu(E_n) = 0$$

Consider functions $f_n = \chi_{E_n} f$ for all $n \ge 1$ By direct inspection

$$\lim_{n\to\infty} f_n(x) = \bar{f}(x) := \begin{cases} 0 & \text{if } x \notin E \\ f(x) & \text{if } x \in E \end{cases}$$

Also note that $|f_n| \le |f| \in L^1_\mu(X)$, hence by the Lebesgue Dominated Convergence

$$\int_{E_n} f\, d\mu \overset{5.8(\mathrm{e})}{=} \int_X f_n\, d\mu \overset{6.9}{\to} \int_X \bar{f}\, d\mu$$

as $n \to \infty$, and

$$\int_X \bar{f}\, d\mu \overset{5.17}{=} \int_{X\setminus E} 0\, d\mu + \int_E f\, d\mu = 0,$$

where the first integral is $0 \cdot \mu(X \setminus E) = 0$, and the second is zero due to Theorem 5.8. □

7 Role of sets of measure zero (null sets)

Let $(X, \mathfrak{M}, \mu)$ be a measure space and Y a topological space.

DEFINITION 7.1. Let $P(x)$ be a property which a point $x \in X$ may or may not have. We say that P holds **almost everywhere (a.e.)** on a set $E \subset X$ if there exists $N \subset X$, $\mu(N) = 0$, such that P holds at every $x \in E \setminus N$.

DEFINITION 7.2. Given two measurable functions $f, g \colon X \to Y$, we say that $f = g$ a.e. if $\mu\{x \in X \colon f(x) \neq g(x)\} = 0$.

- $f = g$ a.e. is an equivalence relation.

DEFINITION 7.3. We say that a sequence of measurable functions $f_n(x)$ converges a.e. to a limit function $f(x)$ if $\mu\{x \in X \colon f_n(x) \nrightarrow f(x)\} = 0$.

A general philosophy of Real Analysis is that sets of measure zero are negligible, what happens in those sets is insignificant. In many instances null sets are handled somewhat casually or ignored altogether.

Theorem 7.4. *Let $f, g \colon X \to [-\infty, \infty]$ or $f, g \colon X \to \mathbb{C}$ be two measurable functions. If $f = g$ a.e., then $\int_E f\, d\mu = \int_E g\, d\mu$ for any $E \in \mathfrak{M}$.*

- In this theorem, either both integrals exist ($\Rightarrow$ equal) or both fail to exist.

Proof. Denote $N = \{x \colon f(x) \neq g(x)\}$. By our assumption, $\mu(N) = 0$.

Case 1: $f, g \colon X \to [0, \infty]$ are nonnegative functions (with possible infinite values). We decompose E as $E = (E \setminus N) \uplus (E \cap N)$. Note that $\mu(E \cap N) \leq \mu(N) = 0$. Therefore

$$\int_E f\, d\mu \overset{5.17}{=} \int_{E \setminus N} f\, d\mu + \int_{E \cap N} f\, d\mu$$

$$\int_E g\, d\mu \overset{5.17}{=} \int_{E \setminus N} g\, d\mu + \int_{E \cap N} g\, d\mu.$$

The integrals over $E \setminus N$ are equal because $f = g$ on $E \setminus N$. The integrals over $E \cap N$ are both equal to zero, because $\mu(E \cap N) = 0$; cf. Theorem 5.8(d).

Case 2: $f, g \colon X \to [-\infty, \infty]$ are real-valued functions (with possible infinite values). By direct inspection, if $f(x) = g(x)$, then $f^+(x) = g^+(x)$ and $f^-(x) = g^-(x)$. Therefore $f^+ = g^+$ a.e. and $f^- = g^-$ a.e. Now by Case 1

$$\int_E f\, d\mu = \int_E f^+\, d\mu - \int_E f^-\, d\mu = \int_E g^+\, d\mu - \int_E g^-\, d\mu = \int_E g\, d\mu.$$

Case 3: $f, g\colon X \to \mathbb{C}$ are complex-valued functions with finite values. Again, by direct inspection, if $f(x) = g(x)$, then $\texttt{Re}\, f(x) = \texttt{Re}\, g(x)$ and $\texttt{Im}\, f(x) = \texttt{Im}\, g(x)$. Therefore $\texttt{Re}\, f = \texttt{Im}\, g$ a.e. and $\texttt{Re}\, f = \texttt{Im}\, g$ a.e. Now by Case 2

$$\int_E f\, d\mu = \int_E \texttt{Re}\, f\, d\mu + \mathbf{i} \int_E \texttt{Im}\, f\, d\mu = \int_E \texttt{Re}\, g\, d\mu + \mathbf{i} \int_E \texttt{Im}\, g\, d\mu = \int_E g\, d\mu.$$

$\square$

- In plain words: a function f can be modified arbitrarily on a set of measure zero, and this will not affect the value of its integral (over any set E).

Therefore, for the purpose of integration we do not even have to define a function f on the whole space X; it is enough to define it on a set of full measure. We will say that such functions are defined *almost everywhere*.

Definition 7.5. Let $(X, \mathfrak{M}, \mu)$ be a measure space and Y a topological space. We say that a function f with values in Y is **measurable and a.e. defined** on X if there exists $N \subset X$ with $\mu(N) = 0$, such that f is defined on $X \setminus N$ and for every open set $V \subset Y$ we have $f^{-1}(V) \setminus N \in \mathfrak{M}$.

- The function f above may also be defined on N or on any part of N.

Exercise 29. Let f be a function as above. Fix a $y \in Y$ and define

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in X \setminus N \\ y & \text{if } x \in N \end{cases}$$

Show that $\tilde{f}\colon X \to Y$ is measurable.

Exercise 30. Let f be a function as above and μ a complete measure. Let $g\colon X \to Y$ be an arbitrary (not necessarily measurable) function. Define

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in X \setminus N \\ g(x) & \text{if } x \in N \end{cases}$$

Show that $\tilde{f}\colon X \to Y$ is measurable.

Proposition 7.6.

(a) *Let $N_1, N_2, \ldots$ be null sets. Then $N = \cup_{i=1}^{\infty} N_i$ is a null set.*

(b) *Let $A_1, A_2, \ldots$ be full measure sets. Then $A = \cap_{i=1}^{\infty} A_i$ is a full measure set.*

Proof. (Proposition 7.6)

(a) By the σ-subadditivity (Theorem 3.11)

$$\mu\Big(\cup_{i=1}^{\infty} N_i\Big) \le \sum_{i=1}^{\infty} \mu(N_i) = \sum_{i=1}^{\infty} 0 = 0,$$

hence $\mu(N) = 0$.

(b) Note that A_i^c are null sets. Now

$$A^c = \Big(\cap_{i=1}^{\infty} A_i\Big)^c = \cup_{i=1}^{\infty} A_i^c$$

which is a null set due to Part (a). □

- If $f_1, f_2, \dots$ are measurable and a.e. defined functions, then there exists a full measure set $A \subset X$ on which *all* of these functions are defined.

"Almost everywhere" principle.

The convergence theorems proven in Sections 5–6 can be extended to functions defined a.e., and the assumptions made on those functions only need to hold a.e. For example, in Lebesgue's Monotone and Dominated Convergence Theorems we only need to assume that $f_n \to f$ a.e., $|f_n(x)| \le g(x)$ a.e., etc. In Theorem 5.13, we only need to assume that $f(x) = \sum_{n=1}^{\infty} f_n(x)$ a.e., etc.

Theorem 7.7. *Let $\{f_n\}$ be a sequence of complex measurable functions defined a.e. on X and*

$$\sum_{n=1}^{\infty} \int_X |f_n|\, d\mu < \infty.$$

Then the series $f(x) = \sum_{n=1}^{\infty} f_n(x)$ converges for a.e. $x \in X$ and

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

Proof. For each $n \ge 1$ the function f_n is supposed to be defined on a full measure set, call it S_n. Then *all* these functions are defined on the intersection $S = \cap_{n=1}^{\infty} S_n$, and S is a full measure set due to Proposition 7.6(b).

For each $x \in S$ denote $\varphi(x) = \sum_{n=1}^{\infty} |f_n(x)|$. By Theorem 5.13 we have

$$\int_S \varphi\, d\mu = \int_S \sum_{n=1}^{\infty} |f_n|\, d\mu \overset{5.13}{=} \sum_{n=1}^{\infty} \int_S |f_n|\, d\mu < \infty.$$

(note: here we can integrate over X, instead of S, due to Theorem 7.4).

According to Theorem 5.8(f), $E = \{x \in S\colon \varphi(x) < \infty\}$ is a full measure set. For every $x \in E$ we have $\sum_{n=1}^{\infty} |f_n(x)| = \varphi(x) < \infty$, in particular, the series $\sum_{n=1}^{\infty} f_n(x)$ converges absolutely. Therefore $f(x) = \sum_{n=1}^{\infty} f_n(x)$ is well defined (and finite) for each $x \in E$. Also, f is measurable on E, as it is the limit of measurable functions $g_N(x) = \sum_{n=1}^{N} f_n(x)$, as $N \to \infty$. To summarize: f is a measurable function defined almost everywhere.

Next, by the triangle inequality

$$|f(x)| = \left|\sum_{n=1}^{\infty} f_n(x)\right| \le \sum_{n=1}^{\infty} |f_n(x)| = \varphi(x).$$

Since $\int_X \varphi\, d\mu < \infty$, this implies $f \in L^1_\mu(X)$.

Now again we need functions $g_N(x) = \sum_{n=1}^{N} f_n(x)$. By the triangle inequality

$$|g_N(x)| = \left|\sum_{n=1}^{N} f_n(x)\right| \le \sum_{n=1}^{N} |f_n(x)| \le \sum_{n=1}^{\infty} |f_n(x)| = \varphi(x),$$

so g_N's are dominated by the integrable function φ. Also note that

$$g_N(x) = \sum_{n=1}^{N} f_n(x) \to \sum_{n=1}^{\infty} f_n(x) = f(x)$$

for all $x \in E$.

Now by Lebesgue's Dominated Convergence Theorem

$$\sum_{n=1}^{N} \int_X f_n\, d\mu \overset{6.7}{=} \int_X \sum_{n=1}^{N} f_n\, d\mu = \int_X g_N\, d\mu \to \int_X f\, d\mu$$

as $N \to \infty$, proving the theorem. $\square$

Corollary 7.8. **Borel-Cantelli Lemma**

If $E_k \in \mathfrak{M}$ and $\sum_{k=1}^{\infty} \mu(E_k) < \infty$, then almost every $x \in X$ belongs to finitely many of the sets E_k.

Proof. Just apply Theorem 7.7 to functions $f_n = \chi_{E_n}$. $\square$

- The Borel-Cantelli Lemma consists of two parts. Above is the *easier* part. The harder part involves probabilistic notion of independence (beyond the scope of this course).

Exercise 31. In the above corollary, let A be the set of points which belong to infinitely many of the sets E_k. Show that

$$A = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} E_k.$$

Use this fact to prove the corollary without any reference to integration. Hint: use Theorem 3.11.

Definition 7.9. A sequence f_n of functions $X \to \mathbb{C}$ converges to a function $f\colon X \to \mathbb{C}$ **uniformly on** X if for any $\varepsilon > 0$ there exists $N > 0$ such that for all $n > N$

$$\sup_{x \in X} |f_n(x) - f(x)| < \varepsilon.$$

Equivalently, $\sup_{x\in X} |f_n(x) - f(x)| \to 0$ as $n \to \infty$.

- Uniform convergence implies pointwise convergence, but not vice versa.

Example 7. Let $X = (0, 1)$. Functions $f_n(x) = x^n$ converge, as $n \to \infty$, to the function $f(x) = 0$ pointwise, but not uniformly.

Exercise 32. Suppose $\mu(X) < \infty$. Let $f_n \in L^1_\mu(X)$ be complex measurable functions uniformly converging to a function $f \in L^1_\mu(X)$. Prove that

$$\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu. \tag{7.1}$$

Show that the assumption $\mu(X) < \infty$ cannot be omitted, i.e., give an example of a sequence of functions $f_n \in L^1_\mu(X)$ uniformly converging to a function $f \in L^1_\mu(X)$ such that (7.1) fails.

Definition 7.10. A sequence f_n of functions $X \to \mathbb{C}$ converges to a function $f\colon X \to \mathbb{C}$ **in measure** if for any $\varepsilon > 0$ and $\delta > 0$ there exists $N > 0$ such that for all $n > N$

$$\mu\{x \in X : |f_n(x) - f(x)| > \varepsilon\} < \delta.$$

Equivalently, $\mu\{x \in X : |f_n(x) - f(x)| > \varepsilon\} \to 0$ as $n \to \infty$.

- Convergence in measure is, generally, weaker than the pointwise convergence.

Exercise 33. Suppose $\mu(X) < \infty$ and f_n are measurable functions defined a.e. on X. Prove that if $f_n \to f$ a.e. on X, then $f_n \to f$ in measure. What happens if $\mu(X) = \infty$?

Example 8. **Amazing shrinking sliding rectangles**
Suppose $X = [0,1]$ and $\mathbf{m}$ is the Lebesgue measure. Let $f_n = \chi_{[\frac{j}{2^k}, \frac{j+1}{2^k}]}$, where $k = [\log_2 n]$ and $j = n - 2^k$. The first nine terms of this sequence are

$$\chi_{[0,1]},\quad \chi_{[0,\frac{1}{2}]},\quad \chi_{[\frac{1}{2},1]},\quad \chi_{[0,\frac{1}{4}]},\quad \chi_{[\frac{1}{4},\frac{1}{2}]},\quad \chi_{[\frac{1}{2},\frac{3}{4}]}\quad \chi_{[\frac{3}{4},1]},\quad \chi_{[0,\frac{1}{8}]},\quad \chi_{[\frac{1}{8},\frac{1}{4}]}$$

The graphs of these functions are vertical rectangles of the same height $(= 1)$ but decreasing widths $(= \frac{1}{2^k})$, and their bases keep moving (sliding) along the interval $[0, 1]$, from left to right, as n increases. The bases of these rectangles cover the interval $[0, 1]$ over and over again, infinitely many times.
This sequence converges to the zero function $f \equiv 0$ in measure, because

$$\mathbf{m}\big(x\colon \{|f_n(x) - f(x)| > \varepsilon\}\big) = \frac{1}{2^k} \to 0 \quad \text{as} \quad n \to \infty$$

but $f_n(x)$ has no limit, as $n \to \infty$ for any point $x \in [0, 1]$.

Exercise 34. Prove that if $f_n \to f$ in measure, then there is a subsequence $\{f_{n_k}\}$ of $\{f_n\}$ such that $f_{n_k} \to f$ a.e. on X. Hint: use Corollary 7.8.

Exercise 35. [Bonus] Suppose $f \in L^1_\mu(X)$. Prove that $\forall\varepsilon > 0\ \exists\delta > 0$ such that $\int_E |f|\,d\mu < \varepsilon$ whenever $\mu(E) < \delta$.

Our next general goal is to show that the values of the integrals for a function f over different sets $E \in \mathfrak{M}$ give plenty of information about f itself.

Theorem 7.11. *Let $(X, \mathfrak{M}, \mu)$ be a measure space.*

(a) *If $f\colon X \to [0, \infty]$ is measurable, $E \in \mathfrak{M}$, and $\int_E f\,d\mu = 0$, then $f = 0$ a.e. on E;*

(b) *If $f \in L^1_\mu(X)$ and $\int_E f\,d\mu = 0$ for every $E \in \mathfrak{M}$, then $f = 0$ a.e. on X;*

Proof. **(a)** We need to prove that $\mu(S) = 0$, where

$$S = \{x \in E\colon f(x) > 0\} = f^{-1}((0, \infty]) \cap E.$$

Note that $S = \cup_{n\geq 1} S_n$, where

$$S_n = \{x \in E\colon f(x) > 1/n\} = f^{-1}((1/n, \infty]) \cap E$$

and $S_1 \subset S_2 \subset \cdots$. Thus by Continuity-II (Theorem 3.13) we have $\mu(S) = \lim_{n\to\infty} \mu(S_n)$. If $\mu(S) > 0$, then $\mu(S_n) > 0$ for some n, and then

$$\int_E f\,d\mu \overset{5.8\text{(b)}}{\geq} \int_{S_n} f\,d\mu \overset{5.8\text{(a)}}{\geq} \int_{S_n} \tfrac{1}{n}\,d\mu = \tfrac{1}{n}\,\mu(S_n) > 0,$$

a contradiction. Thus $\mu(S) = 0$ as claimed.

(b) If $f\colon X \to \mathbb{C}$ is a complex-valued function, then

$$\begin{aligned}\int_E f\,d\mu = 0 \quad &\Longrightarrow \quad \int_E \mathtt{Re}\,f\,d\mu + \mathbf{i}\int_E \mathtt{Im}\,f\,d\mu = 0\\ &\Longrightarrow \quad \int_E \mathtt{Re}\,f\,d\mu = 0 \quad \text{and} \quad \int_E \mathtt{Im}\,f\,d\mu = 0\end{aligned}$$

thus it is enough to prove the claim for real-valued functions $f\colon X \to \mathbb{R}$. Denote

$$E_1 = \{x\colon f(x) \geq 0\} = f^{-1}([0,\infty))$$
$$E_2 = \{x\colon f(x) \leq 0\} = f^{-1}((-\infty,0])$$

Since $\int_{E_1} f\,d\mu = 0$ by our assumption, Part (a) implies $f = 0$ a.e. on E_1. Similarly, $\int_{E_2} f\,d\mu = 0$ by our assumption, then $\int_{E_2}(-f)\,d\mu = 0$, and again Part (a) implies $f = 0$ a.e. on E_2. Thus $f = 0$ a.e. on $E_1 \cup E_2 = X$. $\square$

Corollary 7.12. *If $f \in L^1_\mu(X)$ and*

$$\left|\int_X f\,d\mu\right| = \int_X |f|\,d\mu,$$

then there exists a constant $\theta \in [0, 2\pi)$ such that $f = e^{\mathbf{i}\theta}|f|$ a.e. on X.

Proof. Recall: in the proof of Theorem 6.8 we showed that $\exists\theta \in [0,2\pi)$ such that

$$\left|\int_X f\,d\mu\right| = \int_X \mathtt{Re}\left(e^{-\mathbf{i}\theta}f\right)d\mu \leq \int_X |f|\,d\mu$$

and mentioned that $|f| - \mathtt{Re}\left(e^{-\mathbf{i}\theta}f\right) \geq 0$. Now due to our assumption

$$\int_X \left(|f| - \mathtt{Re}\left(e^{-\mathbf{i}\theta}f\right)\right)d\mu = 0$$

thus

$$|f| - \mathtt{Re}\left(e^{-\mathbf{i}\theta}f\right) = 0 \qquad \text{a.e.}$$

This also implies that $\mathtt{Im}\left(e^{-\mathbf{i}\theta}f\right) = 0$ a.e. Therefore $|f| = e^{-\mathbf{i}\theta}f$ a.e. $\square$

Exercise 36.

(a) Let $f\colon X \to (0,\infty]$ and $\mu(E) > 0$. Show that $\int_E f\, d\mu > 0$.

(b) Let $f, g \in L^1_\mu(X)$ be real-valued functions and $f < g$. Assuming $\mu(X) > 0$ show

$$\int_E f\, d\mu < \int_E g\, d\mu$$

Definition 7.13. Let $(X, \mathfrak{M}, \mu)$ be a measure space and $f\colon X \to \mathbb{C}$ a measurable function. The **average value** of $f(x)$ over a set $E \in \mathfrak{M}$ is

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

provided the integral exists and $\mu(E) \in (0, \infty)$.

Lemma 7.14.

(a) *If $f\colon X \to \mathbb{R}$ is a real-valued function and $m \le f(x) \le M$ for all $x \in E$, then $m \le A_E(f) \le M$;*

(b) *If $f\colon X \to \mathbb{C}$ is a complex-valued function and $f(x) \in R$ for all $x \in E$, where $R \subset \mathbb{C}$ is a closed rectangular domain with horizontal and vertical sides, then $A_E(f) \in R$.*

Proof.

(a) By Theorem 5.8(a), $m\mu(E) \le \int_E f(x)\, d\mu \le M\mu(E)$.

(b) just apply Part (a) to $\mathrm{Re}\, f$ and $\mathrm{Im}\, f$ separately. □

Theorem 7.15. *Let $(X, \mathfrak{M}, \mu)$ be a measure space and $\mu(X) < \infty$. Let $f \in L^1_\mu(X)$ and $S \subset \mathbb{C}$ a closed set. Suppose for every $E \in \mathfrak{M}$ with $\mu(E) > 0$ we have $A_E(f) \in S$. Then $f(x) \in S$ a.e. on X.*

Proof. For any closed rectangle $R \subset \mathbb{C} \setminus S$ let $E_R = \{x\colon f(x) \in R\} = f^{-1}(R)$. If $\mu(E_R) > 0$, we would have $A_{E_R}(f) \in R$ by Lemma 7.14(b), which contradicts our assumption. Therefore $\mu(E_R) = 0$. Now every open set in the plane is a countable union of some closed rectangles. In particular, $\mathbb{C} \setminus S = \cup_{i=1}^{\infty} R_i$, hence

$$f^{-1}(S) = \cup_{i=1}^{\infty} f^{-1}(R_i) \quad \Longrightarrow \quad \mu(f^{-1}(S)) \le \sum_{i=1}^{\infty} \mu(f^{-1}(R_i)) = 0$$

thus $f(x) \in S$ a.e. $\square$

Corollary 7.16. *In particular, if $\int_E f\, d\mu$ is real-valued for every E, then f is real-valued a.e. on X.*

8 Regularity of the Lebesgue measure

Theorem 8.1. *For any measurable set $E \subset \mathbb{R}$ and any $\varepsilon > 0$ there is an open covering set $V \supset E$ and a closed subset $D \subset E$ such that $\mathbf{m}(V \setminus D) < \varepsilon$. In particular,*

$$\mathbf{m}(V \setminus E) < \varepsilon \qquad \text{and} \qquad \mathbf{m}(E \setminus D) < \varepsilon$$

Proof.
(i) Suppose first that E is bounded (i.e. $E \subset I = [a, b]$ for some $a < b$).
Then $\mathbf{m}(E) < \infty$ and using the outer measure (Definition 1.13) we get

$$\mathbf{m}(E) = \mu^*(E) = \inf \sum_{i=1}^{\infty} |I_i|$$

where the infimum is taken over all countable covers of E by intervals, i.e., such that $E \subset \cup_{i=1}^{\infty} I_i$. Thus for any $\varepsilon > 0$ there exist covering intervals $\cup_{i=1}^{\infty} I_i \supset E$ such that

$$\sum_{i=1}^{\infty} |I_i| - \mathbf{m}(E) < \varepsilon/2$$

Extending I_i's slightly we can make them open and increase their total length by no more than $\varepsilon/2$. More precisely, for each $i \geq 1$ we find an open interval $I_i' \supset I_i$ such that $|I_i'| - |I_i| < \varepsilon/2^{i+1}$. Then we have $E \subset \cup_{i=1}^{\infty} I_i'$ and

$$\sum_{i=1}^{\infty} |I_i'| - \mathbf{m}(E) < \sum_{i=1}^{\infty} |I_i| + \varepsilon/2 - \mathbf{m}(E) < \varepsilon$$

The open set $V = \cup_{i=1}^{\infty} I_i'$ covers E and has the desired property:

$$\mathbf{m}(V \setminus E) = \mathbf{m}(V) - \mathbf{m}(E) \leq \sum_{i=1}^{\infty} |I_i'| - \mathbf{m}(E) < \varepsilon$$

(ii) Let E be unbounded. Then we can represent $\mathbb{R}$ as a disjoint union of some intervals, i.e., $\mathbb{R} = \uplus_{k=1}^{\infty} I_k$. Each set $E_k = E \cap I_k$ is bounded and we can find an open set $V_k \supset E_k$ such that $\mathbf{m}(V_k \setminus E_k) < \varepsilon/2^k$. Now the set $V = \cup_{k=1}^{\infty} V_k$ is open and

$$\mathbf{m}(V \setminus E) \leq \sum_{k=1}^{\infty} \mathbf{m}(V_k \setminus E_k) < \varepsilon.$$

(iii) The set $E^c = \mathbb{R} \setminus E$ is measurable, so due to (i)–(ii) there exists an open set $U \supset E^c$ such that $\mathbf{m}(U \setminus E^c) < \varepsilon$. Its complement $D = U^c$ is a closed set, $D \subset E$, and $E \setminus D = U \setminus E^c$, hence $\mathbf{m}(E \setminus D) = \mathbf{m}(U \setminus E^c) < \varepsilon$. □

In other words, any measurable set can be arbitrarily well approximated by open sets 'from outside' and by closed sets 'from inside'.

The following exercise demonstrates that approximation by open sets 'from inside' or by closed sets 'from outside' is a bad idea:

EXERCISE 37. Find examples of Lebesgue measurable sets $E_1, E_2 \subset \mathbb{R}$ such that

$$\mathbf{m}(E_1) < \inf\{\mathbf{m}(A) \colon E_1 \subset A, \quad A \text{ closed}\}$$

$$\mathbf{m}(E_2) > \sup\{\mathbf{m}(V) \colon V \subset E_2, \quad V \text{ open}\}.$$

The next theorem uses compact sets instead of closed sets:

Theorem 8.2.

(i) *For any measurable set $E \subset \mathbb{R}$ we have*

$$\mathbf{m}(E) = \inf\{\mathbf{m}(V) \colon E \subset V, \quad V \text{ open}\} \tag{8.1}$$

and

$$\mathbf{m}(E) = \sup\{\mathbf{m}(K) \colon K \subset E, \quad K \text{ compact}\} \tag{8.2}$$

(ii) *If $\mathbf{m}(E) < \infty$, then for any $\varepsilon > 0$ there exist an open covering set $V \supset E$ and a compact subset $K \subset E$ such that $\mathbf{m}(V \setminus K) < \varepsilon$. In particular,*

$$\mathbf{m}(V \setminus E) < \varepsilon \qquad \text{and} \qquad \mathbf{m}(E \setminus K) < \varepsilon$$

Proof. Put $\mathbb{R} = \cup_{n=1}^{\infty} I_n$, where $I_n = [-n, n]$.

(a) First suppose that $\mathbf{m}(E) = \infty$. Then (8.1) is trivial: for any open cover $V \supset E$ we have $\mathbf{m}(V) \geq \mathbf{m}(E) = \infty$. To prove (8.2), we use Theorem 8.1 by which there exists a closed set $D \subset E$ such that $\mathbf{m}(E \setminus D) < 1$. In that case $\mathbf{m}(D) \geq \mathbf{m}(E) - 1 = \infty$, hence $\mathbf{m}(D) = \infty$. By standard Definition 2.1, $\mathbf{m}(D) = \lim_{n\to\infty} \mathbf{m}(D \cap I_n)$, hence $\sup_n \mathbf{m}(D \cap I_n) = \infty$. Each set $D \cap I_n \subset E$ is compact, which proves (8.2).

(b) Now suppose that $\mathbf{m}(E) < \infty$. Now claim (i) follows from the stronger claim (ii), so it is enough to prove (ii). Due to Theorem 8.1, there exist an open covering set $V \supset E$ and a closed subset $D \subset E$ such that $\mathbf{m}(V \setminus D) < \varepsilon/2$. Since $\mathbf{m}(D) \leq \mathbf{m}(E) < \infty$ and $\mathbf{m}(D) = \lim_{n\to\infty} \mathbf{m}(D \cap I_n)$, we can find $n \geq 1$ such that $\mathbf{m}(D \cap I_n) > \mathbf{m}(D) - \varepsilon/2$. Now the set $K = D \cap I_n$ is compact, $K \subset E$, and $\mathbf{m}(V \setminus K) = \mathbf{m}(V \setminus D) + \mathbf{m}(D \setminus K) < \varepsilon$. □

Lebesgue measurable sets may be very complicated and "ugly", but the above theorems say that they are "approximately open" and "approximately closed". Sets with finite measure are "approximately compact".

The following approximation is also useful:

Corollary 8.3. *If $\mathbf{m}(E) < \infty$, then for every $\varepsilon > 0$ there exist a finite union of disjoint bounded intervals $J = \biguplus_{n=1}^{N} I_n$ such that $\mathbf{m}(E \Delta J) < \varepsilon$.*

Proof. Due to Theorem 8.2, for any $\varepsilon > 0$ there exists an open set $V \supset E$ such that $\mathbf{m}(V \setminus E) < \varepsilon/2$. Any open set is a finite or countable union of disjoint open intervals, thus $V = \biguplus_{n=1}^{\infty} I_n$. Note that $\mathbf{m}(V) = \sum_{n=1}^{\infty} |I_n| < \infty$, i.e., this series converges. Thus there is $N \geq 1$ such that $\sum_{n>N} |I_n| < \varepsilon/2$. Set $J = \biguplus_{n=1}^{N} I_n$. Now

$$E \Delta J = (E \setminus J) \cup (J \setminus E) \subset (V \setminus J) \cup (V \setminus E),$$

hence

$$\mathbf{m}(E \Delta J) \leq \mathbf{m}(V \setminus J) + \mathbf{m}(V \setminus E) < \varepsilon.$$

□

Theorem 8.4. *For every measurable set $E \in \mathfrak{M}$*

(a) *there exists a G_δ-set G such that $E \subset G$ and $\mathbf{m}(G \setminus E) = 0$;*

(b) *there exists an F_σ-set F such that $F \subset E$ and $\mathbf{m}(E \setminus F) = 0$.*

Proof. Due to Theorem 8.1, there are open covering sets $V_n \supset E$ and closed subsets $D_n \subset E$ such that $\mathbf{m}(V_n \setminus D_n) < 1/n$. Put $G = \cap_{n=1}^{\infty} V_n$ and $F = \cup_{n=1}^{\infty} D_n$. Clearly, G is G_δ and F is F_σ, and $G \supset E \supset F$. Also note that $G \setminus F \subset V_n \setminus D_n$ for every $n \geq 1$, hence $\mathbf{m}(G \setminus F) \leq \mathbf{m}(V_n \setminus D_n) < 1/n$, thus $\mathbf{m}(G \setminus F) = 0$. □

Thus every measurable set is 'almost' G_δ (and 'almost' F_σ), up to a null set.

Regularity in $\mathbb{R}^k$.

All the above theorems and proofs extend, almost verbatim, to the Lebesgue measure $\mathbf{m}$ in $\mathbb{R}^k$, $k \geq 2$. Instead of intervals, one needs to use rectangles in $\mathbb{R}^2$, rectangular boxes in $\mathbb{R}^3$, etc.

More generally, we can consider other measures in $\mathbb{R}^k$:

Definition 8.5. A measure μ defined on a σ-algebra $\mathfrak{M}$ in $\mathbb{R}^k$ is called a **Borel measure** if it is defined on all Borel sets (i.e. $\mathfrak{M}$ contains all the Borel sets).

Definition 8.6. A Borel measure μ is said to be **outer regular** if

$$\mu(E) = \inf\{\mu(V)\colon E \subset V, \quad V \text{ open}\} \quad \forall E \in \mathfrak{M}.$$

A Borel measure μ is said to be **inner regular** if

$$\mu(E) = \sup\{\mu(K)\colon K \subset E, \quad K \text{ compact}\} \quad \forall E \in \mathfrak{M}.$$

A Borel measure μ is said to be **regular** if it is both outer regular and inner regular.

For regular measures, the above approximation results hold true:

Exercise 38. Let $f\colon \mathbb{R} \to [0,\infty]$ be a Borel function, $f \in L^1_{\mathbf{m}}(\mathbb{R})$, and consider the measure $\rho(E) = \int_E f \, d\mathbf{m}$, where $\mathbf{m}$ is the Lebesgue measure. Prove that ρ is regular. Hint: use the result of Exercise 35.

Theorem 8.7. *If $\mu(E) < \infty$ for any bounded set $E \subset \mathbb{R}^k$, then the outer regularity and inner regularity are equivalent.*

Proof. It is basically a repetition of our arguments in the proofs of Theorem 8.1 and Theorem 8.2, so we omit it. □

Exercise 39. Show that the counting measure in $\mathbb{R}^k$ is inner regular, but not outer regular.

9 Approximation of Lebesgue measurable functions

In the previous section we showed Lebesgue measurable sets can be arbitrarily well approximated by open sets ('from outside') and by compact sets ('from inside'). In this section, we will approximate Lebesgue measurable functions arbitrarily well by step and continuous functions.

DEFINITION 9.1. A function $f\colon \mathbb{R} \to \mathbb{C}$ is **Lebesgue measurable** if for any open set $V \subset \mathbb{C}$ its preimage $f^{-1}(V)$ is a Lebesgue measurable set, i.e., belongs to the Lebesgue σ-algebra.

Previously, the term *measurable functions* was always understood in the sense of Borel functions (see Definition 4.3). Now we have two types of measurable functions – Borel measurable and Lebesgue measurable. Note that Borel measurable functions are Lebesgue measurable, but not vice versa.

$$f : \text{Borel} \overset{\Leftarrow\!\!\!\!/}{\Rightarrow} f : \text{Lebesgue}$$

DEFINITION 9.2. A function $\varphi\colon \mathbb{R} \to \mathbb{C}$ is called a **step function** if $\varphi = \sum_{i=1}^{n} \alpha_i \chi_{I_i}$ for some $\alpha_i \in \mathbb{C}$ and disjoint finite intervals $I_i \in \mathbb{R}$.

Do not confuse step functions with simple functions. Every step function is simple, but not vice versa. For example, Dirichlet function (Example 6) is a simple function but not a step function.

Roughly speaking, simple functions have simple range (a finite set) but may have arbitrarily complicated domains (a measurable set). Step functions have simple range (a finite set) and simple domain (a finite set of intervals).

EXERCISE 40. Let $s\colon [a,b] \to \mathbb{R}$ be a simple Lebesgue measurable function. Show that for every $\varepsilon > 0$ there is a step function $\varphi\colon [a,b] \to \mathbb{R}$ and a Lebesgue measurable set $E \subset [a,b]$ such that $s(x) = \varphi(x)$ on E and $\mathbf{m}\big([a,b] \setminus E\big) < \varepsilon$. Hint: use the regularity of $\mathbf{m}$.

EXERCISE 41. Let $f\colon [a,b] \to \mathbb{R}$ be a Lebesgue measurable function. Show that for every $\varepsilon > 0$ there is a step function $g\colon [a,b] \to \mathbb{R}$ such that

$$\mathbf{m}\big\{x \in [a,b] \colon |f(x) - g(x)| \geq \varepsilon\big\} < \varepsilon.$$

Hint: use approximation by simple functions and then the previous exercise.

EXERCISE 42. Let $f \in L^1_{\mathbf{m}}(\mathbb{R})$. Prove that there is a sequence $\{g_n\}$ of step functions such that

$$\lim_{n\to\infty} \int_{\mathbb{R}} |f - g_n| \, d\mathbf{m} = 0.$$

Hint: use the previous exercise.

Definition 9.3. Let X be a topological space and $f\colon X \to \mathbb{C}$. The **support** of f is defined by

$$\operatorname{supp} f = \overline{\{x \in X\colon f(x) \neq 0\}}.$$

Also, $C_c(X)$ denotes the set of all continuous functions with compact support.

Theorem 9.4. (Lusin)

Let $f\colon \mathbb{R} \to \mathbb{C}$ be a Lebesgue measurable function, $A \subset \mathbb{R}$ a Lebesgue measurable set with $\mathbf{m}(A) < \infty$ such that $f(x) = 0$ for all $x \notin A$. Then

(a) *For every $\varepsilon > 0$ there exists a compact set $K \subset A$ such that $\mathbf{m}(A \setminus K) < \varepsilon$ and the restriction of f to K is a continuous function.*

(b) *There exists $g \in C_c(\mathbb{R})$ such that $\mathbf{m}\{x\colon f(x) \neq g(x)\} < \varepsilon$. Furthermore, we may arrange it so that* $\sup|g(x)| \leq \sup|f(x)|$.

Proof.

We can assume that $f\colon \mathbb{R} \to \mathbb{R}$, as otherwise we can deal with the real part and imaginary part of f separately.

(a) Let $V_1, V_2, \ldots \subset \mathbb{R}$ denote all open intervals with rational endpoints.

By Theorem 8.2, $\exists$ compact sets $K_n \subset f^{-1}(V_n) \cap A$ and $K_n' \subset f^{-1}(V_n^c) \cap A$ s.t.

$$\mathbf{m}\big((f^{-1}(V_n) \cap A) \setminus K_n\big) < \tfrac{\varepsilon}{2^{n+1}} \quad \text{and} \quad \mathbf{m}\big((f^{-1}(V_n^c) \cap A) \setminus K_n'\big) < \tfrac{\varepsilon}{2^{n+1}}$$

hence

$$\mathbf{m}\big(A \setminus (K_n \cup K_n')\big) < \varepsilon/2^n$$

Now define $K = \cap_{n=1}^{\infty}(K_n \cup K_n')$. It is a compact set, $K \subset A$, and $\mathbf{m}(A \setminus K) < \varepsilon$

Let $f_K = f|_K$ denote the restriction of f to K. Note that $f_K^{-1}(V_n) = K \cap (\mathbb{R} \setminus K_n')$, which is an open subset of K, hence f_K is continuous.

