Graduate Texts in Mathematics

GTM

Christopher Heil

Introduction to Real Analysis

Springer

Graduate Texts in Mathematics 280

Graduate Texts in Mathematics

Graduate Texts in Mathematics bridge the gap between passive study and creative understanding, offering graduate-level introductions to advanced topics in mathematics. The volumes are carefully written as teaching aids and highlight characteristic features of the theory. Although these books are frequently used as textbooks in graduate courses, they are also suitable for individual study.

More information about this series at http://www.springer.com/series/136

Christopher Heil

Introduction to Real Analysis

Springer

Christopher Heil
School of Mathematics
Georgia Institute of Technology
Atlanta, GA, USA

ISSN 0072-5285 ISSN 2197-5612 (electronic)
Graduate Texts in Mathematics
ISBN 978-3-030-26901-2 ISBN 978-3-030-26903-6 (eBook)
https://doi.org/10.1007/978-3-030-26903-6

Mathematics Subject Classification (2010): 28-XX, 26-XX, 42-XX, 46-XX, 47-XX, 54-XX

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

For Alex, Andrew, and Lea

Contents

Preface

This text grew out of lecture notes that I developed over the years for the "Real Analysis" graduate sequence here at Georgia Tech. This two-semester sequence is taken by first-year mathematics graduate students, well-prepared undergraduate mathematics majors, and graduate students from a wide variety of engineering and scientific disciplines. Covered in this book are the topics that are taught in the first semester: Lebesgue measure, the Lebesgue integral, differentiation and absolute continuity, the Lebesgue spaces $L^p(E)$, and Hilbert spaces and $L^2(E)$. This material not only forms the basis of a core subject in pure mathematics, but also has wide applicability in science and engineering. A text covering the second semester topics in analysis, including abstract measure theory, signed and complex measures, operator theory, and functional analysis, is in development.

This text is an introduction to real analysis. There are several classic analysis texts that I keep close by on my bookshelf and refer to often. However, I find it difficult to use any of these as the textbook for teaching a first course on analysis. They tend to be dense and, in the classic style of mathematical elegance and conciseness, they develop the theory in the most general setting, with few examples and limited motivation. These texts are valuable resources, but I suggest that they should be the *second* set of books on analysis that you pick up.

I hope that this text will be the analysis text that you read first. The definitions, theorems, and other results are motivated and explained; the why and not just the what of the subject is discussed. Proofs are completely rigorous, yet difficult arguments are motivated and discussed. Extensive exercises and problems complement the presentation in the text, and provide many opportunities for enhancing the student's understanding of the material.

Audience

This text is aimed at students who have taken a standard (proof-based) undergraduate mathematics course on the basics of analysis. A brief review of the needed background material is presented in the **Preliminaries** section of the text. This includes:

- sequences, series, limits, suprema and infima, and limsups and liminfs,
- functions,
- cardinality,
- basic topology of Euclidean space (open, closed, and compact sets),
- continuity and differentiability of real-valued functions,
- the Riemann integral.

Online Resources

A variety of resources are available on the author's website,

`http://people.math.gatech.edu/~heil/`

These include the following.

- A **Chapter 0**, which contains a greatly expanded version of the material that appears in the **Preliminaries** section of this text, along with discussions and exercises.

- An **Alternative Chapter 1**, which is an expanded version of the material presented in **Chapter 1**, including detailed discussion, motivation, and exercises, focused on the setting of normed spaces.

- A **Chapter 10**, which provides an introduction to abstract measure theory.

- An **Instructor's Guide**, with a detailed course outline, commentary, remarks, and extra problems. The exposition and problems in this guide may be useful for students and readers as well as instructors.

- **Selected Solutions for Students**, containing approximately one worked solution of a problem or exercise from each section of the text.

- An **Errata List** that will be updated as I become aware of typographical or other errors in the text.

Additionally, a **Solutions Manual** is available to instructors upon request; instructions for obtaining a copy are given on the Birkhäuser website for this text.

Outline

Chapter 1 presents a short review of metric and normed spaces. Students who have completed an undergraduate analysis course have likely encountered much of this material, although possibly only in the context of the Euclidean space $\mathbb{R}^d$ (or $\mathbb{C}^d$) instead of abstract metric spaces. The instructor has the option of beginning the course here or proceeding directly to **Chapter 2.** The online **Alternative Chapter 1** presents a significantly expanded version of this chapter focused on normed spaces. (A detailed introduction to the more general setting of metric spaces is available in the first chapters of the author's text *Metrics, Norms, Inner Products, and Operator Theory* [Heil18].)

In **Chapter 2** we begin the study of Lebesgue measure. The fundamental question that motivates this chapter is: Can we assign a "volume" or "measure" to every subset of $\mathbb{R}^d$ in such a way that all of the properties that we expect of a "volume" function are satisfied? For example, we want the measure of a cube or a ball in $\mathbb{R}^d$ to coincide with the standard definition of the volume of a cube or ball, and if we translate an object rigidly in space then we want its measure to always remain the same. If we break an object into countably many disjoint pieces, then we want the measure of the original object to be the sum of the measures of the pieces. Surprisingly (at least to me!), this simply can't be done (more precisely, the Axiom of Choice implies that it is impossible). However, if we relax this goal somewhat then we find that we can define a measure that obeys the correct rules for a "large" class of sets (the *Lebesgue measurable sets*). **Chapter 2** constructs and studies this measure, which we call the *Lebesgue measure* of subsets of $\mathbb{R}^d$.

In **Chapters 3 and 4** we define the integral of real-valued and complex-valued functions whose domain is a measurable subset of $\mathbb{R}^d$. Unfortunately, we cannot define the *Lebesgue integral* of every function. **Chapter 3** introduces the class of *measurable functions* and deals with issues related to convergence of sequences of measurable functions, while **Chapter 4** defines and studies the Lebesgue integral of a measurable function. The Lebesgue integral extends the Riemann integral, but is far more general. We can define the Lebesgue integral for functions whose domain is any measurable set. We prove powerful results that allow us, in a large family of cases, to make conclusions about the convergence of a sequence of Lebesgue integrals, or to interchange the order of iterated integrals of functions of more than one variable.

The Fundamental Theorem of Calculus (FTC) is, as its name suggests, central to analysis. **Chapters 5 and 6** explore issues related to differentiation and the FTC in detail. We see that there are surprising examples of *nonconstant* functions whose derivatives are zero at "almost every" point (and therefore fail the FTC). In our quest to fully understand the FTC we define functions of *bounded variation* and study averaging operations in **Chapter 5**. Then in **Chapter 6** we introduce the class of *absolutely continuous functions*, which turn out to be the functions for which the FTC holds. The

Banach–Zaretsky Theorem plays a prominent role in **Chapter 6**, and it is central to our understanding of absolute continuity and its impact.

In **Chapter 7** our focus turns from individual functions to spaces of functions. The *Lebesgue spaces* $L^p(E)$ group functions by integrability properties, giving us a family of spaces indexed by an extended real number p with $0 < p \leq \infty$. For $p \geq 1$ these are normed vector spaces of functions, while for $0 < p < 1$ they are metric spaces whose metric is not induced from a norm. The case $p = 2$ is especially important, because we can define an *inner product* on $L^2(E)$, which makes it a *Hilbert space.* This topic is explored in **Chapter 8**. In a metric space, all that we can do is define the *distance* between points in the space. In a normed space we can additionally define the *length* of each vector in the space. But in a Hilbert space, we furthermore have a notion of *angles between vectors* and hence can define orthogonality. This leads to many powerful results, including the existence of an *orthonormal basis* for every separable Hilbert space. Even though a Hilbert space can be infinite-dimensional, in many respects our intuitions from Euclidean space hold when we deal with a Hilbert space.

Chapter 9 contains "extra" material that is usually not covered in our real analysis sequence here at Georgia Tech, but which has many striking applications of the techniques developed in the earlier chapters. First we define the operation of *convolution.* Then we introduce and study the *Fourier transform* and *Fourier series.* These results form the core of the field of *harmonic analysis*, which has wide applicability throughout mathematics, physics, and engineering. Convolution is a generalization of the averaging operations that were used in **Chapters 5** and **6** to characterize the class of functions for which the Fundamental Theorem of Calculus holds. The Fourier transform and Fourier series allow us to both construct and deconstruct a wide class of functions, signals, or operators in terms of much simpler building blocks based on complex exponentials (or sines and cosines in the real case). Although **Chapter 9** presents only a taste of the theorems of harmonic analysis (which deserves another course, and a future text, to do it justice), we do get to see many applications of all of the tools that we derived in earlier chapters, including convergence of sequences of integrals (via the Dominated Convergence Theorem), interchange of iterated integrals (via Fubini's Theorem), and the Fundamental Theorem of Calculus (via the Banach–Zaretsky Theorem).

Many exercises and problems appear in each section of the text. The *Exercises* are directly incorporated into the development of the theory in each section, while the additional *Problems* given at the end of each section provide further practice and opportunities to develop understanding.

Course Options

There are many options for building a course around this text. The course that I teach at Georgia Tech is fast-paced, but covers most of the text in one semester. Here is a brief outline of such a one-semester course; a more detailed outline with much additional information (and extra problems) is contained in the **Instructor's Guide** that is available on the author's website.

Chapter 1: Assign for student reading, not covered in lecture.
Chapter 2: Sections 2.1–2.4.
Chapter 3: Sections 3.1–3.5. Omit Section 3.6.
Chapter 4: Sections 4.1–4.6.
Chapter 5: Sections 5.1–5.2, and selected portions of Sections 5.3–5.5.
Chapter 6: Sections 6.1–6.4. Omit Sections 6.5-6.6.
Chapter 7: Sections 7.1–7.4.
Chapter 8: Sections 8.1–8.4 (as time allows).
Chapter 9: Bonus material, not covered in lecture.

Another option is to begin the course with **Chapter 1** (or the online **Alternative Chapter 1**). A fast-paced course could cover most of **Chapters 1–8**. A moderately paced course could cover the first half of the text in detail in one semester, while a moderately paced two-semester course could cover all of **Chapters 1–9** in considerable detail.

Acknowledgments

Every text builds on those that have come before it, and this one is no exception. Many classic and recent volumes have influenced the writing, the choice of topics, the proofs, and the selection of problems. Among those that have had the most profound influence on my writing are Benedetto and Czaja [BC09], Bruckner, Bruckner, and Thomson [BBT97], Folland [Fol99], Rudin [Rud87], Stein and Shakarchi [SS05], and Wheeden and Zygmund [WZ77]. I greatly appreciate all of these texts and encourage the reader to consult them. Additional texts and papers are listed in the references.

Various versions of the material in this volume have been used over the years in the real analysis courses that were taught at Georgia Tech, and I thank all of the *many* students and colleagues who have provided feedback. Special thanks are due to Shahaf Nitzan, who taught the course out of earlier versions of the text and provided invaluable feedback.

Christopher Heil
Atlanta, Georgia
April 27, 2019

Preliminaries

We use the symbol $\square$ to denote the end of a proof, and the symbol $\diamondsuit$ to denote the end of a definition, remark, example, or exercise. We also use $\diamondsuit$ to indicate the end of the statement of a theorem whose proof will be omitted. A few problems are marked with an asterisk *; this indicates that they may be more challenging. A detailed index of symbols employed in the text can be found at the end of the volume.

Numbers

The set of natural numbers is denoted by $\mathbb{N} = \{1, 2, 3, \dots\}$. The set of integers is $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$, $\mathbb{Q}$ denotes the set of rational numbers, $\mathbb{R}$ is the set of real numbers, and $\mathbb{C}$ is the set of complex numbers. We often refer to $\mathbb{R}$ as the *real line*, and to $\mathbb{C}$ as the *complex plane*.

Complex Numbers. The *real part* of a complex number $z = a+ib$ (where $a, b \in \mathbb{R}$) is $\mathrm{Re}(z) = a$, and its *imaginary part* is $\mathrm{Im}(z) = b$. We say that z is *rational* if both its real and imaginary parts are rational numbers. The *complex conjugate* of z is $\overline{z} = a - ib$. The *modulus*, or *absolute value*, of z is

$$|z| \;=\; \sqrt{z\overline{z}} \;=\; \sqrt{a^2 + b^2}.$$

If $z \neq 0$ then its *polar form* is $z = re^{i\theta}$ where $r = |z| > 0$ and $\theta \in [0, 2\pi)$. In this case the *argument* of z is $\arg(z) = \theta$. Given any $z \in \mathbb{C}$, there is a complex number α such that $|\alpha| = 1$ and $\alpha z = |z|$. If $z \neq 0$ then α is uniquely given by $\alpha = e^{-i\theta} = \overline{z}/|z|$, while if $z = 0$ then α can be any complex number that has unit modulus.

Extended Real Numbers. The set of *extended real numbers* $[-\infty, \infty]$ is

$$[-\infty, \infty] \;=\; \mathbb{R} \cup \{-\infty, \infty\}.$$

We extend many of the normal arithmetic and order notations and operations to $[-\infty, \infty]$. For example, if $a \in [-\infty, \infty]$ then a is a real number if and only if $-\infty < a < \infty$. If $-\infty < a \le \infty$ then we set $a + \infty = \infty$. However, $\infty - \infty$ and $-\infty + \infty$ are undefined, and are referred to as *indeterminate forms.* If $0 < a \le \infty$, then we define

$$a \cdot \infty = \infty, \quad (-a) \cdot \infty = -\infty, \quad a \cdot (-\infty) = -\infty, \quad (-a) \cdot (-\infty) = \infty.$$

We also adopt the following conventions:

$$0 \cdot (\pm\infty) \;=\; 0 \qquad \text{and} \qquad \frac{1}{\pm\infty} \;=\; 0.$$

The Dual Index. Let p be an extended real number in the range $1 \le p \le \infty$. The *dual index* to p is the unique extended real number p' that satisfies

$$\frac{1}{p} + \frac{1}{p'} \;=\; 1.$$

We have $1 \le p' \le \infty$, and $(p')' = p$. If $1 < p < \infty$, then we can write p' explicitly as

$$p' \;=\; \frac{p}{p-1}.$$

Some examples are $1' = \infty$, $\left(\frac{3}{2}\right)' = 3$, $2' = 2$, $3' = \frac{3}{2}$, and $\infty' = 1$.

The Notation $\overline{\mathbf{F}}$. In order to deal simultaneously with the complex plane and the extended real line, we let the symbol $\overline{\mathbf{F}}$ denote a choice of either $[-\infty, \infty]$ or $\mathbb{C}$. Associated with this choice, we declare that:

- if $\overline{\mathbf{F}} = [-\infty, \infty]$, then the word *scalar* means a *finite real number* $c \in \mathbb{R}$;
- if $\overline{\mathbf{F}} = \mathbb{C}$, then the word *scalar* means a *complex number* $c \in \mathbb{C}$.

Note that a *scalar* cannot be $\pm\infty$; instead, a scalar is always a real or complex number.

Sets

The notation $x \in X$ means that x is an element of the set X. We often refer to an element of X as a *point in* X.

We write $A \subseteq B$ to denote that A is a subset of a set B. If $A \subseteq B$ and $A \neq B$ then we say that A is a *proper subset* of B, and we write $A \subsetneq B$.

The *empty set* is denoted by $\varnothing$.

A collection of sets $\{X_i\}_{i \in I}$ is *disjoint* if $X_i \cap X_j = \varnothing$ whenever $i \neq j$. The collection $\{X_i\}_{i \in I}$ is a *partition* of X if it is disjoint and $\cup_{i \in I} X_i = X$.

If X is a set, then the *complement* of $S \subseteq X$ is $X \backslash S = \bigl\{x \in X : x \notin S\bigr\}$. We sometimes abbreviate $X \backslash S$ as S^{C} if the set X is understood. If A and B

are subsets of X, then the *relative complement* of A in B is

$$B \backslash A \;=\; B \cap A^{\mathrm{C}} \;=\; \{x \in B : x \notin A\}.$$

The *power set* of X is $\mathcal{P}(X) = \big\{S : S \subseteq X\big\}$, the set of all subsets of X.

The *Cartesian product* of sets X and Y is $X \times Y = \{(x,y) : x \in X, y \in Y\}$, the set of all ordered pairs of elements of X and Y. The Cartesian product of finitely many sets $X_1, \dots, X_N$ is

$$\prod_{j=1}^{N} X_j \;=\; X_1 \times \cdots \times X_N \;=\; \big\{(x_1, \dots, x_N) : x_k \in X_k,\, k = 1, \dots, N\big\}.$$

Equivalence Relations

Informally, we say that $\sim$ is a *relation* on a set X if for each choice of x and y in X we have only one of the following two possibilities:

$$x \sim y \;\; (x \text{ is related to } y) \qquad \text{or} \qquad x \not\sim y \;\; (x \text{ is not related to } y).$$

An *equivalence relation* on a set X is a relation $\sim$ that satisfies the following conditions for all x, y, $z \in X$.

- Reflexivity: $x \sim x$.
- Symmetry: If $x \sim y$ then $y \sim x$.
- Transitivity: If $x \sim y$ and $y \sim z$ then $x \sim z$.

For example, if we declare that $x \sim y$ if and only if $x - y$ is rational, then $\sim$ is an equivalence relation on $\mathbb{R}$.

If $\sim$ is an equivalence relation on X, then the *equivalence class* of $x \in X$ is the set $[x]$ that contains all elements that are related to x:

$$[x] \;=\; \{y \in X \,:\, x \sim y\}.$$

Any two equivalence classes are either identical or disjoint. That is, if x and y are two elements of X, then either $[x] = [y]$ or $[x] \cap [y] = \varnothing$. The union of all equivalence classes $[x]$ is X. Consequently, the set of *distinct* equivalence classes forms a partition of X.

Intervals

An *interval* in the real line $\mathbb{R}$ is any one of the following sets:

- (a,b), $[a,b)$, $(a,b]$, $[a,b]$ where a, $b \in \mathbb{R}$ and $a < b$, or

- (a, ∞), $[a, \infty)$, $(-\infty, a)$, $(-\infty, a]$ where $a \in \mathbb{R}$, or
- $\mathbb{R} = (-\infty, \infty)$.

An *open interval* is an interval of the form (a, b), (a, ∞), $(-\infty, a)$, or $(-\infty, \infty)$. A *closed interval* is an interval of the form $[a, b]$, $[a, \infty)$, $(-\infty, a]$, or $(-\infty, \infty)$. We refer to $[a, b]$ as a *finite closed interval*, a *bounded closed interval*, or a *compact interval*.

The empty set $\varnothing$ and a singleton $\{a\}$ are not intervals, but even so we adopt the notational conventions

$$[a, a] \;=\; \{a\} \qquad \text{and} \qquad (a, a) \;=\; [a, a) \;=\; (a, a] \;=\; \varnothing.$$

We also consider *extended intervals*, which are any of the following sets:

- $(a, \infty] = (a, \infty) \cup \{\infty\}$ or $[a, \infty] = [a, \infty) \cup \{\infty\}$, where $a \in \mathbb{R}$,
- $[-\infty, b) = (-\infty, b) \cup \{-\infty\}$ or $[-\infty, b] = (-\infty, b] \cup \{-\infty\}$, where $b \in \mathbb{R}$, or
- $[-\infty, \infty] = \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$.

An extended interval is not an interval—whenever we refer to an "interval" without qualification we implicitly *exclude* the extended intervals.

Euclidean Space

We let $\mathbb{R}^d$ denote d-dimensional real Euclidean space, the set of all ordered d-tuples of real numbers. Similarly, $\mathbb{C}^d$ is d-dimensional complex Euclidean space, the set of all ordered d-tuples of complex numbers.

The *zero vector* is $0 = (0, \dots, 0)$. We use the same symbol "0" to denote the zero vector and the number zero; the intended meaning should be clear from context.

The *dot product* of vectors $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d)$ in $\mathbb{R}^d$ or $\mathbb{C}^d$ is

$$x \cdot y \;=\; x_1 \overline{y_1} + \cdots + x_d \overline{y_d},$$

and the *Euclidean norm* of x is

$$\|x\| \;=\; (x \cdot x)^{1/2} \;=\; \bigl(|x_1|^2 + \cdots + |x_d|^2\bigr)^{1/2}.$$

The *translation* of a set $E \subseteq \mathbb{R}^d$ by a vector $h \in \mathbb{R}^d$ (or a set $E \subseteq \mathbb{C}^d$ by a vector $h \in \mathbb{C}^d$) is $E + h = \{x + h : x \in E\}$.

Sequences

Let I be a fixed set. Given a set X and points $x_i \in X$ for $i \in I$, we write $\{x_i\}_{i\in I}$ to denote the sequence of elements x_i indexed by the set I. We call I an *index set* in this context, and refer to x_i as the *ith component* of the sequence $\{x_i\}_{i\in I}$. If we know that the x_i are *scalars* (real or complex numbers), then we often write $(x_i)_{i\in I}$ instead of $\{x_i\}_{i\in I}$. Technically, a sequence $\{x_i\}_{i\in I}$ is shorthand for the mapping $x\colon I \to X$ given by $x(i) = x_i$ for $i \in I$, and therefore the components x_i of a sequence need not be distinct. If the index set I is understood then we may write $\{x_i\}$ or $\{x_i\}_i$, or if the x_i are scalars then we may write (x_i) or $(x_i)_i$.

Often the index set I is countable. If $I = \{1, \dots, d\}$ then we sometimes write a sequence in list form as

$$\{x_n\}_{n=1}^d = \{x_1, \dots, x_d\},$$

or if the x_n are scalars then we often write

$$(x_n)_{n=1}^d = (x_1, \dots, x_d).$$

Similarly, if $I = \mathbb{N}$ then we may write

$$\{x_n\}_{n\in\mathbb{N}} = \{x_1, x_2, \dots\},$$

or if each x_n is a scalar then we usually write

$$(x_n)_{n\in\mathbb{N}} = (x_1, x_2, \dots).$$

A *subsequence* of a countable sequence $\{x_n\}_{n\in\mathbb{N}} = \{x_1, x_2, \dots\}$ is a sequence of the form $\{x_{n_k}\}_{k\in\mathbb{N}} = \{x_{n_1}, x_{n_2}, \dots\}$ where $n_1 < n_2 < \cdots$.

We say that a countable sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ is *monotone increasing* if $x_n \le x_{n+1}$ for every n, and *strictly increasing* if $x_n < x_{n+1}$ for every n. We define *monotone decreasing* and *strictly decreasing* sequences similarly.