(b) By Corollary 8.3, there exists a finite union of disjoint bounded open intervals $J = \uplus_{n=1}^{N} I_n$ such that $\mathbf{m}(A \Delta J) < \varepsilon$. By the part (a), for each $n = 1, \ldots, N$ there exists a compact subset $K_n \subset A_n\colon = A \cap I_n$ such that $\mathbf{m}(A_n \setminus K_n) < \varepsilon/N$ and the restriction of f to K_n is continuous.

Now we define g on each interval $I_n = (a_n, b_n)$ as follows:

(i) $g(x) = f(x)$ for all $x \in K_n$

(ii) $g(a_n) = g(b_n) = 0$

(iii) $g(x)$ is extended linearly and continuously to the rest of I_n

We clarify the meaning of (iii). The set $V = I_n \setminus K_n$ is open so it is a finite or countable union of disjoint open intervals $I_{n,i} = (a_{n,i}, b_{n,i})$, $i \geq 1$, such that the endpoints $a_{n,i}, b_{n,i}$ belong to $K_n \cup \{a_n\} \cup \{b_n\}$, and on this union g is already defined.

Now we set

$$g(x) = g(a_{n,i}) + \frac{g(b_{n,i}) - g(a_{n,i})}{b_{n,i} - a_{n,i}} (x - a_{n,i}). \qquad \forall x \in I_{n,i}$$

Lastly, we set $g(x) = 0$ for all $x \in \mathbb{R} \setminus J$. One can verify by direct inspection that g is now continuous on $\mathbb{R}$ and has compact support (i.e. $g \in C_c(\mathbb{R})$).

We now have

$$\begin{aligned} \mathbf{m}\big\{x\colon f(x) \neq g(x)\big\} &\leq \mathbf{m}(A \setminus J) + \sum_{n=1}^{N} \mathbf{m}(I_n \setminus K_n) \\ &\leq \mathbf{m}(A \Delta J) + \sum_{n=1}^{N} \mathbf{m}(A_n \setminus K_n) < 2\varepsilon, \end{aligned}$$

□

Corollary 9.5. *Assume that the hypotheses of Lusin's theorem are satisfied and that $|f| \leq M$. Then there is a sequence $\{g_n\}$ such that $g_n \in C_c(\mathbb{R})$ and $|g_n| \leq M$ for each $n \geq 1$, and $g_n(x) \to f(x)$ a.e.*

Proof. By Lusin Theorem 9.4, for each $n \geq 1$ there exists $g_n \in C_c(\mathbb{R})$ such that

$$\mathbf{m}(E_n) < 2^{-n}, \qquad E_n = \big\{x\colon f(x) \neq g_n(x)\big\}.$$

Since $\sum_{n=1}^{\infty} \mathbf{m}(E_n) < \infty$, the analogue of the Borel-Cantelli Lemma (Corollary 7.8) implies that a.e. point $x \in \mathbb{R}$ belongs to finitely many E_n's. That is, for a.e. $x \in \mathbb{R}$ there exists $N(x)$ such that $x \notin E_n$ for all $n > N(x)$.

This implies $g_n(x) = f(x)$ for all $n > N(x)$, in particular $g_n(x) \to f(x)$. □

Definition 9.6. Let X be a topological space and $f\colon X \to \mathbb{R}$ (or $[-\infty, \infty]$). We say that f is **lower semicontinuous** if $\{x\colon f(x) > a\}$ is open for every $a \in \mathbb{R}$. We say that f is **upper semicontinuous** if $\{x\colon f(x) < a\}$ is open for every $a \in \mathbb{R}$.

Proposition 9.7. *Simple properties of semicontinuous functions:*

(a) *f is continuous $\Leftrightarrow$ f is both upper and lower semicontinuous.*

(b) *If f is semicontinuous, then it is Borel.*

(c) *f is upper semicontinuous $\Leftrightarrow$ $-f$ is lower semicontinuous.*

(d) *If f is upper (lower) semicontinuous and $c > 0$, then cf is upper (lower) semicontinuous.*

(e) *If f, g are upper (lower) semicontinuous, then $f + g$ is upper (lower) semicontinuous.*

(f) *If $\{f_\gamma\}$ is a family of upper (lower) semicontinuous functions, then $\inf f_\gamma$ (resp., $\sup f_\gamma$)is upper (lower) semicontinuous.*

(g) *$V \subset X$ is open $\Leftrightarrow$ χ_V is lower semicontinuous.*

(h) *$A \subset X$ is closed $\Leftrightarrow$ χ_A is upper semicontinuous.*

(i) *If $f_n \geq 0$ are lower semicontinuous, then $\sum_{n=1}^{\infty} f_n$ is lower semicontinuous.*

• Proof is straightforward, we leave it as an exercise.

Exercise 43. Prove that if $f\colon X \to \mathbb{R}$ is upper (lower) semicontinuous and X is compact, then f is bounded above (below) and attains its maximum (minimum).

Theorem 9.8. (Vitali-Caratheodory)
Let $f\colon \mathbb{R} \to \mathbb{R}$ be Lebesgue integrable function (i.e. $f \in L^1$). Then for every $\varepsilon > 0$ there exist functions $u, v\colon \mathbb{R} \to \mathbb{R}$, $u \leq f \leq v$, where u is upper semicontinuous and bounded above, v is lower semicontinuous and bounded below, and

$$\int_{\mathbb{R}} (v - u)\, d\mathbf{m} < \varepsilon.$$

Proof. **Vitali-Caratheodory** Theorem 9.8

First, assume $f \geq 0$. By Theorem 4.22, there exist simple functions $0 \leq s_1 \leq s_2 \leq \cdots$ converging to f pointwise. Then $f = \sum_{n=1}^{\infty} t_n$, where $t_n = s_n - s_{n-1}$ (taking $t_0 = 0$) are nonnegative simple functions. Thus

$$f = \sum_{i=1}^{\infty} \alpha_i \chi_{A_i} \tag{9.1}$$

where $\alpha_i > 0$ and $A_i \subset \mathbb{R}$ are measurable sets (not necessarily disjoint).

We cam assume that $\mathbf{m}(A_i) < \infty$ for each i, as otherwise we can partition A_i into countably many pieces $A_{ij} = A_i \cap [j, j+1)$, each of finite measure, and replace $\alpha_i \chi_{A_i}$ with $\sum_{j=1}^{\infty} \alpha_i \chi_{A_{ij}}$ in (9.1).

Note that $\int_{\mathbb{R}} f \, d\mathbf{m} = \sum_{i=1}^{\infty} \alpha_i \mathbf{m}(A_i) < \infty$.

By Theorem 8.2 there exist compact sets K_i and open sets V_i such that $K_i \subset A_i \subset V_i$ and $\mathbf{m}(V_i \setminus K_i) < \frac{\varepsilon}{\alpha_i 2^i}$ Now we define

$$v = \sum_{i=1}^{\infty} \alpha_i \chi_{V_i} \quad \text{and} \quad u = \sum_{i=1}^{N} \alpha_i \chi_{K_i}$$

where N is chosen so that

$$\sum_{i>N} \alpha_i \mathbf{m}(A_i) < \varepsilon$$

Then u is upper semicontinuous, by Proposition 9.7 (e),(h), and bounded above, v is lower semicontinuous, by Proposition 9.7 (g),(i), and bounded below (by zero).

Thus,

$$\begin{aligned} v - u &= \sum_{i=1}^{N} \alpha_i \chi_{V_i \setminus K_i} + \sum_{i=N+1}^{\infty} \alpha_i \chi_{V_i} \\ &\leq \sum_{i=1}^{\infty} \alpha_i \chi_{V_i \setminus K_i} + \sum_{i=N+1}^{\infty} \alpha_i \chi_{A_i} \end{aligned}$$

so that $\int_{\mathbb{R}} (v - u) \, d\mathbf{m} < 2\varepsilon$.

In the general case, $f = f^+ - f^-$, and we attach u_1 and v_1 to f^+ and u_2 and v_2 to f^-, as above, and put $u = u_1 - v_2$ and $v = v_1 - u_2$. Note that $-v_2$ is upper semicontinuous and $-u_2$ is lower semicontinuous (see Proposition 9.7 (c)). $\square$

Corollary 9.9. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a Lebesgue measurable function. Then there exist Borel functions $g, h\colon \mathbb{R} \to \mathbb{R}$ such that $g(x) = h(x)$ a.e. and $g(x) \le f(x) \le h(x)$ for every $x \in X$.*

Proof. First, we prove the theorem for a function restricted on a finite interval.

Let $N > 0$ and let us restrict our function f onto the interval $[-N, N]$.
For any $M \ge 1$, let

$$f_M(x) = \begin{cases} f(x) & \text{if } |f(x)| \le M \\ M & \text{if } f(x) > M \\ -M & \text{if } f(x) < -M \end{cases}$$

be the function f truncated at levels M and $-M$. Since f_M is bounded on the finite interval $[-N, N]$, it is integrable.

By Vitali-Caratheodory Theorem 9.8, For all $n \ge 1$ there exist the respective functions $u_{M,n} \le f_M \le v_{M,n}$ such that

$$\int_{[-N,N]} (v_{M,n} - u_{M,n})\, d\mathbf{m} < 1/n.$$

Let $u_M = \sup_n u_{M,n}$ and $v_M = \inf_n v_{M,n}$. Then $u_M \le f_M \le v_M$ and

$$\int_{[-N,N]} (v_M - u_M)\, d\mathbf{m} = 0.$$

This implies $u_M = v_M$ a.e. on $[-N, N]$, i.e., $\mathbf{m}(E_M) = 0$, where

$$E_M = \{x \in [-N, N] \colon u_M(x) \ne v_M(x)\}.$$

Taking the limit as $M \to \infty$ we obtain $f = \lim f_M$; in fact $f(x) = f_M(x)$ for all $M > M_0(x)$ (the sequence $f_M(x)$ "stabilizes" as $M \to \infty$).

Now define two Borel functions

$$g_N = \limsup_{M\to\infty} u_M \qquad \text{and} \qquad h_N = \liminf_{M\to\infty} v_M.$$

By direct inspection, $g_N \le f \le h_N$ on $[-N, N]$. Also, $g_N(x) = h_N(x)$ for all $x \in [-N, N] \setminus \cup_M E_M$, hence $g_N(x) = h_N(x)$ a.e. on $[-N, N]$.

Finally, we define $g(x) = \limsup g_N(x)$ and $h(x) = \liminf h_N(x)$ on the entire $\mathbb{R}$. □

Theorem 9.10. (Egorov)
Let $\mu(X) < \infty$ and $\{f_n\}$ a sequence of complex measurable functions on X such that $\lim_{n\to\infty} f_n(x) = f(x)$ for a.e. $x \in X$. Then for every $\varepsilon > 0$ there is $E \subset X$ such that $\mu(X \setminus E) < \varepsilon$ and $f_n \to f$ uniformly on E.

Proof. For all $i \geq 1$ and $k \geq 1$ consider the "bad" set

$$B_{i,k} = \{x \in X \colon |f_i(x) - f(x)| \geq 1/k\}.$$

Due to a.e. convergence, for each $k \geq 1$ a.e. point can only belong to finitely many of $B_{i,k}$, thus $\mu\big(\cap_{n\geq 1} \cup_{i\geq n} B_{i,k}\big) = 0$. By Continuity-I (Theorem 3.12) we have $\mu\big(\cup_{i\geq n} B_{i,k}\big) \to 0$ as $n \to \infty$

Given $\varepsilon > 0$, for each $k \geq 1$ there exists $n_{k,\varepsilon} \geq 1$ such that $\mu\big(\cup_{i\geq n_{k,\varepsilon}} B_{i,k}\big) < \varepsilon/2^k$
Let

$$B_\varepsilon = \bigcup_{k\geq 1} \bigcup_{i\geq n_{k,\varepsilon}} B_{i,k}.$$

Note that

$$\mu(B_\varepsilon) \leq \sum_{k\geq 1} \mu\Big(\bigcup_{i\geq n_{k,\varepsilon}} B_{i,k}\Big) < \varepsilon$$

and

$$X \setminus B_\varepsilon = B_\varepsilon^c = \bigcap_{k\geq 1} \bigcap_{i\geq n_{k,\varepsilon}} B_{i,k}^c,$$

i.e., for all $x \in B^c$ we have $|f_i(x) - f(x)| < 1/k$ for all $i \geq n_{k,\varepsilon}$, which means precisely a uniform convergence on $E = B_\varepsilon^c$. $\square$

Littlewood's three principles of real analysis.

Littlewood stated three principles in his 1944 Lectures on the Theory of Functions. They can be roughly expressed as follows:

- Every measurable set is nearly a finite union of intervals
- Every convergent sequence of functions is nearly uniformly convergent
- Every integrable function (of class L^p) is nearly continuous

The first principle is given by Corollary 8.3. The second one by Egorov's Theorem (Theorem 9.10). The third one, by Lusin's Theorem (Theorem 9.4)

(Another version will be given in Section 11).

10 Lebesgue integral versus Riemann integral

Riemann integral (review).

Recall the definition of Riemann integral in calculus: given a function on a bounded interval $f\colon [a,b] \to \mathbb{R}$, we consider all finite partitions $I = \uplus_{i=1}^{n} I_i$ of $I = [a,b]$ into subintervals I_i and denote $\Delta = \{I_i\}$ and $|\Delta| = \max_{1\le i\le n} |I_i|$. Now the Riemann integral is defined by

$$\int_a^b f(x)\, dx = \lim_{|\Delta|\to 0} \sum_{i=1}^{n} f(x_i)|I_i|,$$

where $x_i \in I_i$, provided the (finite) limit exists and does not depend on the details of the partitions $\Delta = \{I_i\}$ and the choice of points $x_i \in I_i$.

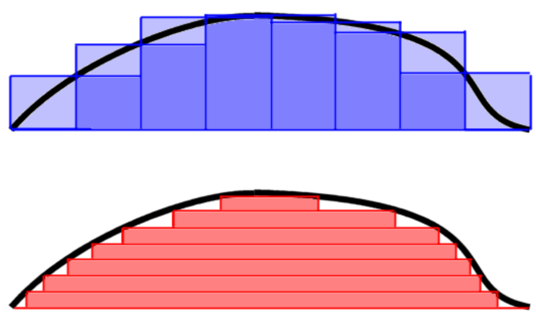


Figure 5: Riemann integration (top, blue) versus Lebesgue integration (bottom, red).

Both integrals, Riemann and Lebesgue, are designed to compute "the area under the graph of the function". In the Riemann integration, we divide the *domain* of the function into small pieces and count the total area of the resulting *vertical* strips. In the Lebesgue integration, we divide the *range* of the function into small pieces and count the total area of the resulting *horizontal* strips.

Roughly speaking, these are different ways of counting money when you've got a pile of paper bills of various denomination. One way is to add up the values of all the bills, one by one...

Like "one dollar plus ten dollars, plus five dollars, plus another one, plus another five..."

This is the idea of Riemann integration.

Another way is to sort the bills according to their denomination first – singles into one pile, fives into another, tens into another, etc., and then count the bills in each pile separately and multiply the number of bills in each pile by their common value. This is the idea of Lebesgue integration.

Riemann integral (Darboux's definition).

A useful definition of Riemann integral is due to Darboux: for each partition $\Delta = \{I_i\}$ denote

$$M_i = \sup_{x \in I_i} f(x) \qquad \text{and} \qquad m_i = \inf_{x \in I_i} f(x).$$

Now consider the "upper" and "lower" sums

$$\bar{\mathcal{J}}(\Delta) = \sum_{i=1}^{n} M_i |I_i| \qquad \text{and} \qquad \underline{\mathcal{J}}(\Delta) = \sum_{i=1}^{n} m_i |I_i|.$$

$$\underline{\mathcal{J}}(\Delta) \leq \bar{\mathcal{J}}(\Delta)$$

Furthermore, for any two partitions $\Delta = \{I_i\}$ and $\Delta' = \{I'_j\}$ we have

$$\underline{\mathcal{J}}(\Delta) \leq \underline{\mathcal{J}}(\Delta^*) \leq \bar{\mathcal{J}}(\Delta^*) \leq \bar{\mathcal{J}}(\Delta')$$

where $\Delta^* = \{I_i \cap I'_j\}$ is the refinement of both Δ and Δ' (see the lemma below).

Therefore
$$\sup_{\Delta} \underline{\mathcal{J}}(\Delta) \leq \inf_{\Delta} \bar{\mathcal{J}}(\Delta).$$

Now the Riemann integral is defined by

$$\int_a^b f(x)\, dx = \sup_{\Delta} \underline{\mathcal{J}}(\Delta) = \inf_{\Delta} \bar{\mathcal{J}}(\Delta),$$

This only applies if the two terms are finite and equal. i.e.

$$\sup_{\Delta} \underline{\mathcal{J}}(\Delta) = \inf_{\Delta} \bar{\mathcal{J}}(\Delta) \in \mathbb{R}.$$

Intuitively, the Riemann integral exists iff the upper an lower sums are finite and can be made arbitrarily close to each other by choosing a fine enough partition.

Lemma 10.1. *Suppose $\Delta = \{I_i\}$ is a partition of the interval I and $\Delta^* = \{I^*_j\}$ is a refinement of Δ, i.e., every interval I^*_j is a part of some interval I_i ($I^*_j \subset I_i$). Then*

$$\underline{\mathcal{J}}(\Delta) \leq \underline{\mathcal{J}}(\Delta^*) \qquad \textit{and} \qquad \bar{\mathcal{J}}(\Delta) \geq \bar{\mathcal{J}}(\Delta^*).$$

Proof. Whenever $I^*_j \subset I_i$, we have

$$m_i = \inf_{x \in I_i} f(x) \leq \inf_{x \in I^*_j} f(x) = m^*_j$$

$$M_i = \sup_{x \in I_i} f(x) \geq \sup_{x \in I^*_j} f(x) = M^*_j$$

Multiplying by $|I^*_j|$ and summing up over j proves the lemma. □

Theorem 10.2. *Let $f\colon [a,b] \to \mathbb{R}$. If the Riemann integral*

$$\mathcal{J} = \int_a^b f(x)\,dx$$

exists, then f is Lebesgue integrable on $[a,b]$ and its Lebesgue integral equals

$$\int_{[a,b]} f\,d\mathbf{m} = \mathcal{J}.$$

The converse is not true (see an example below).

Proof. Let Δ_n denote a sequence of increasingly finer partitions of $I = [a,b]$, for example let $\Delta_n = \{I_{n,i}\}$, $1 \le i \le 2^n$, be the partition of I into 2^n equal subintervals.

Note that each subinterval $I_{n,i}$ in a part of one of the longer subintervals $I_{n-1,j}$, i.e., Δ_n is a refinement of Δ_{n-1}. Denote $M_{n,i} = \sup_{x \in I_{n,i}} f(x)$ and $m_{n,i} = \inf_{x \in I_{n,i}} f(x)$.

Now define "upper" and "lower" functions $\bar{f}_n$ and $\underline{f}_n$ by

$$\bar{f}_n(x) = M_{n,i} \quad \text{and} \quad \underline{f}_n(x) = m_{n,i} \qquad \forall x \in I_{n,i}.$$

Note that

$$\int_a^b \bar{f}_n(x)\,dx = \bar{\mathcal{J}}(\Delta_n) \qquad \text{and} \qquad \int_a^b \underline{f}_n(x)\,dx = \underline{\mathcal{J}}(\Delta_n).$$

Also note that

$$\bar{f}_1 \ge \bar{f}_2 \ge \cdots \ge f \qquad \text{and} \qquad \underline{f}_1 \le \underline{f}_2 \le \cdots \le f$$

hence there exist limits

$$\lim_{n\to\infty} \bar{f}_n = \bar{f} \ge f \qquad \text{and} \qquad \lim_{n\to\infty} \underline{f}_n = \underline{f} \le f.$$

Now by Lebesgue's Dominated Convergence Theorem

$$\int_{[a,b]} \bar{f}\,d\mathbf{m} = \lim_{n\to\infty} \int_{[a,b]} \bar{f}_n\,d\mathbf{m} = \lim_{n\to\infty} \bar{\mathcal{J}}(\Delta_n) = \int_a^b f(x)\,dx$$

and

$$\int_{[a,b]} \underline{f}\,d\mathbf{m} = \lim_{n\to\infty} \int_{[a,b]} \underline{f}_n\,d\mathbf{m} = \lim_{n\to\infty} \underline{\mathcal{J}}(\Delta_n) = \int_a^b f(x)\,dx.$$

Therefore

$$\implies \quad \int_{[a,b]} (\bar{f} - \underline{f})\,d\mathbf{m} = 0 \quad \implies \quad \int_{[a,b]} |\bar{f} - \underline{f}|\,d\mathbf{m} = 0$$

(recall that $\bar{f} \ge f \ge \underline{f}$, hence $\bar{f} - \underline{f} \ge 0$). Theorem 7.11(a) implies $\bar{f}(x) = \underline{f}(x)$ a.e., thus $\bar{f}(x) = \underline{f}(x) = f(x)$ a.e. Now Theorem 7.4 implies

$$\int_{[a,b]} \bar{f}\,d\mathbf{m} = \int_{[a,b]} \underline{f}\,d\mathbf{m} = \int_{[a,b]} f\,d\mathbf{m}$$

so our theorem follows. $\square$

EXAMPLE 9. Let $f\colon [0,1] \to \mathbb{R}$ be the Dirichlet function (Example 6) restricted to the interval $[0,1]$, i.e., $f(x) = 1$ for x rational and $f(x) = 0$ for x irrational. It is not Riemann integrable because $\underline{\mathcal{J}}(\Delta) = 0$ and $\bar{\mathcal{J}}(\Delta) = 1$ for every partition Δ. On the other hand, f is Lebesgue integrable and $\int_{[0,1]} f\, d\mathbf{m} = 0$.

Theorem 10.3. (Riemann-Lebesgue)
The Riemann integral

$$\int_a^b f(x)\, dx$$

exists if and only if f is bounded and almost everywhere continuous.

- See the proof on the next page.
- The last condition means that the set

$$E_f = \{x \in [a,b] \colon f \text{ is discontinuous at } x\}$$

has Lebesgue measure zero, i.e., $\mathbf{m}(E_f) = 0$. The set E_f may be complicated.

EXAMPLE 10. A modified Dirichlet function $f\colon [0,1] \to \mathbb{R}$ is defined as follows: at every irrational number x we set $f(x) = 0$ and at every rational number $x = p/q$, where p/q is an irreducible fraction, we set $f(x) = 1/q$. This function is continuous at every irrational number but discontinuous at every rational number. Thus $E_f = \mathbb{Q} \cap [0,1]$ is a countable and dense set. This function meets the conditions of the above theorem, thus its Riemann integral exists. In fact, $\int_0^1 f(x)\, dx = 0$.

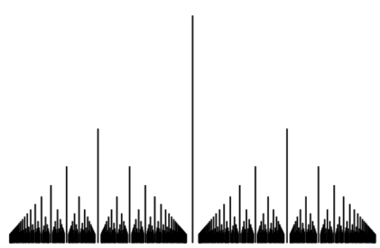


Figure 6: Modified Dirichlet function

Proof. **Riemann-Lebesgue** Theorem 10.3

$\boxed{\Rightarrow}$ Suppose the Riemann integral exists.

If f was unbounded, then for each partition Δ there would be a subinterval $I_i \subset I$ on which either $M_i = \infty$ or $m_i = -\infty$. Hence either $\bar{\mathcal{J}}(\Delta) = \infty$ or $\underline{\mathcal{J}}(\Delta) = -\infty$, which is impossible by the Darboux definition of the Riemann integral.

Next for each subinterval $I' \subset I$ define the *oscillation* of f on I' by

$$\omega_f(I') = \sup_{x \in I'} f(x) - \inf_{x \in I'} f(x).$$

The oscillation of f at a point $x \in I$ is defined by

$$\omega_f(x) = \lim_{\varepsilon \to 0} \omega_f([x - \varepsilon, x + \varepsilon]).$$

Note: f is continuous at x if and only if $\omega_f(x) = 0$.

For each $\delta > 0$ denote

$$E_\delta = \{x \in I \colon \omega_f(x) \geq \delta\}.$$

The set of points where f is discontinuous is $E = \cup_{\delta > 0} E_\delta$.

If $\mathbf{m}(E) > 0$, then there exists a $\delta_0 > 0$ such that $\mathbf{m}(E_{\delta_0}) > 0$. Now consider an arbitrary partition $\Delta = \{I_i\}$ of I. If $\mathrm{int} I_i \cap E_{\delta_0} \neq \emptyset$, then $M_i - m_i \geq \delta_0$. All such subintervals I_i's cover E_{δ_0}, hence their total length is $\geq \mathbf{m}(E_{\delta_0})$. This implies $\bar{\mathcal{J}}(\Delta) - \underline{\mathcal{J}}(\Delta) \geq \delta_0 \mathbf{m}(E_{\delta_0}) > 0$.

Thus the upper sum $\bar{\mathcal{J}}(\Delta)$ and the lower sum $\underline{\mathcal{J}}(\Delta)$ cannot get arbitrarily close to each other, contradicting the Darboux definition of the Riemann integral.

$\boxed{\Leftarrow}$ Suppose f is bounded, $|f| \leq M$, and a.e. continuous.

We use the sequence of partitions Δ_n from the proof of Theorem 10.2. Suppose $x \in [0, 1]$ is <u>not</u> an endpoint of the intervals $I_{n,i} \in \Delta_n$ (i.e., $x \neq p/2^q \quad \forall p, q \in \mathbb{N}$), and f is continuous at x (i.e. $\omega_f(x) = 0$). Then $\bar{f}(x) = \underline{f}(x)$.

Therefore $\bar{f} = \underline{f}$ a.e., hence

$$\lim_{n \to \infty} \bar{\mathcal{J}}(\Delta_n) = \int_{[a,b]} \bar{f}\, d\mathbf{m} = \int_{[a,b]} \underline{f}\, d\mathbf{m} = \lim_{n \to \infty} \underline{\mathcal{J}}(\Delta_n),$$

which implies the existence of the Riemann integral. ∎

Example 11. The Dirichlet function is bounded, but discontinuous at every point. Again we see that it is not Riemann integrable.

Example 12. The function $f\colon [0,1] \to \mathbb{R}$ defined by $f(x) = 1/\sqrt{x}$ is Lebesgue integrable, but it is unbounded hence not Riemann integrable. However, an improper Riemann integral

$$\int_0^1 f(x)\,dx = \lim_{a\to 0^+} \int_a^1 f(x)\,dx$$

exists and is equal to the Lebesgue integral $\int_{[0,1]} f(x)\,d\mathbf{m}$

Example 13. The function $f\colon [0,\infty) \to \mathbb{R}$ defined by $f(x) = (-1)^{n+1}/n$ for $n-1 \le x < n$, $n = 1, 2, \ldots$, has a finite improper Riemann integral

$$\int_0^\infty f(x)\,dx = \lim_{A\to 0^+} \int_0^A f(x)\,dx = \sum_{n=1}^\infty \frac{(-1)^{n+1}}{n} < \infty.$$

But it is not Lebesgue integrable because

$$\int_{[0,\infty]} |f(x)|\,d\mathbf{m} = \sum_{n=1}^\infty \frac{1}{n} = \infty.$$

Improper Riemann integral vs Lebesgue integral.

The existence of an improper Riemann integral does not imply Lebesgue integrability. More precisely, if f has a finite improper Riemann integral, then it is Lebesgue integrable if and only if $|f|$ has a finite improper Riemann integral. In that case the Lebesgue integral is equal to the improper Riemann integral. (Note: if $f \ge 0$, then the existence of an improper Riemann integral is equivalent to Lebesgue integrability.)

Exercise 44. Find a function $f(x)$ on $[0,1]$ such that the improper Riemann integral $\int_0^1 f(x)\,dx$ exists (is finite), but f is not Lebesgue integrable.

11 L^p spaces

Definition 11.1. Let $(X, \mathfrak{M}, \mu)$ be a measure space. For each $p \in (0, \infty)$, $L^p_\mu(X)$ denotes the set of functions $f\colon X \to \mathbb{C}$ such that $|f|^p$ is integrable, i.e.

$$L^p_\mu(X) = \{f\colon \int_X |f|^p \, d\mu < \infty\}$$

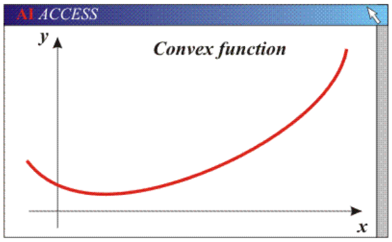
- Note that $|f(x)| \in [0, \infty)$ for each $x \in X$, thus $|f(x)|^p$ is well defined for every $0 < p < \infty$, and $|f|^p$ is a non-negative function.
- For $p = 1$, this definition agrees with our earlier Definition 6.1 of $L^1_\mu(X)$.
- The set $L^p_\mu(X)$ is actually a vector space with a norm, and we need certain tools to introduce this norm.

Definition 11.2. A function $\varphi\colon (a, b) \to \mathbb{R}$, where $-\infty \le a < b \le \infty$ is said to be **convex** if

$$\varphi(px + qy) \le p\varphi(x) + q\varphi(y) \tag{11.1}$$

for all $a < x < y < b$ and $p, q > 0$, $p + q = 1$.

- If the inequality in (11.1) is strict, we have a **strictly convex** function.
- Geometrically, this means that the secant line joining the points $(x, \varphi(x))$ and $(y, \varphi(y))$ lies above the graph of φ between x and y. A parabola $y = x^2$ is a good example.



Lemma 11.3. *A function $f\colon (a, b) \to \mathbb{R}$ is convex if and only if for any $a < s < t < u < b$*

$$\frac{\varphi(t) - \varphi(s)}{t - s} \le \frac{\varphi(u) - \varphi(t)}{u - t}.$$

Proof. Geometrically, this is obvious. Algebraically, one can use (11.1) and set

$$x = s, \quad y = u, \quad p = \frac{u - t}{u - s}, \quad q = \frac{t - s}{u - s}$$

Then $t = px + qy$. The rest is straightforward calculation. $\square$

- If $\varphi(x)$ is strictly convex, then the inequality in Lemma 11.3 is strict.

Lemma 11.4. *If φ is differentiable, then it is convex if and only if φ' is monotonically increasing, i.e. $\varphi'(x) \leq \varphi'(y)$ for all $x < y$.*

Proof. This follows from the previous lemma. □

- If the second derivative exists, then φ is convex if and only if $\varphi''(x) \geq 0$.

Theorem 11.5. *If φ is convex on (a, b), then it is continuous on (a, b).*

Proof. We prove that f is continuous at any point $c \in (a, b)$. Choose $a_1 \in (a, c)$ and $b_1 \in (c, b)$. Then by Lemma 11.3 we have for any $\varepsilon > 0$

$$\frac{\varphi(c) - \varphi(a_1)}{c - a_1} \leq \frac{\varphi(c + \varepsilon) - \varphi(c)}{\varepsilon}$$

which implies $\liminf_{x \to c^+} \varphi(x) \geq \varphi(c)$. Similarly,

$$\frac{\varphi(c + \varepsilon) - \varphi(c)}{\varepsilon} \leq \frac{\varphi(b_1) - \varphi(c + \varepsilon)}{b_1 - c - \varepsilon},$$

which implies $\liminf_{x \to c^+} \varphi(x) \leq \varphi(c)$. Hence $\lim_{x \to c^+} \varphi(x) = \varphi(c)$. The left hand limit is treated similarly. □

- This is not true on closed intervals $[a, b]$.

Theorem 11.6. (Jensen's inequality)
Let $(X, \mathfrak{M}, \mu)$ be a measure space and $\mu(X) = 1$. Let $f\colon X \to (a, b) \subset \mathbb{R}$ be a Lebesgue integrable function, i.e., $f \in L^1_\mu$. Then for any convex function $\varphi\colon (a, b) \to \mathbb{R}$ we have

$$\varphi\Big(\int_X f \, d\mu\Big) \leq \int_X (\varphi \circ f) \, d\mu.$$

Note: the cases $a = -\infty$ and $b = \infty$ are not excluded.

- See the proof on the next page.
- A measure μ on X such that $\mu(X) = 1$ is often called *probability measure*.

An important remark: the integral $\int_X (\varphi \circ f) \, d\mu$ here is understood in the "extended sense" (cf. Section 6). More precisely, the "negative" part of the integrand belongs in $L^1_\mu(X)$, i.e.,

$$\int_X (\varphi \circ f)^- \, d\mu < \infty, \qquad \text{(see the proof below)}$$

but the positive part may not be in $L^1_\mu(X)$, i.e., we can only claim that

$$\int_X (\varphi \circ f)^+ \, d\mu \le \infty,$$

which possibly takes value ∞ (see an example below). Therefore the above integral

$$\int_X (\varphi \circ f) \, d\mu \overset{6.3}{=} \int_X (\varphi \circ f)^+ \, d\mu - \int_X (\varphi \circ f)^- \, d\mu$$

is either finite, or equal to ∞ (but not $-\infty$).

EXAMPLE 14. Let $X = (0,1)$ and μ be the Lebesgue measure on X; note that $\mu(X) = 1$. Suppose $f(x) = x$ and $\varphi(x) = 1/x$. Note that $f \in L^1_\mu(X)$ and φ is convex on $(0,1)$. Now $(\varphi \circ f)(x) = 1/x$. The Riemann integral of this function

$$\int_0^1 \frac{1}{x} \, dx = \ln x \,|_0^1 = \infty$$

is infinite (by elementary Calculus). Thus its Lebesgue integral is infinite, too (recall the material in the end of Section 10).

Proof. **Jensen's inequality** Theorem 11.6

First, note that

$$a = a\mu(X) < \int_X f \, d\mu < b\mu(X) = b$$

(due to the result of Exercise 36(b) and the assumption $\mu(X) = 1$), hence

$$t \colon = \int_X f \, d\mu \in (a, b). \tag{11.2}$$

Next, by Lemma 11.3 for any $u \in (t, b)$

$$\beta \colon = \sup_{s \in (a,t)} \frac{\varphi(t) - \varphi(s)}{t - s} \le \frac{\varphi(u) - \varphi(t)}{u - t} < \infty. \tag{11.3}$$

Therefore, for any $s \in (a, b)$ we have

$$\varphi(s) \ge \varphi(t) + \beta(s - t)$$

(for $s < t$ this follows from the definition of β in (11.3); for $s > t$ this follows from the middle inequality in (11.3), where u must be is replaced with s). Thus for any $x \in X$

$$\varphi(f(x)) \ge \varphi(t) + \beta(f(x) - t). \tag{11.4}$$

Note in particular that

$$(\varphi \circ f)^- \le |\varphi(t) + \beta(f - t)| \le |\varphi(t)| + |\beta|(|f| + |t|),$$

hence

$$\int_X (\varphi \circ f)^- \, d\mu \le |\varphi(t)| + |\beta| \Big(\int_X |f| \, d\mu + |t| \Big) < \infty,$$

as we promised above.

Now we have two cases. First, if $\int_X (\varphi \circ f)^+ \, d\mu = \infty$, then $\int_X (\varphi \circ f) \, d\mu = \infty$, and Jensen's inequality is trivial (a finite number $\le$ infinity).

Second, if $\int_X (\varphi \circ f)^+ \, d\mu < \infty$, then $\varphi \circ f \in L^1_\mu(X)$. Now taking the integral of (11.4) over X gives

$$\int_X \big(\varphi \circ f - \varphi(t) - \beta(f - t) \big) \, d\mu \ge 0.$$

hence by the linearity (Theorem 6.7) we get

$$\int_X (\varphi \circ f) \, d\mu \ge \varphi(t) + \beta \Big(\int_X f \, d\mu - t \Big) = \varphi(t),$$

where the expression in the parentheses vanishes due to (11.2). □

- If $\varphi(x)$ is strictly convex, then Jensen's inequality turns into an equality if and only if f is constant a.e.

Example 15. Let $\varphi(x) = e^x$. Then we get

$$e^{\int_X f \, d\mu} \le \int_X e^f \, d\mu.$$

Example 16. Let $X = \{1, 2, \dots, n\}$ and $\mu(\{i\}) = 1/n$ for every $i = 1, \dots, n$. Let $f(i) = \ln a_i$ for some $a_1, \dots, a_n > 0$ and again $\varphi(x) = e^x$. Then Jensen's inequality implies

$$(a_1 a_2 \cdots a_n)^{1/n} \le \frac{a_1 + a_2 + \cdots + a_n}{n}$$

(geometric mean is $\le$ arithmetic mean). More general: if $\mu(\{i\}) = p_i$ for $i = 1, \dots, n$ and $p_1 + \cdots + p_n = 1$, then

$$a_1^{p_1} a_2^{p_2} \cdots a_n^{p_n} \le p_1 a_1 + p_2 a_2 + \cdots + p_n a_n$$

(geometric weighted mean is $\le$ arithmetic weighted mean).

Exercise 45. Prove that the supremum of any collection of convex functions on (a, b) is convex on (a, b) (assume that the supremum is finite). Prove that a pointwise limit of a sequence of convex functions is convex.

Exercise 46. Let φ be convex on (a, b) and ψ convex and nondecreasing on the range of φ. Prove that $\psi \circ \varphi$ is convex on (a, b). For $\varphi > 0$, show that the convexity of $\log \varphi$ implies the convexity of φ, but not vice versa.

EXERCISE 47. Assume that φ is a continuous real function on (a, b) such that

$$\varphi\Big(\frac{x+y}{2}\Big) \leq \frac{1}{2}\,\varphi(x) + \frac{1}{2}\,\varphi(y)$$

for all $x, y \in (a, b)$. Prove that φ is convex. (The conclusion does not follow if continuity is omitted from the hypotheses.)

DEFINITION 11.7. If $p, q > 1$ and

$$\frac{1}{p} + \frac{1}{q} = 1,$$

then we say that p and q are **conjugate exponents**.

- The most important case is $p = q = 2$.
- If $p \to 1$, then $q \to \infty$; thus $p = 1$ and $q = \infty$ are considered as conjugate exponents.

Theorem 11.8. *Let $(X, \mathfrak{M}, \mu)$ be a measure space and $f, g\colon X \to [0, \infty]$ measurable functions. Let p and q be conjugate exponents.*
Then we have ***Hölder inequality****:*

$$\int_X fg\,d\mu \leq \left[\int_X f^p\,d\mu\right]^{1/p} \left[\int_X g^q\,d\mu\right]^{1/q}$$

Equality holds here if and only if there exist $\alpha, \beta \geq 0$, not both equal to 0, such that $\alpha f(x)^p = \beta g(x)^q$ a.e. (i.e., $f(x)^p$ and $q(x)^q$ are proportional to each other).

Proof. Denote

$$A = \left[\int_X f^p\,d\mu\right]^{1/p}, \qquad B = \left[\int_X g^q\,d\mu\right]^{1/q}$$

If $A = 0$, then $f = 0$ a.e. (by Theorem 7.11a) and the case is trivial. Similarly, if $B = 0$, then $g = 0$ a.e., and the case is trivial.

If $A = \infty$ and $B > 0$, then the Hölder inequality is trivial, as its right hand side is ∞. The same happens if $B = \infty$ and $A > 0$.

Thus we can assume that $A \in (0, \infty)$ and $B \in (0, \infty)$. Consider two functions:

$$F(x) = f(x)/A \qquad \text{and} \qquad G(x) = g(x)/B$$

("normalized" versions of f and g). Note that

$$\int_X F^p\,d\mu = \int_X G^q\,d\mu = 1.$$

It is enough to prove

$$\int_X FG\,d\mu \le 1, \tag{11.5}$$

and then complete the proof of the Hölder inequality as follows:

$$\int_X fg\,d\mu = AB\int_X FG\,d\mu \le AB = \left[\int_X f^p\,d\mu\right]^{1/p}\left[\int_X g^q\,d\mu\right]^{1/q}.$$

We are now proving (11.5). Note that $F < \infty$ and $G < \infty$ a.e., due to Theorem 5.8(f).

The following calculus lemma is the key step:

Lemma. *If $F, G \in [0,\infty)$ are two nonnegative real numbers, then*

$$FG \le \tfrac{1}{p}F^p + \tfrac{1}{q}G^q.$$

Proof of Lemma: If $F = 0$ or $G = 0$, the lemma is trivial. If $F, G > 0$, we can choose $s, t \in \mathbb{R}$ such that $F = e^{s/p}$ and $G = e^{t/q}$. Since $\varphi(x) = e^x$ is a convex function and $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$e^{\frac{s}{p}+\frac{t}{q}} \le \tfrac{1}{p}e^s + \tfrac{1}{q}e^t.$$

In terms of F and G, the above inequality is exactly what Lemma claims. □

Due to this lemma, we have for a.e. $x \in X$

$$F(x)G(x) \le \tfrac{1}{p}F(x)^p + \tfrac{1}{q}G(x)^q.$$

Integrating over X gives

$$\int_X FG\,d\mu \le \frac{1}{p}\int_X F^p\,d\mu + \frac{1}{q}\int_X G^q\,d\mu = \frac{1}{p} + \frac{1}{q} = 1$$

proving (11.5). This completes the proof of the Hölder inequality.

We now turn to the characterization of equality. First, suppose that $\alpha f(x)^p = \beta g(x)^q$ and, without loss of generality, $\beta \neq 0$. Then $g(x) = c^{1/q}f(x)^{p/q}$, where $c = \alpha/\beta$. Now

$$\int_X fg\,d\mu = c^{1/q}\int_X f^{1+\frac{p}{q}}\,d\mu = c^{1/q}\int_X f^p\,d\mu$$

because $\frac{1}{p} + \frac{1}{q} = 1$. On the other hand,

$$\left[\int_X f^p\,d\mu\right]^{1/p}\left[\int_X g^q\,d\mu\right]^{1/q} = \left[\int_X f^p\,d\mu\right]^{1/p}\left[c\int_X f^p\,d\mu\right]^{1/q} = c^{1/q}\int_X f^p\,d\mu$$

again because $\frac{1}{p} + \frac{1}{q} = 1$. Hence the Hölder inequality turns into an equality.

Now suppose that the Hölder inequality turns into an equality. Then it follows from our proof of the Hölder inequality that we must have

$$F(x)G(x) = \tfrac{1}{p}F(x)^p + \tfrac{1}{q}G(x)^q \qquad \text{a.e.}$$

Now we look back into the proof of our lemma. Since $\varphi(x) = e^x$ is a strictly convex function, the equality in the lemma is only possible when $s = t$, i.e., $e^s = e^t$, i.e., $F^p = G^q$. Thus we have

$$F(x)^p = G(x)^q \qquad \text{a.e.,}$$

which implies that $f(x)^p$ and $q(x)^q$ are proportional to each other. □

Corollary 11.9. *If $p = q = 2$, the Hölder inequality is the* ***Schwarz inequality****:*

$$\left[\int_X fg\,d\mu\right]^2 \leq \left[\int_X f^2\,d\mu\right] \cdot \left[\int_X g^2\,d\mu\right].$$

Equality holds here if and only if there exist $\alpha, \beta \geq 0$, not both equal to 0, such that $\alpha f(x) = \beta g(x)$ a.e. (i.e., $f(x)$ and $q(x)$ are proportional to each other).

- If we replace integration with summation we obtain the familiar Cauchy-Schwarz inequality from linear algebra:

$$(\mathbf{f}\cdot\mathbf{g})^2 = \left[\sum f_i g_i\right]^2 \leq \left[\sum f_i^2\right] \cdot \left[\sum g_i^2\right] = \|\mathbf{f}\|^2\|\mathbf{g}\|^2.$$

where $\mathbf{f} = (f_1, f_2, \ldots)$, $\mathbf{g} = (g_1, g_2, \ldots)$ are vectors and $\mathbf{f}\cdot\mathbf{g}$ is their inner product.

Theorem 11.10. *Let $(X, \mathfrak{M}, \mu)$ be a measure space and $f, g\colon X \to [0, \infty]$ measurable functions. Let $p \geq 1$. Then we have* ***Minkowski inequality****:*

$$\left[\int_X (f+g)^p\,d\mu\right]^{1/p} \leq \left[\int_X f^p\,d\mu\right]^{1/p} + \left[\int_X g^p\,d\mu\right]^{1/p}.$$

- *If $p = 1$, this is actually an equality.*
- *If $p > 1$, the equality holds if and only if there exist $\alpha, \beta \geq 0$, not both equal to 0, such that $\alpha f(x) = \beta g(x)$ a.e. (i.e., $f(x)$ and $q(x)$ are proportional to each other).*

Proof. If $p = 1$, we just use the linearity of Lebesgue integral (Theorem 5.11).

Assume $p > 1$. We can also assume that

$$\int_X (f+g)^p\,d\mu > 0, \qquad \int_X f^p\,d\mu < \infty, \qquad \int_X q^p\,d\mu < \infty \tag{11.6}$$

as otherwise the Minkowski inequality is trivial.

Note that

$$(f+q)^p = f(f+g)^{p-1} + g(f+g)^{p-1}.$$

Due to the Hölder inequality

$$\int_X f(f+g)^{p-1}\,d\mu \leq \left[\int_X f^p\,d\mu\right]^{1/p}\left[\int_X (f+g)^{(p-1)q}\,d\mu\right]^{1/q}.$$

and

$$\int_X g(f+g)^{p-1}\,d\mu \leq \left[\int_X g^p\,d\mu\right]^{1/p}\left[\int_X (f+g)^{(p-1)q}\,d\mu\right]^{1/q}.$$

Also note that

$$\frac{1}{p} + \frac{1}{q} = 1 \quad \Longrightarrow \quad p + q = pq \quad \Longrightarrow \quad (p-1)q = p$$

Adding the above two integral inequalities gives

$$\int_X (f+g)^p \, d\mu \le \left(\left[\int_X f^p \, d\mu \right]^{1/p} + \left[\int_X g^p \, d\mu \right]^{1/p} \right) \left[\int_X (f+g)^p \, d\mu \right]^{1/q}.$$

Now if the last factor is finite, we can just divide by it and get

$$\left[\int_X (f+g)^p \, d\mu \right]^{1-1/q} \le \left[\int_X f^p \, d\mu \right]^{1/p} + \left[\int_X g^p \, d\mu \right]^{1/p}$$

which is exactly the desired Minkowski inequality because $1 - \frac{1}{q} = \frac{1}{p}$.

So all we need is to check that

$$\int_X (f+g)^p \, d\mu < \infty. \tag{11.7}$$

The following simple lemma will do the job:

Lemma. *If $f, g \in [0, \infty)$ are two nonnegative real numbers and $p > 1$, then*

$$\left(\frac{f+g}{2} \right)^p \le \frac{f^p + g^p}{2}$$

Proof of Lemma: If $f = 0$ or $g = 0$, the lemma is trivial. If $f, g > 0$, the lemma follows from the convexity of the function $\varphi(x) = x^p$ on the interval $(0, \infty)$. □

Now note that $f(x) < \infty$ and $g(x) < \infty$ a.e., due to Theorem 5.8(f). Thus due to the above lemma, we have for a.e. $x \in X$

$$\left(\frac{f(x) + g(x)}{2} \right)^p \le \frac{f(x)^p + g(x)^p}{2}$$

Integrating over X and using our assumptions (11.6) proves (11.7).

We now turn to the characterization of equality. First, suppose that $\alpha f(x) = \beta g(x)$ and, without loss of generality, $\beta \ne 0$. Then $g(x) = cf(x)$, where $c = \alpha/\beta$, and by direct calculation we can verify that the Minkowski inequality turns into an equality.

Now suppose that the Minkowski inequality turns into an equality. The main step of the above proof of the Minkowski inequality is application of the Hölder inequality twice: once to the functions f and $(f+g)^{p-1}$ and then once again to the functions g and $(f+g)^{p-1}$. Then the equality requires

$$f^p \qquad \text{be proportional to} \qquad (f+g)^{(p-1)q}$$

and

$$g^p \qquad \text{be proportional to} \qquad (f+g)^{(p-1)q}$$

We see clearly that f must be proportional to g, as claimed. □

- If we replace integration in the Minkowski inequality with summation, we obtain the familiar triangle inequality for the p-norm of vectors in linear algebra:

$$\|\mathbf{f}+\mathbf{g}\|_p = \left[\sum (f_i+g_i)^p\right]^{1/p} \le \left[\sum f_i^p\right]^{1/p} + \left[\sum g_i^p\right]^{1/p} = \|\mathbf{f}\|_p + \|\mathbf{g}\|_p,$$

where $\mathbf{f} = (f_1, f_2, \ldots)$ and $\mathbf{g} = (g_1, g_2, \ldots)$ are vectors.

Exercise 48. Suppose $\mu(X) = 1$ and suppose f and g are two positive measurable functions on X such that $fg \ge 1$. Prove that $\int_X f\,d\mu \cdot \int_X g\,d\mu \ge 1$.

Exercise 49. Suppose $\mu(X) = 1$ and $h\colon X \to [0,\infty]$ is measurable. Denote $A = \int_X h\,d\mu$. Prove that

$$\sqrt{1+A^2} \le \int_X \sqrt{1+h^2}\,d\mu \le 1 + A.$$

Exercise 50. [Bonus] If $\mathbf{m}$ is Lebesgue measure on $[0,1]$ and if h is a continuous function on $[0,1]$ such that $h = f'$, then the inequalities in the previous exercise have a geometric interpretation. From this, conjecture (for general X) under what conditions on h equality can hold in either of the above inequalities, and prove your conjecture.

Definition 11.11. For a measurable function $f\colon X \to \mathbb{C}$ and $p > 0$ we define the **"p-norm"** of f by

$$\|f\|_p = \left[\int_X |f|^p\,d\mu\right]^{1/p}.$$

- Definition 11.1 can be now stated as $L^p_\mu(X) = \{f\colon \|f\|_p < \infty\}$
- If μ is the Lebesgue measure on $\mathbb{R}^k$, we will simply write $L^p(\mathbb{R}^k)$.

Definition 11.12. A special case: $\mathbb{N}$ and μ is a counting measure. Then functions $f\colon \mathbb{N} \to \mathbb{C}$ can be regarded as complex-valued sequences $\xi = \{\xi_n\}$. Then

$$\|\xi\|_p = \left[\sum_{n=1}^{\infty} |\xi_n|^p\right]^{1/p}$$

and

$$\ell^p = \{\xi\colon \|\xi\|_p < \infty\}$$

Theorem 11.13. *If $1 \le p < \infty$, then $L^p_\mu(X)$ is a complex vector space.*

Proof. Given $f, g \in L^p_\mu(X)$ and $\alpha, \beta \in \mathbb{C}$ we have, by the Minkowski inequality

$$\begin{aligned}
\|\alpha f + \beta g\|_p &= \left[\int_X |\alpha f + \beta g|^p \, d\mu\right]^{1/p} \\
&\le \left[\int_X \big(|\alpha f| + |\beta g|\big)^p \, d\mu\right]^{1/p} && \text{(By triangle inequality for complex numbers)} \\
&\le \left[\int_X |\alpha f|^p \, d\mu\right]^{1/p} + \left[\int_X |\beta g|^p \, d\mu\right]^{1/p} && \text{(By Minkowski inequality)} \\
&= |\alpha| \, \|f\|_p + |\beta| \, \|g\|_p < \infty && \text{(By the linearity of the Lebesgue integral)}
\end{aligned}$$

Thus $L^p_\mu(X)$ is closed under additions and multiplications by scalars. □

Next we investigate the case $p = \infty$.

Definition 11.14. Let $f\colon X \to [0, \infty]$ be measurable. A number $s \ge 0$ is an **essential upper bound** for f if

$$\mu\{x\colon f(x) > s\} = \mu(f^{-1}(s, \infty]) = 0$$

- The set of all essential upper bounds is either $\emptyset$ or a closed infinite interval $[S, \infty)$.

Definition 11.15. The smallest essential upper bound for a given function f is called the **essential supremum** and denoted by ess-sup f. If f has no essential upper bounds, we set ess-sup $f = \infty$.

Definition 11.16. The **infinity-norm** of a function $f\colon X \to \mathbb{C}$ is defined by

$$\|f\|_\infty = \text{ess-sup}\, |f|.$$

We also denote by

$$L^\infty_\mu(X) = \big\{f\colon \|f\|_\infty < \infty\big\}$$

the space of **essentially bounded functions**.

- A special case: ℓ^∞ denotes the space of bounded sequences.

Example 17. The Dirichlet function (Example 6) takes two values, 0 and 1. Its essential supremum is zero, because the value 1 is taken on a null set.

Exercise 51. Let $f, g \colon X \to [0,\infty]$. Show that $\text{ess-sup}\,|f+g| \le \text{ess-sup}\,|f| + \text{ess-sup}\,|g|$.

Theorem 11.17. *L^∞_μ is a complex vector space.*

Proof.
$$\begin{aligned}\|\alpha f+\beta g\|_\infty &= \text{ess-sup}\,|\alpha f+\beta g| \\ &\le \text{ess-sup}\,|\alpha f| + \text{ess-sup}\,|\beta g| \\ &= |\alpha|\,\text{ess-sup}\,|f| + |\beta|\,\text{ess-sup}\,|g| \\ &= |\alpha|\,\|f\|_\infty + |\beta|\,\|g\|_\infty\end{aligned}$$
□

Theorem 11.18. (Hölder inequality for norms) *Let p and q be complex conjugate exponents, including the limiting cases $p=1$, $q=\infty$ and $p=\infty$, $q=1$. Then for every $f\in L^p_\mu(X)$ and $g\in L^q_\mu(X)$ we have $fg\in L^1_\mu(X)$ and*
$$\|fg\|_1 \le \|f\|_p\,\|g\|_q$$

Proof. If $1<p,q<\infty$, we can just apply Hölder inequality to $|f|$ and $|g|$. Let $p=\infty$ and $q=1$. Now
$$|f(x)g(x)| \le \|f\|_\infty |g(x)| \qquad \text{for a.e. } \; x\in X$$
hence
$$\|fg\|_1 = \int_X |fg|\,d\mu \le \|f\|_\infty \int_X |g|\,d\mu = \|f\|_\infty \|g\|_1.$$
The case $p=1$ and $q=\infty$ is handled similarly. □

Exercise 52. When does one get equality in $\|fg\|_1 \le \|f\|_\infty\,\|g\|_1$?

Proposition 11.19. *Let $1\le p\le\infty$. Then for every $f,g\in L^p_\mu(X)$ we have*
$$\|f+q\|_p \le \|f\|_p + \|g\|_p$$
and
$$\|cf\|_p = |c|\,\|f\|_p.$$

Proof. See the proofs of Theorem 11.13 and Theorem 11.17. □

- Thus, $\|\cdot\|_p$ has two properties of a norm. But it is not a norm, see the next example.

Example 18. Let $X = \mathbb{R}$ and μ be the Lebesgue measure. The Dirichlet function $\chi_{\mathbb{Q}}$ (Example 6) satisfies $\|\chi_{\mathbb{Q}}\|_p = 0$ for every $p \in [1, \infty]$.

- $\|\cdot\|_p$ is not a norm, as there may be nonzero functions $0 \neq f \in L^p_\mu(X)$ with $\|f\|_p = 0$.

Lemma 11.20. *Let $f\colon X \to \mathbb{C}$ be measurable. Then*

$$\exists p \in [1, \infty]\colon \|f\|_p = 0 \quad \Longleftrightarrow \quad f = 0 \ \text{ a.e.} \quad \Longleftrightarrow \quad \forall p \in [1, \infty]\colon \|f\|_p = 0.$$

Proof. Direct inspection. For $p < \infty$ apply Theorem 7.11(a). □

Definition 11.21. Let V be a complex vector space and $W \subset V$ a linear subspace. For two vectors $v_1, v_2 \in V$ we say that $v_1 \sim v_2$ iff $v_1 - v_2 \in W$. This is an equivalence relation. Denote by

$$[v] = \{v_1 \in V : v - v_1 \in W\}$$

the equivalence class containing v and

$$V/W = \{[v] : v \in V\}$$

the set of equivalence classes. Then V/W is a vector space with

$$[v] + [w] = [v + w], \qquad c[v] = [cv].$$

We call V/W the **quotient space**.

Lemma 11.22. *Let $(X, \mathfrak{M}, \mu)$ be a measure space and*

$$\mathcal{N}_\mu = \{f\colon X \to \mathbb{C} \ \ \textit{measurable}, \quad f = 0 \ \ \textit{a.e.}\}$$

Then $\mathcal{N}_\mu$ is a vector space.

Proof. Let $f, g \in \mathcal{N}_\mu$ and $\alpha \in \mathbb{C}$. Then

$$\{x\colon f + g \neq 0\} \subset \{x\colon f \neq 0\} \cup \{x\colon g \neq 0\}$$

thus $\mu(\{x\colon f + g \neq 0\}) = 0$, and so $f + g \in \mathcal{N}_\mu$. We also have

$$\{x\colon \alpha f \neq 0\} \subset \{x\colon f \neq 0\}$$

hence $\alpha f \in \mathcal{N}_\mu$. □

Definition 11.23. Let $(X, \mathfrak{M}, \mu)$ be a measure space and

$$\mathcal{N}_\mu = \{f\colon X \to \mathbb{C} \ \text{ measurable}, \quad f = 0 \ \text{ a.e.}\}$$

For every $1 \le p \le \infty$ we define the **reduced $\mathcal{L}^p_\mu(X)$ space** by

$$\mathcal{L}^p_\mu = L^p_\mu / \mathcal{N}_\mu.$$

- $\mathcal{L}^p_\mu(X)$ is obtained from $L^p_\mu(X)$ by identifying functions that coincide a.e.

Corollary 11.24. *Let $1 \le p \le \infty$. Then $\mathcal{L}^p_\mu(X)$ is a vector space with norm $\|\cdot\|_p$.*

Exercise 53. Suppose $f\colon X \to \mathbb{C}$ is measurable and $\|f\|_\infty > 0$. Define

$$\varphi(p) = \int_X |f|^p \, d\mu = \|f\|_p^p \qquad (0 < p < \infty)$$

and consider the set $E = \{p\colon \varphi(p) < \infty\}$. Each of the following questions is graded as a separate exercise. Question (c) and (e) are **bonus** problems.

(a) Let $r < p < s$ and $r, s \in E$. Prove that $p \in E$. Thus E is a connected set.

(b) Prove that $\log \varphi$ is convex in the interior of E and that φ is continuous on E.

(c) Is E necessarily open? Closed? Can E consist of a single point? Can E be any connected subset of $(0, \infty)$?

(d) If $r < p < s$, prove that $\|f\|_p \le \max\bigl(\|f\|_r, \|f\|_s\bigr)$. Show that this implies the inclusion $L^r_\mu(X) \cap L^s_\mu(X) \subset L^p_\mu(X)$.

(e) Assume that $\|f\|_r < \infty$ for some $r < \infty$ and prove that $\|f\|_p \to \|f\|_\infty$ as $p \to \infty$.

Distance in a normed vector space.

Recall if V is a vector space with norm $\|\cdot\|$, the distance (metric) on V is defined

$$d(u, v) = \|u - v\|.$$

Definition 11.25. The space $\mathcal{L}^p_\mu(X)$ with norm $\|\cdot\|_p$ becomes a metric space. We say that a sequence of functions f_n converges to f in the L^p norm (metric) if $\|f_n - f\|_p \to 0$ as $n \to \infty$.

Proposition 11.26. *Suppose $f, f_1, f_2, \ldots \in L^p_\mu(X)$ and $f_n \to f$ in the L^p norm. Then $f_n \to f$ in measure.*

Types of convergence.

We have seen uniform convergence, pointwise convergence, convergence in measure, and now – the convergence in the L^p norm (metric). For $p < \infty$, it is stronger than the convergence in measure, but generally a little weaker than pointwise convergence. For $p = \infty$, the convergence in the L^∞ norm is equivalent to the uniform convergence a.e.

EXAMPLE 19. Recall the 'Amazing shrinking sliding rectangles' functions in Example 8. This sequence converges to the zero function $f \equiv 0$ in the L^p norm for every $p \in (0, \infty)$, but it does not converge at any point x. Note also that this sequence does not converge in the L^∞ norm.

DEFINITION 11.27. A metric is **complete** if every Cauchy sequence converges to a limit. A vector space V with norm $\|\cdot\|$ that induces a complete metric on it is called **Banach space**.

- Reminder: a Cauchy sequence $\{a_n\}$ has the following property:

$$\forall \varepsilon > 0\ \exists N \geq 1 \text{ such that } \forall m, n \geq N \text{ we have } |a_n - a_m| < \varepsilon$$

Theorem 11.28. *Let $1 \leq p \leq \infty$. Then $\mathcal{L}^p_\mu(X)$ is a Banach space.*

Proof. We consider separately two cases:

Case I: $1 \leq p < \infty$.

Let $\{f_n\}$ be a Cauchy sequence in $L^p_\mu(X)$. We need to find $f \in L^p_\mu(X)$ such that $f_n \to f$ in the L^p norm. Example 19 warns us that the functions f_n need not converge pointwise, hence we cannot construct f as a pointwise limit of f_n's. Instead, we will find a subsequence $\{f_{n_k}\}$ that converges pointwise a.e., and then define $f(x)$ to be the pointwise limit of $f_{n_k}(x)$.

Since $\{f_n(x)\}$ is a Cauchy sequence, for any $k \geq 1$ there exists $n_k \geq 1$ such that for all $n', n'' \geq n_k$ we have $\|f_{n'} - f_{n''}\|_p < 2^{-k}$. We can assume that $n_1 \leq n_2 \leq \cdots$. The subsequence $f_{n_k}(x)$ has the following property: $\|f_{n_{k+1}} - f_{n_k}\|_p < 2^{-k}$.

Define functions $g_k = \sum_{i=1}^{k} |f_{n_{i+1}} - f_{n_i}|$. Note that $0 \le g_1 \le g_2 \le \cdots$, hence there is the limit function $g(x)$ defined by

$$g = \lim_{k\to\infty} g_k = \sum_{i=1}^{\infty} |f_{n_{i+1}} - f_{n_i}|.$$

By Fatou's Lemma

$$\int_X g^p \, d\mu = \int_X \liminf g_k^p \, d\mu \le \liminf \int_X g_k^p \, d\mu.$$

Now

$$\begin{aligned} \int_X g_k^p \, d\mu = \|g_k\|_p^p &= \left\| \sum_{i=1}^{k} |f_{n_{i+1}} - f_{n_i}| \right\|_p^p \\ &\le \left[\sum_{i=1}^{k} \|f_{n_{i+1}} - f_{n_i}\|_p \right]^p \qquad \text{(By Proposition 11.19)} \\ &\le \left[\sum_{i=1}^{k} 2^{-i} \right]^p \le 1 \qquad \text{(By our choice of } n_i\text{'s)} \end{aligned}$$

Therefore $\int_X g^p \, d\mu \le 1$. According to Theorem 5.8(f), $g(x) < \infty$ a.e., thus the series $\sum_{i=1}^{\infty} (f_{n_{i+1}} - f_{n_i})$ converges absolutely a.e. We define the function $f(x)$ by

$$f(x) = f_{n_1}(x) + \sum_{i=1}^{\infty} (f_{n_{i+1}(x)} - f_{n_i}(x))$$

wherever the series converges (which happens a.e.) and set $f(x) = 0$ elsewhere. Thus for a.e. $x \in X$

$$f(x) = \lim_{k\to\infty} \left[f_{n_1}(x) + \sum_{i=1}^{k-1} (f_{n_{i+1}(x)} - f_{n_i}(x)) \right] = \lim_{k\to\infty} f_{n_k}(x).$$

In particular, $f(x)$ is measurable.

Next we will show that $f_n \to f$ in the L^p-metric and, in particular, $f \in L^p_\mu(X)$.

Let $\varepsilon > 0$. Since our sequence $\{f_n\}$ is Cauchy, there exists $N_\varepsilon \ge 1$ such that $\forall m, n \ge N_\varepsilon$ we have $\|f_n - f_m\|_p < \varepsilon$. By Fatou's Lemma for each $m \ge N_\varepsilon$

$$\int_X |f - f_m|^p \, d\mu = \int_X \liminf_{k\to\infty} |f_{n_k} - f_m|^p \, d\mu \le \liminf_{k\to\infty} \int_X |f_{n_k} - f_m|^p \, d\mu \le \varepsilon^p.$$

In particular, $f - f_m \in L^p_\mu(X)$ and hence $f \in L^p_\mu(X)$, because $L^p_\mu(X)$ is a vector space. We also have $\|f - f_m\|_p \le \varepsilon$ for all $m \ge N_\varepsilon$. This means $f_n \to f$ in the L^p metric.

Case II: $p = \infty$**.**

Let $\{f_n(x)\}$ be a Cauchy sequence in $L^\infty_\mu(X)$. Note that $\|f_n\|_\infty$ is a bounded sequence, i.e., $\exists M < \infty$ such that $\|f_n\|_\infty \leq M$ for all $n \geq 1$. Define sets

$$A_k = \{x \colon |f_k(x)| > \|f_k\|_\infty\}$$

and

$$B_{m,n} = \{x \colon |f_n(x) - f_m(x)| > \|f_n - f_m\|_\infty\}$$

All of these are null sets, thus their union is a null set:

$$E = \left(\cup_{k=1}^\infty A_k\right) \cup \left(\cup_{m,n=1}^\infty B_{m,n}\right), \qquad \mu(E) = 0$$

For each $x \in X \setminus E$ we have $|f_n(x) - f_m(x)| < \|f_n - f_m\|_\infty$, thus $f_n(x)$ is a Cauchy sequence (of complex numbers), hence it has a limit.

Define $f(x) = \lim_{n\to\infty} f_n(x)$ on $X \setminus E$ and $f(x) = 0$ on E. Next, we have $|f_n(x)| \leq M$ for all $x \in X \setminus E$ and all $n \geq 1$, so $|f(x)| \leq M$ for all $x \in X \setminus E$. Thus $f \in L^\infty_\mu(X)$.

Finally, $f_n \to f$ converges is uniformly, thus $\|f_n - f\|_\infty = \text{ess-sup}\,(f_n - f) \to 0$ $(n \to \infty)$

$\square$

Corollary 11.29. *Let $1 \leq p < \infty$. If $f_n \to f$ in the $L^p_\mu(X)$ metric, then there is a subsequence $\{f_{n_k}\}$ such that $f_{n_k} \to f$ a.e.*

Proof. Since $f_n \to f$ in the $L^p_\mu(X)$ metric, it is a Cauchy sequence, hence the proof of Case I above applies.

$\square$

Exercise 54. Let $\mu(X) = 1$. Each of the following questions is graded as a separate exercise.

(a) Prove that $\|f\|_r \leq \|f\|_s$ if $0 < r < s \leq \infty$.

(b) Under what condition does it happen that $0 < r < s \leq \infty$ and $\|f\|_r = \|f\|_s < \infty$?

(c) Prove that $L^r_\mu \supset L^s_\mu$ if $0 < r < s$. If $X = [0, 1]$ and $\mathbf{m}$ is the Lebesgue measure, show that $L^r_{\mathbf{m}} \neq L^s_{\mathbf{m}}$.

Exercise 55. [Bonus] For some measures, the relation $r < s$ implies $L^r(\mu) \subset L^s(\mu)$; for others, the inclusion is reversed; and there are some for which $L^r(\mu)$ does not contain $L^s(\mu)$ if $r \neq s$. Give examples of these situations, and find conditions on μ under which these situations will occur.

Exercise 56. (a) Show that $\int_0^{\frac{\pi}{2}} \sqrt{x \sin x}\, dx < \frac{\pi}{2\sqrt{2}}$;

(b) Show that $\left[\int_0^1 x^{\frac{1}{2}}(1-x)^{-\frac{1}{3}}\, dx\right]^3 \leq \frac{8}{5}$.

Next we approximate $L^p_\mu(X)$ by simple functions and continuous functions.

Definition 11.30. **Simple functions with a finite-measure support** are defined as follows:

$$S = \big\{s\colon X \to \mathbb{C} \;\text{ simple, measurable,} \quad \mu(s \neq 0) < \infty\big\}.$$

Lemma 11.31. *Let $s = \sum_{i=1}^n \alpha_i \chi_{A_i}$ be a simple function. We can assume that $\alpha_i \neq 0$ for all $1 \leq i \leq n$. Let $p < \infty$. Then we have*

$$s \in L^p_\mu(X) \quad \Longleftrightarrow \quad s \in S$$

Proof. Note that $\int_X |f|^p \, d\mu = \sum_{i=1}^n |\alpha_i|^p \mu(A_i)$. Then $\int_X |f|^p \, d\mu < \infty$ if and only if $\mu(A_i) < \infty$ for every $i = 1, \ldots, n$, which means exactly that $s \in S$. □

Theorem 11.32. *S is dense in $L^p_\mu(X)$ for every $1 \leq p < \infty$.*

Proof.

First, let $f\colon X \to [0, \infty)$. Then by Theorem 4.22 there exist simple functions $s_n\colon X \to \mathbb{R}$ such that

$$0 \leq s_1 \leq s_2 \leq \cdots \leq f$$

and $s_n(x) \to f(x)$ as $n \to \infty$ for every $x \in X$. Since $f \in L^p_\mu(X)$, we also have $s_n \in L^p_\mu(X)$, hence $s_n \in S$ by the above lemma. We also have $|f - s_n| \to 0$ as $n \to \infty$ and $|f - s_n|^p \leq |f|^p \in L^1_\mu(X)$ for all $n \geq 1$. Thus by Lebesgue's Dominated Convergence theorem $\int_X |f - s_n|^p \, d\mu \to 0$, hence $\|f - s_n\|_p \to 0$.

Second, let $f\colon X \to \mathbb{R}$. Then we approximate f^+ and f^-, separately, by simple functions $s_n \in S$, as above.

Third, let $f\colon X \to \mathbb{C}$. Then we approximate $\mathtt{Re}\, f$ and $\mathtt{Im}\, f$, separately, by simple functions $s_n \in S$, as above. □

- This is not true for $p = \infty$ (for example, let $X = \mathbb{R}$ and $f(x) \equiv 1$).
- Recall $C_c(\mathbb{R})$ denotes the space of cont. functions $f\colon \mathbb{R} \to \mathbb{C}$ with compact support.

Theorem 11.33. *$C_c(\mathbb{R})$ is dense in $L^p_{\mathbf{m}}(\mathbb{R})$ for every $1 \le p < \infty$.*

Proof. Due to Theorem 11.32 it is enough to show that for any simple function $s \in S$ and any $\varepsilon > 0$ there exists a continuous function $g \in C_c(\mathbb{R})$ such that $\|s - g\|_p < \varepsilon$. By Lusin's Theorem 9.4(b) there exists $g \in C_c(\mathbb{R})$ such that

$$\mathbf{m}\{x\colon s(x) \neq g(x)\} < \varepsilon$$

and

$$\sup|g(x)| \le \sup|s(x)| =: s_{\max}.$$

Therefore

$$\|s - g\|_p^p = \int_X |s - g|^p \, d\mu = \int_{\{x\colon s(x) \neq g(x)\}} |s - g|^p \, d\mu < [2s_{\max}]^p \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the theorem is proved. □

- This is not true for $p = \infty$ (for example, let $f(x) \equiv 1$).
- We can say that the $L^p_{\mathbf{m}}(\mathbb{R})$ space, for $1 < p < \infty$, is the completion (in the p-metric) of the space $C_c(\mathbb{R})$ of continuous functions with compact support.
- The previous statement is not true for $p = \infty$.
- If we restrict our domain to a finite interval $[a, b]$, then much simpler functions – polynomials – can be used to approximate integrable functions. The set of polynomials on $[a, b]$ is dense in $L^p_{\mathbf{m}}([a, b])$ for any $1 \le p < \infty$. Indeed, the above theorem implies that the space of continuous functions $C([a, b])$ is dense in $L^p_{\mathbf{m}}([a, b])$, and then **Weierstrass approximation theorem** ensures that polynomials are dense in $C([a, b])$. More precisely, the Weierstrass theorem states that every continuous function on $[a, b]$ can be uniformly approximated as closely as desired by a polynomial function. This last theorem is not a part of our course, though.

- Recall that a step function $\varphi\colon \mathbb{R} \to \mathbb{C}$ is defined by $\varphi = \sum_{i=1}^n \alpha_i \chi_{I_i}$ for some $\alpha_i \in \mathbb{C}$ and disjoint finite intervals $I_i \subset \mathbb{R}$.

Theorem 11.34. *Step functions $f\colon \mathbb{R} \to \mathbb{C}$ are dense in $L^p_{\mathbf{m}}(\mathbb{R})$.*

Proof. Due to Theorem 11.33 it is enough to show that for any continuous function $f \in C_c(\mathbb{R})$ with compact support and any $\varepsilon > 0$ there exists a step function φ such that $\|f - \varphi\|_p < \varepsilon$. Let $[-A, A] \subset \mathbb{R}$ be a large finite interval containing the support of f, i.e., $f(x) = 0$ for $|x| > A$. Since f is continuous on the compact interval $[-A, A]$, it is uniformly continuous, hence for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|x' - x''| < \delta \qquad \Longleftrightarrow \qquad |f(x') - f(x'')| < \varepsilon.$$

Let $\Delta = \{I_i\}_{i=1}^n$ be a finite partition of $[-A, A]$ into small intervals of length $\le \delta$ and let $x_i \in I_i$ be arbitrary points. We define the step function by $\varphi(x) =$

$\sum_{i=1}^n f(x_i)\chi_{I_i}(x)$. Then $|f(x)-\varphi(x)| \le \varepsilon$ for all $x \in [-A, A]$, hence

$$\|f-\varphi\|_p^p = \int_X |f-\varphi|^p \, d\mu \le 2A\varepsilon^p.$$

Since $\varepsilon > 0$ is arbitrary, the theorem is proved. □

- We can say that the $L^p_{\mathbf{m}}(\mathbb{R})$ space, for $1 < p < \infty$, is the completion (in the p-metric) of the space of step functions.
- The previous statement is not true for $p = \infty$.
- The above proof shows $\|f-\varphi\|_\infty \le \varepsilon$ (i.e. cont. functions $f \in C_c(\mathbb{R})$ with compact support can be approximated arbitrarily well by step functions in the L^∞ metric).

Next we describe the completion of the space $C_c(\mathbb{R})$ of continuous functions with compact support in the L^∞ metric.

Definition 11.35. A function $f\colon \mathbb{R} \to \mathbb{C}$ is said to **vanish at infinity** if $f(x) \to 0$ as $|x| \to \infty$. The space of continuous functions vanishing at infinity is denoted by $C_0(\mathbb{R})$.