The Kronecker Delta and the Standard Basis Vectors

Given indices i and j in an index set I (typically $I = \mathbb{N}$), the *Kronecker delta* of i and j is the number δ_{ij} defined by the rule

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \ne j. \end{cases}$$

For each integer $n \in \mathbb{N}$, we let δ_n denote the sequence

$$\delta_n = (\delta_{nk})_{k\in\mathbb{N}} = (0,\dots,0,1,0,0,\dots).$$

That is, the nth component of the sequence δ_n is 1, while all other components are zero. We call δ_n the *nth standard basis vector*, and we refer to the family $\{\delta_n\}_{n\in\mathbb{N}}$ as the *sequence of standard basis vectors*, or simply the *standard basis*.

Functions

Let X and Y be sets. We write $f\colon X \to Y$ to mean that f is a function with domain X and codomain Y. We usually write $f(x)$ to denote the image of x under f, but if $L\colon X \to Y$ is a linear map from one vector space X to another vector space Y then we may write Lx instead of $L(x)$. We also use the following notation.

- The *direct image* of a set $A \subseteq X$ under f is $f(A) = \{f(x) : x \in A\}$.
- The *inverse image* of a set $B \subseteq Y$ under f is

$$f^{-1}(B) = \{x \in X : f(x) \in B\}.$$

- The *range* of f is $\mathrm{range}(f) = f(X) = \{f(x) : x \in X\}$.
- f is *injective*, or *one-to-one*, if $f(x) = f(y)$ implies $x = y$.
- f is *surjective*, or *onto*, if $\mathrm{range}(f) = Y$.
- f is *bijective* if it is both injective and surjective. The *inverse function* of a bijection $f\colon X \to Y$ is the function $f^{-1}\colon Y \to X$ defined by $f^{-1}(y) = x$ if $f(x) = y$.
- Given $S \subseteq X$, the *restriction* of a function $f\colon X \to Y$ to the domain S is the function $f|_S\colon S \to Y$ defined by $(f|_S)(x) = f(x)$ for $x \in S$.
- The *zero function on* X is the function $0\colon X \to \mathbb{R}$ defined by $0(x) = 0$ for every $x \in X$. We use the same symbol 0 to denote the zero function and the number zero.
- The *characteristic function* of $A \subseteq X$ is the function $\chi_A\colon X \to \mathbb{R}$ given by

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

- If the domain of a function f is $\mathbb{R}^d$, then the *translation* of f by a vector $a \in \mathbb{R}^d$ is the function $T_a f$ defined by $T_a f(x) = f(x-a)$ for $x \in \mathbb{R}^d$.

Cardinality

A set X is *finite* if either X is empty or there exists a positive integer n and a bijection $f: \{1, \ldots, n\} \to X$. In the latter case we say that X *has n elements*.

A set X is *denumerable* or *countably infinite* if there exists a bijection $f: \mathbb{N} \to X$.

A set X is *countable* if it is either finite or denumerable. In particular, $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ are all denumerable and hence are countable.

A set X is *uncountable* if it is not countable. In particular, $\mathbb{R}$ and $\mathbb{C}$ are uncountable.

Extended Real-Valued Functions

A function that maps a set X into the real line $\mathbb{R}$ is called a *real-valued function*, and a function that maps X into the extended real line $[-\infty, \infty]$ is an *extended real-valued function.* Every real-valued function is extended real-valued, but an extended real-valued function need not be real-valued. An extended real-valued function f is *nonnegative* if $f(x) \geq 0$ for every x.

Let $f: X \to [-\infty, \infty]$ be an extended real-valued function. We associate to f the two extended real-valued functions f^+ and f^- defined by

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

We call f^+ the *positive part* and f^- the *negative part* of f. They are each nonnegative extended real-valued functions, and for every x we have

$$f(x) = f^+(x) - f^-(x) \qquad \text{and} \qquad |f(x)| = f^+(x) + f^-(x).$$

Given $f: X \to [-\infty, \infty]$, to avoid multiplicities of parentheses, brackets, and braces, we often write $f^{-1}(a,b) = f^{-1}((a,b))$, $f^{-1}[a,\infty) = f^{-1}([a,\infty))$, and so forth. We also use shorthands such as

$$\begin{aligned}
\{f \geq a\} &= \{x \in X : f(x) \geq a\}, \\
\{f = a\} &= \{x \in X : f(x) = a\}, \\
\{a < f < b\} &= \{x \in X : a < f(x) < b\}, \\
\{f \geq g\} &= \{x \in X : f(x) \geq g(x)\},
\end{aligned}$$

and so forth.

If $f: S \to [-\infty, \infty]$ is an extended real-valued function on a domain $S \subseteq \mathbb{R}$, then f is *monotone increasing* on S if for all $x, y \in S$ we have

$$x \leq y \implies f(x) \leq f(y).$$

We say that f is *strictly increasing* on S if for all $x, y \in S$,

$$x < y \implies f(x) < f(y).$$

Monotone decreasing and *strictly decreasing* functions are defined similarly.

Notation for Extended Real-Valued and Complex-Valued Functions

A function of the form $f\colon X \to \mathbb{C}$ is said to be *complex-valued.* We have the inclusions $\mathbb{R} \subseteq [-\infty, \infty]$ and $\mathbb{R} \subseteq \mathbb{C}$, so every real-valued function is both an extended real-valued and a complex-valued function. However, neither $[-\infty, \infty]$ nor $\mathbb{C}$ is a subset of the other, so an extended real-valued function need not be a complex-valued function, and a complex-valued function need not be an extended real-valued function. Hence there are usually two separate cases that we need to consider:

- extended real-valued functions of the form $f\colon X \to [-\infty, \infty]$, and
- complex-valued functions of the form $f\colon X \to \mathbb{C}$.

To consider both cases together, we use the notation $\overline{\mathbf{F}}$ introduced earlier, which stands for a choice of either the extended real line $[-\infty, \infty]$ or the complex plane $\mathbb{C}$. Thus, if we write $f\colon X \to \overline{\mathbf{F}}$ then we mean that f could either be an extended real-valued function or a complex-valued function on the domain X. Both possibilities include real-valued functions as a special case. As we declared earlier that, the word *scalar* means a *finite real number* (if $\overline{\mathbf{F}} = [-\infty, \infty]$) or a *complex number* (if $\overline{\mathbf{F}} = \mathbb{C}$). Thus, a *scalar-valued function* cannot take the values $\pm\infty$.

Suprema and Infima

A set of real numbers S is *bounded above* if there exists a real number M such that $x \le M$ for every $x \in S$. Any such number M is called an *upper bound* for S. The definition of *bounded below* is similar, and we say that S is *bounded* if it is bounded both above and below.

A number $x \in \mathbb{R}$ is the *supremum*, or *least upper bound*, of S if

- x is an upper bound for S, and
- if y is any upper bound for S, then $x \le y$.

We denote the supremum of S, if one exists, by $x = \sup(S)$. The *infimum*, or greatest lower bound, of S is defined in an entirely analogous manner, and is denoted by $\inf(S)$.

It is not obvious that every set that is bounded above has a supremum. We take the existence of suprema as the following axiom.

Axiom (Supremum Property of $\mathbb{R}$). Let S be a nonempty subset of $\mathbb{R}$. If S is bounded above, then there exists a real number $x = \sup(S)$ that is the supremum of S. $\diamondsuit$

We extend the definition of supremum to sets that are not bounded above by declaring that $\sup(S) = \infty$ if S is not bounded above. We also declare that $\sup(\varnothing) = -\infty$. Using these conventions, every set $S \subseteq \mathbb{R}$ has a supremum in the extended real sense.

If $S = (x_n)_{n\in\mathbb{N}}$ is countable, then we often write $\sup_n x_n$ or $\sup x_n$ to denote the supremum instead of $\sup(S)$, and similarly we may write $\inf_n x_n$ or $\inf x_n$ instead of $\inf(S)$.

If $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are two sequences of real numbers, then

$$\inf_n x_n + \inf_n y_n \le \inf_n (x_n + y_n) \le \sup_n (x_n + y_n) \le \sup_n x_n + \sup_n y_n.$$

Any or all of the inequalities on the preceding line can be strict. If $c > 0$ then

$$\sup_n cx_n = c \sup_n x_n \qquad \text{and} \qquad \sup_n (-cx_n) = -c \inf_n x_n.$$

Convergent and Cauchy Sequences of Scalars

Convergence of sequences will be discussed in the more general setting of metric spaces in Section 1.1.1. Here we will only consider sequences $(x_n)_{n\in\mathbb{N}}$ of real or complex numbers. We say that a sequence of scalars $(x_n)_{n\in\mathbb{N}}$ *converges* if there exists a scalar x such that for every $\varepsilon > 0$ there is an $N > 0$ such that

$$n \ge N \quad \Longrightarrow \quad |x - x_n| < \varepsilon.$$

In this case we say that x_n *converges to x as $n \to \infty$*, and we write

$$x_n \to x \qquad \text{or} \qquad \lim_{n\to\infty} x_n = x \qquad \text{or} \qquad \lim x_n = x.$$

We say that $(x_n)_{n\in\mathbb{N}}$ is a *Cauchy sequence* if for every $\varepsilon > 0$ there exists an integer $N > 0$ such that

$$m, n \ge N \quad \Longrightarrow \quad |x_m - x_n| < \varepsilon. \qquad \diamondsuit$$

An important consequence of the Supremum Property is that the following equivalence holds for any sequence of scalars:

$$(x_n)_{n\in\mathbb{N}} \text{ is convergent} \quad \Longleftrightarrow \quad (x_n)_{n\in\mathbb{N}} \text{ is Cauchy}.$$

Convergence in the Extended Real Sense

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of *real* numbers. We say that the sequence $(x_n)_{n\in\mathbb{N}}$ *diverges to* ∞ as $n \to \infty$ if for each real number $R > 0$ there is an integer $N > 0$ such that $x_n > R$ for all $n \geq N$. In this case we write

$$\lim_{n\to\infty} x_n \;=\; \infty.$$

We define *divergence to* $-\infty$ similarly.

We say that $\lim_{n\to\infty} x_n$ *exists* or that $(x_n)_{n\in\mathbb{N}}$ *converges in the extended real sense* if

- x_n converges to a real number x as $n \to \infty$, or
- x_n diverges to ∞ as $n \to \infty$, or
- x_n diverges to $-\infty$ as $n \to \infty$.

For example, every monotone increasing sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ converges in the extended real sense, and in this case $\lim x_n = \sup x_n$. Similarly, a monotone decreasing sequence of real numbers converges in the extended real sense and its limit equals its infimum.

Limsup and Liminf

The *limit superior*, or *limsup*, of a sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ is

$$\limsup_{n\to\infty} x_n \;=\; \inf_{n\in\mathbb{N}} \sup_{m\geq n} x_m \;=\; \lim_{n\to\infty} \sup_{m\geq n} x_m.$$

Likewise, the *limit inferior*, or *liminf*, of $(x_n)_{n\in\mathbb{N}}$ is

$$\liminf_{n\to\infty} x_n \;=\; \sup_{n\in\mathbb{N}} \inf_{m\geq n} x_m \;=\; \lim_{n\to\infty} \inf_{m\geq n} x_m.$$

The liminf and limsup of every sequence of real numbers exists in the extended real sense. Further,

$$\begin{array}{c}(x_n)_{n\in\mathbb{N}} \text{ converges in} \\ \text{the extended real sense}\end{array} \quad\Longleftrightarrow\quad \liminf_{n\to\infty} x_n \;=\; \limsup_{n\to\infty} x_n,$$

and in this case $\lim x_n = \liminf x_n = \limsup x_n$.

If $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are two sequences of real numbers, then

$$\begin{aligned}\liminf_{n\to\infty} x_n \;+\; \liminf_{n\to\infty} y_n \;&\leq\; \liminf_{n\to\infty}\,(x_n + y_n) \\ &\leq\; \limsup_{n\to\infty} x_n \;+\; \liminf_{n\to\infty} y_n\end{aligned}$$

$$\leq \limsup_{n\to\infty} (x_n + y_n)$$

$$\leq \limsup_{n\to\infty} x_n + \limsup_{n\to\infty} y_n,$$

as long as none of the sums above takes an indeterminate form $\infty - \infty$ or $-\infty + \infty$. Strict inequality can hold on any line above. If the sequence $(x_n)_{n\in\mathbb{N}}$ converges, then

$$\liminf_{n\to\infty} (x_n + y_n) = \lim_{n\to\infty} x_n + \liminf_{n\to\infty} y_n,$$

and likewise

$$\limsup_{n\to\infty} (x_n + y_n) = \lim_{n\to\infty} x_n + \limsup_{n\to\infty} y_n.$$

If $(x_n)_{n\in\mathbb{N}}$ is a sequence of real numbers, then there exist subsequences $(x_{n_k})_{k\in\mathbb{N}}$ and $(x_{m_j})_{j\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} x_{n_k} = \limsup_{n\to\infty} x_n \qquad \text{and} \qquad \lim_{j\to\infty} x_{m_j} = \liminf_{n\to\infty} x_n.$$

In fact, if $(x_n)_{n\in\mathbb{N}}$ is bounded above then $\limsup x_n$ is the largest possible limit of a subsequence $(x_{n_k})_{k\in\mathbb{N}}$, and likewise if $(x_n)_{n\in\mathbb{N}}$ is bounded below then $\liminf x_n$ is the smallest possible limit of a subsequence. Consequently,

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

On occasion we deal with real-parameter versions of liminf and limsup. Given a real-valued function f whose domain includes an interval centered at a point $x \in \mathbb{R}$, we define

$$\limsup_{t\to x} f(t) = \inf_{\delta>0} \sup_{|t-x|<\delta} f(t) = \lim_{\delta\to 0} \sup_{|t-x|<\delta} f(t),$$

and $\liminf_{t\to x} f(t)$ is defined analogously. The properties of these real-parameter versions of liminf and limsup are similar to those of the sequence versions.

Infinite Series

Infinite series in the general setting of normed spaces will be discussed in Section 1.2.3; here we restrict our attention to infinite series of scalars. If $(c_n)_{n\in\mathbb{N}}$ is a sequence of real or complex numbers, then we say that the infinite series $\sum_{n=1}^{\infty} c_n$ *converges* if there exists a scalar s such that the *partial sums* $s_N = \sum_{n=1}^{N} c_n$ converge to s as $N \to \infty$. In this case $\sum_{n=1}^{\infty} c_n$ is assigned the value s:

$$\sum_{n=1}^{\infty} c_n \;=\; \lim_{N\to\infty} s_N \;=\; \lim_{N\to\infty} \sum_{n=1}^{N} c_n \;=\; s.$$

Series of Real Numbers. Assume that *every c_n is a real number.* Then we say that the series $\sum_{n=1}^{\infty} c_n$ *converges in the extended real sense*, or simply that the series *exists*, if

- s_N converges to a real number s as $N \to \infty$, or
- s_N diverges to ∞ as $N \to \infty$, or
- s_N diverges to $-\infty$ as $N \to \infty$.

Nonnegative Series. *If every c_n is a nonnegative real number* (that is, $c_n \ge 0$ for every n), then the series $\sum_{n=1}^{\infty} c_n$ converges in the extended real sense. Moreover, there are only two possibilities: Either the series converges to a nonnegative real number or it diverges to infinity. We indicate which possibility holds as follows:

$$\sum_{n=1}^{\infty} c_n \;<\; \infty \quad \text{means that the series converges (to a finite real number),}$$

while

$$\sum_{n=1}^{\infty} c_n \;=\; \infty \quad \text{means that the series diverges to infinity.}$$

Pointwise Convergence of Functions

If X is a set and $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of extended real-valued or complex-valued functions whose domain is X, then we say that f_n *converges pointwise to a function f* if

$$f(x) \;=\; \lim_{n\to\infty} f_n(x) \quad \text{for all } x \in X.$$

In this case we write $f_n(x) \to f(x)$ *for every $x \in X$* or $f_n \to f$ *pointwise.* Note that this convergence can be in the extended real sense.

If $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of extended real-valued functions whose domain is a set X, then we say that $\{f_n\}_{n\in\mathbb{N}}$ is a *monotone increasing sequence* if $\{f_n(x)\}_{n\in\mathbb{N}}$ is monotone increasing for each x, i.e., if

$$f_1(x) \;\le\; f_2(x) \;\le\; \cdots \qquad \text{for all } x \in X.$$

In this case $f(x) = \lim_{n\to\infty} f_n(x)$ exists for each $x \in X$ in the extended real sense, and we say that f_n *increases pointwise to f*. We denote this by writing

$$f_n \nearrow f \quad \text{on } X.$$

Continuity

Continuity for the general setting of functions on metric spaces will be discussed in Section 1.1.4. Here we define continuity for scalar-valued functions whose domain is a subset E of $\mathbb{R}^d$. We say that $f\colon E \to \mathbb{C}$ is *continuous on the set* E if whenever we have points $x_n, x \in E$ such that $x_n \to x$, it follows that $f(x_n) \to f(x)$.

Derivatives and Everywhere Differentiability

Let f be a scalar-valued function whose domain includes an open interval centered at a point $x \in \mathbb{R}$. We say that f is *differentiable at* x if the limit

$$f'(x) \;=\; \lim_{y\to x} \frac{f(y)-f(x)}{y-x}$$

exists and is a scalar.

Let $[a,b]$ be a closed interval in the real line. A function f is *everywhere differentiable* or *differentiable everywhere* on $[a,b]$ if it is differentiable at each point in the interior (a,b) and if the appropriate one-sided derivatives exist at the endpoints a and b. That is, f is everywhere differentiable on $[a,b]$ if

$$f'(x) \;=\; \lim_{y\to x,\, y\in[a,b]} \frac{f(y)-f(x)}{y-x}$$

exists and is a scalar for each $x \in [a,b]$.

We use similar terminology if f is defined on other types of intervals in $\mathbb{R}$. For example, $x^{3/2}$ is differentiable everywhere on $[0,1]$ and $x^{1/2}$ is differentiable everywhere on $(0,1]$, but $x^{1/2}$ is not differentiable everywhere on $[0,1]$.

The Riemann Integral

Let $f\colon [a,b] \to \mathbb{R}$ be a bounded, real-valued function on a finite, closed interval $[a,b]$. A *partition* of $[a,b]$ is a choice of finitely many points x_k in $[a,b]$ such that $a = x_0 < x_1 < \cdots < x_n = b$. If we wish to give this partition a name then we will write:

$$\textit{Let } \Gamma = \bigl\{a = x_0 < \cdots < x_n = b\bigr\} \textit{ be a partition of } [a,b].$$

The *mesh size* of Γ is $|\Gamma| = \max\bigl\{x_j - x_{j-1} \,:\, j = 1,\dots,n\bigr\}$.

Given a partition $\Gamma = \{a = x_0 < \cdots < x_n = b\}$, for each $j = 1, \dots, n$ let m_j and M_j denote the infimum and supremum of f on the interval $[x_{j-1}, x_j]$:

$$m_j = \inf_{x \in [x_{j-1}, x_j]} f(x) \qquad \text{and} \qquad M_j = \sup_{x \in [x_{j-1}, x_j]} f(x).$$

The numbers

$$L_\Gamma = \sum_{j=1}^{n} m_j\,(x_j - x_{j-1}) \qquad \text{and} \qquad U_\Gamma = \sum_{j=1}^{n} M_j\,(x_j - x_{j-1}),$$

are called *lower and upper Riemann sums for* f, respectively. We say that f *is Riemann integrable on* $[a,b]$ if there exists a real number I such that

$$\sup_\Gamma L_\Gamma = \inf_\Gamma U_\Gamma = I,$$

where the supremum and infimum are taken over all partitions Γ. In this case, the number I is the *Riemann integral of* f *over* $[a,b]$, and we write $\int_a^b f(x)\,dx = I$.

Here is an equivalent definition of the Riemann integral. Given a partition $\Gamma = \{a = x_0 < \cdots < x_n = b\}$, choose any points $\xi_j \in [x_{j-1}, x_j]$. We call

$$R_\Gamma = \sum_{j=1}^{n} f(\xi_j)\,(x_j - x_{j-1})$$

a *Riemann sum for* f (note that R_Γ implicitly depends on both the partition Γ *and* the choice of points ξ_j). Then f is Riemann integrable on $[a,b]$ if and only if there is a real number I such that $I = \lim_{|\Gamma|\to 0} R_\Gamma$, where this means that for every $\varepsilon > 0$, there is a $\delta > 0$ such that for any partition Γ with $|\Gamma| < \delta$ and any choice of points $\xi_j \in [x_{j-1}, x_j]$ we have $|I - R_\Gamma| < \varepsilon$. In this case, I is the Riemann integral of f over $[a,b]$, and we write $\int_a^b f(x)\,dx = I$.

We declare that a complex-valued function f on $[a,b]$ is Riemann integrable if its real and imaginary parts are both Riemann integrable.

Every continuous function $f\colon [a,b] \to \mathbb{C}$ is Riemann integrable. However, there exist discontinuous functions that are Riemann integrable. We will characterize the Riemann integrable functions on $[a,b]$ in Section 4.5.5.

If $g\colon [a,b] \to \mathbb{C}$ is continuous, then g is Riemann integrable on the interval $[a,x]$ for each $a \le x \le b$, so we can consider the *indefinite integral* of g, defined by

$$G(x) = \int_a^x g(t)\,dt, \qquad x \in [a,b].$$

The *Fundamental Theorem of Calculus* implies that G is differentiable on the interval $[a,b]$, and $G'(x) = g(x)$ for each $x \in [a,b]$. We will prove a more general form of the Fundamental Theorem of Calculus in Section 6.4.

Chapter 1
Metric and Normed Spaces

Much of real analysis centers on issues of *convergence* or *approximation*. In this preliminary chapter we briefly review metric spaces and normed spaces, which are sets on which we can define a notion of distance or length that allows us to quantify the meaning of closeness or convergence. The results in this chapter are presented in a compressed form, without the more extensive motivation and discussion that is provided in the rest of the text. Some proofs are assigned as exercises, and a few longer proofs are omitted. For complete details and proofs of this material we refer to undergraduate real analysis texts such as [Rud76], [BS11], or Chapters 2 and 3 of [Heil18].

1.1 Metric Spaces

A metric provides us with a notion of the distance between points in a set.