- We have $C_c(\mathbb{R}) \subset C_0(\mathbb{R})$, but not vice versa.

Theorem 11.36. *The completion of $C_c(\mathbb{R})$ in the L^∞-metric is $C_0(\mathbb{R})$ (modulo the identification of equivalent functions).*

Proof. Let $f_1, f_2 \ldots \in C_c(\mathbb{R})$ be a Cauchy sequence (in the L^∞ norm).

Due to Theorem 11.28 it converges to a limit function, $f_n \to f \in L^\infty$. We will show

(i) $f \in C_0(\mathbb{R})$;

(ii) every function $f \in C_0(\mathbb{R})$ is a limit of a sequence of functions $f_n \in C_c(\mathbb{R})$.

Proof of (i): For any continuous function $g\colon \mathbb{R} \to \mathbb{C}$ we have

$$\sup|g| = \text{ess-sup}\,|g|.$$

Therefore

$$\sup|f_n - f_m| = \text{ess-sup}\,|f_n - f_m| = \|f_n - f_m\|_\infty.$$

Thus $\{f_n\}$ is a Cauchy sequence with respect to the usual sup norm on $\mathbb{R}$. Therefore $f_n \to f$ uniformly everywhere on $\mathbb{R}$ (not just a.e.). A uniform limit of continuous functions is a continuous function, hence $f \in C(\mathbb{R})$. Now for any $\varepsilon > 0$ there exist $n \ge 1$ such that $\sup|f_n - f| < \varepsilon$ and $A > 0$ such that $f_n(x) = 0$ for all $|x| > A$. This implies $|f(x)| < \varepsilon$ for all $|x| > A$. Thus $f \in C_0(\mathbb{R})$.

Proof of (ii): For any function $f \in C_0(\mathbb{R})$ and $n \geq 1$ define f_n as follows:

$$f_n(x) = \begin{cases} f(x) & \text{if } x \in [-n, n] \\ 0 & \text{if } |x| \geq n+1 \\ f(n) & \text{if } x \in (n, n+1) \\ -f(n) & \text{if } x \in (-n-1, -n) \end{cases}$$

In other words, $f_n(x)$ agrees with $f(x)$ on the interval $[-n, n]$, is zero beyond the slightly larger interval $[-n-1, n+1]$, and is extended linearly and continuously to the two small intervals $[n, n+1]$ and $[-n-1, -n]$. See Figure 7. Also note that

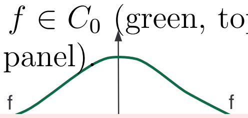
$$\sup_{[n,n+1]} |f_n(x)| \leq |f(n)| \qquad \text{and} \qquad \sup_{[-n-1,-n]} |f_n(x)| \leq |f(-n)|.$$

Therefore

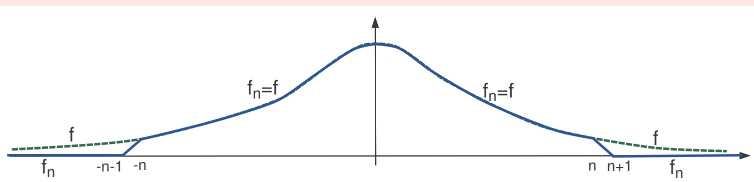
$$\sup |f_n - f| \leq 2 \sup_{|x| \geq n} |f(x)|.$$

Since $f \in C_0(\mathbb{R})$, we have $\sup_{|x| \geq n} |f(x)| \to 0$ as $n \to \infty$, thus $f_n \to f$ in the L^∞ norm. □

Figure 7: The original function $f \in C_0$ (green, top panel) and the approximating function $f_n \in C_c$ (blue, bottom panel).



Corollary 11.37. *$C_0(\mathbb{R})$ with the L^∞ norm is a Banach space.*



EXERCISE 57. [Bonus] Suppose $1 < p < \infty$ and $f \in L^p_{\mathbf{m}}((0,\infty))$ relative to the Lebesgue measure. Define

$$F(x) = \frac{1}{x}\int_0^x f(t)\,dt \qquad (0 < x < \infty).$$

Prove Hardy's inequality

$$\|F\|_p \leq \frac{p}{p-1}\|f\|_p$$

which shows that the mapping $f \to F$ carries $L^p_{\mathbf{m}}((0,\infty))$ into $L^p_{\mathbf{m}}((0,\infty))$.

[Hint: assume first that $f \geq 0$ and $f \in C_c((0,\infty))$, i.e. the support of f is a finite closed interval $[a,b] \subset (0,\infty)$. Then use integration by parts:

$$\int_\varepsilon^A F^p(x)\,dx = -p\int_\varepsilon^A F^{p-1}xF'(x)\,dx$$

where $\varepsilon < a$ and $A > b$. Note that $xF' = f - F$, and apply Hölder inequality to $\int F^{p-1}f\,dx$.]

12 Complex measures

The following simple fact will serve as a motivation for this section:

Lemma 12.1. *Let μ_1 and μ_2 be measures on $(X, \mathfrak{M})$, and $c > 0$*

(a) *We can define the constant positve multiple of a measure $\nu = c\mu_1$ by*

$$\nu(E) = c \cdot \mu_1(E) \qquad \forall\ E \in \mathfrak{M}$$

(b) *We can define the sum of two measures $\mu = \mu_1 + \mu_2$ by*

$$\mu(E) = \mu_1(E) + \mu_2(E) \qquad \forall\ E \in \mathfrak{M}$$

Proof. **(a)** By direct inspection (recall a remark made right before Theorem 3.18)

(b) Clearly, $\mu(A) \geq 0$ is a non-negative function on $\mathfrak{M}$.

For any $E = \uplus_{n=1}^{\infty} E_n$:

$$\begin{aligned}\mu(E) = \mu_1(E) + \mu_2(E) &= \sum_{n=1}^{\infty} \mu_1(E_n) + \sum_{n=1}^{\infty} \mu_2(E_n) \\ &= \sum_{n=1}^{\infty} [\mu_1(E_n) + \mu_2(E_n)] = \sum_{n=1}^{\infty} \mu(E_n),\end{aligned}$$

Thus, μ is a σ-additive function, which makes it a measure. $\square$

For part (b) above, the essential step for the proof was the rearranging of an infinite sum of non-negative numbers. We remember from Calculus that an infinite series of non-negative numbers either converges or diverges to infinity, and its sum is independent of the order in which its terms are added. This principe applies to more than one series due to Corollary 5.14.

- *However*, this does not hold for infinite series with real or complex values (Example 26)

Measures make a vector space?.

We see that we can add measures and multiply them by positive constants (scalars). Thus the collection of measures on $(X, \mathfrak{M})$ is 'almost' a vector space, except we cannot subtract measures or multiply them by negative scalars (at least not yet...). Next we will extend the notion of measure and define *complex measures* that will make a vector space.

Another motivation for complex measures.

Recall that given a measure μ on $(X, \mathfrak{M})$ and a measurable function $f \geq 0$ on X we can define a measure ν by $\nu(E) = \int_E f\, d\mu$ for all $E \in \mathfrak{M}$. (We write $d\nu = f\, d\mu$ and call f the density of ν) But what if f takes negative values, or even complex values? Would such a definition of ν give something like 'complex measure'?

Definition 12.2. A function $\lambda\colon \mathfrak{M} \to \mathbb{C}$ is a **complex measure** if it is σ-additive.

i.e. for any countable collection of pairwise disjoint measurable sets $\{E_i\}_{i=1}^{\infty}$

$$\lambda\big(\uplus_{i=1}^{\infty} E_i\big) = \sum_{i=1}^{\infty} \lambda(E_i).$$

- The measures defined in earlier sections (denoted μ) are called **positive measures**.
- Observe *not every positive measure is a complex measure.* This is because positive measures can take infinite values and complex measures cannot.
- Observe that the convergence of the series in Definition 12.2 is now part of the requirement (unlike positive measures, where the corresponding series could either converge or diverge to infinity).

Reminder:
- ∘ A complex series $\sum_{n=1}^{\infty} a_n$ converges absolutely if $\sum_{n=1}^{\infty} |a_n| < \infty$
- ∘ A series is absolutely convergent if and only if its sum is finite and does not change after a rearrangement of its terms.

Lemma 12.3. *The series $\sum_{i=1}^{\infty} \lambda(E_i)$ in Definition 12.2 must converge absolutely.*

Proof. Indeed, the union of the sets E_i is not changed if the subscripts are permuted, thus every rearrangement of the above series must converge to the same value. □

Definition 12.4. A complex measure λ on $(X, \mathfrak{M})$ is **dominated** by a positive measure μ if $|\lambda(E)| \leq \mu(E)$ for all $E \in \mathfrak{M}$.

- Every complex measure has a dominating measure.

The **smallest dominating measure** can be constructed as follows:

Definition 12.5. Let λ be a complex measure on $(X, \mathfrak{M})$. Define $|\lambda|$ on $\mathfrak{M}$ by

$$|\lambda|(E) = \sup \sum_{n=1}^{\infty} |\lambda(E_n)|$$

where the supremum is taken over all measurable partitions $E = \uplus_{n=1}^{\infty} E_n$
Then $|\lambda|$ is called the **total variation** of λ.

- The total variation $|\lambda|$ actually *is* a positive measure (see the next theorem).

• If λ is a positive measure, then of course $|\lambda| = \lambda$

Theorem 12.6. *Let λ be a complex measure on $(X, \mathfrak{M})$.*

(a) *The total variation $|\lambda|$ is a positive measure on $(X, \mathfrak{M})$ and dominates λ i.e. $|\lambda(E)| \leq |\lambda|(E)$ $\forall E \in \mathfrak{M}$*

(b) *If μ is a positive measure on $(X, \mathfrak{M})$ dominating λ: $|\lambda|(E) \leq \mu(E)$ $\forall E \in \mathfrak{M}$*

Proof. Note: (a) and (b) imply $|\lambda(E)| \leq |\lambda|(E) \leq \mu(E)$ $\forall E \in \mathfrak{M}$

(a) Let $E = \biguplus_{n=1}^{\infty} E_n$. First we show that

$$|\lambda|(E) \geq \sum_{n=1}^{\infty} |\lambda|(E_n). \tag{12.1}$$

Let $\varepsilon > 0$. We can find a partition $E_n = \biguplus_{m=1}^{\infty} E_{nm}$ of E_n such that

$$|\lambda|(E_n) \leq \frac{\varepsilon}{2^n} + \sum_{m=1}^{\infty} |\lambda(E_{nm})|$$

Adding these inequalities up gives

$$\sum_{n=1}^{\infty} |\lambda|(E_n) \leq \varepsilon + \sum_{m,n=1}^{\infty} |\lambda(E_{nm})| \leq \varepsilon + |\lambda|(E)$$

(because $E = \biguplus_{m,n=1}^{\infty} E_{nm}$ is a countable partition of E). Since $\varepsilon > 0$ is arbitrary, we get (12.1). Now we prove that

$$|\lambda|(E) \leq \sum_{n=1}^{\infty} |\lambda|(E_n). \tag{12.2}$$

For any $\varepsilon > 0$ we can find a partition $E = \biguplus_{m=1}^{\infty} A_m$ such that

$$|\lambda|(E) \leq \varepsilon + \sum_{m=1}^{\infty} |\lambda(A_m)|$$

Denote $E_{nm} = E_n \cap A_m$. Then $\lambda(A_m) = \sum_{n=1}^{\infty} \lambda(E_{nm})$ and by the triangle inequality

$$|\lambda|(E) \leq \varepsilon + \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} |\lambda(E_{nm})| = \varepsilon + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} |\lambda(E_{nm})| \leq \varepsilon + \sum_{n=1}^{\infty} |\lambda|(E_n)$$

(because $E_n = \biguplus_{m=1}^{\infty} E_{nm}$ is a partition of E_n). Since $\varepsilon > 0$ is arbitrary, we get (12.2). Now (12.1) and (12.2) together imply that $|\lambda|$ is a σ-additive. Thus Part (a).

(b) We have

$$|\lambda|(E) = \sup \sum_{n=1}^{\infty} |\lambda(E_n)| \leq \sup \sum_{n=1}^{\infty} \mu(E_n) = \mu(E)$$

the supremum is taken over all partitions of $E \in \mathfrak{M}$ into measurable subsets E_n □

Our next goal is to show that the total variation $|\lambda|$ is not only a positive measure, but a *finite* positive measure (i.e. $|\lambda|(X) < \infty$).

First we need a technical lemma about complex numbers:

Lemma 12.7. *If $z_1, \dots, z_N \in \mathbb{C}$, then there exists $S \subset \{1, \dots, N\}$ such that:*

$$\left|\sum_{k\in S} z_k\right| \geq \frac{1}{\pi}\sum_{k=1}^{N} |z_k|.$$

Proof. We use polar representation for complex numbers: $z_k = |z_k| e^{\mathbf{i}\alpha_k}$

For each $\theta \in [-\pi, \pi]$, let $S(\theta)$ denote the set of all $k \in \{1, \dots, N\}$ for which $\cos(\alpha_k - \theta) > 0$ (this condition means, geometrically, that the point z_k lies in the half-plane determined by the normal direction θ). Then

$$\begin{aligned}
\left|\sum_{k\in S(\theta)} z_k\right| = \left|\sum_{k\in S(\theta)} z_k e^{-\mathbf{i}\theta}\right| &\geq \texttt{Re} \sum_{k\in S(\theta)} z_k e^{-\mathbf{i}\theta} \\
&= \texttt{Re} \sum_{k\in S(\theta)} |z_k| e^{\mathbf{i}(\alpha_k - \theta)} \\
&= \sum_{k\in S(\theta)} |z_k| \cos(\alpha_k - \theta) \\
&= \sum_{k=1}^{N} |z_k| \cos^+(\alpha_k - \theta) :\, = g(\theta)
\end{aligned}$$

where we adopt notation $\cos^+ \gamma :\, = \max\{0, \cos\gamma\}$. The last line defines $g(\theta)$.

Now $g(\theta)$ is a function on $[-\pi, \pi]$, and it is easy to see that $g(\theta)$ is continuous. Thus it takes a maximum at some point θ_0. That maximum is at least as large as the average of $g(\theta)$ over $[-\pi, \pi]$. Thus we have

$$\left|\sum_{k\in S(\theta_0)} z_k\right| \geq g(\theta_0) \geq \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\theta)\, d\theta = \sum_{k=1}^{N} |z_k| \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos^+(\alpha_k - \theta)\, d\theta = \frac{1}{\pi}\sum_{k=1}^{N} |z_k|$$

the last identity is based on the following elementary fact:

$$\forall\alpha: \qquad \int_{-\pi}^{\pi} \cos^+(\alpha - \theta)\, d\theta = \int_{-\pi}^{\pi} \cos^+\theta\, d\theta = \int_{-\pi/2}^{\pi/2} \cos\theta\, d\theta = 2$$

$\square$

Theorem 12.8.
If λ is a complex measure on $(X, \mathfrak{M})$, then $|\lambda|$ is a finite measure (*i.e.* $|\lambda|(X) < \infty$)

Proof. First we need certain preparations.

Suppose that some set E has $|\lambda|(E) = \infty$. Put $t = \pi(1 + |\lambda(E)|)$. Note that t is a finite positive number. Since $|\lambda|(E) > t$, there is a countable partition $E = \uplus E_n$ of E such that $\sum_{n=1}^{\infty} |\lambda(E_n)| > t$. Furthermore, $\sum_{i=1}^{N} |\lambda(E_n)| > t$ for some $N \geq 1$.

Apply Lemma 12.7 with $z_n = \lambda(E_n)$ to conclude that there is a subcollection $S \subset \{1, \ldots, N\}$ such that

$$\left|\sum_{n\in S} \lambda(E_n)\right| \geq \frac{1}{\pi} \sum_{n=1}^{n} |\lambda(E_n)| > \frac{t}{\pi}$$

Consider the sets $A = \uplus_{n\in S} E_n$ and $B = E \setminus A$. We have

$$|\lambda(A)| = \left|\sum_{n\in S} \lambda(E_n)\right| > \frac{t}{\pi} \geq 1$$

and

$$|\lambda(B)| = |\lambda(E) - \lambda(A)| \geq |\lambda(A)| - |\lambda(E)| > \frac{t}{\pi} - |\lambda(E)| = 1.$$

In other words, every set E with $|\lambda|(E) = \infty$ can be split into two subsets $E = A \uplus B$ such that $|\lambda(A)| > 1$ and $|\lambda(B)| > 1$. Since $|\lambda|(E) = |\lambda|(A) + |\lambda|(B)$, at least one of $|\lambda|(A)$ and $|\lambda|(B)$ has to be infinite, and without loss of generality we can assume that $|\lambda|(B) = \infty$.

We now turn to the main part of the proof. Suppose $|\lambda|(X) = \infty$. Then we can split $X = A_1 \uplus B_1$ so that $|\lambda(A_1)| > 1$ and $|\lambda|(B_1) = \infty$. Then inductively we split each B_{k-1} into two parts, A_k and B_k, such that $|\lambda(A_k)| > 1$ and $|\lambda|(B_k) = \infty$. In the end we get a countable sequence of disjoint sets A_k such that $|\lambda(A_k)| > 1$ for all k. Now for the set $A = \uplus_{k=1}^{\infty} A_k$ we have

$$\lambda(A) = \sum_{k=1}^{\infty} \lambda(A_k)$$

where the series must converge absolutely by Lemma 12.3. But it cannot converge absolutely because $|\lambda(A_k)| > 1$ for all k, a contradiction. $\square$

Remark: Observe that the range of λ is bounded, since $|\lambda(E)| \leq |\lambda|(E) \leq |\lambda|(X)$

i.e. All values of λ lie in a closed disk D of radius $R = |\lambda|(X)$

i.e. $\lambda(E) \in D$ for all $E \in \mathfrak{M}$.

EXERCISE 58. Let $(X, \mathcal{M}, \mu)$ be a positive measure space. For $f \in L^1_\mu$ define

$$\mu_f(E) = \int_E f \, d\mu \qquad (\forall E \in \mathfrak{M})$$

Show:
(a) μ_f is a complex measure;
(b) $|\mu_f| = \mu_{|f|}$, assuming that f is real-valued;
(c) $|\mu_f| = \mu_{|f|}$, now for a general $f \in L^1_\mu(X)$.

DEFINITION 12.9. Denote $\underline{\mathbb{M}(X, \mathfrak{M})}$ the set of all complex measures on $(X, \mathfrak{M})$

- For $\lambda_1, \lambda_2 \in \mathbb{M}(X, \mathfrak{M})$ we define a measure $\lambda = \lambda_1 + \lambda_2$ by
$$\lambda(E) = \lambda_1(E) + \lambda_2(A) \quad {\scriptstyle \forall E \in \mathfrak{M}}$$
- For $\lambda_1 \in \mathbb{M}(X, \mathfrak{M})$ on $(X, \mathfrak{M})$ and $c \in \mathbb{C}$ we define a measure $\lambda = c\lambda_1$ by
$$\lambda(E) = c\lambda_1(E) \quad {\scriptstyle \forall E \in \mathfrak{M}}$$

- It is easy to check the λ's defined above are measures.

Theorem 12.10. $\mathbb{M}(X, \mathfrak{M})$ *is a vector space with a norm* $\|\mu\| = |\mu|(X)$

Proof. The verification is routine.
To show that $\|\lambda\| = |\lambda|(X)$ is a norm, we need to check the following:

- $\|\lambda\| = 0$ implies $\lambda = 0$. Indeed, $|\lambda(E)| \leq |\lambda|(E) \leq |\lambda|(X) = 0 \quad (\forall E \in \mathfrak{M})$
- $\lambda = 0$ implies $\|\lambda\| = 0$. Indeed, $|\lambda|(X) = \sup \sum |\lambda(E_n)| = 0$ where $X = \uplus E_n$
- $|c\lambda|(X) = |c||\lambda|(X)$. Indeed, for all partitions $X = \uplus E_n$
$$|c\lambda|(X) = \sup \sum |c\lambda(E_n)| = \sup |c| \sum |\lambda(E_n)| = |c| \sup \sum |\lambda(E_n)| = |c||\lambda|(X)$$
- $|\lambda_1 + \lambda_2|(X) \leq |\lambda_1|(X) + |\lambda_2|(X)$. Indeed, for all partitions $X = \uplus E_n$
$$\begin{aligned}
|\lambda_1 + \lambda_2|(X) &= \sup \sum |(\lambda_1 + \lambda_2)(E_n)| \\
&= \sup \sum |\lambda_1(E_n) + \lambda_2(E_n)| \\
&\leq \sup \sum |\lambda_1(E_n)| + |\lambda_2(E_n)| \\
&\leq \sup \sum |\lambda_1(E_n)| + \sum |\lambda_2(E_n)| \\
&\leq \sup \sum |\lambda_1(E_n)| + \sup \sum |\lambda_2(E_n)| \\
&= |\lambda_1|(X) + |\lambda_2|(X)
\end{aligned}$$

$\square$

EXERCISE 59. **[Bonus]** Prove that the space $\mathbb{M}(X, \mathfrak{M})$ with the norm $\|\lambda\|$ is a Banach space i.e. it is a complete metric space (every Cauchy sequence converges to a limit).

Hint: given a Cauchy sequence of complex measures $\{\lambda_n\}$ you need to construct the limit measure μ and prove that $\|\lambda_n - \lambda\| \to 0$ as $n \to \infty$.

A particular class of complex measure consists of those taking *real values* (positive or negative). Such measures are called **signed measures** or **charges**.

DEFINITION 12.11. Let λ be a signed measure on $(X, \mathfrak{M})$ (i.e. a complex measure with real values). Define

$$\lambda^+ = \tfrac{1}{2}(|\lambda| + \lambda) \qquad \text{and} \qquad \lambda^- = \tfrac{1}{2}(|\lambda| - \lambda)$$

These are called **positive and negative variations** of λ.

Proposition 12.12. *λ^+ and λ^- are finite positive measures.*

Proof. Since $|\lambda|$ is a finite measure (Theorem 12.8), it is a complex measure. Thus λ^+ and λ^- are complex measures (being sums of complex measures).

Next, for every $E \in \mathfrak{M}$ we have $|\lambda(E)| \leq |\lambda|(E)$, hence

$$\lambda^+(E) = \tfrac{1}{2}\big(|\lambda|(E) + \lambda(E)\big) \geq 0 \qquad \text{and} \qquad \lambda^-(E) = \tfrac{1}{2}\big(|\lambda|(E) - \lambda(E)\big) \geq 0$$

thus λ^+ and λ^- are positive measures. □

Corollary 12.13. *Every signed measure μ satisfies*

$$\lambda = \lambda^+ - \lambda^- \qquad \textit{and} \qquad |\lambda| = \lambda^+ + \lambda^-$$

The first formula above is called **Jordan decomposition** of λ:

$$\lambda = \lambda^+ - \lambda^-$$

Every signed measure is the difference between two finite positive measures.

- We will prove λ^+ and λ^- are *minimal* measures that satisfy Jordan decomposition.

Definition 12.14.

Let μ be a positive measure and λ a complex measure on $(X, \mathfrak{M})$

(a) λ is **concentrated on** $A \in \mathfrak{M}$ if

$$\lambda(E) = \lambda(E \cap A) \quad \forall E \in \mathfrak{M}$$

(b) Two complex measures λ_1 and λ_2 are **mutually singular** (written $\underline{\lambda_1 \perp \lambda_2}$) if there exist disjoint subsets $A_1, A_2 \subset X$ s.t λ_1 is concentrated on A_1 and λ_2 is concentrated on A_2.

(c) λ is **absolutely continuous** with respect to μ (written $\underline{\lambda \ll \mu}$) if

$$\mu(E) = 0 \;\Rightarrow\; \lambda(E) = 0 \quad \forall E \in \mathfrak{M}$$

Note: The set A in (a) and the sets A_1, A_2 in (b) are not unique.

λ is concentrated on A if and only if

$$E \subset A^c = \emptyset \implies \lambda(E) = 0 \quad \forall E \in \mathfrak{M}$$

If $\lambda \ll \mu$ and μ is concentrated on A, then λ is concentrated on A as well.

Suppose that $0 \neq c \in \mathbb{C}$; then

(i) $\lambda \ll \mu \iff c\lambda \ll \mu$

(ii) λ is concentrated on $A \iff c\lambda$ is concentrated on A.

(iii) $\lambda_1 \perp \lambda_2 \iff \exists A\colon\; \lambda_1$ is conc. on A and λ_2 is conc. on A^c

Example 20.

Let μ be a positive measure on $(X, \mathfrak{M})$ and $f \in L^1_\mu$. Then the measure μ_f defined by

$$\mu_f(E) = \int_E f \, d\mu \qquad \text{(recall Exercise 58)}$$

is absolutely continuous with respect to μ, i.e. $\mu_f \ll \mu$.

Remark: We will see that any complex measure $\lambda \ll \mu$ is actually one of the above type. (i.e. $\lambda = \mu_f$ for some $f \in L^1_\mu$)

Proposition 12.15.
Let μ's be positive measures and λ's complex measures on $(X, \mathfrak{M})$

(a) *λ is concentrated on A, then $|\lambda|$ is concentrated on A*

(b) *If $\lambda_1 \perp \lambda_2$, then $|\lambda_1| \perp |\lambda_2|$*

(c) *If $\lambda_1 \perp \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 + \lambda_2 \perp \mu$*

(d) *If $\lambda_1 \ll \mu$ and $\lambda_2 \ll \mu$, then $\lambda_1 + \lambda_2 \ll \mu$*

(e) *If $\lambda \ll \mu$, then $|\lambda| \ll \mu$*

(f) *If $\lambda_1 \ll \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 \perp \lambda_2$*

(g) *If $\lambda \ll \mu$ and $\lambda \perp \mu$, then $\lambda = 0$*

Proof. Straightforward verification. □

Exercise 60. Let μ be a positive measure on $(X, \mathfrak{M})$ and $f, g \in L^1_\mu$.
Define μ_f and μ_g (as in Exercise 58). Prove the following:

(a) μ_f is concentrated on A if and only if $\mu\{x \in A^c : f(x) \neq 0\} = 0$;

(b) $\mu_f \perp \mu_g$ if and only if $\mu\{x \in X : f(x)g(x) \neq 0\} = 0$;

(c) if $f \geq 0$, then
$$\mu \ll \mu_f \iff f(x) > 0 \text{ for } \mu-\text{a.e. } x \in X.$$

Exercise 61. Let λ be a positive measure on $(X, \mathfrak{M})$

(a) Prove λ is concentrated on A if and only if $\lambda(A^c) = 0$

(b) Give a counterexample to this statement in the case of a complex measure λ

Definition 12.16. A positive measure μ on $(X, \mathfrak{M})$ is said to be **σ-finite** if there exists a countable sequence $\{X_n\}_{n=1}^\infty \subset \mathfrak{M}$ s.t. $X = \cup_n X_n$ and $\mu(X_n) < \infty \ \forall n$

- The sets X_n can be made disjoint (i.e. we may assume $X = \uplus_{n=1}^\infty X_n$)

Theorem 12.17. **Lebesgue Decomposition**

Let μ be a positive σ-finite measure and λ a complex measure on $(X, \mathfrak{M})$.

Then there is a unique pair of complex measures λ_{a} and λ_{s} such that

$$\lambda = \lambda_{\mathrm{a}} + \lambda_{\mathrm{s}}$$

and $\lambda_{\mathrm{a}} \ll \mu$ and $\lambda_{\mathrm{s}} \perp \mu$

Theorem 12.18. **Radon-Nikodym**

Let μ be a positive σ-finite measure and $\lambda_{\mathrm{a}} \ll \mu$ a complex measure absolutely continuous with respect to μ on $(X, \mathfrak{M})$.

Then there is a unique $h \in L^1_\mu$ (up to μ-equivalence) such that

$$\lambda_{\mathrm{a}}(E) = \int_E h \, d\mu$$

(i.e. $\lambda_{\mathrm{a}} = \mu_h$) We can also write $d\lambda_{\mathrm{a}} = h \, d\mu$, or $h = d\lambda_{\mathrm{a}} \,/\, d\mu$

*Note: h is called the **Radon-Nikodym derivative** of λ_{a} w.r.t. μ.*

Proof. **Theorem 12.17 and Theorem 12.18**

These two theorems will be proved together. First we address the uniqueness.

Uniqueness of Lebesgue Decomposition. Suppose $(\lambda'_{\mathrm{a}}, \lambda'_{\mathrm{s}})$ is another pair of measures satisfying the requirements. Then $\lambda_{\mathrm{a}} + \lambda_{\mathrm{s}} = \lambda'_{\mathrm{a}} + \lambda'_{\mathrm{s}}$, hence

$$\lambda'_{\mathrm{a}} - \lambda_{\mathrm{a}} = \lambda_{\mathrm{s}} - \lambda'_{\mathrm{s}}$$

where $\lambda'_{\mathrm{a}} - \lambda_{\mathrm{a}} \ll \mu$ and $\lambda_{\mathrm{s}} - \lambda'_{\mathrm{s}} \perp \mu$ due to Proposition 12.15, parts (d) and (c), respectively. Hence both differences are zero due to Proposition 12.15 (g).

Uniqueness of Radon-Nikodym derivative. Suppose $h' \in L^1_\mu$ is another function satisfying the requirements. Then $\forall E \in \mathfrak{M}$

$$\int_E (h - h') \, d\mu = \int_E h \, d\mu - \int_E h' \, d\mu = \lambda_{\mathrm{a}}(E) - \lambda_{\mathrm{a}}(E) = 0$$

Hence $h - h' = 0$ a.e. due to Theorem 7.11 (b).

Existence. First assume that μ is a *finite* measure, i.e., $\mu(X) < \infty$. Then $\sigma = \mu + |\lambda|$ is a finite positive measure on X. Note that if $\sigma(E) = 0$ for some $E \in \mathfrak{M}$, then $\mu(E) = 0$ and $|\lambda|(E) = 0$, implying $\lambda(E) = 0$. Thus μ and λ are both absolutely continuous with respect to σ.

Now let $X = A_1 \uplus A_2 \uplus \cdots \uplus A_n$ be a finite partition of X into measurable sets. We denote this partition by P and define a function h_P on X by

$$h_P = \sum_{i=1}^{n} \frac{\lambda(A_i)}{\sigma(A_i)} \chi_{A_i},$$

i.e., h_P takes value $\lambda(A_i)/\sigma(A_i)$ on A_i (in the exceptional case where $\sigma(A_i) = 0$ we set $h_P \equiv 0$ on A_i). Note that h_P plays the role of a rough approximation to the density of λ w.r.t. σ. Note that $0 \le |h_P| \le 1$, because for each i we have $|\lambda(A_i)| \le |\lambda|(A_i) \le \sigma(A_i)$.

Further, we have

$$\int_X h_P \, d\sigma = \sum_{i=1}^{n} \frac{\lambda(A_i)}{\sigma(A_i)} \sigma(A_i) = \sum_{i=1}^{n} \lambda(A_i) = \lambda(X).$$

Also note that

$$\int_X |h_P|^2 \, d\sigma = \sum_{i=1}^{n} \frac{|\lambda(A_i)|^2}{\sigma(A_i)} \le \sum_{i=1}^{n} |\lambda(A_i)| \le |\lambda|(X) < \infty,$$

so the following upper bound is finite:

$$H = \sup_P \int_X |h_P|^2 \, d\sigma < \infty.$$

Let another partition $Q = \{B_1, \ldots, B_m\}$ of X be a *refinement* of P (i.e., every B_j is contained in one of the A_i's). It is important to note that

$$\int_X h_Q \bar{h}_P \, d\sigma = \sum_{A_i} \sum_{B_j \subset A_i} \frac{\lambda(B_j)}{\sigma(B_j)} \frac{\bar{\lambda}(A_i)}{\sigma(A_i)} \sigma(B_j) = \sum_{A_i} \frac{\lambda(A_i)}{\sigma(A_i)} \frac{\bar{\lambda}(A_i)}{\sigma(A_i)} \sigma(A_i) = \int_X h_P \bar{h}_P \, d\sigma$$

Therefore

$$\begin{aligned}
\int_X |h_Q|^2 \, d\sigma &= \int_X \big(h_P + (h_Q - h_P)\big)\big(\bar{h}_P + (\bar{h}_Q - \bar{h}_P)\big) \, d\sigma \\
&= \int_X |h_P|^2 \, d\sigma + \int_X |h_Q - h_P|^2 \, d\sigma \\
&\quad + \int_X h_P(\bar{h}_Q - \bar{h}_P)\big) \, d\sigma + \int_X \bar{h}_P(h_Q - h_P)\big) \, d\sigma \\
&\quad \text{(the above two integrals vanish due to the previous identity)} \\
&= \int_X |h_P|^2 \, d\sigma + \int_X |h_Q - h_P|^2 \, d\sigma \\
&\ge \int_X |h_P|^2 \, d\sigma
\end{aligned}$$

Now for each $n \geq 1$ let P_n be a partition of X such that

$$H - \frac{1}{4^n} \leq \int_X |h_{P_n}|^2 \, d\sigma \leq H$$

Let Q_n be the refinement of the partitions $P_1, \ldots, P_n$ (obtained by taking intersections of the elements of $P_1, \ldots, P_n$). Then

$$H - \frac{1}{4^n} \leq \int_X |h_{P_n}|^2 \, d\sigma \leq \int_X |h_{Q_n}|^2 \, d\sigma \leq H$$

Note that Q_{n+1} is a refinement of Q_n, thus

$$\int_X |h_{Q_{n+1}} - h_{Q_n}|^2 \, d\sigma = \int_X |h_{Q_{n+1}}|^2 \, d\sigma - \int_X |h_{Q_n}|^2 \, d\sigma \leq \frac{1}{4^n}$$

By the Schwarz inequality

$$\int_X |h_{Q_{n+1}} - h_{Q_n}| \, d\sigma \leq \left[\int_X |h_{Q_{n+1}} - h_{Q_n}|^2 \, d\sigma\right]^{1/2} \left[\int_X d\sigma\right]^{1/2} \leq \frac{1}{2^n} \sqrt{\sigma(X)}$$

Therefore

$$\int_X \left(\sum_{n=1}^{\infty} |h_{Q_{n+1}} - h_{Q_n}|\right) d\sigma = \sum_{n=1}^{\infty} \int_X |h_{Q_{n+1}} - h_{Q_n}| \, d\sigma \leq \sqrt{\sigma(X)} < \infty$$

This implies that the series $\sum_{n=1}^{\infty} |h_{Q_{n+1}} - h_{Q_n}|$ converges a.e. (with respect to σ), thus the series

$$h(x) = h_{Q_1}(x) + \sum_{n=1}^{\infty} h_{Q_{n+1}}(x) - h_{Q_n}(x)$$

converges absolutely at almost every point $x \in X$. Its partial sum is $h_{Q_n}(x)$, thus

$$h(x) = \lim_{n \to \infty} h_{Q_n}(x)$$

is defined a.e. with respect to the measure σ.

Our next goal is to show that h is the density of λ, i.e. $\lambda(A) = \int_A h \, d\sigma$ for any $A \in \mathfrak{M}$. Let R_n denote the refinement of Q_n and the two-set partition $\{A, X \setminus A\}$. Then, as before,

$$\int_X |h_{R_n} - h_{Q_n}|^2 \, d\sigma = \int_X |h_{R_n}|^2 \, d\sigma - \int_X |h_{Q_n}|^2 \, d\sigma \leq \frac{1}{4^n}$$

and by the Schwarz inequality

$$\int_X |h_{R_n} - h_{Q_n}| \, d\sigma \leq \left[\int_X |h_{R_n} - h_{Q_n}|^2 \, d\sigma\right]^{1/2} \left[\int_X d\sigma\right]^{1/2} \leq \frac{1}{2^n} \sqrt{\sigma(X)}$$

Therefore

$$\int_A |h_{R_n} - h_{Q_n}| \, d\sigma \to 0$$

Now we have

$$\lambda(A) = \int_A h_{R_n}\, d\sigma = \int_A h_{Q_n}\, d\sigma + \int_A (h_{R_n} - h_{Q_n})\, d\sigma$$

Taking the limit $n \to \infty$ we get

$$\left| \int_A (h_{R_n} - h_{Q_n})\, d\sigma \right| \leq \int_A |h_{R_n} - h_{Q_n}|\, d\sigma \to 0,$$

thus

$$\lambda(A) = \lim_{n\to\infty} \int_A h_{Q_n}\, d\sigma$$

Applying Lebesgue Dominated Convergence Theorem gives

$$\lambda(A) = \int_A \lim_{n\to\infty} h_{Q_n}\, d\sigma = \int_A h\, d\sigma$$

as desired. We will denote $h = h_\lambda$.

Similarly, we can find a function h_μ (again, defined a.e.) s.t. for any $A \in \mathfrak{M}$:

$$\mu(A) = \int_A h_\mu\, d\sigma$$

Since μ is a positive measure, we have $h_\mu \geq 0$. Now define X_0 and X_1 by

$$X_0 = \{x \colon h_\mu(x) = 0 \quad \text{or not defined}\}$$

$$X_1 = X_0^c = \{x \colon h_\mu(x) > 0\}$$

Then we define two complex measures λ_a and λ_s by

$$\lambda_\mathrm{a}(A) = \lambda(A \cap X_1)$$

$$\lambda_\mathrm{s}(A) = \lambda(A \cap X_0)$$

for all $A \in \mathfrak{M}$. Then it is clear that $\lambda_\mathrm{a} + \lambda_\mathrm{s} = \lambda$ and $\lambda_\mathrm{s} \perp \mu$.