Definition 1.1.1 (Metric Space). Let X be a nonempty set. A *metric* on X is a function $\mathrm{d}\colon X \times X \to \mathbb{R}$ such that for all $x, y, z \in X$ we have:

(a) Nonnegativity: $0 \le \mathrm{d}(x,y) < \infty$,

(b) Symmetry: $\mathrm{d}(x,y) = \mathrm{d}(y,x)$,

(c) The Triangle Inequality: $\mathrm{d}(x,z) \le \mathrm{d}(x,y) + \mathrm{d}(y,z)$, and

(d) Uniqueness: $\mathrm{d}(x,y) = 0$ if and only if $x = y$.

If these conditions are satisfied, then X is a called a *metric space*. The number $\mathrm{d}(x,y)$ is called the *distance* from x to y. $\diamondsuit$

For example,

$$\mathrm{d}(x,y) \;=\; \|x-y\| \;=\; \left(\sum_{k=1}^{d} |x_k - y_k|^2\right)^{1/2}, \qquad x, y \in \mathbb{C}^d, \tag{1.1}$$

C. Heil, *Introduction to Real Analysis*, Graduate Texts in Mathematics 280,
https://doi.org/10.1007/978-3-030-26903-6_1

is a metric on $\mathbb{C}^d$, called the *Euclidean metric*. The Euclidean metric on $\mathbb{R}^d$ is the restriction of equation (1.1) to $x, y \in \mathbb{R}^d$. Unless otherwise specified, we always assume that the metric on $\mathbb{R}^d$ or $\mathbb{C}^d$ is the Euclidean metric.

1.1.1 Convergence and Completeness

If d is a metric, then the number $\mathrm{d}(x,y)$ represents the distance from the point x to the point y. We will say that points x_n are *converging* to a point x if the distance from x_n to x shrinks to zero as n increases. Closely related is the idea of a *Cauchy sequence*, which is a sequence where the distance $\mathrm{d}(x_m, x_n)$ between two points in the sequence decreases as m and n increase.

Definition 1.1.2 (Convergent and Cauchy Sequences). Let X be a metric space.

(a) A sequence of points $\{x_n\}_{n\in\mathbb{N}}$ in X *converges* to a point $x \in X$ if

$$\lim_{n\to\infty} \mathrm{d}(x_n, x) \;=\; 0.$$

That is, for every $\varepsilon > 0$ there must exist some integer $N > 0$ such that

$$n \ge N \quad \Longrightarrow \quad \mathrm{d}(x_n, x) \,<\, \varepsilon.$$

In this case, we write $x_n \to x$.

(b) A sequence of points $\{x_n\}_{n\in\mathbb{N}}$ in X is a *Cauchy sequence* if for every $\varepsilon > 0$ there exists an integer $N > 0$ such that

$$m, n \ge N \quad \Longrightarrow \quad \mathrm{d}(x_m, x_n) \,<\, \varepsilon. \qquad \diamondsuit$$

Convergence implicitly depends on the choice of metric for X, so if we want to emphasize that we are using a particular metric, we may write $x_n \to x$ *with respect to the metric* d.

By applying the Triangle Inequality, we immediately obtain the following relation between convergent and Cauchy sequences.

Lemma 1.1.3 (Convergent Implies Cauchy). *If $\{x_n\}_{n\in\mathbb{N}}$ is a convergent sequence in a metric space X, then $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in X.* $\diamondsuit$

Some metric spaces have the property that every Cauchy sequence in the space converges to an element of the space. Since we can test whether a sequence is Cauchy without having the limit vector x in hand, this is often very useful. We give such spaces the following name.

Definition 1.1.4 (Complete Metric Space). Let X be a metric space. If every Cauchy sequence in X converges to an element of X, then we say that X is *complete*. $\diamondsuit$

For example, the real line $\mathbb{R}$ and the complex plane $\mathbb{C}$ are complete (with respect to the usual metric $\mathrm{d}(x,y) = |x-y|$), and it follows from this that $\mathbb{R}^d$ and $\mathbb{C}^d$ are complete with respect to the Euclidean metric. In contrast, the set of rational numbers $\mathbb{Q}$ is not complete with respect to the metric $\mathrm{d}(x,y) = |x-y|$. For example, if we set $x_1 = 3.1$, $x_2 = 3.14$, $x_3 = 3.141$, $x_4 = 3.1415$, and so forth (truncating the decimal expansion of $\pi = 3.14159\ldots$), then $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{Q}$, but it does not converge *to an element of* $\mathbb{Q}$ (it does converge to π, but π does not belong to $\mathbb{Q}$). An example of an incomplete infinite-dimensional normed space is given in Problem 1.3.8.

1.1.2 Topology in Metric Spaces

Since a metric space has a notion of distance, we can define an open ball to be the set of all points that lie within a distance r of a point x. Using open balls we then define open and closed sets, accumulation points, boundary points, and other useful notions.

Definition 1.1.5. Let X be a metric space.

- Given $x \in X$ and $r > 0$, the *open ball in X centered at x with radius r* is

$$B_r(x) \;=\; \bigl\{y \in X : \mathrm{d}(x,y) < r\bigr\}.$$

- A set $E \subseteq X$ is *bounded* if $E \subseteq B_r(x)$ for some $x \in X$ and $r > 0$.
- A set $U \subseteq X$ is *open* if for each $x \in U$ there exists an $r > 0$ such that $B_r(x) \subseteq U$. Equivalently, U is open if and only if U can be written as a union of open balls.
- The *topology* of X is the set of all open subsets of X.
- The *interior* of a set $E \subseteq X$ is the largest open set E° that is contained in E. Explicitly, $E^\circ = \bigcup\bigl\{U \subseteq X : U \text{ is open and } U \subseteq E\bigr\}$.
- A set $E \subseteq X$ is *closed* if $X \setminus E$ is open.
- The *closure* of a set $E \subseteq X$ is the smallest closed set $\overline{E}$ that contains E. Explicitly, $\overline{E} = \bigcap\bigl\{F \subseteq X : F \text{ is closed and } E \subseteq F\bigr\}$.
- A set $E \subseteq X$ is *dense* in X if $\overline{E} = X$.
- X is *separable* if there exists a *countable* subset of X that is dense.
- A point $x \in X$ is an *accumulation point* or *cluster point* of a set $E \subseteq X$ if there exist $x_n \in E$ with all $x_n \neq x$ such that $x_n \to x$.
- A point $x \in X$ is a *boundary point* of a set $E \subseteq X$ if for every $r > 0$ we have both $B_r(x) \cap E \neq \varnothing$ and $B_r(x) \cap E^{\mathrm{C}} \neq \varnothing$. The set of all boundary points of E is called the *boundary* of E, and it is denoted by ∂E. $\diamondsuit$

The reader should check that the empty set $\varnothing$ and the entire space X are open, the union of any collection of open subsets of X is open, and the intersection of finitely many open sets is open (it is these three properties that are the inspiration for the definition of a topology in an abstract setting).

The following exercise gives an equivalent characterization of closed sets in terms of limits of points of E.

Exercise 1.1.6. Let E be a subset of a metric space X. Prove that E is closed if and only if the following statement holds:

$$\text{If } x_n \in E \text{ and } x_n \to x \in X, \text{ then } x \in E. \qquad \diamondsuit$$

Here are some further useful facts.

Exercise 1.1.7. Given a subset E of a metric space X, prove the following statements.

(a) $\overline{E} = \bigl\{y \in X : \text{there exist } x_n \in E \text{ such that } x_n \to y\bigr\}$.

(b) E is dense in X if and only if for every point $x \in X$ there exist points $x_n \in E$ such that $x_n \to x$. $\diamondsuit$

To summarize Exercises 1.1.6 and 1.1.7:

- E is closed if and only if it contains every limit of points from E,
- the closure of E is the set of all limits of points from E, and
- E is dense in X if and only if every point in X is a limit of points from E.

For example, the set of rationals $\mathbb{Q}$ is not closed in $X = \mathbb{R}$ because a limit of rational points need not be rational; the closure of $\mathbb{Q}$ is $\mathbb{R}$ because every point in $\mathbb{R}$ can be written as a limit of rational points; and $\mathbb{Q}$ is dense in $\mathbb{R}$ because every point in $\mathbb{R}$ can be written as a limit of rational points.

1.1.3 Compact Sets in Metric Spaces

Next we introduce compact sets, which are defined in terms of "coverings" of a set by open sets. By a *cover* of a set S, we mean a collection of sets $\{E_i\}_{i\in I}$ whose union contains S. If each set E_i is open, then we call $\{E_i\}_{i\in I}$ an *open cover* of S. The index set I may be finite or infinite (even uncountable). If I is finite then we call $\{E_i\}_{i\in I}$ a *finite cover* of S. Thus a *finite open cover* of S is a collection of finitely many open sets whose union contains S.

Definition 1.1.8 (Compact Set). A subset K of a metric space X is *compact* if every covering of K by open sets has a finite subcovering. That is, K is compact if it is the case that whenever

$$K \subseteq \bigcup_{i\in I} U_i,$$

where $\{U_i\}_{i \in I}$ is any collection of open subsets of X, then there exist *finitely* many indices $i_1, \dots, i_N \in I$ such that $K \subseteq U_{i_1} \cup \cdots \cup U_{i_N}$. $\diamondsuit$

In order to give an equivalent reformulation of compactness, we introduce the following terminology.

Definition 1.1.9 (Sequentially Compact Set). A subset K of a metric space X is *sequentially compact* if every sequence $\{x_n\}_{n \in \mathbb{N}}$ of points of K contains a convergent subsequence $\{x_{n_k}\}_{k \in \mathbb{N}}$ whose limit belongs to K. $\diamondsuit$

In an abstract topological space the notions of compactness and sequential compactness need not be the same. However, they do coincide *in metric spaces*. We state this as the following theorem; for one proof see [Heil18, Thm. 2.8.9].

Theorem 1.1.10. *If K is a subset of a metric space X, then*

$$K \text{ is compact} \iff K \text{ is sequentially compact.} \qquad \diamondsuit$$

We prove that compact sets in metric spaces are both closed and bounded.

Lemma 1.1.11. *If K is a compact subset of a metric space X, then K is closed and bounded.*

Proof. Suppose K is compact, and fix $x \in X$. The union of the open balls $B_n(x)$ over all $n \in \mathbb{N}$ covers X, so this cover must have a finite subcover $\{B_{n_1}(x), \dots, B_{n_M}(x)\}$. Choosing the ball of largest radius from this finite subcover, we see that K is contained in a single open ball and hence is bounded.

Now we show that K is closed. If $K = X$ then K is closed and we are done, so assume that $K \neq X$. Choose any point $y \in K^{\mathrm{C}} = X \backslash K$. If $x \in K$ then $x \neq y$, so by the *Hausdorff property* stated in Problem 1.1.19 there exist disjoint open sets U_x and V_x such that $x \in U_x$ and $y \in V_x$. The collection $\{U_x\}_{x \in K}$ is an open cover of K, so it must contain some finite subcover, say

$$K \subseteq U_{x_1} \cup \cdots \cup U_{x_N}.$$

Each V_{x_j} is disjoint from U_{x_j}, so $V = V_{x_1} \cap \cdots \cap V_{x_N}$ is entirely contained in the complement of K. Thus, V is an open set and $y \in V \subseteq K^{\mathrm{C}}$. This implies that K^{C} is open, and therefore K is closed. $\square$

The converse of Lemma 1.1.11 need not hold. That is, in some metric spaces there exist sets that are closed and bounded but not compact; Problem 1.3.10 gives an example. However, for Euclidean space we have the following classical result (for one proof, see [Heil18, Thm. 2.8.4]).

Theorem 1.1.12 (Heine–Borel Theorem). *If K is a subset of $\mathbb{R}^d$ or $\mathbb{C}^d$, then K is compact if and only if K is closed and bounded.* $\diamondsuit$

1.1.4 Continuity for Functions on Metric Spaces

In abstract topological spaces, continuity is defined in terms of inverse images of open sets. We give that definition next, for the setting of functions on metric spaces.

Definition 1.1.13 (Continuous Function). Let X and Y be metric spaces. We say that a function $f\colon X \to Y$ is *continuous* if for every open set $V \subseteq Y$, its inverse image $f^{-1}(V)$ is an open subset of X. $\diamondsuit$

In contrast, the *direct* image of an open set under a continuous function need not be open (for example, if $f(x) = \sin x$ then $f(0,2\pi) = [-1,1]$). Likewise, the direct image of a closed set under a continuous function need not be closed. Even so, the following exercise shows that *a continuous functions maps compact sets to compact sets.*

Exercise 1.1.14. Let X and Y be metric spaces, and assume that $f\colon X \to Y$ is continuous. Prove that if K is a compact subset of X, then $f(K)$ is a compact subset of Y. $\diamondsuit$

The next exercise gives a useful reformulation of continuity for functions on metric spaces in terms of preservation of limits.

Exercise 1.1.15. Let X be a metric space with metric d_X, and let Y be a metric space with metric d_Y. Given a function $f\colon X \to Y$, prove that the following three statements are equivalent.

(a) f is continuous.

(b) If x is any point in X, then for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $y \in X$ we have

$$\mathrm{d}_X(x,y) < \delta \quad \Longrightarrow \quad \mathrm{d}_Y\bigl(f(x), f(y)\bigr) < \varepsilon.$$

(c) If $x \in X$ and $\{x_n\}_{n\in\mathbb{N}}$ is any sequence of points in X, then

$$x_n \to x \text{ in } X \quad \Longrightarrow \quad f(x_n) \to f(x) \text{ in } Y. \qquad \diamondsuit$$

The number δ that appears in statement (b) of Exercise 1.1.15 depends on both the point x and the number ε. If δ can be chosen independently of x, then we say that f is uniformly continuous.

Definition 1.1.16 (Uniform Continuity). Let X be a metric space with metric d_X, and let Y be a metric space with metric d_Y. If $E \subseteq X$, then we say that a function $f\colon E \to Y$ is *uniformly continuous* on E if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all x and y in E we have

$$\mathrm{d}_X(x,y) < \delta \quad \Longrightarrow \quad \mathrm{d}_Y\bigl(f(x),\, f(y)\bigr) < \varepsilon. \qquad \diamondsuit$$

According to the next result, a continuous function whose domain is a compact set is uniformly continuous on that set (for one proof, see [Heil18, Lemma 2.9.6]).

Theorem 1.1.17. *Let X and Y be metric spaces. If $K \subseteq X$ is compact and $f\colon K \to Y$ is continuous, then f is bounded and uniformly continuous on K.* $\diamondsuit$

Problems

1.1.18. Given that $\mathbb{R}$ and $\mathbb{C}$ are complete, prove that $\mathbb{R}^d$ and $\mathbb{C}^d$ are complete with respect to the Euclidean metric.

1.1.19. Let X be a metric space.

(a) Prove that X is *Hausdorff*, i.e., if $x \neq y$ are two distinct elements of X, then there exist *disjoint open sets* U and V such that $x \in U$ and $y \in V$.

(b) Prove that the limit of a convergent sequence in X is unique, i.e., if $x_n \to y$ and $x_n \to z$ then $y = z$.

1.1.20. Assume $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in a metric space X, and there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ that converges to $x \in X$. Prove that $x_n \to x$.

1.1.21. Given a sequence $\{x_n\}_{n\in\mathbb{N}}$ in a metric space X, prove the following statements.

(a) If $\mathrm{d}(x_n, x_{n+1}) < 2^{-n}$ for every $n \in \mathbb{N}$, then $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy (and therefore converges if X is complete).

(b) If $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy, then there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ such that $\mathrm{d}(x_{n_k}, x_{n_{k+1}}) < 2^{-k}$ for each $k \in \mathbb{N}$.

1.1.22. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of points in a metric space X. Prove that $x_n \to x$ if and only if for every subsequence $\{y_n\}_{n\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ there exists a subsequence $\{z_n\}_{n\in\mathbb{N}}$ of $\{y_n\}_{n\in\mathbb{N}}$ such that $z_n \to x$.

1.1.23. Let X be a metric space. Extend the definition of convergence to families indexed by a real parameter by declaring that if $x \in X$ and $x_t \in X$ for t in the interval $(0, c)$, where $c > 0$, then $x_t \to x$ as $t \to 0^+$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $\mathrm{d}(x_t, x) < \varepsilon$ whenever $0 < t < \delta$. Show that $x_t \to x$ as $t \to 0^+$ if and only if $x_{t_k} \to x$ for every sequence of real numbers $\{t_k\}_{k\in\mathbb{N}}$ in $(0, c)$ that satisfy $t_k \to 0$.

1.1.24. We say that a function $f\colon \mathbb{R}^d \to \mathbb{R}$ is *upper semicontinuous* (abbreviated usc) at a point $x \in \mathbb{R}^d$ if $\limsup_{y\to x} f(y) \le f(x)$. Explicitly, this means that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$|x - y| < \delta \quad\implies\quad f(y) \le f(x) + \varepsilon.$$

An analogous definition is made for *lower semicontinuity* (lsc). Prove the following statements.

(a) If $g\colon \mathbb{R}^d \to \mathbb{R}$ and $r > 0$, then $h(x) = \inf\bigl\{g(y) : y \in B_r(x)\bigr\}$ is usc at every point where $h(x) \neq -\infty$.

(b) If $f\colon \mathbb{R}^d \to \mathbb{R}$, then f is continuous at x if and only if f is both usc and lsc at x.

(c) If $\{f_i\}_{i\in I}$ is a family of functions that are each usc at a point x, then $g = \inf_{i\in I} f_i$ is usc at x.

(d) $f\colon \mathbb{R}^d \to \mathbb{R}$ is usc at every point $x \in \mathbb{R}^d$ if and only if the set $f^{-1}[a,\infty) = \{x \in \mathbb{R}^d : f(x) \geq a\}$ is closed for each $a \in \mathbb{R}$. Likewise, f is lsc at every point x if and only if $f^{-1}(a,\infty) = \{x \in \mathbb{R}^d : f(x) > a\}$ is open for each $a \in \mathbb{R}$.

(e) If K is a compact subset of $\mathbb{R}^d$ and $f\colon \mathbb{R}^d \to \mathbb{R}$ is usc at every point of K, then f is bounded above on K.

1.2 Normed Spaces

1.2.1 Vector Spaces

We assume that the reader is familiar with vector spaces. The scalar field associated with the vector spaces in this volume will always be either the real line $\mathbb{R}$ or the complex plane $\mathbb{C}$. The elements of the scalar field are referred to as *scalars*. If X is a vector space and we choose the scalar field to be $\mathbb{R}$ then we say that X is a *real* vector space, while if we choose the scalar field to be $\mathbb{C}$ then we say that X is a *complex* vector space.

We recall the definition of a spanning set and an independent set in a vector space.

Definition 1.2.1 (Span and Independence). Let X be a vector space, let I be an index set, and let $\mathcal{F} = \{x_i\}_{i\in I}$ be a sequence of vectors in X.

(a) The *finite linear span* of $\mathcal{F} = \{x_i\}_{i\in I}$, or simply the *span* for short, is the set of all finite linear combinations of elements of $\mathcal{F}$:

$$\operatorname{span}(\mathcal{F}) \;=\; \operatorname{span}\{x_i\}_{i\in I} \;=\; \biggl\{\sum_{n=1}^{N} c_n x_{i_n} \;:\; N > 0,\; i_n \in I,\; c_n \text{ is a scalar}\biggr\}.$$

(b) We say that $\mathcal{F} = \{x_i\}_{i\in I}$ is *finitely linearly independent*, or simply *independent* for short, if for every choice of finitely many distinct indices $i_1, \dots, i_N \in I$, we have

$$\sum_{n=1}^{N} c_n x_{i_n} \;=\; 0 \quad\Longrightarrow\quad c_1 \;=\; \cdots \;=\; c_N \;=\; 0. \qquad \diamondsuit$$

Next we recall the definition of a basis for a vector space. To distinguish this from the related of notion of a *Schauder basis* for a Banach space (which will be discussed in Chapter 8), we will refer to the usual vector space notion of a basis as a *Hamel basis*.

Definition 1.2.2 (Hamel Basis). Let X be a vector space. A *Hamel basis*, *vector space basis*, or simply a *basis* for X is a set $\mathcal{B} \subseteq X$ such that $\mathcal{B}$ is finitely linearly independent and $\operatorname{span}(\mathcal{B}) = X$. $\diamondsuit$

The *standard basis* for $\mathbb{R}^d$ or $\mathbb{C}^d$ is the Hamel basis $\mathcal{B} = \{e_1, \dots, e_d\}$, where $e_k = (0, \dots, 0, 1, 0, \dots, 0)$ has a 1 in the kth component and zeros elsewhere.

1.2.2 Seminorms and Norms

While a metric provides us with a notion of the *distance between points* in a space, a norm gives us a notion of the *length* of an individual vector. A norm can only be defined on a vector space, while a metric can be defined on any set.

Definition 1.2.3 (Seminorms and Norms). Let X be a vector space. A *seminorm* on X is a function $\|\cdot\| \colon X \to \mathbb{R}$ such that for all vectors $x, y \in X$ and all scalars c we have:

(a) Nonnegativity: $0 \le \|x\| < \infty$,

(b) Homogeneity: $\|cx\| = |c|\,\|x\|$, and

(c) The Triangle Inequality: $\|x + y\| \le \|x\| + \|y\|$.

A seminorm is a *norm* if we also have:

(d) Uniqueness: $\|x\| = 0$ if and only if $x = 0$.

A vector space X together with a norm $\|\cdot\|$ is called a *normed vector space*, a *normed linear space*, or simply a *normed space*. We refer to the number $\|x\|$ as the *length* of the vector x, and we say that $\|x - y\|$ is the *distance* between the vectors x and y. $\diamondsuit$

If X is a normed space, then it follows directly that

$$\mathrm{d}(x, y) \;=\; \|x - y\|, \qquad x, y \in X,$$

defines a metric on X (called the *metric on X induced from* $\|\cdot\|$, or simply the *induced metric* on X). Consequently, whenever we are given a normed space X, we have a metric on X as well. Therefore all of the definitions we made for metric spaces also apply to normed spaces, using the induced norm $\mathrm{d}(x, y) = \|x - y\|$. For example, convergence in a normed space is defined by

$$x_n \to x \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} \|x - x_n\| \;=\; 0.$$

It may be possible to place a metric on X other than the induced metric, but unless we explicitly state otherwise, all metric-related statements on a normed space are taken with respect to the induced metric.

The Euclidean norm $\|x\| = \big(|x_1|^2 + \cdots + |x_d|^2\big)^{1/2}$ is a norm on $\mathbb{R}^d$ and $\mathbb{C}^d$. The metric induced from the Euclidean norm is the Euclidean metric defined in equation (1.1).

Here are some properties of norms.

Exercise 1.2.4. Let X be a normed space, and let x, y, x_n, and y_n denote elements of X. Prove that the following statements hold.