It remains to show that $\lambda_\mathrm{a} \ll \mu$. We construct the density of λ_a as

$$h(x) = \begin{cases} \dfrac{h_\lambda(x)}{h_\mu(x)} & \text{if } \ x \in X_1 \\ 0 & \text{if } \ x \in X_0 \end{cases}$$

Then for any measurable set $A \subset X_1$ we have

$$\lambda_\mathrm{a}(A) = \lambda(A) = \int_A h_\lambda\, d\sigma = \int_A h h_\mu\, d\sigma = \int_A h\, d\mu$$

which implies $\lambda_\mathrm{a} \ll \mu$.

Last case: σ-finite measures. It remains to extent our proof of existence to σ-finite measures μ. Thus assume that $X = \uplus_{n=1}^{\infty} X_n$ so that $\mu(X_n) < \infty$ for each n.

Each X_n can be treated as a measurable space with σ-algebra $\mathfrak{M}_n = \{A \in \mathfrak{M} \colon A \subset X_n\}$. Then the restrictions of μ and λ to $\mathfrak{M}_n$ will be a positive and a complex measure on X_n, respectively, we will denote them by μ_n and λ_n. Each μ_n is finite, so our proof applies and gives us decompositions

$$X_n = X_{n,0} \uplus X_{n,1}$$

as above and

$$\lambda_n = \lambda_{n,\mathrm{a}} + \lambda_{n,\mathrm{s}}$$

and a density $h_n = d\lambda_{n,\mathrm{a}}/d\mu_n$ on X_n.

We want to define two complex measures λ_{a} and λ_{s} on X by

$$\lambda_{\mathrm{a}}(A) = \sum_{n=1}^{\infty} \lambda_{n,\mathrm{a}}(A \cap X_n) = \sum_{n=1}^{\infty} \lambda_n(A \cap X_{n,1}) = \sum_{n=1}^{\infty} \lambda(A \cap X_{n,1})$$
$$\lambda_{\mathrm{s}}(A) = \sum_{n=1}^{\infty} \lambda_{n,\mathrm{a}}(A \cap X_n) = \sum_{n=1}^{\infty} \lambda_n(A \cap X_{n,0}) = \sum_{n=1}^{\infty} \lambda(A \cap X_{n,0})$$

Note that for $i = 0, 1$ we have

$$\sum_{n=1}^{\infty} |\lambda(A \cap X_{n,i})| \leq \sum_{n=1}^{\infty} |\lambda|(A \cap X_{n,i}) = |\lambda|\big(A \cap \uplus X_{n,i}\big) \leq |\lambda|(X)$$

thus the above series converge absolutely and their values are within the bounded disk of radius $|\lambda|(X)$. Hence the above definitions give us indeed two complex measures, λ_{a} and λ_{s}.

The density $h = d\lambda_{\mathrm{a}}/d\mu$ simply coincides with the corresponding h_n on each X_n because

$$\lambda_{\mathrm{a}}(A) = \sum_{n=1}^{\infty} \lambda_{n,\mathrm{a}}(A \cap X_n) = \sum_{n=1}^{\infty} \int_{A \cap X_n} h_n \, d\mu_n = \int_A h \, d\mu$$

and we have $h \in L^1_\mu$ because

$$\int_X |h| \, d\mu = \sum_{n=1}^{\infty} \int_{X_n} |h_n| \, d\mu_n = \sum_{n=1}^{\infty} |\lambda_{\mathrm{a}}|(X_n) = |\lambda_{\mathrm{a}}|(X) < \infty$$

$\square$

- The existence is usually proved by referring to Riesz Representation Theorem. Our proof is more elementary and constructive. It is modeled on the journal article: Bradley, R. C. *An Elementary Treatment of the Radon-Nikodym Derivative*, Amer. Math. Monthly **96**(5) (1989), 437–440.

Corollary 12.19. *If* $\lambda \ll \mu$ *then* $\lambda = h \, d\mu$ *for some* $h \in L^1_\mu$.

Extensions.

The Lebesgue Decomposition and the Radon-Nikodym theorem can be extended to the case where both μ and λ are positive σ-finite measures. But we cannot go beyond σ-finiteness, as the following exercise shows.

EXERCISE 62. Let $X = [0,1]$, μ the Lebesgue measure on X, ν the counting measure on X. Show:

(a) ν is not σ-finite.

(b) ν has no Lebesgue decomposition $\nu_{\mathrm{a}} + \nu_{\mathrm{s}}$ with respect to μ.

(c) A Lebesgue-measurable function $h\colon X \to \mathbb{C}$ is in L^1_ν if and only if $A\colon = \{x \in X : h(x) \neq 0\}$ is countable, and $\sum_{x\in A} |h(x)| < \infty$. In this case $\int_E h\, d\nu = \sum_{x \in E\cap A} h(x)$ for all E.

(d) $\mu \ll \nu$ but there is no $h \in L^1_\nu$ such that $d\mu = h\, d\nu$.

Theorem 12.20. **Characterization of absolute continuity**

Let μ be a positive measure and λ a complex measure on $(X, \mathfrak{M})$. TFAE:

(a) $\lambda \ll \mu$

(b) $\forall \varepsilon > 0 \;\; \exists \delta_\varepsilon > 0 \quad s.t. \quad \mu(E) < \delta_\varepsilon \Longrightarrow |\lambda(E)| < \varepsilon \quad {\scriptstyle (\forall E \in \mathfrak{M})}$

Proof. • The property (b) justifies the name *absolute continuity.*

• If λ is a positive but not finite measure, then (a), (b) are *not equivalent* (see Example 21). Thus (b) should *not* be used as the definition

$\boxed{\text{(b)}\Rightarrow\text{(a)}}$ $\mu(E) = 0 \;\Rightarrow\; \mu(E) < \delta_\varepsilon \;\; {\scriptstyle (\forall \delta_\varepsilon > 0)} \;\Rightarrow\; |\lambda(E)| < \varepsilon \;\; {\scriptstyle (\forall \varepsilon > 0)} \;\Rightarrow\; \lambda(E) = 0$

$\boxed{\text{(a)}\Rightarrow\text{(b)}}$ BWOC, suppose (b) is false. Then

$\exists \varepsilon > 0 : \;\; \forall n \geq 1 : \;\; \exists E_n \in \mathfrak{M} : \;\; \mu(E_n) < 2^{-n}$ and $|\lambda(E_n)| > \varepsilon$ (thus $|\lambda|(E_n) > \varepsilon$)

Put
$$A_n = \cup_{i=n}^\infty E_i \qquad A = \cap_{n=1}^\infty A_n$$

On the one hand, $\mu(A_n) \leq \sum_{i=1}^\infty 2^{-i} = 2^{-n+1}$, and $A_1 \supset A_2 \supset \cdots$, implying $\mu(A) = \lim_{n\to\infty} \mu(A_n) = 0$

On the other hand, $|\lambda|(A) = \lim_{n\to\infty} |\lambda|(A_n) \geq \limsup_{n\to\infty} |\lambda|(E_n)| \geq \varepsilon$

This contradicts the assumption $\lambda \ll \mu$, due to Proposition 12.15(e). $\square$

EXAMPLE 21. Let $X = (0, 1)$, μ be the Lebesgue measure, and $\lambda(E) = \int_E t^{-1}\,dt$. Then λ is an infinite positive measure and $\lambda \ll \mu$, but the property (b) does not hold. Indeed, just consider the sets $E_n = (\frac{1}{n}, \frac{2}{n})$...

Theorem 12.21. <u>**Polar representation**</u>

Let λ be a complex measure on $(X, \mathfrak{M})$. Then there is a measurable function $h\colon X \to \mathbb{C}$ such that $|h| = 1$ and $d\lambda = h\,d|\lambda|$.

Proof. Since $\lambda \ll |\lambda|$ (Proposition 12.15 a), we have $d\lambda = h\,d|\lambda|$ for some $h \in L^1_{|\lambda|}$. All we need is to show that $|h| = 1$. It is enough to show that $|h| = 1$ a.e. with respect to $|\lambda|$, as then h can be redefined on the null set $\{x\colon |h(x)| \neq 1\}$.

(i) First we show that $|h| \geq 1$ a.e. For any $0 < r < 1$ denote $A_r = \{x\colon |h(x)| < r\}$. For any partition $A_r = \uplus_{n=1}^{\infty} E_n$ we have

$$\sum_n |\lambda(E_n)| = \sum_n \left| \int_{E_n} h\,d|\lambda| \right| \leq \sum_n r|\lambda|(E_n) = r|\lambda|(A_r)$$

hence $|\lambda|(A_r) \leq r|\lambda|(A_r)$, which is only possible if $\lambda(A_r) = 0$. Lastly,

$$|\lambda|\big(\{x\colon |h(x) < 1\}\big) = |\lambda|\Big(\cup_n A_{1-\frac{1}{n}}\Big) \leq \sum_n |\lambda|\Big(A_{1-\frac{1}{n}}\Big) = 0.$$

(ii) Second we show that $|h| \leq 1$ a.e. For any $E \in \mathfrak{M}$ with $|\lambda|(E) > 0$ we have

$$\left| \frac{1}{|\lambda|(E)} \int_E h\,d|\lambda| \right| = \left| \frac{1}{|\lambda|(E)}\,\lambda(E) \right| = \frac{|\lambda(E)|}{|\lambda|(E)} \leq 1$$

i.e. all the $|\lambda|$-averages of h are in the unit disk. Due to Theorem 7.15 all the values $h(x)$ are in the unit disk as well (i.e. $|h(x)| \leq 1$ a.e.) □

Theorem 12.22. *Let μ be a positive measure on $(X, \mathfrak{M})$. Suppose $g \in L^1_\mu$ and $d\lambda = g\,d\mu$ Then $d|\lambda| = |g|\,d\mu$*

Proof. Due to Polar representation, $h\,d|\lambda| = d\lambda = g\,d\mu$ for some measurable h such that $|h| = 1$. This implies, for any bounded measurable f

$$\int_X fh\,d|\lambda| = \int_X fg\,d\mu$$

Now for any $E \in \mathfrak{M}$ let $f = \chi_E \bar{h}$, then

$$|\lambda|(E) = \int_E d|\mu| = \int_E \frac{g}{h}\,d\mu = \int_E \bar{h}g\,d\mu.$$

Thus $d|\lambda| = \bar{h}g\,d\mu$. Since $|\lambda|$ is a positive measure, $\bar{h}g \geq 0$ a.e. w.r.t. μ, hence

$$\bar{h}g = |\bar{h}g| = |\bar{h}| \cdot |g| = |g|$$

a.e. w.r.t. μ. This implies $d|\lambda| = |g|\,d\mu$, as claimed. □

Theorem 12.23. **Hahn decomposition**

Let λ be a real-valued measure on $(X, \mathfrak{M})$. Then there is a decomposition $X = A \uplus B$ of the space X into two measurable parts A and B such that for any $E \in \mathfrak{M}$

$$\lambda^+(E) = \lambda(E \cap A), \qquad \lambda^-(E) = -\lambda(E \cap B)$$

Proof. Due to Polar representation, $d\lambda = h\,d|\lambda|$, where $|h| = 1$. Since λ is real, it follows that h is real (first, a.e., and therefore everywhere, by redefining on a null set). Hence $h = \pm 1$. Put

$$A = \{x\colon h(x) = 1\}, \qquad B = \{x\colon h(x) = -1\}$$

Since $\lambda^+ = \frac{1}{2}(|\lambda| + \lambda)$ and since

$$\frac{1}{2}(1+h) = \begin{cases} h & \text{on } A \\ 0 & \text{on } B \end{cases}$$

we have, for any $E \in \mathfrak{M}$,

$$\lambda^+(E) = \frac{1}{2}\int_E (1+h)\,d|\lambda| = \int_{E\cap A} h\,d|\lambda| = \lambda(E \cap A).$$

Since $\lambda(E) = \lambda(E \cap A) + \lambda(E \cap B)$ and since $\lambda = \lambda^+ - \lambda^-$, we conclude that $\lambda^-(E) = -\lambda(E \cap B)$, as claimed. □

Corollary 12.24. *Let λ be a real-valued measure on $(X, \mathfrak{M})$. If $\lambda = \lambda_1 - \lambda_2$, where λ_1 and λ_2 are positive measures, then $\lambda_1 \geq \lambda^+$ and $\lambda_2 \geq \lambda^-$.*

Proof. Since $\lambda \leq \lambda_1$, we have for any $E \in \mathfrak{M}$

$$\lambda^+(E) = \lambda(E \cap A) \leq \lambda_1(E \cap A) \leq \lambda_1(E)$$

Next, since $\lambda_1 - \lambda_2 = \lambda^+ - \lambda^-$, we have $\lambda^- \leq \lambda_2$. □

- Thus, in the Jordan decomposition the measures λ^+ and λ^- are minimal.

• The rest of this section is presented without proofs.

Let X and Y be vector spaces with norms and $\Lambda\colon X \to Y$ a linear mapping.

DEFINITION 12.25. The **norm** of Λ is

$$\begin{aligned}\|\Lambda\| &= \sup\{\|\Lambda x\|\colon x \in X, \quad \|x\| = 1\}\\ &= \sup\{\|\Lambda x\|/\|x\|\colon x \in X, \quad x \neq 0\}\end{aligned}$$

We say that Λ is **bounded** if $\|\Lambda\| < \infty$.

Theorem 12.26. *The following conditions are equivalent:*

(a) Λ *is bounded;*

(b) Λ *is continuous on* X*;*

(c) Λ *is continuous at some point* $x \in X$*.*

DEFINITION 12.27. In a special case, where $Y = \mathbb{C}$, we deal with linear **functionals** $L\colon X \to \mathbb{C}$ and call

$$X^* = \{\text{all bounded linear functionals} \quad L\colon X \to \mathbb{C}\}$$

the **dual space** (to X). It is a vector space with norm $\|L\|$.

Now consider $L^p_\mu(X)$ on a measure space $(X, \mathfrak{M}, \mu)$ with $1 \leq p \leq \infty$. Let q be the exponent conjugate to p, i.e., $\frac{1}{p} + \frac{1}{q} = 1$. Given a function $g \in L^q_\mu(X)$, we construct a linear functional $\Phi_g\colon L^p_\mu(X) \to \mathbb{C}$ by

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

It is bounded (by Hölder inequality) and $\|\Phi_g\| \leq \|g\|_q$ (here $\|g\|_q$ denotes the norm in the $L^q_\mu(X)$ space).

Theorem 12.28. *Let μ be a finite or σ-finite measure and $1 \le p < \infty$. Then for every bounded linear functional $\Phi\colon L^p_\mu(X) \to \mathbb{C}$ there exists a unique $g \in L^q_\mu(X)$ (up to equivalence) such that*

$$\Phi(f) = \int_X fg\,d\mu \qquad \forall f \in L^p_\mu(X),$$

i.e. $\Phi = \Phi_g$. Furthermore, $\|\Phi\| = \|g\|_q$.

- In other words, the dual space $L^p_\mu(X)^*$ can be identified with $L^q_\mu(X)$, they are isometrically equivalent to each other.

- For $1 < p < \infty$, the theorem holds without σ-finiteness.

- This theorem does not hold for $p = \infty$, see the next exercise.

Exercise 63. [Bonus] Let $X = [0,1]$ and $\mathbf{m}$ the Lebesgue measure. Show that $L^\infty_{\mathbf{m}}(X)^* \supset L^1_{\mathbf{m}}(X)$, but $L^\infty_{\mathbf{m}}(X)^* \neq L^1_{\mathbf{m}}(X)$ (in the sense $g \to \Phi_g$). (Hint: Use the following consequence of the Hahn-Banach theorem: If X is a Banach space (i.e. a complete metric space, in which every Cauchy sequence converges to a limit) and $A \subset X$ is a closed subspace of X, with $A \neq X$, then there exists $f \in X^*$ with $f \neq 0$, and $f(x) = 0$ for all $x \in A$.)

13 Differentiation of measures

Differentiation of functions.
A function $f\colon \mathbb{R} \to \mathbb{C}$ is **differentiable** at $x \in \mathbb{R}$ and $f'(x) = A$ if the following limit exists:

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = A$$

Equivalently, we can rewrite this limit as

$$\lim_{a\to x-,\ b\to x+} \frac{f(b) - f(a)}{b-a} = A$$

In the last version, it is essential that $a \le x \le b$, i.e., $x \in I = [a,b]$ and both a and b converge to x, i.e., $|I| \to 0$. So we can rewrite the above limit once again as

$$\lim_{x\in I=[a,b],\ |I|\to 0} \frac{f(b) - f(a)}{|I|} = A$$

In the epsilon-delta language, we can rewrite this as

$$\forall \varepsilon > 0 \quad \exists \delta_\varepsilon > 0\colon \quad \left| \frac{f(b) - f(a)}{\mathbf{m}(I)} - A \right| < \varepsilon$$

for every closed interval $I = [a,b] \subset \mathbb{R}$ s.t. $x \in I$ and $\mathbf{m}(I) < \delta_\varepsilon$.

Definition 13.1. Let μ be a complex measure on $\mathbb{R}$ (with Borel σ-algebra). Then $F(x) = \mu\big((-\infty, x)\big)$ is called the **distribution function** of the measure μ.

- In this section, μ will usually denote a complex measure.
- Reminder: $\mathbf{m}$ always denotes the Lebesgue measure on $\mathbb{R}$ (and on $\mathbb{R}^k$).

Theorem 13.2. *The following two conditions are equivalent:*

(a) *The distribution function F is differentiable at $x \in \mathbb{R}$ and $F'(x) = A$;*

(b) $\forall \varepsilon > 0\ \exists \delta_\varepsilon > 0\colon\ \left|\frac{\mu(I)}{\mathbf{m}(I)} - A\right| < \varepsilon$ *for every open interval $I \subset \mathbb{R}$ s.t. $x \in I$, $\mathbf{m}(I) < \delta_\varepsilon$.*

Proof. First we need to verify that both (a) and (b) imply $\mu(\{x\}) = 0$. Then we can easily relate the statement (b) to the above definition of differentiability (open intervals can be replaced with closed ones and vice versa because $\mu(\{x\}) = 0$). □

Motivated by this we will define derivatives of measures in $\mathbb{R}^k$.

Definition 13.3. Denote the **open ball** of radius $r > 0$ centered on $x \in \mathbb{R}^k$ by

$$B(x,r) = \{y \in \mathbb{R}^k \colon |y-x| < r\}$$

If μ is a complex measure on $\mathbb{R}^k$ (with Borel σ-algebra), then we denote

$$(Q_r\mu)(x) = \frac{\mu\big(B(x,r)\big)}{\mathbf{m}\big(B(x,r)\big)}$$

and define the **symmetric derivative** of μ at x (if the limit exists) by

$$(D\mu)(x) = \lim_{r\to 0}(Q_r\mu)(x)$$

Exercise 64. Let μ be a complex Borel measure on $\mathbb{R}$ and assume that its symmetric derivative $(D\mu)(x)$ exists at some $x_0 \in \mathbb{R}$. Does it follow that its distribution function $F(x) = \mu((-\infty, x))$ is differentiable at x_0?

Definition 13.4. If μ is a complex measure on $\mathbb{R}^k$ (with Borel σ-algebra), then the **maximal function** $M\mu \colon \mathbb{R}^k \to [0,\infty]$ defined by

$$(M\mu)(x) = \sup_{0<r<\infty}(Q_r|\mu|)(x) = \sup_{0<r<\infty}\frac{|\mu|\big(B(x,r)\big)}{\mathbf{m}\big(B(x,r)\big)}.$$

- Note that the numerator is bounded because $|\mu|\big(B(x,r)\big) \le |\mu|(\mathbb{R}^k) < \infty$, while the denominator grows to infinity as $r \to \infty$.

Lemma 13.5. *The maximal function $M\mu$ is lower semicontinuous.*
i.e. the set $\{x \colon M\mu(x) > a\}$ is open for every $a \in \mathbb{R}$.

Proof. If $\mu = 0$, then $(M\mu)(x) \equiv 0$ (trivial case). If $\mu \ne 0$, then $|\mu| > 0$, so $(M\mu)(x) > 0$ for all $x \in \mathbb{R}^k$. Now for all $a \le 0$ we have $\{x \colon M\mu(x) > a\} = \mathbb{R}^k$, an open set. For $a > 0$, note that $M\mu(x) > a$ means

$$\exists r > 0, t > a\colon \quad |\mu|(B(x,r)) = t\mathbf{m}(B(x,r))$$

Now choose $\delta > 0$ such that $(r+\delta)^k < \frac{r^k t}{a}$. Then for any $y \in B(x,\delta)$ we have $B(y, r+\delta) \supset B(x,r)$, so

$$\begin{aligned}|\mu|(B(y,r+\delta)) &\ge |\mu|(B(x,r)) = t\mathbf{m}(B(x,r)) = t\Big[\frac{r}{r+\delta}\Big]^k\mathbf{m}(B(y,r+\delta)) \\ &> a\mathbf{m}(B(y,r+\delta))\end{aligned}$$

hence $(M\mu)(y) > a$ for all $y \in B(x,\delta)$. $\square$

EXAMPLE 22. The maximal function $M\mu$ is not necessarily upper semicontinuous. For instance, let μ be a uniform measure on the unit interval $[0,1]$, i.e., let $\frac{d\mu}{d\mathbf{m}} = \chi_{[0,1]}$. Then $(M\mu)(x) = 1$ for all $0 < x < 1$ and $(M\mu)(x) = \frac{1}{2}$ for $x = 0$ and $x = 1$.

Lemma 13.6. **The Covering Lemma**

Let $W = \cup_{i=1}^{N} B(x_i, r_i)$ be a finite union of open balls in $\mathbb{R}^k$.
Then there is $S \subset \{1, 2, \ldots, N\}$ such that

(a) *the balls $B(x_i, r_i)$, $i \in S$, are disjoint;*

(b) *$W \subset \cup_{i\in S} B(x_i, 3r_i)$;*

(c) *$\mathbf{m}(W) \leq 3^k \sum_{i\in S} \mathbf{m}\big(B(x_i, r_i)\big)$.*

Proof. Order the balls by their size, so that $r_1 \geq r_2 \geq \cdots \geq r_N$.

Put $i_1 = 1$, and discard all the balls intersecting $B(x_{i_1}, r_{i_1})$. Let $B(x_{i_2}, r_{i_2})$ be the first (biggest) ball among the remaining ones (if there are any), and so on, as long as possible. Then (a) is obvious.

To prove (b) note that all the balls discarded right after $B(x_{i_k}, r_{i_k})$ is selected and before the next one, $B(x_{i_{k+1}}, r_{i_{k+1}})$, is selected are not larger than $B(x_{i_k}, r_{i_k})$ and intersect it, thus they lie within $B(x_{i_k}, 3r_{i_k})$. That proves (b).

Finally, (c) follows from (b) since $\mathbf{m}(B(x, 3r)) = 3^k \mathbf{m}(B(x, r))$ for any ball $B(x, r)$. $\square$

(1) Order the balls by their radii (largest to smallest)

(2) Start with the largest ball and disregard any other balls intersecting it, repeat using the next largest ball, etc. We obtain a set of disjoint balls.

(3) The balls formed with $3\times$radii on the selected balls form a covering of W

Theorem 13.7. *Let μ be a complex measure on $\mathbb{R}^k$. Then $\forall z > 0$*

$$\mathbf{m}(M\mu > z) \le \frac{3^k \|\mu\|}{z}$$

where $\|\mu\| = |\mu|(\mathbb{R}^k)$.

- Roughly speaking, the maximal function cannot be large on a large set.

Proof. Fix $z > 0$ and let $K \subset \{M\mu > z\}$ be a compact set. For any $x \in K$ there exists an open ball $B(x)$ such that $|\mu|(B(x)) > z\,\mathbf{m}(B(x))$. Since $K \subset \cup_x B(x)$, there is a finite subcover, $K \subset \cup_{i=1}^N B(x_i)$. By the covering lemma, there is a disjoint subcollection $B_1, \dots, B_M$ such that

$$\mathbf{m}(K) \le 3^k \sum_{j=1}^M \mathbf{m}(B_j) \le \frac{3^k}{z} \sum_{j=1}^M |\mu|(B_j) \le \frac{3^k}{z}\,\|\mu\|$$

(at the last step we used the disjointness of the B_j's. Lastly, by the regularity of the Lebesgue measure (Theorem 8.2),

$$\mathbf{m}(M\mu > z) = \sup_{K \subset \{M\mu > z\}} \mathbf{m}(K) \le \frac{3^k}{z}\,\|\mu\|$$

□

Definition 13.8. **Weak $\mathbf{L}^1$ space** is the space of measurable functions

$$L^1_W(\mathbb{R}^k) = \left\{ f \colon \mathbb{R}^k \to \mathbb{C} \;\middle|\; z\,\mathbf{m}\{|f| > z\} \text{ is bounded on } z \in (0, \infty) \right\}$$

The value

$$\sup_{0<z<\infty} z\,\mathbf{m}\{|f| > z\}$$

may be called the "weak L^1 norm" of f. (Though it is not a norm by any means.)

Proposition 13.9. *For all $f \in L^1_{\mathbf{m}}(\mathbb{R}^k)$ and $z > 0$ we have $\mathbf{m}(|f| > z) \le z^{-1}\|f\|_1$. Thus $L^1_{\mathbf{m}}(\mathbb{R}^k) \subset L^1_W(\mathbb{R}^k)$.*

Exercise 65. Prove this proposition.

Example 23. There are functions in $L^1_W(\mathbb{R}^k)$ that are not in $L^1_{\mathbf{m}}(\mathbb{R}^k)$. For $k = 1$, such a function is $f = 1/x$. For $k \ge 2$, take $f(x) = 1/\|x\|^k$.

Definition 13.10. For every $f \in L^1_{\mathbf{m}}(\mathbb{R}^k)$ define the **maximal function** $Mf\colon \mathbb{R}^k \to [0,\infty]$ by

$$(Mf)(x) = \sup_{0<r<\infty} \frac{1}{\mathbf{m}(B(x,r))} \int_{B(x,r)} |f| \, d\mathbf{m}.$$

Note that $\mathbf{m}(B(x,r)) = \mathbf{m}(B(0,r))$ does not depend on x.

- We have $Mf = M\mu_f$, if μ_f is defined by $d\mu_f = f\, d\mathbf{m}$.
- According to Theorem 13.7, the 'maximal function' M induces an operator $L^1_{\mathbf{m}}(\mathbb{R}^k) \to L^1_W(\mathbb{R}^k)$. It is bounded in the sense that the weak L^1 norm of Mf is $\leq 3^k \|f\|_1$.

Definition 13.11. If $f \in L^1_{\mathbf{m}}(\mathbb{R}^k)$, then any point $x \in \mathbb{R}^k$ for which

$$\lim_{r\to 0} \frac{1}{\mathbf{m}(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| \, d\mathbf{m}(y) = 0$$

is called a **Lebesgue point** of f.

- Roughly speaking, x is a Lebesgue point if f does not oscillate too much near x, in an average sense.
- This definition depends on the representative of f in the equivalence class. That is, changing f on a null set may affect its Lebesgue points.

Lemma 13.12. *If f is continuous at x, then x is a Lebesgue point of f.*

Lemma 13.13. *If x is a Lebesgue point of f, then*

$$f(x) = \lim_{r\to 0} \frac{1}{\mathbf{m}(B(x,r))} \int_{B(x,r)} f \, d\mathbf{m}$$

(but the converse is not true).

Exercise 66. Let $f \in L^1_{\mathbf{m}}(\mathbb{R}^k)$. Show that $|f(x)| \leq (Mf)(x)$ at every Lebesgue point x of f.

Exercise 67. Let

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \text{ or } x > 1 \\ 1 & \text{if } 0 < x < 1 \end{cases}$$

Is it possible to define $f(0)$ and $f(1)$ such that 0 and 1 become Lebesgue points of f?

Exercise 68. [Bonus] Construct a function $f\colon \mathbb{R} \to \mathbb{R}$ such that $f(0) = 0$ and 0 is a Lebesgue point of f, but for every $\varepsilon > 0$

$$\mathbf{m}\big\{x \in \mathbb{R}\colon |x| < \varepsilon \text{ and } |f(x)| \geq 1\big\} > 0,$$

i.e. f is essentially discontinuous at 0.

Theorem 13.14. *If $f \in L^1(\mathbb{R}^k)$, then almost every point $x \in \mathbb{R}^k$ is a Lebesgue point of f.*

Proof. Define

$$(T_r f)(x) = \frac{1}{\mathbf{m}(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| \, d\mathbf{m}(y)$$

and put

$$(Tf)(x) = \limsup_{r \to 0} (T_r f)(x)$$

We have to prove that $Tf = 0$ a.e. in $\mathbb{R}^k$.

Pick $z > 0$ and $n \geq 1$. Since $C_c(\mathbb{R}^k)$ is dense in $L^1_{\mathbf{m}}(\mathbb{R}^k)$ (Theorem 11.33 extended to $\mathbb{R}^k$), there exists $g \in C_c(\mathbb{R}^k)$ such that $\|f - g\|_1 < 1/n$. Put $h = f - g$.

Since g is continuous, $Tg = 0$. By triangle inequality

$$(T_r h)(x) \leq \left[\frac{1}{\mathbf{m}(B(x,r))} \int_{B(x,r)} |h| \, d\mathbf{m}\right] + |h(x)|$$

Taking the limit $r \to 0$ gives

$$Th \leq Mh + |h|$$

Since $T_r f \leq T_r g + T_r h$, taking the limit $r \to 0$ gives

$$Tf \leq Tg + Th \leq Mh + |h|$$

Therefore

$$\{Tf > 2z\} \subset \{Mh > z\} \cup \{|h| > z\}$$

Now we have

$$\begin{aligned} \mathbf{m}\big(\{Tf > 2z\}\big) &\leq \mathbf{m}(Mh > z) + \mathbf{m}(|h| > z) \\ &\leq \frac{3^k \|h\|_1}{z} + \frac{\|h\|_1}{z} \leq \frac{3^k + 1}{zn} \end{aligned}$$

where we used Theorem 13.7, Proposition 13.9, and in the end – our assumption $\|h\|_1 < 1/n$. Since $\varepsilon > 0$ is arbitrary, we have $\mathbf{m}\big(\{Tf > 2z\}\big) = 0$ for any $z > 0$. This implies $Tf = 0$ a.e. $\square$

Theorem 13.15. *Let μ be a complex measure on $\mathbb{R}^k$ such that $\mu \ll \mathbf{m}$. Then the Radon-Nikodym derivative $f = d\mu/d\mathbf{m}$ is almost everywhere equal to the symmetric derivative of μ, i.e.*

$$\frac{d\mu}{d\mathbf{m}} = D\mu \qquad \mathbf{m}\text{-}a.e.$$

Equivalently, for any Borel set $E \subset \mathbb{R}^k$

$$\mu(E) = \int_E (D\mu)\, d\mathbf{m}$$

Proof. Since $f \in L^1_{\mathbf{m}}(\mathbb{R}^k)$, Theorem 13.14 guarantees that a.e. $x \in \mathbb{R}^k$ is a Lebesgue point of f, which means

$$f(x) = \lim_{r\to 0} \frac{1}{\mathbf{m}(B(x,r))} \int_{B(x,r)} f\, d\mathbf{m} = \lim_{r\to 0} \frac{\mu(B(x,r))}{\mathbf{m}(B(x,r))} = (D\mu)(x)$$

□

Next we explore some other ways of computing symmetric derivatives.

Definition 13.16. We say that Borel sets $E_i \subset \mathbb{R}^k$ **shrink nicely** to a point $x \in \mathbb{R}^k$ if there exists $\alpha > 0$ such that for some sequence of balls $B(x, r_i)$ with $r_i \to 0$ we have

$$E_i \subset B(x, r_i)$$

and

$$\mathbf{m}(E_i) \geq \alpha\, \mathbf{m}(B(x, r_i)) \qquad \forall i \geq 1$$

(Note: it is not required that E_i's contain x.)

Example 24.

(a) Any sequence of intervals $I_i \ni x$ in $\mathbb{R}$ such that $\mathbf{m}(I_i) \to 0$ shrinks to x nicely

(b) Any sequence of balls or cubes in $\mathbb{R}^k$ containing x and whose sizes converge to zero shrinks nicely to x

(c) The sequence of rectangles $[-\frac{1}{i}, \frac{1}{i}] \times [-\frac{1}{i^2}, \frac{1}{i^2}]$ does <u>not</u> shrink nicely to 0.

Theorem 13.17. *If $x \in \mathbb{R}^k$ is a Lebesgue point of $f \in L^1_{\mathbf{m}}(\mathbb{R}^k)$, then for any sequence $\{E_i\}$ that shrinks nicely to x we have*

$$f(x) = \lim_{i\to\infty} \frac{1}{\mathbf{m}(E_i)} \int_{E_i} f \, d\mathbf{m}.$$

Proof. Applying the definition of nicely shrinking sets gives

$$\begin{aligned} 0 &\le \frac{\alpha}{\mathbf{m}(E_i)} \int_{E_i} |f(y) - f(x)| \, d\mathbf{m}(y) \\ &\le \frac{1}{\mathbf{m}(B(x, r_i))} \int_{B(x,r_i)} |f(y) - f(x)| \, d\mathbf{m}(y) \to 0 \end{aligned}$$

as $i \to \infty$. Now

$$\frac{1}{\mathbf{m}(E_i)} \int_{E_i} f \, d\mathbf{m} = \frac{1}{\mathbf{m}(E_i)} \int_{E_i} f(x) \, d\mathbf{m} + \frac{1}{\mathbf{m}(E_i)} \int_{E_i} (f - f(x)) \, d\mathbf{m}$$

The fist integral on the right hand side is simply equal to $f(x)$, and the second one converges to zero as $i \to \infty$.

□

Definition 13.18. Let $E \subset \mathbb{R}^k$ be a measurable set and $x \in \mathbb{R}^k$. Then

$$\lim_{r\to 0} \frac{\mathbf{m}(E \cap B(x, r))}{\mathbf{m}(B(x, r))}$$

is called the **metric density** of E at x (if the limit exists).

- The metric density is a number in the interval $[0, 1]$. It shows how "solid" the set is near the point x.
- If E is open, then its metric density is 1 at every point $x \in E$.
- If E is a closed ball, then its metric density is 1 at every interior point $x \in \operatorname{int} E$, it is $\frac{1}{2}$ at every boundary point $x \in \partial E$, and it is 0 at every outside point $x \in E^c$.

Definition 13.19. Points $x \in \mathbb{R}^k$ where the given set $E \subset \mathbb{R}^k$ has metric density 1 are called **Lebesgue points** of E.

Theorem 13.20. *For every measurable set $E \subset \mathbb{R}^k$ the metric density is 1 at a.e. point $x \in E$ and 0 at a.e. point $x \in E^c$.*

Proof. If $\mathbf{m}(E) < \infty$, then $\chi_E \in L^1$ and we use Theorem 13.14 along with Lemma 13.13. If $\mu(E) = \infty$, we use the sequence of finite-measure subsets $E_N = E \cap B(0, N)$ that 'exhaust' E as $N \to \infty$. □

- Thus, every measurable set E is "solid" at almost every point $x \in E$.

Corollary 13.21.

(a) *If $\varepsilon > 0$, then there is no measurable set $E \subset \mathbb{R}$ such that*

$$\epsilon < \frac{\mathbf{m}(E \cap I)}{\mathbf{m}(I)} < 1 - \varepsilon$$

for every finite nontrivial interval $I \subset \mathbb{R}$.

(b) *If $\varepsilon > 0$ and*

$$\frac{\mathbf{m}(E \cap I)}{\mathbf{m}(I)} > \epsilon$$

for every finite nontrivial interval $I \subset \mathbb{R}$, then $\mathbf{m}(E^c) = 0$.

(c) *If $\varepsilon > 0$ and*

$$\frac{\mathbf{m}(E \cap I)}{\mathbf{m}(I)} < 1 - \varepsilon$$

for every finite nontrivial interval $I \subset \mathbb{R}$, then $\mathbf{m}(E) = 0$.

14 Functions of bounded variation and absolute continuity

Let us recall the main fact in standard calculus:

Fundamental Theorem of Calculus (FTC).

- **(Easy Part)** If $f\colon [a,b] \to \mathbb{R}$ is continuous and
$$F(x) = \int_a^x f(t)\,dt$$
then $F'(x) = f(x)$ for all $x \in (a,b)$.
- **(Hard Part)** If $F\colon [a,b] \to \mathbb{R}$ is continuously differentiable, then
$$F(x) - F(a) = \int_a^x F'(t)\,dt$$
for all $x \in [a,b]$.

Note that in both parts of the FTC the Riemannian integral is used.

Our goal is to extend this theorem to Lebesgue integrable functions.

Theorem 14.1. **FTC in L^1 (the easy part)**

If $f \in L^1_{\mathbf{m}}(\mathbb{R})$ and
$$F(x) = \int_{(-\infty,x]} f\,d\mathbf{m}$$
for $x \in \mathbb{R}$, then $F'(x) = f(x)$ at every Lebesgue point x of f.