(a) Reverse Triangle Inequality: $\big|\|x\| - \|y\|\big| \le \|x - y\|$.

(b) Convergent implies Cauchy: If $x_n \to x$, then $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy.

(c) Boundedness of Cauchy sequences: If $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence, then $\sup \|x_n\| < \infty$.

(d) Continuity of the norm: If $x_n \to x$, then $\|x_n\| \to \|x\|$.

(e) Continuity of vector addition: If $x_n \to x$ and $y_n \to y$, then $x_n + y_n \to x + y$.

(f) Continuity of scalar multiplication: If $x_n \to x$ and $c_n \to c$ (where c_n and c are scalars), then $c_n x_n \to cx$. $\diamondsuit$

Every convergent sequence is Cauchy, but the converse need not hold. Still, in *some* normed spaces it happens that every Cauchy sequence in the space converges to an element of the space. We give such spaces the following name.

Definition 1.2.5 (Banach Space). Let X be a normed space. If every Cauchy sequence in X converges to an element of X, then we say that X is *complete*, and in this case we also say that X is a *Banach space*. $\diamondsuit$

The real line and the complex plane are complete, and likewise $\mathbb{R}^d$ and $\mathbb{C}^d$ are Banach spaces with respect to the Euclidean norm.

1.2.3 Infinite Series in Normed Spaces

We define infinite series in a normed space as follows.

Definition 1.2.6 (Convergent Series). Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of vectors in a normed space X. We say that the series $\sum_{n=1}^\infty x_n$ *converges and equals* $x \in X$ if the *partial sums* $s_N = \sum_{n=1}^N x_n$ converge to x, i.e., if

$$\lim_{N\to\infty} \|x - s_N\| \;=\; \lim_{N\to\infty} \biggl\| x - \sum_{n=1}^N x_n \biggr\| \;=\; 0.$$

In this case, we write $x = \sum_{n=1}^\infty x_n$, and we also use the shorthands $x = \sum x_n$ or $x = \sum_n x_n$. $\diamondsuit$

In order for an infinite series to converge in X, the *norm of the difference* between x and the partial sum s_N must converge to zero. If we wish to emphasize which norm we are referring to, we may write that $x = \sum x_n$ *converges with respect to* $\|\cdot\|$, or we may say that $x = \sum x_n$ *converges in* X.

If $\{x_n\}_{n\in\mathbb{N}}$ is a sequence of vectors in X, then $\{\|x_n\|\}_{n\in\mathbb{N}}$ is a sequence of real scalars. What connection, if any, is there between the convergence of the series $\sum x_n$ in X (which is a series of *vectors*) and convergence of the series $\sum \|x_n\|$ (which is a series of *scalars*)? In order to address this, we introduce the following terminology.

Definition 1.2.7. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in a normed space X. We say that the series $\sum_{n=1}^{\infty} x_n$ is *absolutely convergent* if $\sum_{n=1}^{\infty} \|x_n\| < \infty$. $\diamondsuit$

A convergent series need not converge absolutely. For example, consider $X = \mathbb{R}$ and $x_n = (-1)^n/n$. The *alternating harmonic series* $\sum_{n=1}^{\infty} (-1)^n/n$ converges, but the *harmonic series* $\sum_{n=1}^{\infty} 1/n$ does not.

Also, a series that converges absolutely need not converge. One example in the incomplete space $X = C_c(\mathbb{R})$ is constructed in Problem 1.3.11. The next theorem states that *if X is complete* then every absolutely convergent series in X must converge. Moreover, the converse also holds: In any incomplete normed space there exists a series that converges absolutely yet does not converge, i.e., there exist vectors $x_n \in X$ such that $\sum \|x_n\| < \infty$ but $\sum x_n$ does not converge.

Theorem 1.2.8. *If X is a normed space, then the following two statements are equivalent.*

(a) *X is complete (i.e., X is a Banach space).*

(b) *Every absolutely convergent series in X converges in X. That is, if $\{x_n\}_{n\in\mathbb{N}}$ is a sequence in X and $\sum \|x_n\| < \infty$, then the series $\sum x_n$ converges in X.*

Proof. (a) $\Rightarrow$ (b). We assign the proof of this implication to the reader.

(b) $\Rightarrow$ (a). Suppose that every absolutely convergent series in X is convergent. Let $\{x_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in X. Appealing to Problem 1.1.21, there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ such that $\|x_{n_{k+1}} - x_{n_k}\| < 2^{-k}$ for every $k \in \mathbb{N}$. Consequently, the series $\sum_{k=1}^{\infty} (x_{n_{k+1}} - x_{n_k})$ is absolutely convergent. Therefore, by hypothesis, this series converges in X. Let $x = \sum_{k=1}^{\infty} (x_{n_{k+1}} - x_{n_k})$. Then, by definition, the partial sums

$$s_M = \sum_{k=1}^{M} (x_{n_{k+1}} - x_{n_k}) = x_{n_{M+1}} - x_{n_1}$$

converge to x as $M \to \infty$. Let $y = x + x_{n_1}$. Then, since n_1 is fixed,

$$x_{n_M} = s_{M-1} + x_{n_1} \to x + x_{n_1} = y \qquad \text{as } M \to \infty.$$

Thus $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence that has a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ that converges to the vector y. Appealing now to Problem 1.1.20, this implies that $x_n \to y$. Hence every Cauchy sequence in X converges, so X is complete. $\square$

1.2.4 Equivalent Norms

A vector space X can have many different norms. Some of these norms may be "comparable" in the following sense.

Definition 1.2.9 (Equivalent Norms). We say that two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on a vector space X are are *equivalent* if there exist constants $C_1, C_2 > 0$ such that

$$C_1\,\|x\|_a \;\le\; \|x\|_b \;\le\; C_2\,\|x\|_a, \qquad \text{for all } x \in X. \qquad \diamondsuit$$

The reader should show that if two norms are equivalent, then they determine the same convergence criterion, i.e.,

$$\lim_{n\to\infty} \|x - x_n\|_a \;=\; 0 \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} \|x - x_n\|_b \;=\; 0. \tag{1.2}$$

Conversely, if equation (1.2) holds, then $\|\cdot\|_a$ and $\|\cdot\|_b$ are equivalent (for one proof of this, see [Heil18, Thm. 3.6.2]).

We have the following important fact for *finite-dimensional spaces* (see [Heil18, Thm. 3.7.2]).

Theorem 1.2.10. *If X is a finite-dimensional vector space, then any two norms on X are equivalent.* $\diamondsuit$

One consequence of Theorem 1.2.10 is that *all finite-dimensional subspaces of a normed space are closed* (see [Heil18, Cor. 3.7.3]).

Problems

1.2.11. Let X be a normed space. Prove that every open ball $B_r(x)$ in X is *convex*, i.e., if $x, y \in B_r(x)$ and $0 \le t \le 1$, then $ty + (1-t)z \in B_r(x)$.

1.2.12. Let Y be a subspace of a Banach space X, and let the norm on Y be the norm on X restricted to the set Y. Prove that Y is a Banach space with respect to this norm if and only if Y is a closed subset of X.

1.2.13. Assume that $\sum_{n=1}^{\infty} x_n$ is a convergent infinite series in a normed space X. Prove that

$$\left\|\sum_{n=1}^{\infty} x_n\right\| \;\leq\; \sum_{n=1}^{\infty} \|x_n\|.$$

Note that the right-hand side of this inequality could be ∞.

1.2.14. Let X be a normed space. We define the *closed span* of a set $S \subseteq X$ to be the closure of the span of S, and we denote this closed span by $\overline{\text{span}}(S)$. Prove that $\overline{\text{span}}(S)$ is the *smallest closed subspace* of X that contains S. That is, $\overline{\text{span}}(S)$ is a closed subspace of X, and if M is any other closed subspace such that $S \subseteq M$, then $\overline{\text{span}}(S) \subseteq M$.

1.3 The Uniform Norm

Let X be a metric space. Recall from the Preliminaries that we let the symbol $\overline{\mathbf{F}}$ denote a choice of $[-\infty, \infty]$ or $\mathbb{C}$. We let $C(X)$ be the vector space that consists of all *continuous, scalar-valued functions* on X. Specifically, if $\overline{\mathbf{F}} = \mathbb{C}$, then $C(X)$ is the set of continuous, complex-valued functions on X, while if $\overline{\mathbf{F}} = [-\infty, \infty]$, then $C(X)$ is the set of continuous, real-valued functions on X (we do not allow functions in $C(X)$ to take the values $\pm\infty$). We let $C_b(X)$ be the subspace of all bounded continuous functions on X:

$$C_b(X) \;=\; \bigl\{f \in C(X) : f \text{ is bounded}\bigr\}.$$

If X is compact, then Theorem 1.1.17 implies that $C_b(X) = C(X)$.

To avoid multiplicities of brackets and parentheses, if $X = (a, b)$ then we usually write $C(a, b)$ instead of $C((a, b))$, if $X = [a, b)$ then we write $C[a, b)$ instead of $C([a, b))$, and so forth.

In order to define a norm on $C_b(X)$, we introduce the following terminology.

Definition 1.3.1 (Uniform Norm). Let X be a metric space. The *uniform norm* of a function $f\colon X \to \overline{\mathbf{F}}$ is

$$\|f\|_{\mathrm{u}} \;=\; \sup_{x \in X} |f(x)|. \qquad \diamondsuit \tag{1.3}$$

Note that $\|f\|_{\mathrm{u}}$ is defined for every function on X, although $\|f\|_{\mathrm{u}} = \infty$ if f is unbounded. Therefore $\|f\|_{\mathrm{u}} < \infty$ for all $f \in C_b(X)$, and the reader should check that $\|\cdot\|_{\mathrm{u}}$ is a norm on $C_b(X)$ in the sense of Definition 1.2.3. Hence $C_b(X)$ is a normed vector space.

Convergence with respect to the uniform norm is called *uniform convergence*. That is, f_n *converges uniformly to* f if

$$\lim_{n\to\infty} \|f - f_n\|_{\mathrm{u}} \;=\; \lim_{n\to\infty} \left(\sup_{x \in X} |f(x) - f_n(x)|\right) \;=\; 0.$$

If $f_n \to f$ uniformly, then for each $x \in X$ we have that $f_n(x) \to f(x)$ as $n \to \infty$. Thus *uniform convergence implies pointwise convergence*. However, pointwise convergence does not imply uniform convergence in general (see Example 3.4.1).

The following exercise shows that the uniform limit of a sequence of bounded continuous functions is itself bounded and continuous.

Exercise 1.3.2. Let X be a metric space. Prove that if functions $f_n \in C_b(X)$ converge uniformly to a function $f\colon X \to \overline{\mathbf{F}}$, then $f \in C_b(X)$. $\diamondsuit$

To illustrate a typical completeness argument, we will prove that $C_b(X)$ is complete with respect to the uniform norm (for a more challenging completeness exercise, see Problem 1.4.5).

Theorem 1.3.3 ($C_b(X)$ Is Complete). *Let X be a metric space. If $\{f_n\}_{n\in\mathbb{N}}$ is a sequence in $C_b(X)$ that is Cauchy with respect to $\|\cdot\|_{\mathrm{u}}$, then there exists a function $f \in C_b(X)$ such that f_n converges uniformly to f. Consequently $C_b(X)$ is a Banach space with respect to the uniform norm.*

Proof. Assume that $\{f_n\}_{n\in\mathbb{N}}$ is Cauchy with respect to the uniform norm. If we fix one particular point $x \in X$, then for all m and n we have $|f_m(x) - f_n(x)| \le \|f_m - f_n\|_{\mathrm{u}}$. It follows that $\{f_n(x)\}_{n\in\mathbb{N}}$ is a *Cauchy sequence of scalars*. Since $\mathbb{R}$ and $\mathbb{C}$ are complete, this sequence of scalars must converge. Define $f(x) = \lim_{n\to\infty} f_n(x)$. By construction, f_n converges *pointwise* to f. We will show that f_n converges *uniformly* to f.

Choose any $\varepsilon > 0$. Then there exists an N such that $\|f_m - f_n\|_{\mathrm{u}} < \varepsilon$ for all $m, n \ge N$. Therefore, if $n \ge N$, then for every $x \in X$ we have

$$|f(x) - f_n(x)| \;=\; \lim_{m\to\infty} |f_m(x) - f_n(x)| \;\le\; \limsup_{m\to\infty} \|f_m - f_n\|_{\mathrm{u}} \;\le\; \varepsilon.$$

Taking the supremum over all $x \in X$, we see that $\|f - f_n\|_{\mathrm{u}} \le \varepsilon$ whenever $n \ge N$, so $f_n \to f$ uniformly. Therefore $f \in C_b(X)$ by Exercise 1.3.2. Thus every uniformly Cauchy sequence in $C_b(X)$ converges uniformly to a function in $C_b(X)$, so we conclude that $C_b(X)$ is complete with respect to the uniform norm. $\square$

1.3.1 Some Function Spaces

We will define several vector spaces of functions whose domain is $\mathbb{R}^d$. We have already seen $C(\mathbb{R}^d)$, the space of continuous functions on $\mathbb{R}^d$, and $C_b(\mathbb{R}^d)$, the space of bounded continuous functions on $\mathbb{R}^d$.

We say that $f\colon \mathbb{R}^d \to \overline{\mathbf{F}}$ *vanishes at infinity* if $\lim_{\|x\|\to\infty} f(x) = 0$. Precisely, this means that if $\varepsilon > 0$ is given, then there exists some $R > 0$ such that $|f(x)| < \varepsilon$ for all x with $\|x\| \ge R$. The space of continuous functions that vanish at infinity is

$$C_0(\mathbb{R}^d) \;=\; \Big\{f \in C(\mathbb{R}^d) : \lim_{\|x\|\to\infty} f(x) = 0\Big\}.$$

The *support* of a continuous function f on $\mathbb{R}^d$ is the *closure* in $\mathbb{R}^d$ of the set of points where f is nonzero:

$$\operatorname{supp}(f) \;=\; \overline{\{x \in \mathbb{R}^d : f(x) \neq 0\}}.$$

We say that f has *compact support* if $\operatorname{supp}(f)$ is a compact set. Since $\operatorname{supp}(f)$ is a closed subset of $\mathbb{R}^d$ (by definition), the Heine–Borel Theorem implies that $\operatorname{supp}(f)$ is compact if and only if it is bounded. Hence,

$$f \textit{ has compact support} \quad\Longleftrightarrow\quad f \textit{ is zero outside of some ball } B_r(0).$$

The space of continuous functions with compact support is

$$C_c(\mathbb{R}^d) \;=\; \big\{f \in C(\mathbb{R}^d) : \operatorname{supp}(f) \text{ is compact}\big\}.$$

We have the inclusions $C_c(\mathbb{R}^d) \subsetneq C_0(\mathbb{R}^d) \subsetneq C_b(\mathbb{R}^d) \subsetneq C(\mathbb{R}^d)$. Theorem 1.3.3 showed that $C_b(\mathbb{R}^d)$ is complete with respect to the uniform norm. According to Problems 1.3.7 and 1.3.8, $C_0(\mathbb{R}^d)$ is also complete with respect to the uniform norm, while $C_c(\mathbb{R}^d)$ is not.

We define some related spaces of differentiable functions. Given an integer $m \geq 0$, we let $C^m(\mathbb{R})$ denote the space of m-times differentiable functions f on $\mathbb{R}$ such that $f, f', \dots, f^{(m)}$ are all continuous. $C_b^m(\mathbb{R})$ denotes the subspace that consists of those functions $f \in C^m(\mathbb{R})$ such that $f, f', \dots, f^{(m)}$ are bounded, and $C_c^m(\mathbb{R})$ is the space of functions $f \in C^m(\mathbb{R})$ that have compact support. $C^\infty(\mathbb{R})$ is the space of infinitely differentiable functions on $\mathbb{R}$, and $C_c^\infty(\mathbb{R})$ is the subspace of infinitely differentiable, compactly supported functions.

We also state a classical result on the approximation of continuous functions by polynomials on a finite interval. There are many different proofs of this theorem; one can be found in [Heil18, Thm. 4.6.2].

Theorem 1.3.4 (Weierstrass Approximation Theorem). *Let $[a,b]$ be a finite closed interval. If $f \in C[a,b]$ and $\varepsilon > 0$, then there exists a polynomial $p(x) = \sum_{k=0}^n c_k x^k$ such that*

$$\|f - p\|_{\mathrm{u}} \;=\; \sup_{x \in [a,b]} |f(x) - p(x)| \;<\; \varepsilon. \qquad \diamondsuit$$

Problems

1.3.5. Let I be an interval in $\mathbb{R}$. For each $k \geq 0$, define $p_k(x) = x^k$. Prove that $\{p_k\}_{k\geq 0}$ is a linearly independent set in $C(I)$.

1.3.6. Prove that $f\colon \mathbb{R}^d \to \mathbb{C}$ is uniformly continuous on $\mathbb{R}^d$ if and only if

$$\lim_{a\to 0} \|T_a f - f\|_{\mathrm{u}} \;=\; 0,$$

where $T_a f(x) = f(x-a)$ denotes the translation of f by $a \in \mathbb{R}^d$.

1.3.7. Prove that $C_0(\mathbb{R})$ is a Banach space with respect to the uniform norm. Show that every function in $C_0(\mathbb{R})$ is uniformly continuous, and exhibit a function in $C_b(\mathbb{R})$ that is not uniformly continuous.

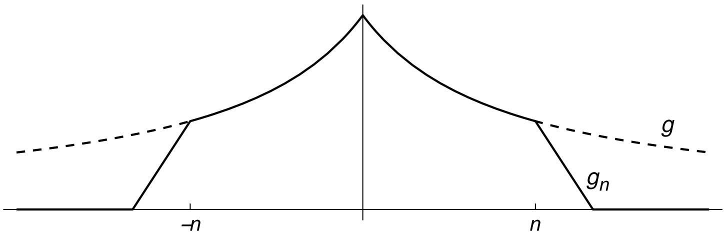


Fig. 1.1 A function g and a compactly supported approximation g_n.

1.3.8. Let $g \in C_0(\mathbb{R})$ be any function that does not belong to $C_c(\mathbb{R})$. For each integer $n > 0$, define a compactly supported approximation to g by setting $g_n(x) = g(x)$ for $|x| \le n$ and $g_n(x) = 0$ for $|x| > n+1$, and let g_n be linear on $[n, n+1]$ and $[-n-1, -n]$ (see Figure 1.1). Show that $\{g_n\}_{n\in\mathbb{N}}$ is Cauchy in $C_c(\mathbb{R})$ with respect to the uniform norm, but it does not converge uniformly to any function in $C_c(\mathbb{R})$. Conclude that $C_c(\mathbb{R})$ is not complete with respect to $\|\cdot\|_{\mathrm{u}}$, and is not a closed subset of $C_0(\mathbb{R})$.

1.3.9. Prove that $C_c(\mathbb{R})$ is a *dense* subspace of $C_0(\mathbb{R})$ with respect to the uniform norm. That is, show that if $g \in C_0(\mathbb{R})$, then there exist functions $g_n \in C_c(\mathbb{R})$ such that $g_n \to g$ uniformly.

1.3.10. The *unit disk* D in $C_b(\mathbb{R})$ is the set of all functions in $C_b(\mathbb{R})$ whose uniform norm is at most 1, i.e., $D = \{f \in C_b(\mathbb{R}) : \|f\|_{\mathrm{u}} \;\le\; 1\}$.

(a) Prove that D is a closed and bounded subset of $C_b(\mathbb{R})$.

(b) The *hat function* or *tent function* on the interval $[-1,1]$ is

$$W(x) \;=\; \max\bigl\{1-|x|, 0\bigr\} \;=\; \begin{cases} 1-x, & \text{if } 0 \le x \le 1, \\ 1+x, & \text{if } -1 \le x \le 0, \\ 0, & \text{if } |x| \ge 1. \end{cases}$$

Let $f_k(x) = W(x-k)$. Observe that $\|f_k\|_{\mathrm{u}} = 1$, so the sequence $\{f_k\}_{k\in\mathbb{N}}$ is contained in the unit disk D. Prove that $\{f_k\}_{k\in\mathbb{N}}$ is not a Cauchy sequence and contains no Cauchy subsequences.

(c) Prove that D is not a compact subset of $C_b(\mathbb{R})$.

1.3.11. Consider $C_c(\mathbb{R})$, which is a normed space with respect to the uniform norm. Let W be the hat function defined in Problem 1.3.10, and let $g_k(x) = 2^{-k}\,W(2^{-k}x)$. Using the uniform norm, prove that the series $\sum_{k=1}^{\infty} g_k$ converges absolutely in $C_c(\mathbb{R})$, but it does not converge in $C_c(\mathbb{R})$. What happens if we replace $C_c(\mathbb{R})$ with $C_0(\mathbb{R})$?

1.4 Hölder and Lipschitz Continuity

Sometimes we deal with functions that are "better than continuous" yet are "not quite differentiable." The next definition gives one way to quantify behavior that lies between continuity and differentiability.

Definition 1.4.1 (Hölder and Lipschitz Continuous Functions). Let I be an interval in the real line, and let $f\colon I \to \mathbb{C}$ be a function on I.

(a) We say that f is *Hölder continuous on I with exponent $\alpha > 0$* if there exists a constant $K \ge 0$ such that

$$|f(x) - f(y)| \;\le\; K\,|x-y|^{\alpha}, \qquad \text{for all } x, y \in I.$$

(b) If f is Hölder continuous with exponent $\alpha = 1$, then we say that f is *Lipschitz continuous on I*, or simply that f is *Lipschitz*. That is, f is Lipschitz if there exists a constant $K \ge 0$ such that

$$|f(x) - f(y)| \;\le\; K\,|x-y|, \qquad \text{for all } x, y \in I.$$

A number K for which this holds is called a *Lipschitz constant* for f. ◇

By using the Mean Value Theorem, we can see that any function $f\colon I \to \mathbb{C}$ that is differentiable everywhere on I and has a bounded derivative f' is Lipschitz on I (this is Problem 1.4.2). However, a Lipschitz function need not be differentiable at every point. For example, $f(x) = |x|$ is Lipschitz on $[-1,1]$ but it is not differentiable at $x = 0$.

Lipschitz functions will appear frequently in the text. In Chapter 5 we will prove that every Lipschitz function on $[a,b]$ has *bounded variation* and is *absolutely continuous*. We will encounter Hölder continuous functions with exponents $\alpha < 1$ less frequently. The *Cantor–Lebesgue function*, which will be introduced in Section 5.1, is one important example of a Hölder continuous function that is not Lipschitz.

Problems

1.4.2. Let I be an interval. Show that if $f\colon I \to \mathbb{C}$ is differentiable everywhere on I and f' is bounded on I, then f is Lipschitz on I.

Remark: The Mean Value Theorem is directly applicable if f is real-valued. However, the MVT does not hold for complex-valued functions, e.g., consider $f(x) = e^{ix}$ on $[0, 2\pi]$.

1.4.3. Define $h\colon [-1,1] \to \mathbb{R}$ by $h(x) = x^2 \sin \frac{1}{x}$ if $x \neq 0$, and $h(0) = 0$. Prove that h is Lipschitz on $[-1,1]$.

1.4.4. Prove the following statements.

(a) If f is Hölder continuous on an interval I for some exponent $\alpha > 0$, then f is uniformly continuous on I.

(b) If f is Hölder continuous on an interval I for some exponent $\alpha > 1$, then f is constant on I.

(c) The function $f(x) = |x|^{1/2}$ is Hölder continuous on $[-1,1]$ for exponents $0 < \alpha \le 1/2$, but not for any exponent $\alpha > 1/2$.

(d) The function g defined by $g(x) = -1/\ln x$ for $x > 0$ and $g(0) = 0$ is uniformly continuous on $[0, 1/2]$, but it is not Hölder continuous for any exponent $\alpha > 0$.

1.4.5. Let I be an interval in $\mathbb{R}$.

(a) Fix $0 < \alpha < 1$, and let $C^\alpha(I)$ be the space of all bounded functions that are Hölder continuous with exponent α on I, i.e.,

$$C^\alpha(I) \;=\; \bigl\{f \in C_b(I) : f \text{ is Hölder continuous with exponent } \alpha\bigr\}.$$

Show that the following is a norm on $C^\alpha(I)$, and $C^\alpha(I)$ is a Banach space with respect to this norm:

$$\|f\|_{C^\alpha} \;=\; \|f\|_{\mathrm{u}} \;+\; \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x-y|^\alpha}.$$

(b) To avoid confusion with the space $C^1(I)$, which consists of those differentiable functions on I whose derivative is continuous, we let $\mathrm{Lip}(I)$ denote the space of bounded functions that are Lipschitz on I. Extend the results of part (a) to to $\mathrm{Lip}(I)$.

Chapter 2
Lebesgue Measure

We know how to determine the volume of cubes, rectangles, spheres, and some other special subsets of $\mathbb{R}^d$. Does every subset of $\mathbb{R}^d$ have a volume? We are tempted to believe that each set $E \subseteq \mathbb{R}^d$ can be assigned a unique "volume" or "measure" $|E|$ in such a way that the following properties hold:

(i) $0 \le |E| \le \infty$,

(ii) the measure of the unit cube $Q = [0,1]^d$ is $|Q| = 1$,

(iii) if $E_1, E_2, \dots$ are finitely or countably many disjoint subsets of $\mathbb{R}^d$, then

$$\Bigl|\bigcup_k E_k\Bigr| \;=\; \sum_k |E_k|,$$

(iv) $|E + h| = |E|$ for all $h \in \mathbb{R}^d$.

We will prove in Section 2.4 that there is *no way* to define $|E|$ so that all four conditions (i)–(iv) simultaneously hold for every set $E \subseteq \mathbb{R}^d$! (This turns out to be a consequence of the Axiom of Choice; see Theorem 2.4.4.) Even so, we will prove in this chapter that if we relax our goal of defining a volume for *every* subset of $\mathbb{R}^d$, then we can create a useful definition of measure that satisfies properties (i)–(iv) for a very large class of subsets of $\mathbb{R}^d$. This class of "good sets," which we will call the *measurable* subsets of $\mathbb{R}^d$, includes almost every set that we ever encounter in practice. The "volume" $|E|$ that we will define is called the *Lebesgue measure* of the set E; we will show that it is well-defined and "nicely behaved" on the class of measurable subsets of $\mathbb{R}^d$.

The creation of Lebesgue measure is a two-step process, broadly outlined as follows. First, we start with a basic class of subsets of $\mathbb{R}^d$ that we know how we want to measure. There are several choices for this class, but perhaps the simplest is the collection of rectangular boxes (rectangular parallelepipeds) in $\mathbb{R}^d$. The volume of a rectangular box is just the product of the lengths of its sides. We attempt to extend the notion of volume to arbitrary subsets of $\mathbb{R}^d$ by covering them with rectangular boxes in all possible ways. For each set $E \subseteq \mathbb{R}^d$, this gives us a number $|E|_e$ that we call the *exterior Lebesgue*

C. Heil, *Introduction to Real Analysis*, Graduate Texts in Mathematics 280,
https://doi.org/10.1007/978-3-030-26903-6_2

measure of E. Every subset of $\mathbb{R}^d$ has a uniquely defined exterior measure, and the function $|\cdot|_e$ satisfies properties (i), (ii), and (iv) from our list above for every set E. However, there exist *disjoint* sets A and B in $\mathbb{R}^d$ such that $|A \cup B|_e < |A|_e + |B|_e$! Thus exterior Lebesgue measure does not satisfy property (iii) for all choices of disjoint subsets of $\mathbb{R}^d$.

Consequently, we take a second step and construct a class $\mathcal{L}$ of "good subsets" of $\mathbb{R}^d$ such that the number $|E| = |E|_e$ satisfies properties (i)–(iv) for all sets *in the class* $\mathcal{L}$. The sets in this class are called the *measurable sets*, and for a measurable set E the number $|E| = |E|_e$ is called the *Lebesgue measure* of E. All open and closed sets turn out to be measurable, the complement of a measurable set is measurable, and the countable union or countable intersection of measurable sets is measurable. Thus, if we begin with some sets that we know are measurable, such as the open and closed sets, and repeatedly apply the operations of complements, countable unions, and countable intersections, then we obtain measurable sets. This is how most of the sets that we encounter in practice are constructed, so in this sense the class of measurable sets is quite satisfactory.

In this chapter we construct Lebesgue measure and examine its properties. Then in Chapters 3 and 4 we develop the theory of integration with respect to Lebesgue measure. Just as we must restrict our attention to measurable sets, we also must restrict to functions that are measurable in a certain sense. Fortunately, this includes most of the functions that we see in practical contexts. We will see numerous applications of the Lebesgue integral in Chapters 5 and 6, when we consider local and global properties of functions related to continuity and differentiation; in Chapter 7, when we discuss the L^p spaces; in Chapter 8 when we specialize to L^2 spaces; and in Chapter 9, when we discuss convolution, the Fourier transform, and Fourier series.

The domains of most of the functions that we will encounter in this chapter will be $\mathbb{R}^d$ or a subset of $\mathbb{R}^d$. We adopt the Euclidean norm as our "default norm" on $\mathbb{R}^d$. As we stated in the Preliminaries, the Euclidean norm of a point $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ will be denoted by

$$\|x\| \;=\; \bigl(|x_1|^2 + \cdots + |x_d|^2\bigr)^{1/2},$$

and the open ball in $\mathbb{R}^d$ centered at x with radius r is

$$B_r(x) \;=\; \bigl\{y \in \mathbb{R}^d : \|x - y\| < r\bigr\}.$$

2.1 Exterior Lebesgue Measure

In this section we take the first step in the construction of Lebesgue measure, which is to define the *exterior Lebesgue measure* of each subset of $\mathbb{R}^d$.

2.1.1 Boxes

We begin with some especially simple sets whose volumes are known. These are intervals in one dimension, rectangles in two dimensions, and rectangular parallelepipeds in higher dimensions. In fact, we will restrict to rectangular parallelepipeds whose sides are parallel to the coordinate axes. For simplicity, we refer to these sets as "boxes." Here is the precise definition of a box and its volume.

Definition 2.1.1 (Boxes).

(a) A *box* in $\mathbb{R}^d$ is a Cartesian product of d finite closed intervals. In other words, a box is a set of the form

$$Q \;=\; [a_1,b_1]\times\cdots\times[a_d,b_d] \;=\; \prod_{j=1}^{d}[a_j,b_j], \tag{2.1}$$

where $a_j < b_j$ for each j.

(b) The *volume* of the box Q defined in equation (2.1) is the product of the lengths of its sides:

$$\operatorname{vol}(Q) \;=\; (b_1-a_1)\cdots(b_d-a_d) \;=\; \prod_{j=1}^{d}(b_j-a_j).$$

(c) The *interior* of the box Q is the Cartesian product

$$Q^\circ \;=\; (a_1,b_1)\times\cdots\times(a_d,b_d) \;=\; \prod_{j=1}^{d}(a_j,b_j),$$

and the *boundary* of Q is $\partial Q = Q\backslash Q^\circ$.

(d) If the sidelengths $b_j - a_j$ of the box Q are all equal, then we call Q a *cube.* $\diamondsuit$

A "box" will always mean a set of the form given in equation (2.1). In one dimension, a box is a finite closed interval and its volume is its length. In $\mathbb{R}^2$ a box is a rectangle whose sides are parallel to the coordinate axes and its volume is its area. All boxes are closed and bounded, and therefore boxes are nonempty compact subsets of $\mathbb{R}^d$. Because we require $a_j < b_j$ for every j, our boxes all have nonempty interiors, and they have strictly positive (and finite) volumes.

We will encounter many different configurations of collections of boxes. Sometimes boxes will be allowed to overlap, sometimes they will be required to be disjoint, and sometimes we will allow them to overlap as long as they only intersect at their boundaries. We use the following terminology to describe this last type of configuration.

Definition 2.1.2 (Nonoverlapping Boxes). We say that a collection of boxes $\{Q_k\}_{k\in I}$ is *nonoverlapping* if their interiors are disjoint, i.e., if

$$j \neq k \in I \quad \Longrightarrow \quad Q_j^\circ \cap Q_k^\circ = \varnothing. \qquad \diamondsuit$$

We will usually only consider collections of *countably many* boxes. A countable collection can be either finite or countably infinite, and we will need to deal with both possibilities simultaneously. Therefore we introduce the following notational convention.

Notation 2.1.3 (Countable Collections of Boxes). When working with boxes, the notations $\{Q_k\}$ or $\{Q_k\}_k$ will implicitly denote *countable* collections of boxes. That is, $\{Q_k\}$ will denote a family that has one of the forms $\{Q_k\}_{k\in\mathbb{N}}$ or $\{Q_k\}_{k=1}^N$, where N is a positive integer. $\diamondsuit$

We will often consider collections of boxes whose union contains a set E. As we specify in the following definition, such a family is called a *cover* of E.

Definition 2.1.4. We say that a set $E \subseteq \mathbb{R}^d$ is *covered* by a collection of boxes $\{Q_k\}$ if

$$E \subseteq \bigcup_k Q_k. \qquad \diamondsuit$$

2.1.2 Some Facts about Boxes

Every open subset of $\mathbb{R}$ can be written as a union of at most countably many disjoint open intervals. Bounded open intervals in $\mathbb{R}$ are one-dimensional open balls, so every bounded open subset of $\mathbb{R}$ can be written as a union of at most countably many disjoint open balls. This fact does not generalize to higher dimensions. For example, the open square $S = (0,1)^2$ in $\mathbb{R}^2$ cannot be written as a union of countably many disjoint open balls.

Although we cannot write open sets as disjoint unions of balls in general, the following lemma provides us with a useful substitute. According to this lemma, every open set in $\mathbb{R}^d$, in any dimension $d \geq 1$, can be written as a union of *countably many nonoverlapping cubes.* Two easy examples in one dimension (where cubes are simply finite closed intervals) are

$$\mathbb{R} = \bigcup_{k\in\mathbb{Z}} [k, k+1] \qquad \text{and} \qquad (0,\infty) = \bigcup_{k\in\mathbb{Z}} \bigl[2^k, 2^{k+1}\bigr].$$

Since any *finite* union of cubes is a compact set, there is no way that we can write an open set as a union of finitely many cubes. On the other hand, the next lemma shows that we will never need more than *countably many* cubes.

Lemma 2.1.5. *If U is a nonempty open subset of $\mathbb{R}^d$, then there exist countably many nonoverlapping cubes $\{Q_k\}_{k\in\mathbb{N}}$ such that $U = \cup Q_k$.*

Proof. Let $Q = [0,1]^d$, and for each $n \in \mathbb{Z}$ and $k \in \mathbb{Z}^d$ set

$$Q_{n,k} = 2^{-n}Q + 2^{-n}k.$$

If we fix an $n \in \mathbb{Z}$, then the collection $\{Q_{n,k}\}_{k \in \mathbb{Z}^d}$ is a cover of $\mathbb{R}^d$ by nonoverlapping cubes that have sidelengths 2^{-n}.

Let U be a nonempty open set. We will choose from the boxes $Q_{n,k}$ to create a set of nonoverlapping cubes whose union is U. First, we identify the cubes $Q_{0,k}$ with sidelength 1 that are completely contained in U. Specifically, we set

$$I_0 = \bigl\{k \in \mathbb{Z}^d : Q_{0,k} \subseteq U\bigr\}.$$

Then we let I_1 consist of all indices $k \in \mathbb{Z}^d$ such that $Q_{1,k}$ is contained in U but $Q_{1,k}$ is not contained in any cube $Q_{0,j}$ with $j \in I_0$. We continue in this way to collect smaller and smaller cubes. This gives us a collection of nonoverlapping cubes $Q_{n,k}$ that are contained in U. Every point $x \in U$ belongs to at least one such cube (why?). Consequently,

$$U = \bigcup_{n \geq 0} \bigcup_{k \in I_n} Q_{n,k}. \qquad \square$$

It seems "obvious" that the volume of a box Q that is the union of finitely many nonoverlapping boxes $Q_1, \dots, Q_n$ must equal the sum of the volumes of $Q_1, \dots, Q_n$. Later we will see several examples of statements that seem "obviously true" yet turn out to be false. Fortunately, when we are only dealing with *finitely many boxes*, most statements that seem obvious are indeed true. This is the case in the next lemma. On the other hand, the *proof* of this "obvious" statement is more technical than might be expected at first glance.

Lemma 2.1.6. *Let $Q = \prod_{j=1}^d [a_j, b_j]$ be a box in $\mathbb{R}^d$. If $Q_1, \dots, Q_n$ are nonoverlapping boxes such that $Q = Q_1 \cup \cdots \cup Q_n$, then*

$$\operatorname{vol}(Q) = \sum_{k=1}^{n} \operatorname{vol}(Q_k). \tag{2.2}$$

Proof. First consider the special case where the boxes $Q_1, \dots, Q_n$ form a grid-like cover of Q of the type shown in Figure 2.1 for dimension $d = 2$.

If $d = 1$, then this grid-like cover simply corresponds to writing

$$[a,b] = [a_1,b_1] \cup \cdots \cup [a_n,b_n],$$

where

$$a = a_1 < b_1 = a_2 < b_2 = \cdots = a_n < b_n = b.$$

In this case the length of $[a,b]$ equals the sum of the lengths of the intervals $[a_j, b_j]$, and the result follows.

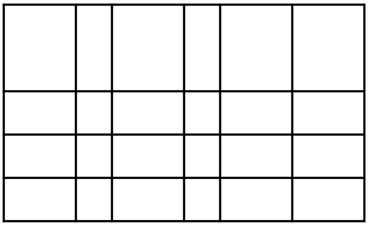


Fig. 2.1 Boxes $Q_1, \dots, Q_n$ that form a grid-like cover of Q.

For $d = 2$, the box Q has the form $I \times J$ for some closed intervals I and J, and the grid-like arrangement in Figure 2.1 corresponds to writing I and J as unions of nonoverlapping closed subintervals, say $I = I_1 \cup \cdots \cup I_M$ and $J = J_1 \cup \cdots \cup J_N$. Then

$$\begin{aligned} \mathrm{vol}(Q) \;=\; \mathrm{vol}(I)\,\mathrm{vol}(J) \;&=\; \left(\sum_{j=1}^{M} \mathrm{vol}(I_j)\right)\left(\sum_{k=1}^{N} \mathrm{vol}(J_k)\right) \\ &=\; \sum_{j=1}^{M}\sum_{k=1}^{N} \mathrm{vol}(I_j)\,\mathrm{vol}(J_k) \\ &=\; \sum_{j=1}^{M}\sum_{k=1}^{N} \mathrm{vol}(I_j \times J_k), \end{aligned}$$

and so equation (2.2) holds. The result then extends to higher dimensions by induction.

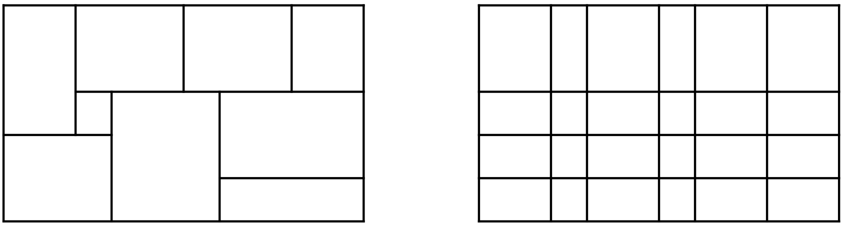


Fig. 2.2 Left: A generic collection of boxes $Q_1, \dots, Q_n$ whose union is a box Q. Right: The sides of the boxes $Q_1, \dots, Q_n$ are extended to form a grid-like cover of Q.

Now let $Q_1, \dots, Q_n$ be any collection of finitely many nonoverlapping boxes whose union is Q. This is the type of arrangement that appears in the left-hand side of Figure 2.2. As in the right-hand side of Figure 2.2, extend the sides of each of the boxes Q_k. This gives us a set of boxes $R_1, \dots, R_m$ (with $m \geq n$) that are in the grid-like configuration discussed before. Applying our previous work, we obtain

$$\operatorname{vol}(Q) = \sum_{j=1}^{m} \operatorname{vol}(R_j).$$

Now, each of the original boxes Q_k is a union of a distinct subset of the boxes $R_1, \dots, R_m$, say $Q_k = \cup_{\ell \in L_k} R_\ell$ where the sets $L_1, \dots, L_n$ form a partition of $\{1, \dots, m\}$. Again applying the argument for grid-like arrangements, for each k we have

$$\operatorname{vol}(Q_k) = \sum_{\ell \in L_k} \operatorname{vol}(R_\ell).$$

Consequently,

$$\sum_{k=1}^{n} \operatorname{vol}(Q_k) = \sum_{k=1}^{n} \sum_{\ell \in L_k} \operatorname{vol}(R_\ell) = \sum_{j=1}^{m} \operatorname{vol}(R_j). \qquad \square$$

An extension of Lemma 2.1.6 shows that the sum of the volumes of finitely many nonoverlapping boxes that cover a box Q must be at least as large as the volume of Q. We assign this proof as the following exercise.

Exercise 2.1.7. Let $Q = \prod_{j=1}^{d} [a_j, b_j]$ be a box in $\mathbb{R}^d$, and assume that $Q_1, \dots, Q_n$ are nonoverlapping boxes such that $Q \subseteq Q_1 \cup \cdots \cup Q_n$. Prove that

$$\operatorname{vol}(Q) \le \sum_{k=1}^{n} \operatorname{vol}(Q_k). \qquad \diamondsuit$$

2.1.3 Exterior Lebesgue Measure

Now we turn from boxes to generic subsets of $\mathbb{R}^d$. In order to define the measure of a set $E \subseteq \mathbb{R}^d$, we will try to approximate it by boxes. Suppose that we cover E by some countable collection of boxes $\{Q_k\}$, so we have

$$E \subseteq \bigcup_k Q_k.$$

We have not yet assigned a measure to either of E or $\cup Q_k$, but whatever those measures are, it seems reasonable to expect that the measure of $\cup Q_k$ should be at least as large as the measure of E. Additionally, it seems reasonable that the measure of a union of boxes should be no more than the sum of the volumes of the boxes Q_k. The measure of the union could be smaller than the sum of the volumes due to overlaps, but we should at least have an inequality. Hence, whatever we decide that the measure of E should be, if we let $|E|_e$ denote that measure then we should have

$$|E|_e \le \sum_k \operatorname{vol}(Q_k).$$

Thus, each covering of E by boxes gives us an upper bound for the measure of E. Some coverings may be "better" than others in some sense, but instead of worrying about how to quantify "better," we will simply take *every possible covering* into account and *declare* that the exterior measure of E is the infimum of $\sum \operatorname{vol}(Q_k)$ over every countable covering of E by boxes (we restrict our attention to coverings by *countably many* boxes because each box has a strictly positive volume). This leads us to the following definition.

Definition 2.1.8 (Exterior Lebesgue Measure). The *exterior Lebesgue measure* (or the *outer Lebesgue measure*) of a set $E \subseteq \mathbb{R}^d$ is

$$|E|_e = \inf\Bigl\{\sum_k \operatorname{vol}(Q_k)\Bigr\},$$

where the infimum is taken over all countable collections of boxes $\{Q_k\}$ such that $E \subseteq \cup Q_k$. $\diamondsuit$

For simplicity, we often abbreviate "exterior Lebesgue measure" just as "exterior measure." *Every* subset E of $\mathbb{R}^d$ has a well-defined exterior measure $|E|_e$ that lies in the range $0 \le |E|_e \le \infty$. By the definition of an infimum, we immediately obtain the following facts.

Lemma 2.1.9. *Let E be any subset of $\mathbb{R}^d$.*

(a) *If $\{Q_k\}$ is* **any** *countable cover of E by boxes, then*

$$|E|_e \le \sum_k \operatorname{vol}(Q_k). \tag{2.3}$$

(b) *If $\varepsilon > 0$, then there exists* **some** *countable cover $\{Q_k\}$ of E by boxes such that*

$$\sum_k \operatorname{vol}(Q_k) \le |E|_e + \varepsilon. \qquad \diamondsuit \tag{2.4}$$

Note that in either of equations (2.3) or (2.4), the exterior measure $|E|_e$ could be infinite. By definition, if E is a *bounded* subset of $\mathbb{R}^d$ then E is contained inside some ball of finite radius. Taking Q to be a box that contains this ball, we see that $\{Q\}$ is a collection of one box that covers E. Part (a) of Lemma 2.1.9 therefore implies that

$$|E|_e \le \operatorname{vol}(Q) < \infty.$$

Thus all bounded sets have finite exterior measure.

Here is an example of an unbounded subset of $\mathbb{R}$ that has finite measure.