Proof. Let $\delta_i \to 0$. Theorem 13.17 with $E_i = [x, x+\delta_i]$ shows that the right-hand derivative of F exists at every Lebesgue point x of f and that it is equal to $f(x)$. If we set $E_i = [x-\delta, x]$ instead, we obtain the same result for the left-hand derivative of F at x. □

Extending the hard part of the FTC to L^1 functions will take a considerable effort. Let us formulate our goals first.

We want to find a class of functions $f\colon [a,b] \to \mathbb{R}$ such that
$$f(x) - f(a) = \int_{[a,x]} f'\,d\mathbf{m} \qquad \forall x \in [a,b] \tag{FTC}$$

Obviously, f must be differentiable, at least almost everywhere. Furthermore, f' must be integrable, i.e., we need $f' \in L^1([a,b])$. But is this enough?

Answer: **No!** See a striking counterexample below.

Cantor function (a.k.a. "Devil's staircase"). Recall page 15 : the middle-third Cantor set $C \subset [0,1]$ was constructed by a recursive removal of open intervals from $[0,1]$ where at each step we remove the middle-third open interval from every closed intervals left at the previous step.

Now we define the Cantor function $f\colon [0,1] \to [0,1]$ as follows. We set $f(0) = 0$ and $f(1) = 1$. On the middle-third open interval $(\frac{1}{3}, \frac{2}{3})$ removed at the first step, $f(x)$ takes constant value $\frac{1}{2}$. On the two middle-third open intervals removed at the second step, it takes constant values $\frac{1}{4}$ and $\frac{3}{4}$, respectively, etc.

Generally, at the n-th step we must remove 2^{n-1} open intervals, each of length $\frac{1}{3\cdot 2^{n-1}}$. The Cantor function takes a constant value on each of those intervals, and those values are set to $\frac{2i-1}{2^n}$ where $i = 1, \dots, 2^{n-1}$ counts the open intervals removed at the nth step, from left to right.

This defines the Cantor function on all the removed intervals, i.e., on $[0,1] \setminus C$. Its graph is shown below. It is monotonically increasing and its values are binary rational numbers $\frac{k}{2^n}$, with all $n \geq 1$ and $1 \leq k \leq 2^n - 1$. These numbers make a dense set in the interval $[0,1]$. Thus we can now define the Cantor function f on the Cantor set C itself by continuity. For every point $x \in C$ we set

$$f(x)\colon = \sup_{y \in [0,1]\setminus C,\ y<x} f(y) = \inf_{y \in [0,1]\setminus C,\ y>x} f(y)$$

These sup and inf coincide due to the denseness of the set $f([0,1] \setminus C)$.

Summary: the Cantor function $f\colon [0,1] \to [0,1]$ is continuous, monotonically increasing (though not strictly). It is constant on every open interval in $[0,1] \setminus C$, so its derivative $f'(x)$ is zero at every point $x \in [0,1] \setminus C$. Since $\mathbf{m}(C) = 0$, we have $f'(x) = 0$ almost everywhere on $[0,1]$.

Exercise 69. Show that $f(C) = [0,1]$, i.e., the Cantor function f maps the Cantor set C (which is a null set!) onto the whole interval $[0,1]$.

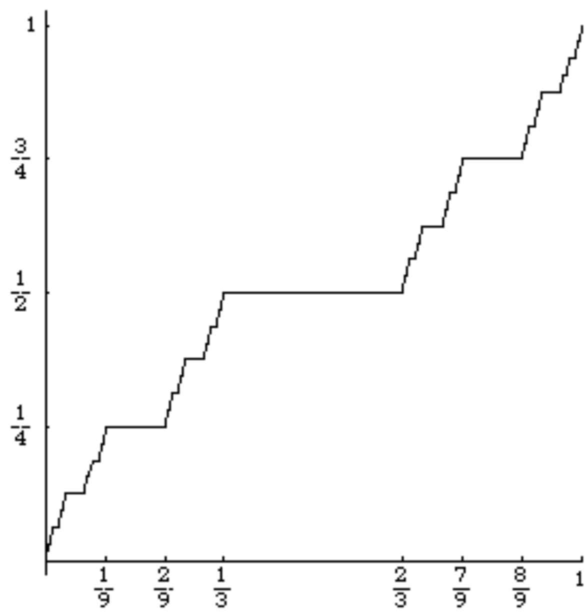


Figure 8: Cantor function ("Devil's staircase").

We now return to our goal: find a class of functions $f\colon [a,b] \to \mathbb{R}$ such that

$$f(x) - f(a) = \int_{[a,x]} f' \, d\mathbf{m} \qquad \forall x \in [a,b] \tag{FTC}$$

The Cantor function f is continuous, differentiable almost everywhere, its derivative f' is integrable, but (FTC) fails because

$$f(1) - f(0) = 1 \neq 0 = \int_{[0,1]} f' \, d\mathbf{m}$$

This example shows that (FTC) may fail on bad (anomalous) functions that experience 'rapid growth on tiny sets'.

Next we define a class of continuous functions that do not have such a pathological behavior.

Definition 14.2. A function $f\colon [a,b] \to \mathbb{C}$ is said to be **absolutely continuous (AC)** on an interval $[a,b]$ if $\forall \varepsilon > 0\ \exists \delta > 0$ such that if $n \in \mathbb{N}$ and

$$(\alpha_1, \beta_1), \dots, (\alpha_n, \beta_n)$$

are disjoint subintervals of $[a,b]$, then

$$\sum_{i=1}^{n} (\beta_i - \alpha_i) < \delta \quad \Longrightarrow \quad \sum_{i=1}^{n} |f(\beta_i) - f(\alpha_i)| < \varepsilon.$$

- Note: the set $\cup_{i=1}^{n} (\alpha_i, \beta_i)$ is small (in the sense of Lebesgue measure), and absolute continuity means that the total change of f on it must be small, too.
- If f is absolutely continuous, then f is continuous, even uniformly continuous.
- But the converse is not true: the Cantor function is continuous (and uniformly continuous), but not absolutely continuous. This will be shown later; see a discussion following Theorem 14.5.

Theorem 14.3. *Suppose that $f\colon [a,b] \to \mathbb{C}$ has the following properties:*

(a) *$f(x)$ is differentiable at almost all $x \in [a,b]$*

(b) *$f' \in L^1([a,b])$*

(c) *(FTC) holds*

Then f must be absolutely continuous.

Proof. Let $\mu = \mu_{f'}$ be the complex measure defined by $d\mu = f'\, d\mathbf{m}$. Then $|\mu|$ is a finite positive measure and $|\mu| \ll \mathbf{m}$. By Theorem 12.20, $\forall \varepsilon > 0\ \exists \delta > 0$: $\mathbf{m}(E) < \delta \Longrightarrow |\mu|(E) < \varepsilon$. Thus

$$\sum_{i=1}^{n} |f(\beta_i) - f(\alpha_i)| = \sum_{i=1}^{n} \left| \int_{(\alpha_i, \beta_i)} f'\, d\mathbf{m} \right| \leq \int_{\cup_{i=1}^{n} (\alpha_i, \beta_i)} |f'|\, d\mathbf{m} = |\mu| \Big(\cup_{i=1}^{n} (\alpha_i, \beta_i) \Big) < \varepsilon.$$

□

Corollary 14.4. *The function $F(x)$ in Theorem 14.1 is absolutely continuous.*

Our next big goal is to show that the converse is also true, i.e., any absolutely continuous function $f(x)$ has properties (a)–(c) of Theorem 14.3.

Theorem 14.5. *Let $f\colon [a,b] \to \mathbb{R}$ be a continuous and monotonically increasing function (not necessarily strictly). Then the following properties are equivalent:*

(a) *f is absolutely continuous*

(b) *f maps sets of measure zero into sets of measure zero;*

(c) *f is differentiable a.e. on $[a,b]$, $f' \in L^1_{\mathbf{m}}([a,b])$ and* (FTC) *holds.*

Proof. We will show that (a)⇒(b)⇒(c).

$\boxed{\textbf{(a)}\Rightarrow\textbf{(b)}}$ Let $E \subset [a,b]$ be a null set, i.e., $\mathbf{m}(E) = 0$. We have to show that $\mathbf{m}\big(f(E)\big) = 0$. Without loss of generality, assume that $a, b \notin E$.

Choose $\varepsilon > 0$ and let $\delta > 0$ be as in Definition 14.2. Due to the regularity of the Lebesgue measure (Theorem 8.1) there is an open cover $V \supset E$ such that $\mathbf{m}(V) < \delta$. The open set V is a finite or countable union of disjoint intervals (α_i, β_i). Then $\sum(\beta_i - \alpha_i) < \delta$ and our choice of δ ensures that

$$\sum_i \big(f(\beta_i) - f(\alpha_i)\big) \le \varepsilon$$

[Definition 14.2 was stated in terms of finite sums; thus the above bound holds for every partial sum of the (possibly) infinite series; taking the limit gives the above bound for the whole series.]

Since $E \subset V$, we have $f(E) \subset f(V) \subset \cup_i [f(\alpha_i), f(\beta_i)]$. Thus

$$\mu^*\big(f(E)\big) \le \sum_i \big(f(\beta_i) - f(\alpha_i)\big) \le \varepsilon$$

Since $\varepsilon > 0$ is arbitrary, we have $\mu^*\big(f(E)\big) = 0$, and due to the completeness of the Lebesgue measure $\mathbf{m}\big(f(E)\big) = 0$ (in particular, $f(E)$ is measurable).

$\boxed{\textbf{(b)}\Rightarrow\textbf{(c)}}$ We want to define a finite positive measure μ on $[a,b]$ by

$$\mu(E) = \mathbf{m}\big(f(E)\big)$$

for any Lebesgue measurable set $E \subset [a,b]$. First of all, we need to make sure that $f(E)$ is a Lebesgue measurable set. Then we need to verify to check that μ is a finite positive measure. This will be done below, now let us finish the proof of (c).

Due to (b), $\mathbf{m}(E) = 0$ implies $\mu(E) = \mathbf{m}\big(f(E)\big) = 0$. Hence $\mu \ll \mathbf{m}$ and by the Radon-Nikodym theorem $d\mu = h\, d\mathbf{m}$ for some function $h \in L^1_{\mathbf{m}}([a,b])$. Thus

$$f(x) - f(a) = \mathbf{m}\big(f([a,x])\big) = \mu([a,x]) = \int_{[a,x]} h\, d\mathbf{m}$$

Due to Theorem 14.1 we have $f'(x) = h(x)$ for every Lebesgue point x of h, i.e., almost everywhere on $[a,b]$. So h can be replaced with f' in the above formula.

$\boxed{\textbf{(c)}\Rightarrow\textbf{(a)}}$ Follows immediately from Corollary 14.4.

It remains to verify that the formula

$$\mu(E) = \mathbf{m}\big(f(E)\big)$$

indeed defines a finite positive measure on $I = [a,b]$. If it is a measure, then it is finite, because $\mu(I) = \mathbf{m}(J) = f(b) - f(a) < \infty$, where $J = f(I) = [f(a), f(b)]$.

$\boxed{\textbf{Case 1}}$ Let the function f be *strictly* monotonically increasing. Then $f\colon I \to J$ is a bijection, so the inverse function f^{-1} is well defined. It is clearly continuous and monotonically increasing. Thus images of Borel sets under f are preimages of Borel sets under f^{-1}, hence they are Borel sets. For any Lebesgue measurable set $E \subset I$ we have $E = E_1 \cup E_0$ where E_1 is a Borel set and E_0 is a null set (Corollary 3.24). Thus $f(E) = f(E_1) \cup f(E_0)$ is the union of a Borel set $f(E_1)$ and a null set $f(E_0)$, as $\mathbf{m}\big(f(E_0)\big) = 0$ due to assumption (b). Thus $f(E)$ is Lebesgue measurable.

Next we verify that μ is a measure. If $E = \uplus_n E_n$ is a countable disjoint union of Lebesgue measurable subsets of I, then $f(E) = \uplus_n f(E_n)$, and therefore

$$\mu(E) = \mathbf{m}\big(f(E)\big) = \mathbf{m}\big(\uplus_n f(E_n)\big) = \sum_n \mathbf{m}\big(f(E_n)\big) = \sum_n \mu(E_n)$$

This completes Case 1. In the future we will only need to use Theorem 14.5 for strictly increasing functions. But for the sake of completeness we outline the proof for non-strictly increasing functions, too.

$\boxed{\textbf{Case 2:}}$ Let the function f be *non-strictly* increasing. Then it still maps $I = [a,b]$ onto $J = [f(a), f(b)]$, but it is no longer one-to one. If $f(x') = f(x'')$ for some $x' < x''$, then due to the monotonicity, $f(x) = f(x')$ for all $x \in [x', x'']$. Thus for each $y \in J$ the preimage $f^{-1}(y)$ is either a single point or a closed interval $I_y = [\alpha, \beta] \subset I$. Clearly, for $y \neq y'$ we have $I_y \cap I_{y'} = \emptyset$, so those intervals are disjoint. Thus there are at most countably many of them. Let $N = \{y \in J \colon \mathbf{m}(I_y) > 0\}$ denote the countable set of points whose preimages are nontrivial intervals.

Now we show that $f(E) \subset J$ is a Borel set for every Borel set $E \subset I$. Let $\mathfrak{G} = \{E \subset I \colon f(E) \text{ is Borel}\}$. We can verify that $\mathfrak{G}$ is a σ-algebra by Lemma 2.4. For its condition (i), note that $f(E^c) = [J \setminus f(E)] \cup N'$ for some $N' \subset N$. Since $f(E)$ is Borel and N' is countable, $f(E^c)$ is also Borel. For (ii), note that $f(\cup_n E_n) = \cup_n f(E_n)$, so if each $f(E_n)$ is Borel, then $f(\cup_n E_n)$ is Borel, too. Next note that for any subinterval $I' \subset I$ the set $f(I') \subset J$ is a subinterval, hence Borel. So $\mathfrak{G}$ is a σ-algebra containing all subintervals, hence it contains all Borel sets.

Next, $f(E) \subset J$ is Lebesgue measurable for any Lebesgue measurable set $E \subset I$ by exactly the same argument as in Case 1.

Lastly, we verify that μ is a measure. If $E = \uplus_n E_n$ is a countable disjoint union of Lebesgue measurable subsets of I, then their images $f(E_n)$ can only intersect at some points $y \in N$, and those countably many points cannot affect the Lebesgue measures of the sets $f(E_n)$, therefore

$$\mu(E) = \mathbf{m}\big(f(E)\big) = \mathbf{m}\big(\cup_n f(E_n)\big) = \sum_n \mathbf{m}\big(f(E_n)\big) = \sum_n \mu(E_n)$$

□

Images of measurable sets under continuous monotonic functions. In the previous proof, we carefully verified that $f(E)$ was a Lebesgue measurable set for any Lebesgue measurable $E \subset I$. One may think that this fact is a triviality. How can a nice continuous function map good (measurable) sets into bad (non-measurable) sets?

This is not a triviality. Some continuous monotonically increasing functions actually map measurable sets into non-measurable sets. Let f be the Cantor function ("Devil's staircase"). Denote

$$N = \left\{ \tfrac{k}{2^n},\ n \geq 1,\ 0 \leq k \leq 2^n \right\}$$

the set of binary rational numbers – these are the values of f on the open intervals whose union is $[0,1] \setminus C$. Let $A \subset [0,1]$ be any non-measurable set. Then the set $A_0 = A \setminus N$ is also non-measurable. Now $B = f^{-1}(A_0) \subset C$. Since the Cantor set C is a null set and B is its subset, B is Lebesgue measurable. So we have $f(B) = A_0$, where B is measurable and A_0 is not.

In our proof that $f(E)$ was a Lebesgue measurable set for any Lebesgue measurable $E \subset I$, the crucial element was assumption (b) saying that f maps null sets into null sets. This assumption, of course, rules out the Cantor function.

As a side result of our discussion: Cantor function is <u>not absolutely continuous</u>.

- Possible exercise for the future: Let $f\colon [0,1] \to [0,1]$ be the Cantor function. Then $h(x) = f(x)+x$ is a continuous *strictly* monotonically increasing function $[0,1] \to [0,2]$. Show that
  (a) $h(C)$ has measure one (here C is the Cantor set)
  (b) there exist measurable sets $A \subset [0,1]$ such that $h(A)$ is not measurable

Next we extend our analysis to non-monotonic functions. This brings us to functions of bounded variation.

Definition 14.6. The **total variation function** V_a^x of a function $f\colon [a,b] \to \mathbb{C}$ is defined by

$$V_a^x = \sup \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|$$

where the supremum is taken over all n and all ordered sequences

$$a = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = x.$$

If $V_a^b < \infty$, then f is said to be of **bounded variation** on $[a,b]$, denoted by $f \in \mathrm{BV}[a,b]$, and the value of V_a^b is called the **total variation** of f over $[a,b]$.

- V_a^x is non-decreasing in x.
- If f is monotonic, then $V_a^x = |f(x) - f(a)|$.
- We have the additivity: $V_a^b = V_a^c + V_c^b$ for every $a < c < b$.
- For the Dirichlet function $f = \chi_{\mathbb{Q}}$ we have $V_a^b = \infty$ for all $a < b$.

Exercise 70. Show that

(a) If $f \in C^1([a,b])$, then $V_a^b \leq \int_a^b |f'(x)|\,dx$

(b) If $f \in C([a,b])$ is continuous on $[a,b]$, differentiable on (a,b) and $|f'(x)|$ is bounded on (a,b), then f is of bounded variation on $[a,b]$ (Hint: use the Mean Value Theorem)

Exercise 71. Let

$$f(x) = \begin{cases} x \cos \frac{\pi}{x} & \text{if } 0 < x \leq 1 \\ 0 & \text{if } x = 0 \end{cases}$$

Show that $V_0^1 = \infty$.

Exercise 72. Let

$$f(x) = \begin{cases} x^2 \cos \frac{\pi}{x} & \text{if } 0 < x \leq 1 \\ 0 & \text{if } x = 0 \end{cases}$$

Show that $V_0^1 < \infty$, i.e., f is of bounded variation on $[0,1]$.

- A function $f\colon [a,b] \to \mathbb{C}$ is of bounded variation if and only if `Re` f and `Im` f are of bounded variation.

Lemma 14.7. *Suppose $f(x)\colon [a,b] \to \mathbb{R}$ is of bounded variation on $[a,b]$. Then $V + f$ and $V - f$ are non-decreasing.*

Proof. For $x < y$ we have

$$V_a^y - V_a^x = V_x^y \geq f(x) - f(y) \quad \Rightarrow \quad V_a^x + f(x) \leq V_a^y + f(y)$$

and

$$V_a^y - V_a^x = V_x^y \geq f(y) - f(x) \quad \Rightarrow \quad V_a^x - f(x) \leq V_a^y - f(y)$$

□

Corollary 14.8. *Let $f\colon [a,b] \to \mathbb{R}$ be of bounded variation on $[a,b]$. Then there are monotonically increasing functions $u, v\colon [a,b] \to \mathbb{R}$ such that $f = u - v$. Moreover, u and v may be chosen strictly monotonically increasing.*

Proof. Set $u = \frac{1}{2}(V + f)$ and $v = \frac{1}{2}(V - f)$. These functions are monotonically increasing and $f = u - v$.

But u and v may not be strictly monotonically increasing, though. Then we can replace them with, say, $u + x$ and $v + x$, which are strictly monotonically increasing and satisfy the same relation $f = u - v$. □

We now return to absolutely continuous functions.

EXERCISE 73. Let f, g be absolutely continuous on $[a, b]$. Show that

(a) $f \pm g$ are absolutely continuous on $[a, b]$

(b) fg is absolutely continuous on $[a, b]$

(c) if $f(x) \neq 0$ for all $x \in [a, b]$, then $1/f$ is absolutely continuous on $[a, b]$

Lemma 14.9. *Suppose $f(x)$ is absolutely continuous on $[a, b]$. Then*

(a) *f is of bounded variation on $[a, b]$;*

(b) *V, $V + f$, and $V - f$ are absolutely continuous on $[a, b]$.*

Proof.
$\boxed{\textbf{(a)}}$ Let $\varepsilon = 1$ and $\delta > 0$ as in Definition 14.2. For any $[c, d] \subset [a, b]$ with $d - c < \delta$ we have $V_c^d \leq 1$. Then we can divide $[a, b]$ into $N \leq \frac{b-a}{\delta} + 1$ subintervals of length $< \delta$ and get $V_a^b \leq N$.

$\boxed{\textbf{(b)}}$ Due to Exercise 73(a) it is enough to show that V is AC, i.e., for any $\varepsilon > 0$ there is a $\delta > 0$

$$\sum_{i=1}^{n} (\beta_i - \alpha_i) < \delta \quad \Longrightarrow \quad \sum_{i=1}^{n} \left| V_a^{\beta_i} - V_a^{\alpha_i} \right| < \varepsilon$$

in the notation of Definition 14.2. But

$$\left| V_a^{\beta_i} - V_a^{\alpha_i} \right| = V_a^{\beta_i} - V_a^{\alpha_i} = V_{\alpha_i}^{\beta_i} = \sup \sum_j |f(t_{i,j+1}) - f(t_{i,j})|$$

where

$$\alpha_i = t_{i,0} < t_{i,1} < \cdots < t_{i,n_i - 1} < t_{i,n_i} = \beta_i$$

is a partition of (α_i, β_i). Note that

$$\sum_i \sum_j (t_{i,j+1} - t_{i,j}) = \sum_i (\beta_i - \alpha_i) < \delta$$

thus by Definition 14.2 applied to the given AC function f we have

$$\sum_i \sum_j |f(t_{i,j+1}) - f(t_{i,j})| < \varepsilon$$

Taking the supremum gives a bound

$$\sum_{i=1}^{n} \left| V_a^{\beta_i} - V_a^{\alpha_i} \right| \leq \varepsilon$$

which shows that V is AC. $\square$

EXERCISE 74. A function $f\colon [a,b] \to \mathbb{C}$ is said to be Lipschitz continuous if $\exists L > 0$ such that $|f(x) - f(y)| \leq L|x - y|$ for all $x, y \in [a,b]$. Prove that if f is Lipschitz continuous, then f is absolutely continuous and $|f'| \leq L$ a.e. Conversely, if f is absolutely continuous and $|f'| \leq L$ a.e., then f is Lipschitz continuous (with that constant L).

Finally we are in a position to prove our main result:

Theorem 14.10. FTC in L^1 (the hard part)

A function $f\colon [a,b] \to \mathbb{C}$ is absolutely continuous if and only if f is differentiable for a.e. $x \in [a,b]$, $f' \in L^1([a,b])$, and (FTC) *holds.*

Proof. The "if" part was already proved in Theorem 14.3.

So let f be AC on $[a,b]$. It is enough to prove the claim for $\texttt{Re}\, f$ and $\texttt{Im}\, f$ separately, so we can assume that $f\colon [a,b] \to \mathbb{R}$ is real-valued.

According to Lemma 14.7 and Corollary 14.8

$$f = u - v, \qquad u = \frac{1}{2}(V + f + x), \quad v = \frac{1}{2}(V - f + x)$$

is the difference between two strictly monotonically increasing functions, each of which is AC due to Lemma 14.9 (note that $f(x) = x$ is AC due to Exercise 74). Now we apply Theorem 14.5 to u and v and get $u', v' \in L^1_{\mathbf{m}}([a,b])$ and

$$u(x) - u(a) = \int_{[a,x]} u' \, d\mathbf{m} \qquad v(x) - v(a) = \int_{[a,x]} v' \, d\mathbf{m} \tag{14.1}$$

Therefore $f' = u' - v' \in L^1_{\mathbf{m}}([a,b])$ and subtracting the second equation in (14.1) from the first one gives (FTC). $\square$

EXERCISE 75. Let

$$f(x) = \begin{cases} x^2 \cos \frac{\pi}{x^2} & \text{if } 0 < x \leq 1 \\ 0 & \text{if } x = 0 \end{cases}$$

(a) Show that $f(x)$ is differentiable at every point $x \in [0,1]$ (including $x = 0$)

(b) Verify that $f'(x)$ *does not* belong to $L^1([0,1])$

Conclude that f is not absolutely continuous.

The following theorem arrives at the hard part of the FTC from a different set of hypotheses:

Theorem 14.11. **FTC in L^1 (the hard part; alternative version)**

Suppose $f\colon [a,b] \to \mathbb{C}$ be differentiable at every point $x \in [a,b]$ and $f' \in L^1([a,b])$. Then (FTC) *holds.*

- Note: the differentiability at *every* point $x \in [a,b]$ is essential.
- This theorem can be given without proof (its proof is in Rudin's book).

EXERCISE 76. In Exercise 72 we showed that the following function had bounded variation:

$$f(x) = \begin{cases} x^2 \cos \frac{\pi}{x} & \text{if } \ 0 < x \le 1 \\ 0 & \text{if } \ x = 0 \end{cases}$$

(a) Show that $f(x)$ is differentiable at every point $x \in [0,1]$ (including $x = 0$)

(b) Verify that $f'(x)$ *does* belong to $L^1([0,1])$

Conclude that f is absolutely continuous.

EXERCISE 77. Let $f\colon [a,b] \to \mathbb{R}$ be absolutely continuous. Prove that $V_a^x \le \int_{[a,x]} |f'| d\mathbf{m}$. For an extra credit: is the equality always true?

EXERCISE 78. [Bonus] Let $f\colon [0,1] \to \mathbb{R}$ be absolutely continuous on $[\delta, 1]$ for each $\delta > 0$, continuous at $x = 0$, and of bounded variation on $[0,1]$. Prove that f is absolutely continuous on $[0,1]$. (Note: you can use the "extra credit" part of the previous exercise only if you properly finish it first.)

Lastly we prove two important facts related to AC and BV functions.

Theorem 14.12. **Lusin N property**

Let $f\colon [a,b] \to \mathbb{R}$ be absolutely continuous.
Then it maps sets of measure zero into sets of measure zero. That is,

$$\forall N \subset [a,b]\colon \quad \mathbf{m}(N) = 0 \quad \Rightarrow \quad \mathbf{m}\big(f(N)\big) = 0$$

- The converse is not true, even if f is continuous and differentiable (Exercise 75).
- Theorem 14.12 does not extend to functions of bounded variation (Exercise 69).

Proof. Let $N \subset [a,b]$ be a null set. Due to the regularity of the Lebesgue measure (Theorem 8.1), $\forall \delta > 0$ there is an open cover $V \supset N$ such that $\mathbf{m}(V) < \delta$. The open set V is a finite or countable union of disjoint intervals (α_i, β_i). On each *closed* interval $[\alpha_i, \beta_i]$ the continuous function f takes a minimum at some $x_i \in [\alpha_i, \beta_i]$ and a maximum at some $y_i \in [\alpha_i, \beta_i]$. Then

$$f\big([\alpha_i, \beta_i]\big) = [f(x_i), f(y_i)]$$

and

$$f(N) \subset f(V) \subset \cup_i f\big([\alpha_i, \beta_i]\big)$$

Now we have

$$\mathbf{m}\Big(f\big([\alpha_i, \beta_i]\big)\Big) = f(y_i) - f(x_i) = \begin{cases} \displaystyle\int_{[x_i, y_i]} f'\, d\mathbf{m} & \text{if } \ x_i \le y_i \\ \displaystyle -\int_{[y_i, x_i]} f'\, d\mathbf{m} & \text{if } \ y_i < x_i \end{cases}$$

Let I_i denote the interval between x_i and y_i (in whichever order they come). Since $f(y_i) - f(x_i) \ge 0$, we can simply write

$$f(y_i) - f(x_i) = \left| \int_{I_i} f'\, d\mathbf{m} \right| \le \int_{I_i} |f'|\, d\mathbf{m} \le \int_{(\alpha_i, \beta_i)} |f'|\, d\mathbf{m}$$

Summing over i gives

$$\mathbf{m}\big(f(N)\big) \le \mathbf{m}\big(f(V)\big) \le \sum_i \mathbf{m}\Big(f\big([\alpha_i, \beta_i]\big)\Big) \le \int_V |f'|\, d\mathbf{m}$$

Since $|f'| \in L^1_{\mathbf{m}}([a,b])$, we have by Theorem 12.20 that for any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that $\int_E |f'|\, d\mathbf{m} < \varepsilon$ whenever $\mathbf{m}(E) < \delta_\varepsilon$. Thus for any $\varepsilon > 0$ we can choose an open cover $V \supset N$ of measure $\mathbf{m}(V) < \delta_\varepsilon$, and that gives us $\mathbf{m}\big(f(N)\big) < \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we have $\mathbf{m}\big(f(N)\big) = 0$. $\square$

Theorem 14.13. <u>**Lebesgue monotone differentiation**</u>

Let $f\colon [a,b] \to \mathbb{R}$ be monotonically increasing.
Then it is differentiable almost everywhere.
Furthermore, $f(b) - f(a) \geq \int_{[a,b]} f' \, d\mathbf{m}$.

- A strict inequality is possible (example: the Cantor function).

Proof. The proof is long but its steps are visually clear.

Dani derivatives. Let us define the following limits:

$$\begin{aligned}
D^+f(x) &= \limsup_{h\to 0+} \frac{f(x+h)-f(x)}{h} && \text{(`upper right')}\\
D_+f(x) &= \liminf_{h\to 0+} \frac{f(x+h)-f(x)}{h} && \text{(`lower right')}\\
D^-f(x) &= \limsup_{h\to 0-} \frac{f(x+h)-f(x)}{h} && \text{(`upper left')}\\
D_-f(x) &= \liminf_{h\to 0-} \frac{f(x+h)-f(x)}{h} && \text{(`lower left')}
\end{aligned}$$

These are called *Dani derivatives*. They always exist, for any function, and take values in $[-\infty, \infty]$. Clearly, $f'(x)$ exists if and only if

$$-\infty < D^+f(x) = D_+f(x) = D^-f(x) = D_-f(x) < \infty$$

Obviously,

$$D^+f(x) \geq D_+f(x), \qquad D^-f(x) \geq D_-f(x). \tag{14.2}$$

Actually, since our function f is increasing, then all of the Dani derivatives must be non-negative, i.e., restricted to $[0, \infty]$.

Main claim. It will be enough to prove that

$$D_-f(x) \geq D^+f(x), \qquad D^+f(x) < \infty \tag{14.3}$$

for almost every point $x \in [a,b]$.

Main claim ⇒ Theorem. If we prove the first inequality in (14.3) for any monotonically increasing function $f(x)$, then it will apply to the function $g(x) = -f(-x)$, too. And it will give

$$D_{-}g(-x) \geq D^{+}g(-x)$$

for almost every x. It is easy to note that $D_{-}g(-x) = D_{+}f(x)$ and $D^{+}g(-x) = D^{-}f(x)$; see the figure. Hence we get

$$D_{+}f(x) \geq D^{-}f(x)$$

for almost every x. Combining this with (14.3) and (14.2) gives

$$D_{-}f(x) \geq D^{+}f(x) \geq D_{+}f(x) \geq D^{-}f(x) \geq D_{-}f(x)$$

which is only possible if all the Dani derivatives are equal:

$$D_{-}f(x) = D^{+}f(x) = D_{+}f(x) = D^{-}f(x)$$

The second bound in (14.3) ensures that they are all finite. So Theorem will be proved once we establish (14.3). After a brief digression, the proof will be continued. □

Definition 14.14. Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function.

(a) A point $x \in [a,b]$ is said to be **invisible from the right** if there exists $x_1 > x$ such that $f(x) < f(x_1)$.

(b) A point $x \in [a,b]$ is said to be **invisible from the left** if there exists $x_1 < x$ such that $f(x) < f(x_1)$.

Lemma 14.15.

(a) *The set of all points invisible from the* <u>*right*</u> *is open, i.e., it is a union of open intervals* (α_n, β_n)*, and for each interval we have* $f(\alpha_n) \leq f(\beta_n)$.

(b) *The set of all points invisible from the* <u>*left*</u> *is open, i.e., it is a union of open intervals* (α_n, β_n)*, and for each interval we have* $f(\alpha_n) \geq f(\beta_n)$.

Proof. The openness follows from the continuity of f. The relation $f(\alpha_n)$ between and $f(\beta_n)$ is easy to verify "by way of contradiction". □

Proof. of Theorem 14.13 (continued). Our argument will be clearer if we assume first that f is a continuous function. In the end we extend it to discontinuous functions. We begin with the proof of the first inequality of (14.3).

First inequality in (14.3): Auxiliary claim. It is enough to show that for any positive real numbers $u < v$ we have

$$\mathbf{m}(N_{uv}) = 0, \qquad N_{uv} = \{x \in [a,b] \colon D_- f(x) < u < v < D^+ f(x)\} \tag{14.4}$$

as then taking the union of $\cup_{u<v} N_{uv}$ over all positive *rationals* $u < v$ will give us

$$\mathbf{m}\big(\{x \in [a,b] \colon D_- f(x) < D^+ f(x)\}\big) = \mathbf{m}\big(\cup_{0<u<v,\ u,v\in\mathbb{Q}} N_{uv}\big) = 0$$

which is exactly the first inequality in (14.3).

The way (14.4) will be proved is that we will show that for any open interval $(\alpha, \beta) \subset [a,b]$ we have

$$\mathbf{m}\big(N_{uv} \cap (\alpha,\beta)\big) \leq \tfrac{u}{v}\,(\beta - \alpha) \tag{14.5}$$

and then due to Corollary 13.21 (c) we will conclude that $\mathbf{m}(N_{uv}) = 0$.

Proof of auxiliary claim (14.5).

If $D_- f(x) < u$ for some $x \in (\alpha, \beta)$, then there is a point $x_1 < x$ such that

$$\frac{f(x_1) - f(x)}{x_1 - x} < u \quad \Rightarrow \quad f(x_1) - ux_1 > f(x) - ux$$

Therefore x is invisible from the left for the function $g(x) = f(x) - ux$. Due to Lemma 14.15 (b), the set of such points inside (α, β) is open and consists of open intervals (α_n, β_n) such that for each interval we have

$$f(\alpha_n) - u\alpha_n = g(\alpha_n) \geq g(\beta_n) = f(\beta_n) - u\beta_n$$

or equivalently

$$f(\beta_n) - f(\alpha_n) \leq u(\beta_n - \alpha_n)$$

Similarly, if $D^+ f(x) > v$ for some $x \in (\alpha_n, \beta_n)$, then there is a point $x_1 > x$ such that

$$\frac{f(x_1) - f(x)}{x_1 - x} > v \quad \Rightarrow \quad f(x_1) - vx_1 > f(x) - vx$$

Therefore x is invisible from the right for the function $g(x) = f(x) - vx$. Due to Lemma 14.15 (a), the set of such points inside (α_n, β_n) is open and consists of open intervals $(\alpha_{nk}, \beta_{nk})$ such that for each interval we have

$$f(\beta_{nk}) - v\alpha_{nk} = g(\alpha_{nk}) \leq g(\beta_{nk}) = f(\beta_{nk}) - v\beta_{nk}$$

or equivalently

$$f(\beta_{nk}) - f(\alpha_{nk}) \geq v(\beta_{nk} - \alpha_{nk})$$

Clearly the set $N_{uv} \cap (\alpha, \beta)$ is covered by the intervals $(\alpha_{nk}, \beta_{nk})$, and it follows from our previous inequalities that

$$\begin{aligned}
\sum_{n,k} (\beta_{nk} - \alpha_{nk}) &\le \tfrac{1}{v} \sum_{n,k} [f(\beta_{nk}) - f(\alpha_{nk})] \\
&\le \tfrac{1}{v} \sum_{n} [f(\beta_n) - f(\alpha_n)] \\
&\le \tfrac{u}{v} \sum_{n} (\beta_n - \alpha_n) \\
&\le \tfrac{u}{v} (\beta - \alpha)
\end{aligned}$$

This implies (14.5), and hence $\mathbf{m}(N_{uv}) = 0$ and the first inequality of (14.3).

Second inequality in (14.3). Let $N = \{x \in [a, b] \colon D^+ f(x) = \infty\}$. If $x \in N$, then for any $C > 0$ there is a point $x_1 > x$ such that

$$\frac{f(x_1) - f(x)}{x_1 - x} > C \quad \Rightarrow \quad f(x_1) - Cx_1 > f(x) - Cx$$

Thus x is invisible from the right for the function $g(x) = f(x) - Cx$. Due to Lemma 14.15 (a), the set of such points is open and consists of open intervals (α_n, β_n) such that for each interval we have

$$f(\beta_n) - C\alpha_n = g(\alpha_n) \le g(\beta_n) = f(\beta_n) - C\beta_n$$

or equivalently

$$f(\beta_n) - f(\alpha_n) \ge C(\beta_n - \alpha_n)$$

Therefore, since $N \subset \cup_n (\alpha_n, \beta_n)$, we obtain

$$\mathbf{m}(N) \le \sum_n (\beta_n - \alpha_n) \le \tfrac{1}{C} \sum_n [f(\beta_n) - f(\alpha_n)] \le \tfrac{1}{C} [f(b) - f(a)]$$

Since this is true for any $C > 0$, we have $\mathbf{m}(N) = 0$.

Discontinuous functions. Things get a little messy. First we note that discontinuous monotonically increasing functions are not too bad. At least for any point x there exist a left limit $f(x-) = \lim_{y \to x-} f(y)$ and a right limit $f(x+) = \lim_{y \to x+} f(y)$. We also have $f(x-) \le f(x) \le f(x+)$. In the above argument we use $g(x) = f(x) - cx$ for various constants $c > 0$; this function also has a left limit and a right limit values at every point. It also satisfies $g(x-) \le g(x) \le g(x+)$.