Example 2.1.10. A box in $\mathbb{R}$ is just a finite closed interval, so $Q_k = [k, k+2^{-k}]$ is a box. Set

$$E \;=\; \bigcup_{k=1}^{\infty} [k, k+2^{-k}].$$

Since E is not contained in any finite interval, it is unbounded. On the other hand, $\{Q_k\}_{k\in\mathbb{N}}$ is a countable covering of E by boxes, so Lemma 2.1.9(a) implies that

$$|E|_e \;\le\; \sum_{k=1}^{\infty} \mathrm{vol}(Q_k) \;=\; \sum_{k=1}^{\infty} 2^{-k} \;=\; 1.$$

Thus E has finite exterior measure, even though it is unbounded. We cannot prove it yet, but later we will see that the exterior measure of E is precisely $|E|_e = 1$. $\diamondsuit$

Next we prove some basic properties of exterior measure.

Lemma 2.1.11. (a) *Exterior Lebesgue measure is* translation-invariant*, i.e., for every set $E \subseteq \mathbb{R}^d$ and every vector $h \in \mathbb{R}^d$ we have*

$$|E+h|_e \;=\; |E|_e.$$

(b) *Exterior Lebesgue measure is* monotonic*, i.e., if $A, B \subseteq \mathbb{R}^d$, then*

$$A \subseteq B \implies |A|_e \;\le\; |B|_e.$$

(c) $|\varnothing|_e = 0$.

(d) *If E is a countable subset of $\mathbb{R}^d$, then $|E|_e = 0$.*

Proof. (a) If $\{Q_k\}_k$ is any countable cover of E by boxes, then $\{Q_k + h\}_k$ is a countable cover of $E+h$ by boxes. Lemma 2.1.11(a) therefore implies that

$$|E+h|_e \;\le\; \sum_k \mathrm{vol}(Q_k + h) \;=\; \sum_k \mathrm{vol}(Q_k).$$

This is true for every covering of E, so we conclude that $|E+h|_e \le |E|_e$. The opposite inequality is entirely symmetric.

(b) Suppose that $A \subseteq B$, and let $\{Q_k\}_k$ be any countable cover of B by boxes. Then $\{Q_k\}_k$ is also a countable cover of A by boxes, so

$$|A|_e \;\le\; \sum_k \mathrm{vol}(Q_k).$$

This is true for every possible covering of B, so

$$|A|_e \;\le\; \inf\Bigl\{\sum_k \mathrm{vol}(Q_k) : \text{all covers of } B \text{ by boxes}\Bigr\} \;=\; |B|_e.$$

(c) If Q is a box, then Q covers $\varnothing$, no matter how small we choose the sides of Q. Therefore $|\varnothing|_e \le \operatorname{vol}(Q)$, and $\operatorname{vol}(Q)$ can be arbitrarily small.

(d) Let $E = \{x_k\}$ be a countable subset of $\mathbb{R}^d$. For each k, let Q_k be a box with volume $\varepsilon/2^k$ that contains x_k. Then $\{Q_k\}_k$ covers E, so

$$|E|_e \;\le\; \sum_k \operatorname{vol}(Q_k) \;\le\; \varepsilon \sum_k \frac{1}{2^k} \;\le\; \varepsilon.$$

Since ε is arbitrary, we conclude that $|E|_e = 0$. $\square$

Since the set of rationals $\mathbb{Q}$ is a countable subset of $\mathbb{R}$, Lemma 2.1.11(d) implies that its exterior measure is zero. Thus $\mathbb{Q}$ is a "very small" part of $\mathbb{R}$ in a measure-theoretic sense. This contrasts with the fact that $\mathbb{Q}$ is dense in $\mathbb{R}$ and therefore is a "very large" part of $\mathbb{R}$ in a topological sense. A set and its closure can have very different exterior measures!

While every countable set has zero exterior measure, there also exist *uncountable* subsets of $\mathbb{R}^d$ whose exterior measure is zero. We will see examples of such sets in Lemma 2.1.21 (for dimensions $d \ge 2$) and in Example 2.1.23 (for dimension $d = 1$).

Remark 2.1.12. We will prove in Theorem 2.1.17 that if Q is a box then $|Q|_e = \operatorname{vol}(Q)$. That is, the exterior measure of a box equals its volume in the usual sense. This is not yet obvious; in fact, a challenge is to try to prove, using only the definition of exterior measure, that the exterior measure of the closed interval $[0,1]$ is 1, or even that it is nonzero. One difficulty in this regard is that Lemma 2.1.6 and Exercise 2.1.7 only apply to finite collections of boxes, whereas the definition of exterior measure involves all possible coverings by *countably many* boxes. $\diamondsuit$

Our next theorem shows that the exterior measure of a countable union of sets is no more than the sum of the exterior measures of these sets (this is called the *countable subadditivity* property of exterior Lebesgue measure). The sets here are not required to be disjoint, so we could very well have strict inequality because of overlaps or duplications of sets. We might expect that if the sets involved are disjoint then the measure of their union will equal the sums of the measures of the sets, but this does not always hold! In particular, we will see in Example 2.4.7 that there exist disjoint sets A and B such that $|A \cup B|_e < |A|_e + |B|_e$.

Theorem 2.1.13 (Countable Subadditivity). *If $E_1, E_2, \dots$ are countably many sets in $\mathbb{R}^d$, then*

$$\left|\bigcup_{k=1}^{\infty} E_k\right|_e \;\le\; \sum_{k=1}^{\infty} |E_k|_e. \tag{2.5}$$

Proof. If any particular set E_k has infinite exterior measure then both sides of equation (2.5) are ∞, so we are done in this case. Therefore, assume that

$|E_k|_e < \infty$ for every k, and fix $\varepsilon > 0$. By Lemma 2.1.9, for each k we can find a covering $\{Q_j^{(k)}\}_j$ of E_k by countably many boxes such that

$$\sum_j \operatorname{vol}(Q_j^{(k)}) \ \le\ |E_k|_e + \frac{\varepsilon}{2^k}. \tag{2.6}$$

Then $\{Q_j^{(k)}\}_{j,k}$ is a covering of $\cup_k E_k$ by countably many boxes, so

$$\begin{aligned} \biggl|\bigcup_{k=1}^{\infty} E_k\biggr|_e &\le \sum_{k=1}^{\infty}\sum_j \operatorname{vol}(Q_j^{(k)}) && \text{(by Lemma 2.1.9)} \\ &\le \sum_{k=1}^{\infty}\Bigl(|E_k|_e + \frac{\varepsilon}{2^k}\Bigr) && \text{(by equation (2.6))} \\ &= \biggl(\sum_{k=1}^{\infty} |E_k|_e\biggr) + \varepsilon. \end{aligned}$$

Since ε is arbitrary, the result follows. $\square$

By setting $E_k = \varnothing$ for $k > N$, a corollary of Theorem 2.1.13 is that exterior Lebesgue measure is *finitely subadditive*, i.e., if $E_1, \dots, E_N$ are finitely many sets in $\mathbb{R}^d$, then

$$\biggl|\bigcup_{k=1}^{N} E_k\biggr|_e \ \le\ \sum_{k=1}^{N} |E_k|_e.$$

However, subadditivity need not hold for *uncountable* collections of sets. For example, the real line is an uncountable union of singletons,

$$\mathbb{R} \ =\ \bigcup_{x\in\mathbb{R}} \{x\},$$

and the exterior measure of each singleton $\{x\}$ is zero, yet we will see in Corollary 2.1.19 that $|\mathbb{R}|_e = \infty$.

The following definition introduces some terminology for sets that we will need later in the text.

Definition 2.1.14 (Limsup and Liminf of Sets). If $\{E_k\}_{k\in\mathbb{N}}$ is a sequence of subsets of $\mathbb{R}^d$, then we define

$$\limsup_{k\to\infty} E_k \ =\ \bigcap_{j=1}^{\infty}\biggl(\bigcup_{k=j}^{\infty} E_k\biggr) \qquad \text{and} \qquad \liminf_{k\to\infty} E_k \ =\ \bigcup_{j=1}^{\infty}\biggl(\bigcap_{k=j}^{\infty} E_k\biggr). \qquad \diamondsuit$$

Exercise 2.1.15. Given sets $E_k \subseteq \mathbb{R}^d$, prove the following statements.

(a) $\limsup E_k$ consists of those points $x \in \mathbb{R}^d$ that belong to *infinitely many* of the E_k.

(b) $\liminf E_k$ consists of those x which belong to *all but finitely many* E_k (i.e., there exists some $k_0 \in \mathbb{N}$ such that $x \in E_k$ for all $k \geq k_0$). $\diamondsuit$

The proof of the following result is an application of countable subadditivity.

Exercise 2.1.16 (Borel–Cantelli Lemma). Suppose that sets $E_k \subseteq \mathbb{R}^d$ satisfy $\sum |E_k|_e < \infty$. Prove that $\liminf E_k$ and $\limsup E_k$ each have exterior measure zero. $\diamondsuit$

2.1.4 The Exterior Measure of a Box

We expect that the exterior measure of a box should coincide with its volume, but we have not proved this yet. Since we can cover a box Q by the collection $\{Q\}$ that contains the single box Q, we do obtain the *inequality* $|Q|_e \leq \mathrm{vol}(Q)$ directly from Definition 2.1.8. However, the opposite inequality is not trivial.

Theorem 2.1.17 (Consistency with Volume). *If Q is a box in $\mathbb{R}^d$, then*

$$|Q|_e = \mathrm{vol}(Q).$$

Proof. As noted above, we have the inequality $|Q|_e \leq \mathrm{vol}(Q)$. To prove the converse inequality, let $\{Q_k\}$ be any covering of Q by countably many boxes, and fix $\varepsilon > 0$. For each $k \in \mathbb{N}$, let Q_k^* be a box that contains Q_k in its interior but is only slightly larger than Q_k in the sense that

$$\mathrm{vol}(Q_k^*) \leq (1+\varepsilon)\,\mathrm{vol}(Q_k).$$

For example, if $Q_k = \prod_{j=1}^d [a_j^k, b_j^k]$, then by choosing $\delta_k > 0$ small enough we can take

$$Q_k^* = \prod_{j=1}^d \bigl[a_j^k - \delta_k, b_j^k + \delta_k\bigr].$$

Since $Q_k \subseteq (Q_k^*)^\circ$, the interiors of the boxes Q_k^* form an open covering of Q:

$$Q \subseteq \bigcup_k Q_k \subseteq \bigcup_k (Q_k^*)^\circ.$$

But Q is compact, so this covering must have a finite subcovering (see Definition 1.1.8). That is, there exists some integer $N > 0$ such that

$$Q \subseteq \bigcup_{k=1}^N (Q_k^*)^\circ \subseteq \bigcup_{k=1}^N Q_k^*.$$

Thus the box Q is covered by the *finitely many boxes* $Q_1^*, \dots, Q_N^*$. It seems obvious that the volume of Q cannot exceed the sum of the volumes of the Q_k^*.

This is true, and furthermore it is a computation that only involves *volumes* of boxes, not exterior measures. In fact, this is precisely the content of Exercise 2.1.7. Applying that exercise, we see that

$$\mathrm{vol}(Q) \;\le\; \sum_{k=1}^{N} \mathrm{vol}(Q_k^*) \;\le\; (1+\varepsilon) \sum_{k=1}^{N} \mathrm{vol}(Q_k) \;\le\; (1+\varepsilon) \sum_{k} \mathrm{vol}(Q_k).$$

In summary, we have shown that $\mathrm{vol}(Q) \le (1+\varepsilon) \sum \mathrm{vol}(Q_k)$ for every covering of Q by countably many boxes. Taking the infimum over all such coverings, we obtain $\mathrm{vol}(Q) \le (1+\varepsilon)\, |Q|_e$. Since ε is arbitrary, the desired inequality $\mathrm{vol}(Q) \le |Q|_e$ follows. $\square$

Remark 2.1.18. The proofs of Theorems 2.1.13 and 2.1.17 illustrate two ways of "getting within ε" when dealing with countable sums. In the proof of Theorem 2.1.17 we introduced a multiplicative $1+\varepsilon$ factor, whereas in the proof of Theorem 2.1.13 we incorporated an additive term of the form $2^{-k}\varepsilon$. Both techniques are useful in practice. $\diamondsuit$

Corollary 2.1.19. $|\mathbb{R}^d|_e = \infty$.

Proof. Let $Q_k = [-k,k]^d$. Then, by monotonicity and Theorem 2.1.17,

$$(2k)^d \;=\; \mathrm{vol}(Q_k) \;=\; |Q_k|_e \;\le\; |\mathbb{R}^d|_e.$$

Letting $k \to \infty$, we see that $|\mathbb{R}^d|_e = \infty$. $\square$

The next result, whose proof we assign to the reader, is an extension of Theorem 2.1.17, and it can be proved in a similar manner. This exercise says that the exterior measure of a union of finitely many nonoverlapping boxes equals the sum of the volumes of those boxes.

Exercise 2.1.20. Show that if $Q_1, \dots, Q_n$ are nonoverlapping boxes in $\mathbb{R}^d$, then

$$|Q_1 \cup \cdots \cup Q_n|_e \;=\; \mathrm{vol}(Q_1) \,+\, \cdots \,+\, \mathrm{vol}(Q_n). \qquad \diamondsuit$$

In dimension $d = 1$, a box is a finite closed interval, and the boundary of a closed interval $Q = [a,b]$ is the two-point set $\partial Q = \{a,b\}$. Since ∂Q is a finite set, Lemma 2.1.11(d) tells us that $|\partial Q|_e = 0$. Combining this with subadditivity and monotonicity, we see that

$$\begin{aligned}
|Q|_e \;&=\; |Q^\circ \cup \partial Q|_e \\
&\le\; |Q^\circ|_e + |\partial Q|_e \quad \text{(by subadditivity)} \\
&=\; |Q^\circ|_e + 0 \\
&\le\; |Q|_e \qquad\qquad\quad\; \text{(by monotonicity)}.
\end{aligned} \tag{2.7}$$

Consequently, at least in dimension $d = 1$, a box Q and its interior Q° have the same exterior measure. The following lemma proves that this equality holds in every dimension (note that ∂Q is not a countable set when $d \ge 2$).

Lemma 2.1.21. *If Q is a box in $\mathbb{R}^d$, then*

$$|\partial Q|_e = 0 \qquad \text{and} \qquad |Q^\circ|_e = |Q|_e.$$

In particular, if $d \geq 2$, then the boundary of box is an uncountable *set that has exterior measure zero.*

Proof. To illustrate the idea, consider the unit square $Q = [0,1]^2$ in $\mathbb{R}^2$. The boundary of Q is a union of four line segments $\ell_1, \ell_2, \ell_3, \ell_4$. Each line segment is an uncountable set, but (as a subset of $\mathbb{R}^2$) it has measure zero since we can cover it with a single rectangle that has arbitrarily small area. For example, for the bottom line segment ℓ_1 we can write

$$\ell_1 = \bigl\{(x,0) : 0 \leq x \leq 1\bigr\} \subseteq [0,1] \times [-\varepsilon, \varepsilon] = Q_\varepsilon,$$

and $\operatorname{vol}(Q_\varepsilon) = 2\varepsilon$. Since we can do this for any $\varepsilon > 0$, the two-dimensional exterior Lebesgue measure of the line segment ℓ_1 is zero. The boundary of Q is the union of four such line segments, so by countable subadditivity we obtain $|\partial Q|_e = 0$. A similar idea works for any box in any dimension; we assign the details as Problem 2.1.36.

Finally, now that we know that $|\partial Q|_e = 0$, we can argue just as we did in equation (2.7) to show that $|Q^\circ|_e = |Q|_e$. $\square$

Corollary 2.1.22. *If $-\infty < a \leq b < \infty$, then*

$$|[a,b]|_e = |[a,b)|_e = |(a,b]|_e = |(a,b)|_e = b - a.$$

Proof. If $a = b$ then the result is immediate. Otherwise $[a,b]$ is a box in $\mathbb{R}$ and its boundary is the finite set $\{a,b\}$, so the equalities follow from Theorem 2.1.17 and Lemma 2.1.21. $\square$

2.1.5 The Cantor Set

In dimensions 2 and greater, the boundary of a box is an uncountable set that has exterior measure zero. It is not as easy to exhibit an uncountable subset of $\mathbb{R}$ that has zero exterior measure, but such sets do exist. We will construct a set C, known as the *Cantor set*, whose exterior measure is zero, and following the construction we give an exercise that sketches a proof that C is uncountable.

Example 2.1.23 (The Cantor Set). Define

$$\begin{aligned}
F_0 &= [0,1], \\
F_1 &= \bigl[0, \tfrac{1}{3}\bigr] \cup \bigl[\tfrac{2}{3}, 1\bigr], \\
F_2 &= \bigl[0, \tfrac{1}{9}\bigr] \cup \bigl[\tfrac{2}{9}, \tfrac{1}{3}\bigr] \cup \bigl[\tfrac{2}{3}, \tfrac{7}{9}\bigr] \cup \bigl[\tfrac{8}{9}, 1\bigr],
\end{aligned}$$

and so forth (see Figure 2.3).

Fig. 2.3 The Cantor set C is the intersection of the sets F_n over all $n \geq 0$.

For a given integer n, the set F_n is the union of 2^n disjoint closed intervals, each of which has length 3^{-n}. Now, a finite closed interval in one dimension is a box, and we know that the exterior measure of a box equals its volume (which in this case is the length of the interval). Subadditivity therefore implies that

$$0 \;\leq\; |F_n|_e \;\leq\; 2^n\,3^{-n} \;=\; (2/3)^n.$$

(In fact, the exterior measure of F_n is precisely $(2/3)^n$, but an upper bound is all that we need here.) We create the set F_{n+1} by removing the middle third from each of the 2^n intervals that comprise F_n. The classical "middle-thirds" *Cantor set* is the intersection of all these sets:

$$C \;=\; \bigcap_{n=0}^{\infty} F_n.$$

The Cantor set is closed because each F_n is closed. Moreover $C \subseteq F_n$, so by monotonicity we have

$$0 \;\leq\; |C|_e \;\leq\; |F_n|_e \;\leq\; (2/3)^n.$$

This is true for every integer $n \geq 0$, so we conclude that the exterior measure of the Cantor set is $|C|_e = 0$. $\diamondsuit$

The following exercise gives one method of showing that the Cantor set is uncountable.

Exercise 2.1.24. The *ternary expansion* of $x \in [0,1]$ is

$$x \;=\; \sum_{n=1}^{\infty} \frac{c_n}{3^n},$$

where each "digit" c_n is either 0, 1, or 2. Every point $x \in [0,1]$ has a unique ternary expansion, except for points of the form $x = m/3^n$ with m, n integer, which have two ternary expansions (one ending with infinitely many 0's, and one with infinitely many 2's). Show that x belongs to C if and only if x has at least one ternary expansion for which every digit c_n is either 0 or 2, and use this to show that C is uncountable. $\diamondsuit$

Thus, although the Cantor set is "small" in terms of measure, it is "large" in terms of cardinality. The Cantor set has many other remarkable properties, some of which are laid out in the next exercise.

Exercise 2.1.25. Prove the following statements about the Cantor set C.

(a) C is closed.

(b) C contains no open intervals.

(c) $C^\circ = \varnothing$ (i.e., the interior of C is empty).

(d) $C = \partial C$ (i.e., every point in C is a boundary point of C).

(e) Every point in C is an accumulation point of C (i.e., if $x \in C$ then there exist points $x_n \in C$ with $x_n \neq x$ such that $x_n \to x$).

(f) Every point in C is an accumulation point of $[0,1]\setminus C$ (i.e., if $x \in C$ then there exist points $x_n \notin C$ such that $x_n \to x$). $\diamondsuit$

A set is *totally disconnected* if it contains no nontrivial connected subsets (in one dimension, connected sets are simply intervals). A nonempty set S is *perfect* if every point $x \in S$ is an accumulation point of S. Using this terminology, the Cantor set is both perfect and totally disconnected. Problem 2.1.45 shows that every perfect subset of $\mathbb{R}^d$ is uncountable.

By slightly changing the process used to construct the Cantor set, we will obtain a set that has some very surprising properties.

Example 2.1.26 (The Fat Cantor Set). Let $F_0 = [0,1]$. To construct the Cantor set, we removed an open interval of length $1/3$ from F_0. Let us instead remove an open interval of length a_1, where a_1 can be different than $1/3$ (although we must have $0 < a_1 < 1$). For simplicity, we center this open interval within F_0, so we are left with a set F_1 that is the union of two closed intervals of equal length. From each of these intervals, remove a centered open interval of length a_2. This gives us a set F_2 that is the union of four closed intervals. From each of these we remove a centered open interval of length a_3, giving us a set F_3. We repeat this process, and set $P = \cap F_n$. Just like the Cantor set, the resulting set P is closed, contains no intervals, and equals its own boundary.

What is the measure of P? We have $P \subseteq F_n$ for every n, but in this construction it need not be the case that $|F_n|_e \to 0$ (depending on how we choose the a_n). So, consider the open set $U = [0,1]\setminus P$. This set is the union of all of the disjoint intervals that were removed from $[0,1]$ during the construction of P. At the first stage, we removed one interval of length a_1. Then we removed two intervals of length a_2 at the second stage, four intervals of length a_3 at the third stage, and so forth. Now, it is not true *in general* that the measure of the union of disjoint sets is the sum of their measures, but we will prove later that this does hold for all *measurable sets*. Open sets are measurable, so if we accept this fact for now, then it follows that the measure of U is

$$|U|_e = \sum_{k=1}^{\infty} 2^{n-1} a_n \qquad \text{(to be justified later).}$$

If the a_n converge to zero rapidly enough, then this sum will be *strictly less than* 1 (for example, consider $a_n = 2^{-2n}$). Since U and P are disjoint *measurable* sets, $|U|_e + |P|_e$ equals $|U \cup P|_e$, which is 1. Consequently,

$$|P|_e = 1 - |U|_e \qquad \text{(still to be proved),}$$

and this can be *strictly positive*. The justification of these results does require facts from Section 2.2, and the details are assigned later as Problem 2.2.42.

The set P is called a *Smith–Volterra–Cantor set* or a *fat Cantor set*. In summary, if we choose a_n that converge rapidly enough to zero, then

P is closed set that has **positive exterior measure**
yet **contains no intervals**!

There are sets—even closed sets—that have empty interiors but still have positive measure. $\diamondsuit$

2.1.6 Regularity of Exterior Measure

Next we prove a "regularity property" of exterior Lebesgue measure. We will show that if E is any subset of $\mathbb{R}^d$ and ε is any positive real number, then we can surround E by an *open set* U whose exterior measure is only ε larger than that of E. By monotonicity we also have $|E|_e \le |U|_e$, so the measure of this set U is very close to the measure of E.

Theorem 2.1.27. *If $E \subseteq \mathbb{R}^d$ and $\varepsilon > 0$, then there exists an open set $U \supseteq E$ such that*

$$|E|_e \le |U|_e \le |E|_e + \varepsilon.$$

Consequently,

$$|E|_e = \inf\bigl\{|U|_e : U \text{ open}, U \supseteq E\bigr\}.$$

Proof. If $|E|_e = \infty$ then we can take $U = \mathbb{R}^d$. So, assume that $|E|_e < \infty$. By Lemma 2.1.9, there exist countably many boxes Q_k such that $E \subseteq \cup Q_k$ and

$$\sum_k \operatorname{vol}(Q_k) \le |E|_e + \frac{\varepsilon}{2}.$$

Let Q_k^* be a larger box that contains Q_k in its interior and satisfies

$$\operatorname{vol}(Q_k^*) \le \operatorname{vol}(Q_k) + 2^{-k-1}\varepsilon.$$

Let $U = \cup (Q_k^*)^\circ$ be the union of the interiors of the boxes Q_k^*. Then $E \subseteq U$, U is open, and

$$|E|_e \le |U|_e \le \sum_k \mathrm{vol}(Q_k^*) \le \sum_k \mathrm{vol}(Q_k) + \frac{\varepsilon}{2} \le |E|_e + \varepsilon. \qquad \square$$

If E has finite exterior measure, then we can refine Theorem 2.1.27 slightly.

Corollary 2.1.28. *If $E \subseteq \mathbb{R}^d$ satisfies $|E|_e < \infty$, then for each $\varepsilon > 0$ there exists an open set $U \supseteq E$ such that*

$$|E|_e \le |U|_e < |E|_e + \varepsilon.$$

Proof. By Theorem 2.1.27, there exists an open set $U \supseteq E$ that satisfies $|U|_e \le |E|_e + \frac{\varepsilon}{2}$. Since $|E|_e$ is finite, we have $|E|_e + \frac{\varepsilon}{2} < |E|_e + \varepsilon$. $\square$

If we apply Theorem 2.1.27 to the set of rationals $\mathbb{Q}$, we see that if $\varepsilon > 0$ then there must exist an open set U that contains $\mathbb{Q}$ and satisfies

$$0 = |\mathbb{Q}|_e \le |U|_e \le |\mathbb{Q}|_e + \varepsilon = \varepsilon.$$

This seems counterintuitive, since it says that even though $\mathbb{Q}$ is dense in $\mathbb{R}$, we can surround it with an open set whose exterior measure is at most ε. To explicitly construct such a set U, let $\mathbb{Q} = \{r_k\}_{k\in\mathbb{N}}$ be an enumeration of the rationals, and for each k let I_k be an open interval of length $2^{-k}\varepsilon$ that contains r_k. Then $U = \cup I_k$ is open, contains *every* rational point, and by subadditivity satisfies

$$|U|_e \le \sum_{k=1}^{\infty} |I_k|_e = \sum_{k=1}^{\infty} 2^{-k}\varepsilon = \varepsilon.$$

Problems

2.1.29. Prove that a countable union of sets that each have exterior measure zero has exterior measure zero. That is, if $Z_k \subseteq \mathbb{R}^d$ and $|Z_k|_e = 0$ for each $k \in \mathbb{N}$, then $\bigl|\cup Z_k\bigr|_e = 0$.

2.1.30. Show that if $Z \subseteq \mathbb{R}^d$ and $|Z|_e = 0$, then $\mathbb{R}^d \setminus Z$ is dense in $\mathbb{R}^d$.

2.1.31. Let Z be a subset of $\mathbb{R}$ such that $|Z|_e = 0$. Set $Z^2 = \{x^2 : x \in Z\}$, and prove that $|Z^2|_e = 0$.

2.1.32. Show that if $f\colon \mathbb{R} \to \mathbb{R}$ is continuous, then its graph

$$\Gamma_f = \bigl\{\bigl(x, f(x)\bigr) : x \in \mathbb{R}\bigr\} \subseteq \mathbb{R}^2$$

has measure zero, i.e., $|\Gamma_f|_e = 0$.

2.1.33. The *symmetric difference* of A, $B \subseteq \mathbb{R}^d$ is $A \triangle B = (A \setminus B) \cup (B \setminus A)$. Prove that if $|A|_e$, $|B|_e < \infty$, then $\big||A|_e - |B|_e\big| \le |A \triangle B|_e$.

2.1.34. Given $E \subseteq \mathbb{R}^d$, prove that $|E|_e = \inf\bigl\{\sum \mathrm{vol}(Q_k)\bigr\}$, where the infimum is taken over all countable collections of boxes $\{Q_k\}$ such that $E \subseteq \cup Q_k^\circ$.

2.1.35. Find the exterior measures of the following sets.

(a) $L = \{(x,x) : 0 \le x \le 1\}$, the diagonal of the unit square in $\mathbb{R}^2$ (this is a special case of part (b), but it may be instructive to work this first).

(b) An arbitrary line segment, ray, or line in $\mathbb{R}^2$.

2.1.36. Prove that the $(d-1)$-dimensional subspace of $\mathbb{R}^d$ defined by

$$S \;=\; \mathbb{R}^{d-1} \times \{0\} \;=\; \bigl\{(x_1, \dots, x_{d-1}, 0) : x_1, \dots, x_{d-1} \in \mathbb{R}\bigr\}$$

has exterior measure $|S|_e = 0$, and consequently every subset of S has exterior measure zero.

2.1.37.* Prove that every subset of every proper subspace of $\mathbb{R}^d$ has exterior measure zero.

2.1.38. (a) Let D be a diagonal matrix with diagonal entries $\delta_1, \dots, \delta_d$. Prove that

$$|D(E)|_e = |\delta_1 \cdots \delta_d|\, |E|_e,$$

where $D(E) = \{Dx : x \in E\} = \bigl\{(\delta_1 x_1, \dots, \delta_d x_d) : x \in E\bigr\}$.

(b) Prove that for each integer $d \ge 1$ there exists some constant C_d such that for every $x \in \mathbb{R}^d$ and $r > 0$ we have $|B_r(x)|_e = C_d r^d$ (an explicit formula for C_d is not required here).

2.1.39. Given a set $E \subseteq \mathbb{R}^d$, show that $|E|_e = 0$ if and only if there exist countably many boxes Q_k such that $\sum \mathrm{vol}(Q_k) < \infty$ and each point $x \in E$ belongs to infinitely many Q_k.

2.1.40. Assume that $Z \subseteq \mathbb{R}$ satisfies $|Z|_e = 0$. Prove that there exists at least one point $h \in \mathbb{R}$ such that the translated set $Z + h$ contains no rational points.

2.1.41.* (a) Let U be a bounded open subset of $\mathbb{R}$, and write U as the union of countably many disjoint open intervals (a_k, b_k). Prove that $|U|_e = \sum_k (b_k - a_k)$.

Remark: If we are allowed to appeal to later results, then this is an immediate consequence of Theorem 2.2.16. The challenge is to find a solution that only uses the tools that have been developed so far in this section.

(b) Prove that the exterior measure of the complement of the Cantor set is $|[0,1] \setminus C|_e = 1$.

2.1.42. Let C be the Cantor set, and let $D = \bigl\{\sum_{n=1}^{\infty} 3^{-n} c_n : c_n = 0, 1\bigr\}$. Show that $D+D = [0,1]$, and use this to show that $C+C = [0,2]$. Therefore $|C+C|_e = 2$, even though $|C|_e = 0$.

2.1.43. Modify the Cantor middle-thirds set construction as follows. Fix a parameter $0 < \alpha < 1$, and at stage n form F_{n+1} by removing a subinterval of relative length α from each of the 2^n intervals whose union is F_n (so $\alpha = \frac{1}{3}$ corresponds to the usual Cantor set). Show that the generalized Cantor set $C_\alpha = \bigcap F_n$ is perfect, has no interior, equals its own boundary, and satisfies $|C_\alpha|_e = 0$.

2.1.44. Let F consist of all numbers $x \in [0,1]$ whose decimal expansion does not contain the digit 4. Find $|F|_e$.

2.1.45. This problem will show that any perfect subset of $\mathbb{R}^d$ must be uncountable. Suppose that $S = \{x_1, x_2, \dots\}$ is a countably infinite perfect subset of $\mathbb{R}^d$. Let $n_1 = 1$ and $r_1 = 1$, and let $U_1 = B_{r_1}(x_{n_1})$. Let n_2 be the first integer greater than n_1 such that $x_{n_2} \in U_1$, and show that we can choose $r_2 > 0$ so that $U_2 = B_{r_2}(x_{n_2})$ satisfies $U_2 \subseteq \overline{U_2} \subseteq U_1$ but $x_{n_1} \notin U_2$. Continue in this way, and then define $K = \bigcap (\overline{U_n} \cap S)$. Prove that the sets $\overline{U_n} \cap S$ are compact and nested decreasing. The *Cantor Intersection Theorem* therefore implies that K is nonempty. Show that no element of S can belong to K.

2.2 Lebesgue Measure

Take another look at Theorem 2.1.27, which says that if E is an arbitrary subset of $\mathbb{R}^d$ and ε is any positive number, then we can find an open set U that contains E and has measure at most ε larger than the measure of E. Thus,

$$|E|_e \ \le\ |U|_e \ \le\ |E|_e + \varepsilon.$$

Since U contains E, we can write U as the union of E and $U \setminus E$:

$$U \ =\ E \cup (U \setminus E). \tag{2.8}$$

Applying countable subadditivity (Theorem 2.1.13), we see that

$$|U|_e \ \le\ |E|_e + |U \setminus E|_e. \tag{2.9}$$

The sets E and $U \setminus E$ in equation (2.8) are actually *disjoint sets*, so we are tempted to believe that the sum of their measures should equal the measure of $E \cup (U \setminus E) = U$. That is, we suspect that

$$|U|_e \ =\ |E|_e + |U \setminus E|_e \qquad \leftarrow \text{WE DO NOT KNOW THIS!}$$

However, as the preceding line emphasizes, we do not know that this equality must hold, and there is nothing that we have proved so far that will allow us to infer that $|U|_e$ and $|E|_e + |U \setminus E|_e$ are equal. In fact, we will see in Example 2.4.7 that equality does *not* always hold! Consequently, in this section we restrict our attention from arbitrary subsets of $\mathbb{R}^d$ to a smaller class of "measurable subsets" on which exterior measure is "well behaved."

2.2.1 Definition and Basic Properties

To motivate the definition of measurability, suppose that U is an open set that contains a set E. As we observed above, we do not know whether $|U|_e$ and $|E|_e + |U \setminus E|_e$ will be equal. *If* it were the case that these quantities were equal, then we could combine this equality with equation (2.9) and infer that $|U \setminus E|_e \le \varepsilon$. The "measurable sets" are precisely the sets for which this inequality can be achieved. Here is the explicit definition.

Definition 2.2.1 (Lebesgue Measure). A set $E \subseteq \mathbb{R}^d$ is *Lebesgue measurable*, or simply *measurable* for short, if

$$\forall \varepsilon > 0, \quad \exists \text{ open } U \supseteq E \ \text{ such that } \ |U \setminus E|_e \le \varepsilon.$$

If E is Lebesgue measurable, then its *Lebesgue measure* is its exterior Lebesgue measure, and in this case we denote this value by $|E| = |E|_e$. $\diamondsuit$

There is no difference between the numeric value of the Lebesgue measure and the exterior Lebesgue measure of a measurable set, but when we know that E is measurable we write $|E|$ instead of $|E|_e$.

Notation 2.2.2. The collection of all Lebesgue measurable subsets of $\mathbb{R}^d$ will be denoted by

$$\mathcal{L} \;=\; \mathcal{L}(\mathbb{R}^d) \;=\; \big\{E \subseteq \mathbb{R}^d : E \text{ is Lebesgue measurable}\big\}. \qquad \diamondsuit$$

We would like to know which types of subsets of $\mathbb{R}^d$ are measurable. A first observation is that $\mathcal{L}$ contains all of the open subsets of $\mathbb{R}^d$.

Lemma 2.2.3 (Open Sets Are Measurable). *If $U \subseteq \mathbb{R}^d$ is open, then U is Lebesgue measurable, and therefore $U \in \mathcal{L}$.*

Proof. If U is open, then U is an open set that contains U, and for each $\varepsilon > 0$ we have $|U \setminus U|_e \;=\; 0 \;<\; \varepsilon$. $\square$

Consequently, from now on we will write the measure of an open set U as $|U|$ instead of $|U|_e$.

Now we show that every set whose exterior measure is zero is measurable. No such set (other than the empty set) can be open, so this gives us examples of measurable sets that are not open.

Lemma 2.2.4 (Null Sets Are Measurable). *If* $Z \subseteq \mathbb{R}^d$ *and* $|Z|_e = 0$, *then* Z *is measurable.*

Proof. Fix any $\varepsilon > 0$. Then, by Theorem 2.1.27, there is an open set $U \supseteq Z$ such that

$$|U|_e \le |Z|_e + \varepsilon = 0 + \varepsilon = \varepsilon.$$

Since $U \setminus Z \subseteq U$, monotonicity implies that $|U \setminus Z|_e \le |U|_e \le \varepsilon$. Therefore Z is measurable. □

We use a variety of phrases to refer to a set Z whose exterior measure is $|Z| = 0$. For example, we may say that Z is a "zero-measure set," a "measure-zero set," a "set of measure zero," and so forth. A set that has measure zero is also called a "null set," and the complement of a null set is sometimes called a set of "full measure." Precisely, if $Z \subseteq E$ and $|Z| = 0$, then we say that Z *is a null set in* E and $E \setminus Z$ *has full measure in* E.

Instead of considering individual sets, let us turn to the family $\mathcal{L}$ of all measurable sets and try to determine what operations this collection is closed under. The next result shows that the union of countably many sets from $\mathcal{L}$ remains in $\mathcal{L}$.

Theorem 2.2.5 (Closure Under Countable Unions). *If* $E_1, E_2, \dots$ *are measurable subsets of* $\mathbb{R}^d$, *then their union* $E = \cup E_k$ *is also measurable, and*

$$|E| \le \sum_{k=1}^{\infty} |E_k|. \tag{2.10}$$

Proof. Fix $\varepsilon > 0$. Since E_k is measurable, there exists an open set $U_k \supseteq E_k$ such that

$$|U_k \setminus E_k|_e \le \frac{\varepsilon}{2^k}.$$

Then $U = \cup U_k$ is an open set, $U \supseteq E$, and

$$U \setminus E = \left(\bigcup_{k=1}^{\infty} U_k\right) \setminus \left(\bigcup_{k=1}^{\infty} E_k\right) \subseteq \bigcup_{k=1}^{\infty} (U_k \setminus E_k).$$

Hence

$$|U \setminus E|_e \le \sum_{k=1}^{\infty} |U_k \setminus E_k|_e \le \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k} = \varepsilon,$$

so E is measurable. Finally, equation (2.10) follows from the countable subadditivity property of Lebesgue measure. □

By setting $E_k = \varnothing$ for $k > N$, a corollary of Theorem 2.2.5 is that a union of finitely many measurable sets is measurable. However, an uncountable union of measurable sets need not be measurable. For example, if N is a nonmeasurable set then we can write $N = \cup_{x \in N} \{x\}$, yet each singleton $\{x\}$ is measurable.

2.2.2 Toward Countable Additivity and Closure under Complements

So far, the only sets that we have explicitly shown to be measurable are open sets and sets whose exterior measure is zero. A box is not open and it has positive measure, so it does not fall into either of these two categories. On the other hand, a box Q is a union of its interior Q° and its boundary ∂Q. The interior is measurable because it is open, and the boundary is measurable because it has exterior measure zero (see Lemma 2.1.21). Theorem 2.2.5 tells us that the union of countably many measurable sets is measurable, so we conclude that $Q = Q^\circ \cup \partial Q$ is measurable. We formalize this as follows.

Corollary 2.2.6 (Boxes Are Measurable). *Every box in $\mathbb{R}^d$ is a Lebesgue measurable set.* $\diamondsuit$

Can we use the same technique to show that every closed set is measurable? After all, if F is a closed set then we can write $F = F^\circ \cup \partial F$, and the interior F° is open and therefore measurable. If $|\partial F|_e = 0$, then ∂F is measurable as well, and so in this case we can conclude that F is measurable. It is hard to imagine a closed set whose boundary does not have measure zero, but such sets do exist! A specific example was constructed in Example 2.1.26. Consequently, it is not obvious whether all closed sets are measurable, and it will take some work to prove that they are.

Since we know that open sets are measurable, if we can prove that the complement of a measurable set is measurable then we will obtain the measurability of closed sets as a corollary. That is one of our goals, and another is to prove that Lebesgue measure is *countably additive* on the measurable sets, i.e., if $E_1, E_2, \dots$ are countably many disjoint measurable sets, then the Lebesgue measure of $\cup E_k$ equals $\sum |E_k|$. We will work simultaneously toward proving closure under complements and countable additivity.

Our first step in this direction considers additivity of two sets, given the extra assumption that these sets are separated by a positive distance. The distance between two nonempty sets $A,\, B \subseteq \mathbb{R}^d$ is

$$\operatorname{dist}(A, B) \;=\; \inf\bigl\{\|x - y\| : x \in A,\, y \in B\bigr\}, \tag{2.11}$$

where, as usual, $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$. We will show that if A and B are any two subsets of $\mathbb{R}^d$ (possibly even nonmeasurable!) that are separated by a strictly positive distance, then the exterior measure of $A \cup B$ equals the sum of the exterior measures of A and B. For this proof, we need to observe that if Q is a box in $\mathbb{R}^d$, then by subdividing each side of Q in two we obtain 2^d nonoverlapping subboxes whose union is Q. Further, the sum of the volumes of these 2^d subboxes is precisely the volume of Q (see Lemma 2.1.6). Consequently, when computing an exterior measure, if we like we can always replace a given box by a finite number of smaller nonoverlapping boxes whose volumes sum to the volume of the original box.

Lemma 2.2.7. *If* $A, B \subseteq \mathbb{R}^d$ *are nonempty and* $\operatorname{dist}(A,B) > 0$, *then*

$$|A \cup B|_e \;=\; |A|_e + |B|_e.$$

Proof. Countable subadditivity implies that $|A \cup B|_e \le |A|_e + |B|_e$. We must prove the opposite inequality.

Fix $\varepsilon > 0$. By Lemma 2.1.9, there exist countably many boxes Q_k such that $A \cup B \subseteq \cup Q_k$ and

$$\sum_k |Q_k| \;\le\; |A \cup B|_e + \varepsilon.$$

As illustrated in Figure 2.4, by dividing each box Q_k into finitely many subboxes if necessary, we can assume that the diameter of Q_k is less than the distance between A and B, i.e.,

$$\operatorname{diam}(Q_k) \;=\; \sup\bigl\{\|x-y\| : x, y \in Q_k\bigr\} \;<\; \operatorname{dist}(A,B).$$

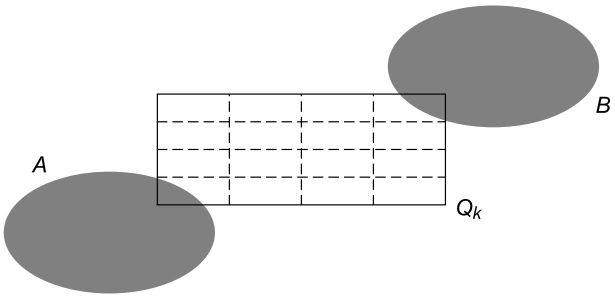


Fig. 2.4 A box Q_k is subdivided into finitely many smaller boxes, each of whose diameter is less than $\operatorname{dist}(A,B)$.

After we have subdivided the boxes in this way, we see that each box Q_k can intersect at most one of A or B. Let $\bigl\{Q_k^A\bigr\}$ be the subsequence of $\{Q_k\}$ that contains those boxes that intersect A, and let $\bigl\{Q_k^B\bigr\}$ be the subsequence of boxes that intersect B. Since $\{Q_k\}$ covers $A \cup B$, it follows that A is covered by $\bigl\{Q_k^A\bigr\}$ and B is covered by $\bigl\{Q_k^B\bigr\}$. Therefore

$$|A|_e + |B|_e \;\le\; \sum_k |Q_k^A| + \sum_k |Q_k^B| \;\le\; \sum_k |Q_k| \;\le\; |A \cup B|_e + \varepsilon.$$

Since ε is arbitrary, we conclude that $|A|_e + |B|_e \le |A \cup B|_e$. $\square$

Any two disjoint nonempty compact subsets of $\mathbb{R}^d$ are separated by a positive distance (this is Problem 2.2.31). Combining Lemma 2.2.7 with an argument by induction, and recalling that the empty set has measure zero, we obtain the following corollary.

Corollary 2.2.8. *If $F_1, \dots, F_N$ are disjoint compact subsets of $\mathbb{R}^d$, then*

$$\left|\bigcup_{k=1}^{N} F_k\right|_e = \sum_{k=1}^{N} |F_k|_e. \qquad \diamondsuit$$

Now we will prove that all compact subsets of $\mathbb{R}^d$ are measurable.

Theorem 2.2.9 (Compact Sets Are Measurable). *Every compact subset of $\mathbb{R}^d$ is Lebesgue measurable.*

Proof. Let F be a nonempty compact subset of $\mathbb{R}^d$, and choose $\varepsilon > 0$. By Theorem 2.1.27, there exists an open set $U \supseteq F$ such that $|U| \le |F|_e + \varepsilon$. Our goal is to show that $|U \setminus F|_e \le \varepsilon$.