Now let f be either continuous or at least have a left limit and a right limit values at every point x such that $f(x-) \le f(x) \le f(x+)$. Then we redefine invisible points as follows: A point $x \in [a, b]$ is *invisible from the right* if there exists $x_1 > x$ such that $f(x+) < f(x_1)$. A point $x \in [a, b]$ is *invisible from the left* if there exists $x_1 < x$ such that $f(x+) < f(x_1)$.

Now in Lemma 14.15 (a) we replace $f(\alpha_n) \le f(\beta_n)$ with $f(\alpha_n+) \le f(\beta_n+)$. Similarly, in Lemma 14.15 (b) we replace $f(\alpha_n) \ge f(\beta_n)$ with $f(\alpha_n+) \ge f(\beta_n-)$. Then Lemma 14.15 extends to discontinuous functions of the above type.

The arguments in the previous two subsections extend to discontinuous functions as well, with one minor adjustment: each point of the sets N_{uv} and N is either *inside* of one of the respective open intervals (α_n, β_n) or *on its boundary* (i.e., it may coincide with either α_n or β_n). All the measure estimates remain valid, though.

Integral inequality. For each $n \geq 1$ define the following approximation to $f'(x)$:

$$g_n(x) = \frac{f(x+\frac{1}{n}) - f(x)}{\frac{1}{n}} = n\big[f(x+\tfrac{1}{n}) - f(x)\big]$$

(if it happens that $x + \frac{1}{n} > b$, we just set $f(x+\frac{1}{n})\colon = f(b)$).

Now we have $g_n \to f'$, as $n \to \infty$, almost everywhere on $[a,b]$. Since f is increasing, we have $g_n \geq 0$ and $f' \geq 0$. Due to Fatou's lemma

$$\begin{aligned} \int_{[a,b]} f'\, d\mathbf{m} &\leq \liminf \int_{[a,b]} g_n\, d\mathbf{m} \\ &= \liminf n \int_{[a,b]} \big[f(x+\tfrac{1}{n}) - f(x)\big]\, d\mathbf{m} \\ &= \liminf n \int_{[b,b+\frac{1}{n}]} f\, d\mathbf{m} - n \int_{[a,a+\frac{1}{n}]} f\, d\mathbf{m} \end{aligned}$$

The first term is equal to $f(b)$ and the second term is $\geq f(a)$, hence

$$\int_{[a,b]} f'\, d\mathbf{m} \leq f(b) - f(a)$$

This completes the long proof of Theorem 14.13... □

Corollary 14.16. *Every function $f\colon [a,b] \to \mathbb{C}$ of bounded variation is differentiable almost everywhere.*

Proof. To differentiate f, we differentiate $\texttt{Re}\, f$ and $\texttt{Im}\, f$ separately, thus it is enough to prove this corollary for real-valued functions. Now we just combine Corollary 14.8 with Theorem 14.13. □

15 Differentiable transformations

Linear transformations. From linear algebra, we know that any linear transformation $T\colon \mathbb{R}^k \to \mathbb{R}^k$ is defined by a $k \times k$ matrix, which we will denote by $\mathbf{A} = (a_{ij})$. It takes a point $(x_1, \dots, x_k) \in \mathbb{R}^k$ to a point $(y_1, \dots, y_k) \in \mathbb{R}^k$ by the rule $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{x} = (x_1, \dots, x_k)^T$ and $\mathbf{y} = (y_1, \dots, y_k)^T$ are column-vectors. Note that the origin $(0, \dots, 0)$ is always mapped to itself.

More generally, a linear transformation *with shift* is defined by $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{a}$, where $\mathbf{a}$ is a fixed vector. Now the origin is shifted to $\mathbf{a}$. If $\mathbf{a} = (0, \dots, 0)^T$, then we get the transformation without shift.

Proposition 15.1. *Let $T\colon \mathbb{R}^k \to \mathbb{R}^k$ be a linear transformation, with or without shift, defined by $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{a}$. Then for any measurable set $E \subset \mathbb{R}^k$ we have*

$$\mathbf{m}\big(T(E)\big) = c\,\mathbf{m}(E)$$

where the scaling factor is $c = |\det \mathbf{A}|$.

Proof. From linear algebra, we know that the map T is a bijection if and only if the matrix $\mathbf{A}$ is not singular, i.e., $\det \mathbf{A} \neq 0$. If it is singular, then T maps $\mathbb{R}^k$ into a lower-dimensional subspace, hence $\mathbf{m}(T(E)) = 0$ for any set $E \subset \mathbb{R}^k$, so we get $c = 0$.

If $\mathbf{A}$ is not singular, then the inverse map T^{-1} is also linear and defined by $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} + \mathbf{a}_1$, where $\mathbf{a}_1 = -\mathbf{A}^{-1}\mathbf{a}$. Hence $T^{-1}\colon \mathbb{R}^k \to \mathbb{R}^k$ is a continuous map. Thus for every Borel set $E \subset \mathbb{R}^k$ the set $T(E)$ is Borel (Proposition 4.4).

We define a new measure μ on the Borel σ-algebra in $\mathbb{R}^k$ by $\mu(E) = \mathbf{m}\big(T(E)\big)$. Now for any Borel set E and any fixed vector $\mathbf{b} \in \mathbb{R}^k$ we have

$$\mu(E + \mathbf{b}) = \mathbf{m}\big(T(E + \mathbf{b})\big) = \mathbf{m}(\mathbf{A}E + \mathbf{A}\mathbf{b} + \mathbf{a}) = \mathbf{m}(\mathbf{A}E + \mathbf{a}) = \mathbf{m}\big(T(E)\big) = \mu(E)$$

(we used the translation invariance of the Lebesgue measure $\mathbf{m}$). Hence the measure μ is also translation invariant. Due to Corollary 3.19 we have $\mu = c\,\mathbf{m}$ with some constant $c \geq 0$.

To find the value of c, it is enough to compute it as $c = \frac{\mathbf{m}(T(E))}{\mathbf{m}(E)}$ for one set E. If we choose E to be the unit box, $E = \{0 \leq x_1 \leq 1, \dots, 0 \leq x_k \leq 1\}$, then the calculation is elementary (see Rudin's book, Section 2.23) and gives $c = |\det \mathbf{A}|$. $\square$

- Every linear transformation with $\det A = \pm 1$ preserves the Lebesgue measure. In particular, every isometry (translation, rotation, reflection) preserves the Lebesgue measure.
- We will denote points in $\mathbb{R}^k$ by x and occasionally by $\mathbf{x}$ when we need to use the coordinates of x as a column-vector. Thus x and $\mathbf{x}$ will be interchangeable in our formulas.

Definition 15.2. Let $T\colon V \to \mathbb{R}^k$ be a (nonlinear) transformation defined on an open set $V \subset \mathbb{R}^k$. Its **derivative** $T'(x)$ at a point $x \in V$ is a $k \times k$ matrix $\mathbf{A}$ that satisfies

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|T(\mathbf{x}+\mathbf{h}) - T(\mathbf{x}) - \mathbf{A}\mathbf{h}\|}{\|\mathbf{h}\|} = 0$$

provided the limit exists. We will write $T'(x) = \mathbf{A}$.

Lemma 15.3. *If the transformation $T\colon V \to \mathbb{R}^k$ is defined coordinate-wise, by k functions of k variables*

$$y_1 = f_1(x_1, \dots, x_k), \ \dots, \ y_k = f_k(x_1, \dots, x_k)$$

then $T'(x)$ is the matrix of their partial derivatives: $\mathbf{A} = (\partial f_i/\partial x_j)$.

Proof. Just use $\mathbf{h} = h\mathbf{e}_i$ in Definition 15.2 and take the limit as $h \to 0$; here $\mathbf{e}_i$ denotes the ith canonical basis vector in $\mathbb{R}^k$. □

Corollary 15.4. *If $T\colon \mathbb{R}^k \to \mathbb{R}^k$ is a linear transformation defined by $\mathbf{y} = \mathbf{A}\mathbf{x}+\mathbf{a}$, then its derivative is $T'(x) = \mathbf{A}$ at every point $x \in \mathbb{R}^k$.*

- Thus $|\det T'(x)| = |\det \mathbf{A}|$ is the factor by which the Lebesgue measure is multiplied by a linear transformation.

Definition 15.5. If $T\colon V \to \mathbb{R}^k$ is differentiable at $x \in V$, then $J_T(x) = \det T'(x)$ is called the **Jacobian** of T at x.

Theorem 15.6. *Let $T\colon V \to \mathbb{R}^k$ be continuous on V and differentiable at $x \in V$. Then*

$$\lim_{r\to 0} \frac{\mathbf{m}(T(B(x,r)))}{\mathbf{m}(B(x,r))} = |\det T'(x)|$$

Proof. The formula in Definition 15.2 can be written as

$$T(\mathbf{x}+\mathbf{h}) - T(\mathbf{x}) = \mathbf{A}\mathbf{h} + \mathbf{g}_{\mathbf{h}}, \qquad \|\mathbf{g}_{\mathbf{h}}\| = o(\|\mathbf{h}\|)$$

For all $\mathbf{x}+\mathbf{h} \in B(\mathbf{x}, r)$ we have $\|\mathbf{h}\| < r$ and so

$$T(\mathbf{x}+\mathbf{h}) = T(\mathbf{x}) + \mathbf{A}\mathbf{h} + \mathbf{g}_{\mathbf{h}}, \qquad \|\mathbf{g}_{\mathbf{h}}\| = o(r)$$

Thus for any $\varepsilon > 0$ there is $r_\varepsilon > 0$ such that for all $r < r_\varepsilon$ we have

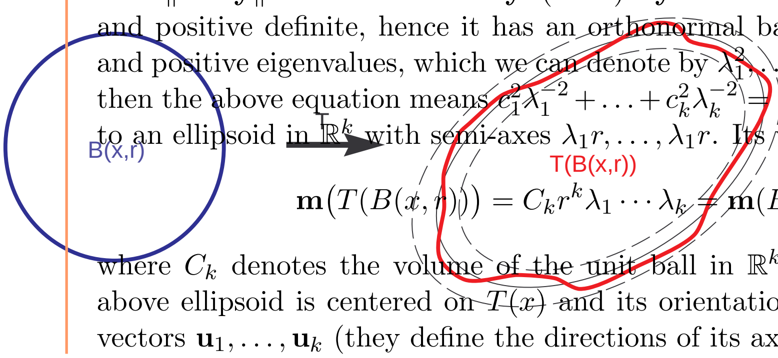
$$T(\mathbf{x}+\mathbf{h}) = T(\mathbf{x}) + \mathbf{A}\mathbf{h} + \mathbf{g}_{\mathbf{h}}, \qquad \|\mathbf{g}_{\mathbf{h}}\| < \varepsilon r \tag{15.1}$$

for all $\mathbf{x}+\mathbf{h} \in B(\mathbf{x}, r)$.

Now suppose for a moment that the transformation T is linear. Then

$$T(\mathbf{x}+\mathbf{h}) = T(\mathbf{x}) + \mathbf{A}\mathbf{h} \qquad \text{(i.e., } \mathbf{g}_{\mathbf{h}} = \mathbf{0}) \tag{15.2}$$

Suppose also that $\det \mathbf{A} \neq 0$. Then the image $T(B(x,r))$ of the ball $B(x,r)$ is an ellipsoid centered on $T(x)$ (see the picture). More precisely, the boundary of $T(B(x,r))$ is made by vectors $T(\mathbf{x}) + \mathbf{A}\mathbf{h}$ with $\|\mathbf{h}\| = r$, i.e., by vectors $T(\mathbf{x}) + \mathbf{y}$ with $\|\mathbf{A}^{-1}\mathbf{y}\| = r$. This means $\mathbf{y}^T(AA^T)^{-1}\mathbf{y} = r^2$. The matrix AA^T is symmetric and positive definite, hence it has an orthonormal basis of eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ and positive eigenvalues, which we can denote by $\lambda_1^2, \dots, \lambda_k^2$. If $\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k$, then the above equation means $c_1^2\lambda_1^{-2} + \ldots + c_k^2\lambda_k^{-2} = r^2$. This equation corresponds to an ellipsoid in $\mathbb{R}^k$ with semi-axes $\lambda_1 r, \dots, \lambda_1 r$. Its volume is



$$\mathbf{m}\big(T(B(x,r))\big) = C_k r^k \lambda_1 \cdots \lambda_k = \mathbf{m}(B(x,r))|\det \mathbf{A}|$$

where C_k denotes the volume of the unit ball in $\mathbb{R}^k$ (a standard constant). The above ellipsoid is centered on $T(x)$ and its orientation in $\mathbb{R}^k$ is determined by the vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ (they define the directions of its axes).

For nonlinear maps we have an additional term $\mathbf{g_h}$ in (15.1) whose norm is $< \varepsilon r$. This means that the figure $T(B(x,r))$ has boundary that is within distance $< \varepsilon r$ from the boundary of the above ellipsoid. Thus $E'_r \subset T(B(x,r)) \subset E''_r$, where E'_r and E''_r are ellipsoids centered on $T(x)$ with the same directions of axes as the above ellipsoid but with the semi-axes given by $r(\lambda_1 \pm \varepsilon), \ldots, r(\lambda_k \pm \varepsilon)$, where $+$ corresponds to E''_r and $-$ to E'_r. The Lebesgue measure of $T(B(x,r))$ is bounded by

$$C_k r^k \prod_{i=1}^{k} (\lambda_i - \varepsilon) \le \mathbf{m}\big(T(B(x,r))\big) \le C_k r^k \prod_{i=1}^{k} (\lambda_i + \varepsilon)$$

Therefore

$$\prod_{i=1}^{k} \Big(1 - \frac{\varepsilon}{\lambda_i}\Big) \le \frac{\mathbf{m}(T(B(x,r)))}{\mathbf{m}(B(x,r))} \le \prod_{i=1}^{k} \Big(1 + \frac{\varepsilon}{\lambda_i}\Big)$$

Since $\varepsilon > 0$ can be made arbitrarily small in the limit $r \to 0$, we obtain the claim of the theorem.

It remains to deal with the singular case $\det \mathbf{A} = 0$. Now the linear transformation (15.2) takes $B(x,r)$ into a subspace of a lower dimension; more precisely - into a lower dimensional ellipsoid centered on $T(x)$ whose dimensions (semi-axes) are proportional to r. The nonlinear transformation (15.1) takes $B(x,r)$ into a figure which is in the εr-neighborhood of that ellipsoid. It can be imagined as a thin plate ('pancake') whose dimensions are proportional to r and whose 'width' ('thickness') is proportional to εr. Thus its volume is $\mathcal{O}(\varepsilon r^k)$. Dividing by $\mathbf{m}(B(x,r)) = C_k r^k$ and taking the limit $r \to 0$ gives us zero, as claimed.

$\square$

- Our proof appeals to geometric intuition and requires some knowledge of linear algebra. But it is not entirely rigorous – it hides some unpleasant technical details that can be found in Rudin's book (Lemma 7.23 and Theorem 7.24).

- The following is a (global) change of variables rule:

Theorem 15.7. *Let $T\colon V \to \mathbb{R}^k$ be continuous and differentiable at every point $x \in V$ and one-to-one. Then for any measurable function $f\colon \mathbb{R}^k \to [0, +\infty]$*

$$\int_{T(V)} f \, d\mathbf{m} = \int_V (f \circ T)\, |\det T'| \, d\mathbf{m}$$

We will prove a lemma first, and then prove the theorem.

Lemma 15.8. *T maps null sets into null sets.*

Proof. Intuitively, the map T 'expands' the Lebesgue measure by a finite factor $|\det T'(x)|$ at every point $x \in V$ (by Theorem 15.6), so it should be impossible to 'expand' a set of measure zero into a set of positive measure.

Since $T'(x)$ exists at every point $x \in V$, we have

$$\limsup_{\mathbf{h}\to\mathbf{0}} \frac{\|T(\mathbf{x}+\mathbf{h}) - T(\mathbf{x})\|}{\|\mathbf{h}\|} < \infty \tag{15.3}$$

For any $m, n \geq 1$ let $V_{m,n} \subset V$ consist of points $x \in V$ such that

$$\|T(\mathbf{x}+\mathbf{h}) - T(\mathbf{x})\| \leq m\|\mathbf{h}\| \qquad \forall \|\mathbf{h}\| \leq 1/n \tag{15.4}$$

Due to (15.3), we have $\cup_{m,n} V_{m,n} = V$.

Let $N \subset V$ be a null set. Then $N_{m,n} = N \cap V_{m,n}$ is also a null set. So for any $\varepsilon > 0$ it can be covered by a countable union of balls $B(x_i, r_i)$ such that $x_i \in N_{m,n}$ and $r_i \leq 1/n$ in such a way that $\sum_i \mathbf{m}(B(x_i, r_i)) < \varepsilon$. [To do this, first cover $N_{m,n}$ by an open set W of a very small measure $\ll \varepsilon$, then partition W into countably many boxes $\{R_i\}$ of a very small diameter $\ll 1/n$, then for each box R_i find a point $x_i \in N_{m,n} \cap R_i$ (if none exists, just remove R_i from W), then set $r_i = \operatorname{diam} R_i$. This gives $B(x_i, r_i) \supset R_i$ and $\mathbf{m}(B(x_i, r_i)) \leq C_k \mathbf{m}(R_i)$. Hence

$$\cup_i B(x_i, r_i) \supset W \supset N_{m,n}$$

and

$$\sum_i \mathbf{m}(B(x_i, r_i)) < C_k \sum_i \mathbf{m}(R_i) \leq C_k \mathbf{m}(W) < \varepsilon$$

where $C_k > 0$ is a constant depending only on the dimensionality k.]

Now each ball $B(x_i, r_i)$ is mapped by T into a ball of radius $\leq m r_i$, due to (15.4), so its volume grows by a factor $\leq m^k$. Thus $T(N_{m,n})$ will be covered by countable many balls of total Lebesgue measure $\leq m^k \varepsilon$, hence $\mathbf{m}(T(N_{m,n})) \leq m^k \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we actually have $\mathbf{m}(T(N_{m,n})) = 0$, and taking the union over all $m, n \geq 1$ gives $\mathbf{m}(T(N)) = 0$. $\square$

Proof. (of Theorem 15.7). This is done in several steps.

Step 1. T maps Lebesgue measurable sets into Lebesgue measurable sets. (Note: this does not simply follow from the continuity of T; recall our discussion on page 138 .)

Indeed, let $E \subset V$ be a Lebesgue measurable set. Due to regularity of the Lebesgue measure (Theorem 8.4) there exists an F_σ-set $F \subset E$ such that $\mathbf{m}(E \setminus F) = 0$. Due to the continuity of T, the image $T(F)$ is also an F_σ-set, thus it is Borel measurable. And the image $T(E \setminus F)$ is a null set due to Lemma 15.8. Hence $T(E)$ is the union of a Borel set and a null set, hence it is Lebesgue measurable.

Step 2. For $n \geq 1$, let $V_n = \{x \in V : \|T(x)\| < n\}$. Note that V_n is open, $V_n \subset V_{n+1}$ and $\cup_{n \geq 1} V_n = V$. For each $n \geq 1$ we define a new measure on V by

$$\mu_n(E) = \mathbf{m}(T(E \cap V_n)) \tag{15.5}$$

Why is it a measure? Because $T(E \cap V_n)$ is Lebesgue measurable due to Step 1, and since T is one-to-one, μ_n is σ-additive. Also μ_n is a finite measure because $\mu_n(V) \leq \mathbf{m}(B(0, n)) < \infty$ (this was the reason for replacing V temporarily by V_n). Now $\mu_n \ll \mathbf{m}$ by another application of Lemma 15.8. Thus due to Theorem 13.15

$$\mu_n(E) = \int_E (D\mu_n)\, d\mathbf{m}$$

Step 3. We claim that for all $x \in V_n$

$$(D\mu_n)(x) = |\det T'(x)| \tag{15.6}$$

Indeed, V_n is open, hence $B(x, r) \subset V_n$ for sufficiently small $r > 0$. Due to (15.5)

$$\mu_n(B(x, r)) = \mathbf{m}\big(T(B(x, r))\big)$$

If we divide by $\mathbf{m}(B(x, r))$ and refer to Theorem 15.6, we obtain (15.6).

Step 4. Thus we can write

$$\mathbf{m}(T(E \cap V_n)) = \int_{V_n} \chi_E |\det T'|\, d\mathbf{m} = \int_V \chi_{E \cap V_n} |\det T'|\, d\mathbf{m}$$

Taking the limit $n \to \infty$ and applying the Lebesgue Monotone Convergence gives

$$\mathbf{m}(T(E)) = \int_V \chi_E |\det T'|\, d\mathbf{m}$$

for every measurable set $E \subset V$.

Step 5. Let $A \subset \mathbb{R}^k$ be a Borel set. Since T is continuous, $E = T^{-1}(A)$ is also a Borel set and $E \subset V$. Note that $T(E) = A \cap T(V)$ and $\chi_E = \chi_A \circ T$. Thus

$$\int_{T(V)} \chi_A\, d\mathbf{m} = \mathbf{m}(A \cap T(V)) = \mathbf{m}(T(E)) = \int_V (\chi_A \circ T)\, |\det T'|\, d\mathbf{m} \tag{15.7}$$

Step 6. Next we extend (15.7) to Lebesgue measurable sets $A \subset \mathbb{R}^k$. First, if N is a null set, then $N \subset A_N$ for a Borel null set A_N. For $A = A_N$, the first integral in (15.7) is zero, hence so is the second one, hence $(\chi_{A_N} \circ T)\,|\det T'| = 0$ a.e. Now $\chi_N \le \chi_{A_N}$, therefore $(\chi_N \circ T)\,|\det T'| = 0$ a.e. Thus

$$\mathbf{m}(N \cap T(V)) = \int_{T(V)} \chi_N \, d\mathbf{m} = \int_V (\chi_N \circ T)\,|\det T'|\, d\mathbf{m} \tag{15.8}$$

(because both integrals are zero). Since every Lebesgue measurable set $E \subset \mathbb{R}^k$ is a disjoint union, $E = A \uplus N$, of a Borel set A and a null set N, and then $\chi_E = \chi_A + \chi_N$. Now adding (15.7) and (15.8) gives

$$\int_{T(V)} \chi_E \, d\mathbf{m} = \int_V (\chi_E \circ T)\,|\det T'|\, d\mathbf{m} \tag{15.9}$$

Step 7. Once we have (15.9), it is clear that

$$\int_{T(V)} s \, d\mathbf{m} = \int_V (s \circ T)\,|\det T'|\, d\mathbf{m}$$

for every *simple* function $s \ge 0$ on $\mathbb{R}^k$. Another application of the Lebesgue Monotone Convergence completes the proof of the theorem. □

- We did not prove that $f \circ T$ is a measurable function. In fact, this is not necessarily true! What our proof does establish is that the product $(f \circ T)\,|\det T'|$ is a measurable function.

Corollary 15.9. *Suppose that*
(i) $\varphi\colon (a, b) \to (\alpha, \beta)$ *is a monotonically increasing differentiable function such that*

$$\lim_{x \to a+} \varphi(x) = \alpha \quad \textit{and} \quad \lim_{x \to b-} \varphi(x) = \beta$$

(ii) $f \ge 0$ *is a Lebesgue measurable function. Then*

$$\int_{[\alpha,\beta]} f \, d\mathbf{m} = \int_{[a,b]} (f \circ \varphi)\,|\varphi'|\, d\mathbf{m}$$

Proof. If φ is strictly monotonically increasing, then it is one-to-one and the result follows from Theorem 15.7 (we just set $k = 1$ and $V = (a, b)$). If φ is not strictly monotonically increasing, then it is constant on some intervals. On those intervals $\varphi' = 0$, so they do not affect the value of the second integral; thus they can be collapsed and removed, after which φ becomes strictly monotonically increasing (we omit the details of this "collapsing" procedure). □

- This corollary easily extends to integrable functions $f \in L^1_{\mathbf{m}}(a,b)$ and monotonically *decreasing* differentiable functions $\varphi\colon (a,b) \to (\alpha,\beta)$ such that

$$\lim_{x\to a+} \varphi(x) = \beta \quad \text{and} \quad \lim_{x\to b-} \varphi(x) = \alpha$$

Exercise 79. Let $f \geq 0$ and $f \in L^1(\mathbb{R})$. Find

$$\lim_{n\to\infty} \int_{[0,1]} f(nx)\, d\mathbf{m}(x)$$

Exercise 80. Let $f, g\colon [a,b] \to \mathbb{C}$ be two AC functions. Prove the integration-by-parts formula

$$\int_{[a,b]} f'g\, d\mathbf{m} = f(b)g(b) - f(a)g(a) - \int_{[a,b]} fg'\, d\mathbf{m}$$

Exercise 81. Let $f \geq 0$ and $f \in L^1([0,\infty))$ and $g(x)\colon = 2xf(x^2)$ for all $x \in [0,\infty)$. Show that $g \in L^1([0,\infty))$ and

$$\int_{[0,\infty)} f\, d\mathbf{m} = \int_{[0,\infty)} g\, d\mathbf{m}$$

Exercise 82. [Bonus] Let $f\colon \mathbb{R} \to \mathbb{R}$ be integrable, i.e., $f \in L^1_{\mathbf{m}}(\mathbb{R})$. Define a function $g\colon \mathbb{R} \to \mathbb{R}$ by

$$g(x) = \begin{cases} f\left(x - \frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Prove that $g(-1/x) = g(x)$ and

$$\int_{\mathbb{R}} f\, d\mathbf{m} = \int_{\mathbb{R}} g\, d\mathbf{m}$$

16 Product spaces

Definition 16.1. Let X and Y be two sets. Its **Cartesian product** $X \times Y$ is the set of all (ordered) pairs (x, y), with $x \in X$ and $y \in Y$. If $A \subset X$ and $B \subset Y$, then the set $A \times B \subset X \times Y$ is called a **rectangle**.

Definition 16.2. Let $(X, \mathfrak{M})$ and $(Y, \mathfrak{N})$ be two measurable spaces. For every $A \in \mathfrak{M}$ and $B \in \mathfrak{N}$ the set $A \times B$ is called a **measurable rectangle** in $X \times Y$. Any finite union $E = R_1 \cup R_2 \cup \cdots \cup R_n$ of disjoint measurable rectangles is called an **elementary set**. The collection of elementary sets is denoted by $\mathcal{E}$.

Proposition 16.3. *$\mathcal{E}$ is an algebra, i.e. finite unions, intersections, differences and complements of elementary sets are elementary sets.*

Proof. A straightforward verification. (Recall also Example 1.) $\square$

- Note: $\mathcal{E}$ is *not* a σ-algebra.

Definition 16.4. We denote by $\mathfrak{M} \times \mathfrak{N}$ the (*minimal*) σ-algebra in $X \times Y$ generated by elementary sets (equivalently, by measurable rectangles).

Definition 16.5. Let $E \subset X \times Y$ and $x \in X$, $y \in Y$. Then

$$E_x = \{y' \colon (x, y') \in E\} \subset Y$$

is called the x-**section** of E and

$$E^y = \{x' \colon (x', y) \in E\} \subset X$$

is called the y-**section** of E.

Theorem 16.6. *If $E \in \mathfrak{M} \times \mathfrak{N}$, then $E_x \in \mathfrak{N}$ and $E^y \in \mathfrak{M}$ for every $x \in X$ and every $y \in Y$.*

Proof. Let $\Omega = \{E \subset X \times Y \colon\ E_x \in \mathfrak{N}\ \ \forall x \in X\}$. We will show that

(a) Ω contains all measurable rectangles

(b) Ω is a σ-algebra.

Then it will follow that $\Omega \supset \mathfrak{M} \times \mathfrak{N}$, because $\mathfrak{M} \times \mathfrak{N}$ is the *minimal* σ-algebra containing all measurable rectangles.

Proof of (a): if $E = A \times B$ is a measurable rectangle, then $E_x = B$ for $x \in A$ and $E_x = \emptyset$ for $x \notin A$, so $E_x \in \mathfrak{N}$ in either case.

Proof of (b): First, if $E \in \Omega$, then for every $x \in X$ we have $E_x \in \mathfrak{N}$, hence $(E^c)_x = (E_x)^c \in \mathfrak{N}$, thus $E^c \in \Omega$. Second, if $E_i \in \Omega$ and $E = \cup_i E_i$, then $E_x = \cup_i (E_i)_x \in \mathfrak{N}$, thus $E \in \Omega$ (σ-additivity).

According to Lemma 2.4, Ω is a σ-algebra. The proof for E^y is similar. $\square$

Definition 16.7. Let $f \colon X \times Y \to Z$ be a function.
For every $x \in X$, the function f_x on Y is defined by $f_x(y) = f(x, y)$.
For every $y \in Y$, the function f^y on X is defined by $f^y(x) = f(x, y)$.
(these can be regarded as **sections** of a function of two variables.)

Theorem 16.8. *Let Z be a topological space and $f \colon X \times Y \to Z$ a measurable function (with respect to the σ-algebra $\mathfrak{M} \times \mathfrak{N}$). Then*

(i) *f_x is $\mathfrak{N}$-measurable on Y for each $x \in X$;*

(ii) *f^y is $\mathfrak{M}$-measurable on X for each $y \in Y$;*

Proof. Let $V \subset Z$ be an open set. The measurability of f means that

$$Q = f^{-1}(V) = \{(x, y) \colon f(x, y) \in V\} \in \mathfrak{M} \times \mathfrak{N}$$

Note that for each $x \in X$

$$Q_x = \{y \colon (x, y) \in Q\} = \{y \colon f_x(y) \in V\}$$

and by Theorem 16.6 we have $Q_x \in \mathfrak{N}$ thus f_x is $\mathfrak{N}$-measurable. Similarly, f^y is $\mathfrak{M}$-measurable. $\square$

Definition 16.9. A **monotone class** $\mathcal{M}$ is a collection of sets with the following properties:

(i) $A_i \in \mathcal{M},\ A_i \subset A_{i+1}\ (i = 1, 2, \ldots) \implies \cup_{i=1}^{\infty} A_i \in \mathcal{M}$;

(ii) $B_i \in \mathcal{M},\ B_i \supset B_{i+1}\ (i = 1, 2, \ldots) \implies \cap_{i=1}^{\infty} B_i \in \mathcal{M}$.

- Note: σ-algebras are monotone classes, but not vice versa.

Example 25. Any finite collection of sets $\{A_1, \ldots, A_n\}$ makes a monotone class (check it out!), but in order to be a σ-algebra it must contain all their unions, intersections, differences, complements, etc.

The following are analogues of Theorem 3.4 and Theorem 3.5:

Lemma 16.10. *Let $\{\mathcal{M}_\alpha\}$ be an arbitrary collection of monotone classes of a set X. Then their intersection $\cap_\alpha \mathcal{M}_\alpha$ is a monotone class of X as well.*

Proof. Direct inspection. Note that the collection of monotone classes here may be finite, countable, or uncountable; its cardinality is not essential. □

Lemma 16.11. *Let $\mathcal{F}$ be any collection of subsets of X. Then there exists a unique monotone class $\mathcal{M}^* \supset \mathcal{F}$ such that for any other monotone class $\mathcal{M} \supset \mathcal{F}$ we have $\mathcal{M}^* \subset \mathcal{M}$. (In other words, $\mathcal{M}^*$ is the minimal monotone class containing $\mathcal{F}$.)*

Proof. $\mathcal{M}^*$ is the intersection of all monotone classes containing $\mathcal{F}$. □

Minimal monotone class containing all rectangles. It is easy to see that if $X \subset \mathbb{R}$ is a finite interval, then the collection of all subintervals $I \subset X$ (including all open, closed, and semiopen intervals) is a monotone class. Similarly, if $X \subset \mathbb{R}^2$ is a rectangle, then the collection of all subrectangles $R \subset X$ is a monotone class. More generally, the collection of all measurable rectangles in $X \times Y$ defined above is a monotone class.

Recall now that the collection of *finite unions* of subintervals $I \subset X$ is an algebra; cf. Example 2. It is *not a monotone class.* The minimal monotone class containing all finite unions of intervals includes all G_δ and all F_σ sets, in particular the Cantor set C, so it is very rich.

The minimal monotone class containing all finite unions of measurable rectangles (i.e., all elementary sets) in $X \times Y$ is very rich:

Theorem 16.12. *The minimal monotone class containing $\mathcal{E}$, i.e., containing all elementary sets in $X \times Y$ is the entire σ-algebra $\mathfrak{M} \times \mathfrak{N}$.*

Proof. Let $\mathfrak{G}$ denote the minimal monotone class containing $\mathcal{E}$. We claim that

(a) $P, Q \in \mathfrak{G} \quad \Rightarrow \quad P \setminus Q \in \mathfrak{G}$;

(b) $P, Q \in \mathfrak{G} \quad \Rightarrow \quad P \cup Q \in \mathfrak{G}$.

Indeed, for any set $P \subset X \times Y$ (measurable or not) let us denote

$$\Omega(P) = \{Q \subset X \times Y : P \setminus Q \in \mathfrak{G},\ Q \setminus P \in \mathfrak{G},\ P \cup Q \in \mathfrak{G}\}$$

One can verify, by direct inspection, that

(i) $Q \in \Omega(P) \quad \Leftrightarrow \quad P \in \Omega(Q) \qquad$ (symmetry)

(ii) $\Omega(P)$ is a monotone class (since so is $\mathfrak{G}$)

Now fix $P \in \mathcal{E}$. Then $\mathcal{E} \subset \Omega(P)$ (because $\mathcal{E}$ is an algebra). Hence $\mathfrak{G} \subset \Omega(P)$.

Now fix $Q \in \mathfrak{G}$. Then $Q \in \Omega(P)$ for all $P \in \mathcal{E}$. By (i) we have $P \in \Omega(Q)$ for all $P \in \mathcal{E}$. Therefore $\mathcal{E} \subset \Omega(Q)$, and then $\mathfrak{G} \subset \Omega(Q)$. This proves our claims (a) and (b).

Note that $\mathfrak{G} \subset \mathfrak{M} \times \mathfrak{N}$, because $\mathfrak{M} \times \mathfrak{N}$ is a monotone class. Thus if we show that $\mathfrak{G}$ is a σ-algebra, it would imply that $\mathfrak{G} = \mathfrak{M} \times \mathfrak{N}$, as desired. Next we check that $\mathfrak{G}$ is a σ-algebra.

First, if $Q \in \mathfrak{G}$, then $Q^c = (X \times Y) \setminus Q \in \mathfrak{G}$ by the property (a) above. Second, if $P_i \in \mathfrak{G}$ and $P = \cup_i P_i$, then $Q_n = P_1 \cup \cdots \cup P_n \in \mathfrak{G}$ for every $n \geq 1$ by property (b) above. Lastly, since $Q_n \subset Q_{n+1}$, the sets $\{Q_n\}$ make a monotone sequence, hence $\cup_n Q_n \in \mathfrak{G}$ because $\mathfrak{G}$ is a monotone class. Thus $P = \cup_n Q_n \in \mathfrak{G}$ proving the σ-additivity of $\mathfrak{G}$. According to Lemma 2.4, $\mathfrak{G}$ is a σ-algebra. $\square$

Our next goal is to construct the product of measures.

Theorem 16.13. *Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \lambda)$ be two measure spaces with σ-finite positive measures μ and λ. Let $Q \in \mathfrak{M} \times \mathfrak{N}$. Define*

$$\begin{aligned} \varphi(x) &= \lambda(Q_x) \qquad \text{for every } x \in X \\ \psi(y) &= \mu(Q^y) \qquad \text{for every } y \in Y \end{aligned}$$

Then φ is $\mathfrak{M}$-measurable, ψ is $\mathfrak{N}$-measurable, and

$$\int_X \varphi \, d\mu = \int_Y \psi \, d\lambda. \tag{16.1}$$

- Note that

$$\lambda(Q_x) = \int_Y \chi_Q(x, y) \, d\lambda(y)$$

and similarly

$$\mu(Q^y) = \int_X \chi_Q(x, y) \, d\mu(x)$$

Thus we can rewrite (16.1) as

$$\int_X \left(\int_Y \chi_Q(x, y) \, d\lambda(y) \right) d\mu(x) = \int_Y \left(\int_X \chi_Q(x, y) \, d\mu(x) \right) d\lambda(y). \tag{16.2}$$

Proof. Let Ω denote the class of all $Q \in \mathfrak{M} \times \mathfrak{N}$ for which the conclusion of the theorem holds. We claim that Ω has the following properties:

(a) Every measurable rectangle belongs in Ω

(b) If $Q_1 \subset Q_2 \subset \cdots$ and each $Q_i \in \Omega$, then $\cup_i Q_i \in \Omega$

(c) If $\{Q_i\}$ are disjoint members of Ω, then $\uplus_i Q_i \in \Omega$

(d) If $\mu(A) < \infty$ and $\lambda(B) < \infty$, and if $A \times B \supset Q_1 \supset Q_2 \supset \cdots$ for some $Q_i \in \Omega$, then $\cap_i Q_i \in \Omega$.

To prove (a) note that if $Q = A \times B$ is a measurable rectangle, then

$$\lambda(Q_x) = \lambda(B)\chi_A(x) \qquad \text{and} \qquad \mu(Q^y) = \mu(A)\chi_B(y)$$

therefore each of the integrals in (16.1) equals $\mu(A)\lambda(B)$.

To prove (b), let φ_i and ψ_i be associated with Q_i in the way in which φ and ψ are associated with Q. The continuity (Theorem 3.13) of μ and λ implies pointwise convergence

$$\varphi_i(x) \to \varphi(x) \qquad \text{and} \qquad \psi_i(y) \to \psi(y)$$

as $i \to \infty$, for every point $x \in X$ and $y \in Y$. Also note that φ_i and ψ_i are monotonically increasing sequences of functions. Since they are assumed to satisfy the conclusion of the theorem, (b) follows from the Lebesgue Monotone Convergence.

The claim (c) is easy for finite disjoint unions, because the characteristic function of the union is the sum of the characteristic functions of the sets. Now for countable disjoint unions, (c) follows from (b).

The proof of (d) is like that of (b), except we use the Lebesgue Dominated Convergence, which is legitimate since $\mu(A) < \infty$ and $\lambda(B) < \infty$.

Now due to σ-finiteness of μ and λ we have $X = \uplus X_n$ with $\mu(X_n) < \infty$ and $Y = \uplus Y_m$ with $\mu(Y_m) < \infty$. Define

$$Q_{mn} = Q \cap (X_n \times Y_m)$$

and let $\mathfrak{G}$ be the class of all $Q \in \mathfrak{M} \times \mathfrak{N}$ such that $Q_{mn} \in \Omega$ for all m, n. Then (b) and (d) show that $\mathfrak{G}$ is a monotone class. Also, (a) and (c) show that $\mathcal{E} \subset \mathfrak{G}$. Now Theorem 16.12 implies that $\mathfrak{G} \supset \mathfrak{M} \times \mathfrak{N}$.

Thus $Q_{mn} \in \Omega$ for every $Q \in \mathfrak{M} \times \mathfrak{N}$ and all m, n. Since $Q = \uplus_{m,n} Q_{mn}$, we conclude from (c) that $Q \in \Omega$. This completes the proof. $\square$

DEFINITION 16.14. Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \lambda)$ be two measure spaces with σ-finite positive measures μ and λ. Then we can define a measure, $\mu \times \lambda$, on $\mathfrak{M} \times \mathfrak{N}$ by

$$(\mu \times \lambda)(Q) = \int_X \lambda(Q_x)\, d\mu(x) = \int_Y \mu(Q^y)\, d\lambda(y).$$

The measure $\mu \times \lambda$ is called the **product** of the measures μ and λ.

- The fact that $\mu \times \lambda$ is indeed a measure (i.e., its σ-additivity) follows immediately from Theorem 5.13.
- Observe that $\mu \times \lambda$ is also a $\underline{\sigma\text{-finite}}$ positive measure.
- Now we can rewrite (16.2) as

$$\begin{aligned}\int_{X \times Y} \chi_Q \, d(\mu \times \lambda) &= \int_X \left(\int_Y \chi_Q(x, y)\, d\lambda(y) \right) d\mu(x) \\ &= \int_Y \left(\int_X \chi_Q(x, y)\, d\mu(x) \right) d\lambda(y). \end{aligned} \tag{16.3}$$

EXERCISE 83. Show that $\mu \times \lambda$ is a unique measure on $\mathfrak{M} \times \mathfrak{N}$ such that

$$(\mu \times \lambda)(A \times B) = \mu(A)\lambda(B)$$

for all measurable rectangles $A \times B$ in $X \times Y$. (Hint: similar to the proof of Theorem 16.13.)

17 Fubini Theorem

Theorem 17.1. <u>**Fubini**</u>

Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \lambda)$ be two measure spaces with σ-finite positive measures. Let f be an $(\mathfrak{M} \times \mathfrak{N})$-measurable function on $X \times Y$. Then

(a) *Suppose $0 \le f \le \infty$. Define, for each $x \in X$ and $y \in Y$*

$$\varphi(x) = \int_Y f_x \, d\lambda, \qquad \psi(y) = \int_X f^y \, d\mu$$

Then φ is $\mathfrak{M}$-measurable, ψ is $\mathfrak{N}$-measurable, and

$$\int_{X \times Y} f \, d(\mu \times \lambda) = \int_X \varphi \, d\mu = \int_Y \psi \, d\lambda. \tag{17.1}$$

(b) *Suppose f is complex-valued and*

$$\int_X \varphi^* \, d\mu < \infty \qquad \textit{where} \qquad \varphi^*(x) = \int_Y |f|_x \, d\lambda$$

Then $f \in L^1_{\mu \times \lambda}$.

(c) *Suppose $f \in L^1_{\mu \times \lambda}$. Then*

$$f_x \in L^1_\lambda \quad \text{for a.e.} \quad x \in X,$$

$$f^y \in L^1_\mu \quad \text{for a.e.} \quad y \in Y,$$

$$\varphi(x) = \int_Y f_x \, d\lambda \in L^1_\mu,$$

$$\psi(y) = \int_X f^y \, d\mu \in L^1_\lambda,$$

and the above equation (17.1) holds.

- Note that equation (17.1) can be written as

$$\begin{aligned} \int_{X \times Y} f \, d(\mu \times \lambda) &= \int_X \left(\int_Y f(x, y) \, d\lambda(y) \right) d\mu(x) \\ &= \int_Y \left(\int_X f(x, y) \, d\mu(x) \right) d\lambda(y) \end{aligned} \tag{17.2}$$

(a **double integral** is represented by **iterated integrals**).

Corollary 17.2. *Suppose $f \ge 0$. If one iterated integral in (17.2) exists, then all the three integrals exist and are equal.*

(For non-positive functions, this is not true, see Example 26 below.)

- The corollary is obtained by a combination of (b) and (c).

Proof. We begin with (a).

Proof of (a). According to Theorem 16.8, f_x and f^y are measurable non-negative functions, thus the definitions of φ and ψ are legitimate. The rest of the proof consists of three steps:

Step 1: Indicator functions. Let $Q \in \mathfrak{M} \times \mathfrak{N}$ and $f = \chi_Q$. Then our target equation (17.2) is exactly (16.3) obtained earlier.

Step 2: Simple functions. Suppose $f = \sum_{n=1}^{N} \alpha_n \chi_{Q_n}$ with some $\alpha_n \geq 0$, where $Q_n \in \mathfrak{M} \times \mathfrak{N}$ and $X \times Y = \biguplus_{n=1}^{N} Q_n$. Then the conclusion is obtained due to Step 1 and the obvious linearity of (17.2)

Step 3: General functions. Given a measurable $f \geq 0$, there is an increasing sequence of simple nonnegative functions $\{s_n\}$ converging to f pointwise (Theorem 4.22):

$$0 \leq s_1 \leq s_2 \leq \cdots \qquad \text{and} \qquad s_n(x,y) \to f(x,y)$$

for every $(x,y) \in X \times Y$. If φ_n is associated with s_n in the same way in which φ is associated with f, then Step 2 implies

$$\int_X \varphi_n \, d\mu = \int_{X \times Y} s_n \, d(\mu \times \lambda)$$

The Lebesgue Monotone Convergence applied on $(Y, \mathfrak{N}, \lambda)$ shows that

$$0 \leq \varphi_1(x) \leq \varphi_2(x) \leq \cdots \qquad \text{and} \qquad \varphi_n(x) \to \varphi(x)$$

for every $x \in X$. Taking the limit $n \to \infty$ and using Lebesgue Monotone Convergence again, now on $(X, \mathfrak{M}, \mu)$, gives the first equality in (17.2). The second one is similar.

Proof of (b). Due to (a), we have

$$\int_{X \times Y} |f| \, d(\mu \times \lambda) = \int_X \left(\int_Y |f|_x \, d\lambda(y) \right) d\mu(x) = \int_X \varphi^* \, d\mu < \infty$$

This implies (b).

Proof of (c). It is enough to prove (c) for *real-valued* f, then the complex case follows directly due to Definition 6.5 and (6.1).

If $f \in L^1_{\mu \times \lambda}$ is a real-valued function, then $f = f^+ - f^-$, where f^+ and f^- are nonnegative *integrable* functions (Lemma 6.2). Let $\varphi^\pm$ be associated with $f^\pm$ in the same way in which φ is associated with f. Then (a) implies

$$\int_X \varphi^\pm \, d\mu = \int_{X \times Y} f^\pm \, d(\mu \times \lambda) < \infty$$

hence $\varphi^{\pm} \in L^1_\mu$. In particular, $\varphi^{\pm}(x) < \infty$ for a.e. $x \in X$. This can be rephrased as $(f^{\pm})_x \in L^1_\lambda$ for a.e. $x \in X$. Therefore

$$f_x = (f^+)_x - (f^-)_x \in L^1_\lambda$$

for a.e. $x \in X$. This is the first claim in (c). The second one is similar.

Next, for a.e. $x \in X$ the above conclusions hold and by the linearity of the Lebesgue integrals we have

$$\varphi(x) = \int_Y f_x \, d\mu = \int_Y (f^+)_x \, d\mu - \int_Y (f^-)_x \, d\mu = \varphi^+(x) - \varphi^-(x)$$

thus $\varphi \in L^1_\mu$. This is the third claim in (c). The fourth one is similar.

Lastly, we can apply (17.1) to f^+ and f^- separately, based on the part (a) which is already proven. This gives

$$\begin{aligned}\int_X \varphi \, d\mu &= \int_X \varphi^+ \, d\mu - \int_X \varphi^- \, d\mu \\ &= \int_{X\times Y} f^+ \, d(\mu \times \lambda) - \int_{X\times Y} f^- \, d(\mu \times \lambda) = \int_{X\times Y} f \, d(\mu \times \lambda)\end{aligned}$$

This is the first equation in (17.1). The second one is similar. □

It is instructive to see a few counterexamples that show that the assumptions of Fubini's theorem cannot be dispensed with.

EXAMPLE 26. Let $X = Y = \mathbb{N}$ and $\mu = \lambda$ be the counting measure. Then $\mu \times \lambda$ is the counting measure on $X \times Y$ (check this!). Define a function $f(x, y)$ on $X \times Y$ as follows: $f(i, j) = a_{ij}$ for $i, j \geq 1$, where

$$a_{ij} = \begin{cases} 1 & \text{if } i = j \\ -1 & \text{if } i = j + 1 \\ 0 & \text{otherwise} \end{cases}$$

The values of f, presented in a matrix form, are

$$\begin{bmatrix} 1 & 0 & \ddots & & \\ -1 & 1 & 0 & \ddots & \\ 0 & -1 & 1 & 0 & \ddots \\ \ddots & 0 & -1 & 1 & \ddots \\ & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

Then we have

$$\int_X \left(\int_Y f(x, y)\, d\lambda(y) \right) d\mu(x) = \sum_i \sum_j a_{ij} = 1$$

On the other hand,

$$\int_Y \left(\int_X f(x, y)\, d\mu(x) \right) d\lambda(y) = \sum_j \sum_i a_{ij} = 0.$$

Incidentally, we found a sequence $\{a_{ij}\}$ of real numbers such that

$$\sum_i \sum_j a_{ij} \neq \sum_j \sum_i a_{ij}$$

The reason why Fubini's theorem does not apply here is that $f \notin L^1_{\mu \times \lambda}$.

See a Lebesgue-measure version of this example in Rudin, page 166.

EXAMPLE 27. Let $X = Y = [0,1]$. Let μ be the Lebesgue measure on $[0,1]$ and λ be the counting measure on $[0,1]$. Define a function $f(x,y)$ on $X \times Y$ (the unit square) as follows:

$$f(x,y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$$

In other words, f is the characteristic function of the diagonal. Then we have

$$\int_X \left(\int_Y f(x,y)\, d\lambda(y) \right) d\mu(x) = 1$$

because the inner integral is equal to 1 for every $x \in X$.
On the other hand,

$$\int_Y \left(\int_X f(x,y)\, d\mu(x) \right) d\lambda(y) = 0$$

because the inner integral is equal to 0 for every $y \in Y$.
The reason why Fubini's theorem does not apply is that λ is *not* a σ-finite measure.

- Note that in the above example the function f is $\mathfrak{M} \times \mathfrak{N}$ measurable. Indeed, the diagonal $D = \{x = y\}$ can be obtained as

$$D = \cap_{n \geq 1} Q_n$$

  where Q_n is the union of small squares placed along the diagonal:

$$Q_n = \bigcup_{i=1}^{n} \left[\tfrac{i-1}{n}, \tfrac{i}{n} \right] \times \left[\tfrac{i-1}{n}, \tfrac{i}{n} \right]$$

- In the above two examples, Fubini's theorem failed because either the function f or one of the measures μ and λ were "too big". A more sophisticated example can be constructed where the measures and the function are "small". More precisely, in that example $\mu(X) = \lambda(Y) = 1$ and f only takes values 0 and 1. In that example the iterated integrals in (17.2) exist but are *different*. The reason why Fubini's theorem fails is that f happens to be *not measurable* with respect to the product σ-algebra $\mathfrak{M} \times \mathfrak{N}$. But this example is too complicated to discuss in class; see Rudin, page 167.

Exercise 84. Let $a_n \geq 0$ for $n = 1, 2, \ldots$, and for $t \geq 0$ let

$$N(t) = \#\{n : a_n > t\}$$

Prove that

$$\sum_{n=1}^{\infty} a_n = \int_0^{\infty} N(t)\, dt$$

Exercise 85. [Bonus] Generalize the previous exercise as follows. Let $\phi\colon [0, \infty) \to [0, \infty)$ be an increasing locally absolutely continuous function (the latter means that ϕ is AC on every finite interval) such that $\phi(0) = 0$. Find a formula for $\sum_{n=1}^{\infty} \phi(a_n)$ in terms of $N(t)$.

Exercise 86. Let $f \in L^1([0,1])$ and $f \geq 0$. Show that

$$\int_0^1 \frac{f(y)}{|x-y|^{1/2}}\, dy$$

is finite for a.e. $x \in [0,1]$ and, as a function of x, integrable with respect to the Lebesgue measure on $[0,1]$.

Exercise 87. Use Fubini's theorem and the relation

$$\frac{1}{x} = \int_0^{\infty} e^{-xt}\, dt \quad \text{for} \quad x > 0$$

to prove that

$$\lim_{A \to \infty} \int_0^A \frac{\sin x}{x}\, dx = \frac{\pi}{2}$$

Completeness of the product measure. The measure $\mu \times \lambda$ is not necessarily complete. In fact, in most cases it is not complete.

Indeed, suppose that there is a non-empty null set $A \subset X$ and a non-measurable set $B \subset Y$. Then a simple application of Theorem 16.6 shows that $A \times B$ is *not* measurable. On the other hand, $A \times B$ is a subset of a null set, $A \times Y$.

A particularly interesting example is $X = Y = \mathbb{R}$ and $\mu = \lambda = \mathbf{m}$ the Lebesgue measure. By the above argument the product measure $\mathbf{m} \times \mathbf{m}$ on $\mathbb{R}^2$ is *not* complete, thus it is *not* the Lebesgue measure on $\mathbb{R}^2$. It turns out, however, that the *completion* of $\mathbf{m} \times \mathbf{m}$ *is* the Lebesgue measure on $\mathbb{R}^2$, see below.

Theorem 17.3. *Let $\mathbf{m}_k$ denote the Lebesgue measure on $\mathbb{R}^k$. Then the completion of $\mathbf{m}_s \times \mathbf{m}_t$ is the Lebesgue measure $\mathbf{m}_{s+t}$.*

Proof. Let $\mathfrak{B}_k$ and $\mathfrak{M}_k$ denote the Borel and Lebesgue σ-algebras in $\mathbb{R}^k$, respectively. Note that $\mathfrak{B}_k \subset \mathfrak{M}_k$. The theorem is proved in five steps.

Step 1: $\mathfrak{B}_{s+t} \subset \mathfrak{B}_s \times \mathfrak{B}_t \big(\subset \mathfrak{M}_s \times \mathfrak{M}_t\big)$. Indeed, $\mathfrak{B}_{s+t}$ is generated by open sets in $\mathbb{R}^{s+t}$, in particular by open rectangular boxes, each of which is in $\mathfrak{B}_s \times \mathfrak{B}_t$.

Step 2: $\mathbf{m}_{s+t}$ and $\mathbf{m}_s \times \mathbf{m}_t$ coincide on $\mathfrak{B}_{s+t}$. Indeed, these two measures agree on rectangles, then we can apply either Theorem 3.17 or Corollary 3.19.

Step 3: $\mathfrak{B}_{s+t} = \mathfrak{B}_s \times \mathfrak{B}_t$. The inclusion "$\subset$" was proved in Step 1. For the other inclusion "$\supset$" it is enough to show that

$$A \in \mathfrak{B}_s,\ B \in \mathfrak{B}_t \quad \Rightarrow \quad A \times B \in \mathfrak{B}_{s+t} \tag{17.3}$$

because the sets $A \times B$ generate $\mathfrak{B}_s \times \mathfrak{B}_t$. So we need to prove (17.3). If $A \in \mathfrak{B}_s$ and $B \in \mathfrak{B}_t$ are *open*, then $A \times B$ is open, hence it is in $\mathfrak{B}_{s+t}$. Now fix an open se $B \in \mathfrak{B}_t$ and consider

$$\mathfrak{G}_s = \{A \subset \mathbb{R}^s \colon A \times B \in \mathfrak{B}_{s+t}\}$$

It is a σ-algebra in $\mathbb{R}^s$ containing open sets, hence $\mathfrak{G}_s \supset \mathfrak{B}_s$. This means $A \times B \in \mathfrak{B}_{s+t}$ provided $A \in \mathfrak{B}_s$ and B is open. Next fix $A \in \mathfrak{B}_s$ and consider

$$\mathfrak{G}'_t = \{B \subset \mathbb{R}^t \colon A \times B \in \mathfrak{B}_{s+t}\}$$

It is a σ-algebra in $\mathbb{R}^t$ containing open sets, hence $\mathfrak{G}'_t \supset \mathfrak{B}_t$. This means $A \times B \in \mathfrak{B}_{s+t}$ provided $A \in \mathfrak{B}_s$ and $B \in \mathfrak{B}_t$.

Step 4: $\mathfrak{M}_s \times \mathfrak{M}_t \subset \mathfrak{M}_{s+t}$. Let $E \in \mathfrak{M}_s$ and $F \in \mathfrak{M}_t$. By Theorem 8.4, there are $A, B \in \mathfrak{B}_s$ such that $B \subset E \subset A$ and $\mathbf{m}_s(A \setminus B) = 0$. Then

$$B \times \mathbb{R}^t \subset E \times \mathbb{R}^t \subset A \times \mathbb{R}^t$$

On the other hand, due to Steps 2 and 3

$$\begin{aligned}\mathbf{m}_{s+t}((A \times \mathbb{R}^t) \setminus (B \times \mathbb{R}^t)) &= \mathbf{m}_{s+t}((A \setminus B) \times \mathbb{R}^t) = (\mathbf{m}_s \times \mathbf{m}_t)((A \setminus B) \times \mathbb{R}^t)\\ &= \mathbf{m}_s(A \setminus B) \cdot \mathbf{m}_t(\mathbb{R}^t) = 0 \cdot \infty = 0\end{aligned}$$

Since $\mathbf{m}_{s+t}$ is complete, we have $E \times \mathbb{R}^t \in \mathfrak{M}_{s+t}$. Similarly, $\mathbb{R}^s \times F \in \mathfrak{M}_{s+t}$. Thus

$$E \times F = (E \times \mathbb{R}^t) \cap (\mathbb{R}^s \times F) \in \mathfrak{M}_{s+t}$$

Since these sets $E \times F$ generate $\mathfrak{M}_s \times \mathfrak{M}_t$, Step 4 is proved.

Step 5: $\mathbf{m}_{s+t}$ and $\mathbf{m}_s \times \mathbf{m}_t$ coincide on $\mathfrak{M}_s \times \mathfrak{M}_t$. Indeed, let $Q \in \mathfrak{M}_s \times \mathfrak{M}_t \subset \mathfrak{M}_{s+t}$ (due to Step 4). By Theorem 8.4, there are $A, B \in \mathfrak{B}_{s+t} \subset \mathfrak{M}_s \times \mathfrak{M}_t$ (due to Step 1) such that $B \subset Q \subset A$ and $\mathbf{m}_{s+t}(A \setminus B) = 0$. Therefore $(\mathbf{m}_s \times \mathbf{m}_t)(A \setminus B) = 0$ due to Step 2. As a result, $(\mathbf{m}_s \times \mathbf{m}_t)(Q \setminus B) = 0$ thus

$$(\mathbf{m}_s \times \mathbf{m}_t)(Q) = (\mathbf{m}_s \times \mathbf{m}_t)(B) \overset{\text{Step 2}}{=} \mathbf{m}_{s+t}(B) = \mathbf{m}_{s+t}(Q)$$

This completes the proof of Step 5.

$\square$

Next we look into measurable functions on complete measure spaces.

Lemma 17.4. *Let $(X, \mathfrak{M}, \mu)$ a measure space and $(X, \overline{\mathfrak{M}}, \overline{\mu})$ be its completion. If f is an $\overline{\mathfrak{M}}$-measurable function, then there exists an $\mathfrak{M}$-measurable function g such that $f = g$ a.e. with respect to μ, i.e., $\{f \neq g\} \subset N$ with $\mu(N) = 0$.*

Proof. It suffices to do this for non-negative functions $f \geq 0$. In that case, due to Theorem 4.22, there are simple $\overline{\mathfrak{M}}$-measurable functions

$$0 = s_0 \leq s_1 \leq s_2 \leq \cdots \qquad \text{and} \qquad s_n \to f$$

Therefore

$$f = \sum_{n=1}^{\infty} (s_n - s_{n-1}) = \sum_{i=1}^{\infty} c_i \chi_{E_i}$$

for some $c_i \geq 0$ and $E_i \in \overline{\mathfrak{M}}$. Due to Theorem 3.22 there are sets $A_i, N_i \in \mathfrak{M}$ such that $A_i \subset E_i \subset A_i \cup N_i$ and $\mu(N_i) = 0$. Now we define $g = \sum_{i=1}^{\infty} c_i \chi_{A_i}$ and complete the proof because $\{f \neq g\} \subset \cup_{i=1}^{\infty} N_i$, which is a null set. □

Lemma 17.5. *Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \lambda)$ be two measure spaces with* complete *σ-finite positive measures. Let h be an $\overline{(\mathfrak{M} \times \mathfrak{N})}$-measurable function on $X \times Y$ such that $h = 0$ a.e. with respect to $\overline{\mu \times \lambda}$. Then for μ-almost every $x \in X$ we have $h_x(y) = 0$ a.e. on Y (in particular, h_x is $\mathfrak{N}$-measurable for μ-almost every $x \in X$). A similar statement holds for h^y.*

Proof. Denote

$$P = \{(x, y) \in X \times Y : h(x, y) \neq 0\}$$

It is assumed that $(\overline{\mu \times \lambda})(P) = 0$. Thus there exists $Q \in \mathfrak{M} \times \mathfrak{N}$ such that $P \subset Q$ and $(\mu \times \lambda)(Q) = 0$. By Theorem 16.13

$$\int_X \lambda(Q_x)\, d\mu = (\mu \times \lambda)(Q) = 0 \tag{17.4}$$

Now let $N = \{x \in X : \lambda(Q_x) > 0\}$. It follows from (17.4) that $\mu(N) = 0$. For every $x \notin N$ we have the following:

(i) $\lambda(Q_x) = 0$

(ii) Since $P_x \subset Q_x$ and λ is a complete measure, we have that $\lambda(P_x) = 0$, too. Moreover, all subsets of P_x belong in $\mathfrak{N}$

(iii) As a result, h_x is $\mathfrak{N}$-measurable and $h_x = 0$ a.e. on Y

This completes the proof of the lemma. □

The following is a version of the Fubini theorem for complete measure spaces. In particular, it applies to the Lebesgue measures on $\mathbb{R}^k$.

Theorem 17.6. *Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \lambda)$ be two measure spaces with* complete *σ-finite positive measures. Let f be an $(\overline{\mathfrak{M} \times \mathfrak{N}})$-measurable function on $X \times Y$. Then*

(a) *f_x is $\mathfrak{N}$-measurable for* almost every $x \in X$

(b) *f^y is $\mathfrak{M}$-measurable for* almost every $y \in Y$

All the other conclusions of the standard Fubini theorem hold.

Proof. Due to Lemma 17.4, we have

$$f = g + h$$

where

(i) g is $(\mathfrak{M} \times \mathfrak{N})$-measurable

(ii) $h = 0$ a.e. with respect to $\overline{\mu \times \lambda}$

Note that $f_x = g_x + h_x$ and $f^y = g^y + h^y$.

Now the standard Fubini Theorem 17.1 applies to g. And Lemma 17.5 shows that $f_x = g_x$ a.e. on Y for μ-almost all $x \in X$. Similarly, $f^y = g^y$ a.e. on X for λ-almost all $y \in Y$. $\square$

Exercise 88. Let E be a Lebesgue measurable subset of $\mathbb{R}^2$. Suppose that for a.e. $x \in \mathbb{R}$ the set $E_x = \{y \in \mathbb{R} : (x, y) \in E\}$ has Lebesgue measure zero. Prove that for a.e. $y \in \mathbb{R}$ the set $E^y = \{x \in \mathbb{R} : (x, y) \in E\}$ has Lebesgue measure zero. Compute $\mathbf{m}_2(E)$.

18 Convolution and layer-cake integration

DEFINITION 18.1. Let $f, g \in L^1_{\mathbf{m}}(\mathbb{R})$ be two Lebesgue integrable functions. Then their **convolution** $h = f * g$ is defined by

$$h(x) = \int_{\mathbb{R}} f(x-t)\, g(t)\, d\mathbf{m}(t) \tag{18.1}$$

(whenever the integral exists).

Lemma 18.2. *We have $f * g = g * f$, i.e., convolution is a symmetric operation.*

Proof. Applying the change of variable $\varphi(t) = x - t$ and Corollary 15.9 gives

$$\int_{\mathbb{R}} f(x-t)\, g(t)\, d\mathbf{m}(t) = \int_{\mathbb{R}} f(t)\, g(x-t)\, d\mathbf{m}(t)$$

(wherever one integral exists, so does the other, and they are equal). $\square$

- For every $x \in \mathbb{R}$, the function $f(x-t)$ can be represented as $f \circ \varphi$, where $\varphi(t) = x - t$. By the change-of-variable rule (Corollary 15.9) we have $f(x-t) \in L^1_{\mathbf{m}}(\mathbb{R})$ and $\int_{\mathbb{R}} f(x-t)\, d\mathbf{m}(t) = \int_{\mathbb{R}} f\, d\mathbf{m}$. Thus in (18.1) we integrate the product of two integrable functions.
- It is not always true that if $f, g \in L^1_{\mathbf{m}}(\mathbb{R})$ then $fg \in L^1(\mathbb{R})$. For example, consider $f = g = x^{-1/2}\chi_{[0,1]}$, then $f, g \in L^1(\mathbb{R})$ but $fg \notin L^1(\mathbb{R})$. Thus the existence of the integral in (18.1) is not guaranteed for any $x \in \mathbb{R}$; actually this fact is far from obvious.

Theorem 18.3. *Let $f, g \in L^1(\mathbb{R})$. Then*

$$\int_{\mathbb{R}} |f(x-t)\, g(t)|\, d\mathbf{m}(t) < \infty$$

for almost every $x \in \mathbb{R}$, thus $h(x)$ in (18.1) exists for almost every $x \in \mathbb{R}$. Moreover, $h \in L^1(\mathbb{R})$ and we have the following ***Young inequality****:*

$$\|h\|_1 \leq \|f\|_1 \|g\|_1 \tag{18.2}$$

Proof. Due to Lemma 17.4, there exist Borel functions f_0 and g_0 such that $f_0 = f$ a.e. and $g_0 = g$ a.e. The integral (18.1) is unchanged if we replace f by f_0 and g by g_0. So we may assume, to begin with, that f and g are Borel functions.

Define a function F on $\mathbb{R}^2$ by

$$F(x,y) = f(x-y)\, g(y)$$

It is a Borel function because $F = (f \circ \varphi) \cdot (g \circ \psi)$, where $\varphi(x,y) = x - y$ and $\psi(x,y) = y$ are obviously Borel functions.

Note that

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |F(x,y)|\, d\mathbf{m}(x) \right) d\mathbf{m}(y) = \int_{\mathbb{R}} |g(y)| \left(\int_{\mathbb{R}} |f(x-y)|\, d\mathbf{m}(x) \right) d\mathbf{m}(y)$$

The inner integral is

$$\int_{\mathbb{R}} |f(x-y)|\, d\mathbf{m}(x) = \int_{\mathbb{R}} |f|\, d\mathbf{m} = \|f\|_1$$

due to the translation invariance of the Lebesgue measure. Therefore

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |F(x,y)|\, d\mathbf{m}(x) \right) d\mathbf{m}(y) = \|f\|_1\, \|g\|_1 < \infty \tag{18.3}$$

Thus by Fubini Theorem 17.1 (b) we have $F \in L^1(\mathbb{R}^2)$ and the clause (c) of the same theorem implies that $h(x)$ exists for almost every $x \in \mathbb{R}$ and $h \in L^1(\mathbb{R})$. Finally,

$$\begin{aligned} \|h\|_1 = \int_{\mathbb{R}} |h|\, d\mathbf{m} &\leq \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x-y)\, g(y)|\, d\mathbf{m}(y) \right) d\mathbf{m}(x) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x-y)\, g(y)|\, d\mathbf{m}(x) \right) d\mathbf{m}(y) = \|f\|_1\, \|g\|_1 \end{aligned}$$

due to (18.3). □

- If we assume that $f, g \geq 0$, then $\|h\|_1 = \|f\|_1 \|g\|_1$.

EXERCISE 89. Suppose $f \in L^1(\mathbb{R})$ and $g \in L^p(\mathbb{R})$ for some $1 \le p \le \infty$. Show that $f * g$ exists at a.e. $x \in \mathbb{R}$ and $f * g \in L^p(\mathbb{R})$, and prove that

$$\|f * g\|_p \le \|f\|_1 \|g\|_p.$$

Hints: the case $p = \infty$ is simple and can be treated separately. If $p < \infty$, then use Hölder inequality and argue as in the proof of the previous theorem.

EXERCISE 90. Let $f \in L^1(\mathbb{R})$ and

$$g(x) = \int_{\mathbb{R}} f(y)\, e^{-(x-y)^2}\, d\mathbf{m}(y).$$

Show that $g \in L^p(\mathbb{R})$, for all $1 \le p \le \infty$, and estimate $\|g\|_p$ in terms of $\|f\|_1$. You can use the following standard fact: $\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$.

EXERCISE 91. [Bonus] Let $E = [1, \infty)$ and $f \in L^2_{\mathbf{m}}(E)$. Also assume that $f \ge 0$ a.e. and define

$$g(x) = \int_E f(y)\, e^{-xy}\, d\mathbf{m}(y).$$

Show that $g \in L^1(E)$ and

$$\|g\|_1 \le c\, \|f\|_2$$

for some $c < 1$. Estimate the minimal value of c the best you can.

Definition 18.4. Let $(X, \mathfrak{M}, \mu)$ be a measure space with a σ-finite positive measure μ. Let $f\colon X \to [0, \infty]$ be a measurable function. Then

$$g(t) = \mu\{f > t\} = \mu(\{x \in X\colon f(x) > t\})$$

is called the **distribution function** of f.

- In probability theory, where $\mu(X) = 1$, the distribution function is defined slightly differently, as $F(t) = \mu\{f \le t\}$. Thus $F(t) = 1 - g(t)$.
- The distribution function $g(t)$ is monotonically decreasing (though not necessarily strictly). Thus it is a Borel measurable function.

Theorem 18.5. *Let f and μ be as above. Suppose $\varphi\colon [0, \infty] \to [0, \infty]$ is a monotonic increasing function that is absolutely continuous on every finite interval $[0, T]$ and satisfies $\varphi(0) = 0$ and $\lim_{t\to\infty} \varphi(t) = \varphi(\infty)$. Then*

$$\int_X (\varphi \circ f)\, d\mu = \int_0^\infty \mu\{f > t\}\varphi'(t)\, dt. \tag{18.4}$$

- Special case: if $\varphi(t) = t$, we obtain a useful formula

$$\int_X f\, d\mu = \int_0^\infty \mu\{f > t\}\, dt,$$

which is sometimes given as a definition of the Lebesgue integral. This formula is known as **layer-cake integration**.

Figure 9: Layer Cake integration.

Proof. (of Theorem 18.5) Let

$$E = \{(x,t) \in X \times [0,\infty] \colon f(x) > t\}$$

denote the region "under the graph" of $f(x)$.

First we check that E is $\mu \times \mathbf{m}$ measurable. Indeed, if f is simple, then E is just a finite union of measurable rectangles. In general, we can approximate f by a sequence of simple functions (recall Theorem 4.22)

$$0 \le s_1 \le s_2 \le \cdots \le f, \qquad s_n \to f$$

Now each s_n corresponds to a measurable set E_n and we have $E = \cup_{n\ge 1} E_n$, thus E is measurable.

Now let $E^t = \{x \in X \colon (x,t) \in E\}$ denote the t-section of E. Then

$$g(t) = \mu\{f > t\} = \mu(E^t) = \int_X \chi_{E^t}\, d\mu = \int_X \chi_E(x,t)\, d\mu(x)$$

The right hand side of (18.4) is therefore

$$\int_0^\infty \mu(E^t)\varphi'(t)\, dt = \int_0^\infty \varphi'(t)\left(\int_X \chi_E(x,t)\, d\mu(x)\right) dt$$

Due to the Fubini theorem, part (a), we can interchange the order of integration:

$$\int_0^\infty \varphi'(t)\left(\int_X \chi_E(x,t)\, d\mu(x)\right) dt = \int_X \left(\int_0^\infty \chi_E(x,t)\varphi'(t)\, dt\right) d\mu(x)$$

Now $\chi_E(x,t) = \chi_{[0,f(x))}(t)$ and by Theorem 14.5

$$\int_0^\infty \chi_{[0,f(x))}(t)\varphi'(t)\, dt = \int_0^{f(x)} \varphi'(t)\, dt = \varphi(t) - \varphi(0) = \varphi(f(x))$$

Therefore the right hand side of (18.4) becomes $\int_X (\varphi \circ f)\, d\mu$, as desired. □

Exercise 92. Let $f \colon \mathbb{R} \to \mathbb{R}$ be Lebesgue measurable.

(a) Prove that the set
$$A = \{(x,y) \in \mathbb{R}^2 \mid y < f(x)\}$$
is Lebesgue measurable (in the two-dimensional sense)

(b) Let $f \ge 0$. Is it always true that $\int_{\mathbb{R}} f\, d\mathbf{m}$ equals the Lebesgue measure of A?

(c) Prove that $\{(x,y) \in \mathbb{R}^2 \mid y = f(x)\}$ is a null set.

Exercise 93. [Bonus] Let $f \colon \mathbb{R}^2 \to \mathbb{R}$ be such that f_x is Borel-measurable for every $x \in \mathbb{R}$ and f^y is continuous for every $y \in \mathbb{R}$. Prove that f is Borel-measurable. (See hint on p. 176 in Rudin's book.)

Definition 18.6. For every Lebesgue measurable function $f\colon \mathbb{R}^k \to \mathbb{C}$ define the **maximal function** $Mf\colon \mathbb{R}^k \to [0,\infty]$ by

$$(Mf)(x) = \sup_{0<r<\infty} \frac{1}{\mathbf{m}(B(x,r))} \int_{B(x,r)} |f|\, d\mathbf{m}$$

whenever the integral exists.

- Earlier we defined this function only for $f \in L^1(\mathbb{R}^k)$, and in that case we proved $Mf \in L^1_W(\mathbb{R}^k)$.

Theorem 18.7. *If $f \in L^1(\mathbb{R}^k)$ and $Mf \in L^1(\mathbb{R}^k)$, then $f = 0$ a.e.*

Proof. Indeed, if $\int_{\mathbb{R}^k} |f|\, d\mathbf{m} > 0$, then there is bounded region $E \subset \mathbb{R}^k$ such that $\int_E |f|\, d\mathbf{m} > 0$. Now

$$(Mf)(x) \geq \sup_{0<r<\infty} \frac{1}{\mathbf{m}(B(x,r))} \int_{E\cap B(x,r)} |f|\, d\mathbf{m}$$

and this gives us $(Mf)(x) \geq c|x|^{-k}$ for some $c > 0$ and all sufficiently large $|x|$. It is a calculus exercise to see that the function $|x|^{-k}$ has infinite Riemann integral over the set $\chi_{\mathbb{R}^k\setminus B(0,r)}$ for any $r > 0$. Thus, for any bounded region $E \subset \mathbb{R}^k$, the function $|x|^{-k}\chi_{\mathbb{R}^k\setminus E}$ is not integrable on $\mathbb{R}^k$. $\square$

- The situation in the L^p spaces with $p > 1$ is different:

Theorem 18.8. [Hardy–Littlewood]

Let $1 < p \leq \infty$. If $f \in L^p(\mathbb{R}^k)$, then $Mf \in L^p(\mathbb{R}^k)$.

Can be given without proof. The proof is in Rudin's book.