Since U is open and F is closed, their relative complement $U \setminus F$ is open. Applying Lemma 2.1.5, there exist countably many nonoverlapping boxes Q_k such that

$$U \setminus F = \bigcup_{k=1}^{\infty} Q_k.$$

For each finite N, let

$$R_N = \bigcup_{k=1}^{N} Q_k. \tag{2.12}$$

This is a compact set, and even though we have not yet proved that generic compact sets are measurable, we know that this set R_N is measurable because it is a finite union of boxes, each of which is measurable. Further, because $Q_1, \dots, Q_N$ are *finitely many nonoverlapping boxes*, Exercise 2.1.20 implies that

$$|R_N| = \sum_{k=1}^{N} |Q_k|. \tag{2.13}$$

Now, R_N and F are *disjoint* compact sets that are each contained in U. Using equation (2.13), Corollary 2.2.8, and monotonicity, we compute that

$$|F|_e + \sum_{k=1}^{N} |Q_k| = |F|_e + |R_N| = |F \cup R_N|_e \le |U| \le |F|_e + \varepsilon.$$

Since all of the quantities that appear on the preceding line are finite, we can subtract $|F|_e$ from both sides to obtain $\sum_{k=1}^{N} |Q_k| \le \varepsilon$. Finally, taking the limit as $N \to \infty$, we see that

$$|U \setminus F|_e = \left|\bigcup_{k=1}^{\infty} Q_k\right| \le \sum_{k=1}^{\infty} |Q_k| = \lim_{N\to\infty} \sum_{k=1}^{N} |Q_k| \le \varepsilon.$$

Therefore F is measurable. $\square$

An arbitrary closed set in $\mathbb{R}^d$ need not be compact, but we can write every closed set E as a countable union of compact sets. There are many ways to do this. For example,

$$E = \bigcup_{k=1}^{\infty} F_k, \qquad \text{where } F_k = E \cap [-k,k]^d.$$

Since the class of measurable sets is closed under countable unions, this gives us the following result.

Corollary 2.2.10 (Closed Sets are Measurable). *Every closed subset of $\mathbb{R}^d$ is Lebesgue measurable.* $\diamondsuit$

Next, we use the measurability of closed sets to prove that $\mathcal{L}$ is closed under complements.

Theorem 2.2.11 (Closure Under Complements). *If $E \subseteq \mathbb{R}^d$ is Lebesgue measurable, then so is $E^{\mathrm{C}} = \mathbb{R}^d \backslash E$.*

Proof. Since E is measurable, Theorem 2.1.27 implies that for each $k \in \mathbb{N}$ we can find an open set $U_k \supseteq E$ such that $|U_k \backslash E|_e < \frac{1}{k}$. Let F_k be the complement of U_k. Then F_k is closed, so it is measurable. Consequently, the set

$$H = \bigcup_{k=1}^{\infty} F_k = \bigcup_{k=1}^{\infty} U_k^{\mathrm{C}}$$

is measurable, and $H \subseteq E^{\mathrm{C}}$. Let $Z = E^{\mathrm{C}} \backslash H$. For each fixed j we have

$$Z = E^{\mathrm{C}} \backslash \bigcup_{k=1}^{\infty} U_k^{\mathrm{C}} \subseteq E^{\mathrm{C}} \backslash U_j^{\mathrm{C}} = U_j \backslash E,$$

and therefore

$$|Z|_e \le |U_j \backslash E|_e < \frac{1}{j}.$$

Since this is true for every $j \in \mathbb{N}$, it follows that $|Z|_e = 0$. Hence Z is measurable, so $E^{\mathrm{C}} = H \cup Z$ is measurable as well. $\square$

As corollaries of Theorem 2.2.11, we immediately obtain two additional closure results. First, the intersection of any countable collection of measurable sets is measurable.

Corollary 2.2.12 (Closure Under Countable Intersections). *If the sets $E_1, E_2, \ldots \subseteq \mathbb{R}^d$ are each Lebesgue measurable, then so is $E = \bigcap E_k$.* $\diamondsuit$

Second, if A and B are both measurable sets, then their relative complement $A \backslash B$ is also measurable.

Corollary 2.2.13 (Closure Under Relative Complements). *If A and B are Lebesgue measurable subsets of $\mathbb{R}^d$, then so is $A \backslash B = A \cap B^{\mathrm{C}}$.* $\diamondsuit$

In summary, the collection $\mathcal{L}$ of Lebesgue measurable subsets of $\mathbb{R}^d$ is closed under both countable unions and under complements. We have a name for collections of sets that satisfy these properties.

Definition 2.2.14 (Sigma Algebra). Let X be a set, and let Σ be a family of subsets of X (in other words, $\Sigma \subseteq \mathcal{P}(X)$, the power set of X). If:

(a) Σ is not empty,

(b) Σ is closed under complements, and

(c) Σ is closed under countable unions,

then Σ is called a *σ-algebra* of subsets of X. ♢

Using this terminology, the set $\mathcal{L}$ of Lebesgue measurable subsets of $\mathbb{R}^d$ is a σ-algebra of subsets of $\mathbb{R}^d$. Abstract σ-algebras are important for the construction of measures other than Lebesgue measure on $\mathbb{R}^d$, and for defining measures on more general domains.

2.2.3 Countable Additivity

It still remains to prove that Lebesgue measure is countably additive on disjoint measurable sets. To do this, we will need the following characterization of measurable sets in terms of approximations from within by closed sets.

Lemma 2.2.15. *A set $E \subseteq \mathbb{R}^d$ is Lebesgue measurable if and only if for each $\varepsilon > 0$ there exists a closed set $F \subseteq E$ such that $|E \backslash F|_e < \varepsilon$.*

Proof. $\Rightarrow$. Suppose that E is measurable. Then $E^{\mathrm{C}} = \mathbb{R}^d \backslash E$ is measurable, so there exists an open set $U \supseteq E^{\mathrm{C}}$ such that $|U \backslash E^{\mathrm{C}}| < \varepsilon$. Then $F = U^{\mathrm{C}}$ is closed and satisfies $E \backslash F = U \backslash E^{\mathrm{C}}$, so $|E \backslash F| < \varepsilon$.

$\Leftarrow$. Suppose that for every $\varepsilon > 0$ there exists a closed set $F \subseteq E$ such that $|E \backslash F|_e < \varepsilon$. Then $U = F^{\mathrm{C}}$ is open, and $U \supseteq E^{\mathrm{C}}$. Further, $U \backslash E^{\mathrm{C}} = E \backslash F$, so $|U \backslash E^{\mathrm{C}}|_e = |E \backslash F|_e < \varepsilon$. Therefore E^{C} is measurable, so E is measurable as well. □

We have now assembled the tools that we need to prove that Lebesgue measure is countably additive on the class of measurable sets.

Theorem 2.2.16 (Countable Additivity). *If $E_1, E_2, \dots$ are disjoint, Lebesgue measurable subsets of $\mathbb{R}^d$, then*

$$\left| \bigcup_{k=1}^{\infty} E_k \right| = \sum_{k=1}^{\infty} |E_k|. \tag{2.14}$$

Proof. Step 1. Assume first that each set E_k is bounded. From subadditivity we obtain

$$\left|\bigcup_{k=1}^{\infty} E_k\right| \;\le\; \sum_{k=1}^{\infty} |E_k|,$$

so our task is to prove the opposite inequality.

Fix $\varepsilon > 0$. By Lemma 2.2.15, there exists a closed set $F_k \subseteq E_k$ such that

$$|E_k \backslash F_k| \;<\; \frac{\varepsilon}{2^k}. \tag{2.15}$$

Since E_k is bounded, F_k is compact. Hence $\{F_k\}_{k\in\mathbb{N}}$ is a collection of disjoint compact sets. Let N be any finite positive integer. Then, by using Corollary 2.2.8 and monotonicity, we see that

$$\sum_{k=1}^{N} |F_k| \;=\; \left|\bigcup_{k=1}^{N} F_k\right| \;\le\; \left|\bigcup_{k=1}^{N} E_k\right| \;\le\; \left|\bigcup_{k=1}^{\infty} E_k\right|.$$

Taking the limit as $N \to \infty$,

$$\sum_{k=1}^{\infty} |F_k| \;=\; \lim_{N\to\infty} \sum_{k=1}^{N} |F_k| \;\le\; \left|\bigcup_{k=1}^{\infty} E_k\right|. \tag{2.16}$$

Therefore

$$\begin{aligned}
\sum_{k=1}^{\infty} |E_k| \;&=\; \sum_{k=1}^{\infty} |F_k \cup (E_k \backslash F_k)| \\
&\le\; \sum_{k=1}^{\infty} \bigl(|F_k| \,+\, |E_k \backslash F_k|\bigr) && \text{(by finite subadditivity)} \\
&\le\; \sum_{k=1}^{\infty} \left(|F_k| \,+\, \frac{\varepsilon}{2^k}\right) && \text{(by equation (2.15))} \\
&=\; \left(\sum_{k=1}^{\infty} |F_k|\right) \,+\, \varepsilon \\
&\le\; \left|\bigcup_{k=1}^{\infty} E_k\right| \,+\, \varepsilon && \text{(by equation (2.16)).}
\end{aligned}$$

Since ε is arbitrary, equation (2.14) follows.

Step 2. Now assume that $E_1, E_2, \dots$ are arbitrary disjoint measurable subsets of $\mathbb{R}^d$. Set

$$E_k^j \;=\; \bigl\{x \in E_k \,:\, j-1 \le \|x\| < j\bigr\}, \qquad \text{for } j, k \in \mathbb{N}.$$

Then $\{E_k^j\}_{k,j}$ is a countable collection of disjoint *bounded* measurable sets. For each fixed $k \in \mathbb{N}$ we have

$$\bigcup_{j=1}^{\infty} E_k^j \;=\; E_k, \tag{2.17}$$

and furthermore

$$\bigcup_{k=1}^{\infty}\bigcup_{j=1}^{\infty} E_k^j \;=\; \bigcup_{k=1}^{\infty} E_k \;=\; E. \tag{2.18}$$

Therefore

$$\begin{aligned}
\left|\bigcup_{k=1}^{\infty} E_k\right| &= \left|\bigcup_{k=1}^{\infty}\bigcup_{j=1}^{\infty} E_k^j\right| && \text{(by equation (2.18))}\\
&= \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} |E_k^j| && \text{(by Step 1)}\\
&= \sum_{k=1}^{\infty}\left|\bigcup_{j=1}^{\infty} E_k^j\right| && \text{(by Step 1)}\\
&= \sum_{k=1}^{\infty} |E_k| && \text{(by equation (2.17)).} \qquad \square
\end{aligned}$$

It is worth noting that what makes Step 2 of the preceding proof possible is the fact that $\mathbb{R}^d$, whose measure is infinite, can be written as the union of countably many measurable sets that each have finite measure (in the language of abstract measure theory, this says that *Lebesgue measure on* $\mathbb{R}^d$ *is* σ*-finite*). While simple, this observation is extremely useful, as it often allows us to reduce issues about generic sets to sets that have finite measure. There are many ways to write $\mathbb{R}^d$ as a countable union of sets that have finite measures; here are a few typical examples.

(a) $\mathbb{R}^d = \cup_{n=1}^{\infty} B_n(0)$.

(b) $\mathbb{R}^d = \cup_{n=1}^{\infty} \bigl\{x \in \mathbb{R}^d : n-1 \le \|x\| < n\bigr\}$.

(c) $\mathbb{R}^d = \cup_{k\in\mathbb{Z}^d} (Q+k)$ where $Q = [0,1]^d$.

The sets $B_n(0)$ in the union in (a) are not disjoint, whereas the sets in the union in (b) are disjoint. Although the sets in the union in (c) are not disjoint, they are nonoverlapping closed cubes.

Combining Theorem 2.2.16 with the fact that the boundary of a box has measure zero, we obtain the following result.

Corollary 2.2.17. *If* $\{Q_k\}$ *is a countable collection of nonoverlapping boxes, then* $\bigl|\cup Q_k\bigr| = \sum |Q_k|$.

Proof. The interiors of the boxes Q_k are disjoint. Further, ∂Q_k has measure zero for every k, so $Z = \cup\, \partial Q_k$ also has measure zero. Applying countable additivity, we conclude that

$$\Big|\bigcup_k Q_k\Big| \;=\; \Big|\Big(\bigcup_k Q_k^\circ\Big) \cup Z\Big| \;=\; \sum_k |Q_k^\circ| \;+\; |Z| \;=\; \sum_k |Q_k| \;+\; 0. \qquad \square$$

2.2.4 Equivalent Formulations of Measurability

As we have seen, the collection $\mathcal{L}$ of all Lebesgue measurable subsets of $\mathbb{R}^d$ is closed under countable unions and complements. Since $\mathcal{L}$ contains all of the open and closed subsets of $\mathbb{R}^d$, it must therefore also contain all of the following types of sets.

Definition 2.2.18 (G_δ-Sets and F_σ-Sets).

(a) A set $H \subseteq \mathbb{R}^d$ is a G_δ*-set* if there exist countably many open sets U_k such that $H = \cap U_k$.

(b) A set $H \subseteq \mathbb{R}^d$ is an F_σ*-set* if there exist countably many closed sets F_k such that $H = \cup F_k$. $\diamondsuit$

The symbol σ in this definition is reminiscent of the word "sums" and hence unions, while δ suggests the word "difference" and hence intersections. More precisely, F_σ is derived from the French words *fermé* (closed) and *somme* (union), while G_δ is derived from the German *Gebiet* (area, neighborhood, open set) and *Durchschnitt* (average, intersection).

The half-open interval $[a,b)$ is neither an open nor a closed subset of $\mathbb{R}$, but it is both a G_δ-set and an F_σ-set because we can write

$$\bigcap_{k=1}^{\infty} \big(a - \tfrac{1}{k}, b\big) \;=\; [a,b) \;=\; \bigcup_{k=1}^{\infty} \big[a, b - \tfrac{1}{k}\big]. \tag{2.19}$$

Here are some additional examples.

Example 2.2.19. (a) Let $\mathbb{Q} = \{r_k\}_{k\in\mathbb{N}}$ be an enumeration of the set of rationals. Since $\mathbb{Q}$ is a countable union of singletons, each of which is closed, $\mathbb{Q}$ is an F_σ-set.

(b) Let r_k be as in part (a), and for each k let U_k be the complement of the point r_k:

$$U_k \;=\; \mathbb{R}\backslash\{r_k\} \;=\; (-\infty, r_k) \;\cup\; (r_k, \infty), \qquad \text{for } k \in \mathbb{N}. \tag{2.20}$$

The set U_k is open and contains every point in $\mathbb{R}$ except r_k. Consequently

$$\bigcap_{k=1}^{\infty} U_k \;=\; \mathbb{R}\backslash\mathbb{Q}.$$

Hence $\mathbb{R}\backslash\mathbb{Q}$, the set of irrationals, is a G_δ-set.

(c) Could the set of rationals be a G_δ-set? If it were, then we could write $\mathbb{Q} = \cap V_k$ where each V_k is open. Since V_k contains $\mathbb{Q}$, it is dense in $\mathbb{R}$. The sets U_k defined in equation (2.20) are also dense in $\mathbb{R}$, and the intersection of all of the U_k and V_k is

$$\biggl(\bigcap_{k=1}^{\infty} V_k\biggr) \cap \biggl(\bigcap_{k=1}^{\infty} U_k\biggr) \;=\; \mathbb{Q} \cap (\mathbb{R}\backslash\mathbb{Q}) \;=\; \varnothing.$$

However, the *Baire Category Theorem* implies that a countable intersection of open, dense subsets of $\mathbb{R}$ must be nonempty (for the statement and a proof of the Baire Category Theorem, see [Heil11, Thm. 2.21] or [Heil18, Thm 2.11.3]). This is a contradiction, so we conclude that $\mathbb{Q}$ cannot be a G_δ-set. $\diamondsuit$

We can keep going and define an $F_{\sigma\delta}$-set to be a countable intersection of F_σ-sets, a $G_{\delta\sigma}$-set to be a countable union of G_δ-sets, an $F_{\sigma\delta\sigma}$-set to be a countable union of $F_{\sigma\delta}$-sets, and so forth. All of these sets are Lebesgue measurable (but the collection of all such sets does not exhaust the family $\mathcal{L}$; see [Fol99, Sec. 1.6]).

Our next lemma shows that every set E, measurable or not, can be surrounded by a G_δ-set that has *exactly the same measure as* E.

Lemma 2.2.20. *Let E be a subset of $\mathbb{R}^d$.*

(a) *There exists a G_δ-set $H \supseteq E$ such that $|E|_e = |H|$.*

(b) *We can arrange the set H in part* (a) *to have the form $H = \cap V_k$ where $V_1 \supseteq V_2 \supseteq \cdots$ is a nested decreasing sequence of open sets.*

Proof. (a) If $|E|_e = \infty$, then we can take $H = \mathbb{R}^d$. Otherwise, applying Theorem 2.1.27, for each $k \in \mathbb{N}$ there exists an open set $U_k \supseteq E$ such that $|U_k| < |E|_e + \frac{1}{k}$. Then $H = \cap U_k$ is a G_δ-set and $E \subseteq H \subseteq U_k$ for every k. Therefore, by monotonicity, $|E|_e \le |H| \le |U_k| \le |E|_e + \frac{1}{k}$. This is true for every k, so $|E|_e = |H|$.

(b) Using the sets U_k from part (a), set $V_k = U_1 \cap \cdots \cap U_k$. $\square$

It does not follow from Lemma 2.2.20 that $H\backslash E$ has measure zero. In fact, this is one of the equivalent conditions for measurability of E given in the next lemma.

Lemma 2.2.21. *If $E \subseteq \mathbb{R}^d$, then the following three statements are equivalent.*

(a) *E is Lebesgue measurable.*

(b) *$E = H\backslash Z$ where H is a G_δ-set and $|Z| = 0$.*

(c) *$E = H \cup Z$ where H is an F_σ-set and $|Z| = 0$.*

Proof. (a) $\Rightarrow$ (b). This argument is a small refinement of the proof of Lemma 2.2.20. Suppose that E is measurable. Then for each $k \in \mathbb{N}$ we can find an open set $U_k \supseteq E$ such that $|U_k \backslash E| < 1/k$. Set $H = \cap U_k$ and let $Z = H \backslash E$. Then H is a G_δ-set, $H \supseteq E$, and $Z = H \backslash E \subseteq U_k \backslash E$ for every k. Hence $|Z|_e \le |U_k \backslash E| < 1/k$ for every k, so $|Z| = 0$.

(b) $\Rightarrow$ (a). If $E = H \backslash Z$ where H is a G_δ-set and $|Z| = 0$, then E is measurable since both H and Z are measurable.

(a) $\Leftrightarrow$ (c). By making use of Lemma 2.2.15, this argument is similar to the proof of (a) $\Leftrightarrow$ (b). $\square$

If $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is a continuous function then, by definition, the *inverse image* of any open subset of $\mathbb{R}^n$ under f is an open subset of $\mathbb{R}^m$. However, the *direct image* of an open set under a continuous function need not be open in general (consider the image of the open interval $U = (0, 2\pi)$ under the continuous function $f(x) = \sin x$). Even so, the following exercise shows that if $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is continuous, then the direct image of a compact set under f is compact, and the direct image of an F_σ set is another F_σ set.

Exercise 2.2.22. Suppose that $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is a continuous function. Prove that the following statements hold.

(a) f maps compact sets to compact sets, i.e.,

$$K \subseteq \mathbb{R}^n \text{ is compact} \implies f(K) \subseteq \mathbb{R}^m \text{ is compact.}$$

(b) f maps F_σ-sets to F_σ-sets, i.e.,

$$E \subseteq \mathbb{R}^n \text{ is an } F_\sigma\text{-set} \implies f(E) \subseteq \mathbb{R}^m \text{ is an } F_\sigma\text{-set.} \qquad \diamondsuit$$

2.2.5 Carathéodory's Criterion

As presented in Definition 2.2.1, our definition of Lebesgue measurable sets is formulated in terms of the existence of surrounding open sets. Lemma 2.2.21 likewise interprets measurability in terms of sets that have other topological properties. In contrast, the equivalent formulation of measurability given in the next theorem does not (directly) involve topology. This criterion says that a set E is measurable if and only if it has the property that when *any other set* A is given, the exterior measures of the two disjoint pieces $A \cap E$ and $A \backslash E$ must precisely sum to the exterior measure of A (see the illustration in Figure 2.5).

Theorem 2.2.23 (Carathéodory's Criterion). *A set $E \subseteq \mathbb{R}^d$ is Lebesgue measurable if and only if*

$$\forall A \subseteq \mathbb{R}^d, \qquad |A|_e = |A \cap E|_e + |A \backslash E|_e. \tag{2.21}$$

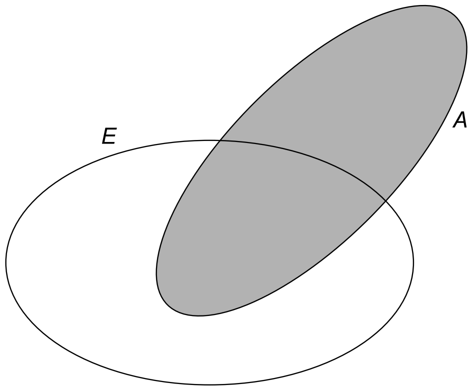


Fig. 2.5 If E is measurable, then $|A \cap E|_e$ and $|A \setminus E|_e$ must sum to $|A|_e$ for *every* set A.

Proof. $\Rightarrow$. Suppose that E is measurable, and fix any set $A \subseteq \mathbb{R}^d$. Since $A = (A \cap E) \cup (A \setminus E)$, subadditivity implies that

$$|A|_e \;\le\; |A \cap E|_e \;+\; |A \setminus E|_e.$$

By Lemma 2.2.20, there exists a G_δ-set $H \supseteq A$ such that $|H| = |A|_e$. We can write H as the disjoint union $H = (H \cap E) \cup (H \setminus E)$. Since Lebesgue measure is countably additive on measurable sets and since H and E are measurable, we conclude that

$$\begin{aligned} |A|_e \;=\; |H| \;&=\; |H \cap E| \;+\; |H \setminus E| && \text{(countable additivity)} \\ &\ge\; |A \cap E|_e \;+\; |A \setminus E|_e && \text{(monotonicity).} \end{aligned}$$

$\Leftarrow$. Let E be any subset of $\mathbb{R}^d$ that satisfies equation (2.21). For each $k \in \mathbb{N}$, let $E_k = E \cap B_k(0)$. Fix $\varepsilon > 0$, and let U be an open set that contains E_k and satisfies

$$|E_k|_e \;\le\; |U| \;\le\; |E_k|_e + \varepsilon.$$

By replacing U with $U \cap B_k(0)$ if necessary, we can assume that $U \subseteq B_k(0)$. Using equation (2.21), we compute that

$$\begin{aligned} |E_k|_e \;+\; |U \setminus E_k|_e \;&=\; |U \cap E_k|_e \;+\; |U \setminus E_k|_e && \text{(since } E_k \subseteq U\text{)} \\ &=\; |U \cap E|_e + |U \setminus E|_e && \text{(since } U \subseteq B_k(0)\text{)} \\ &=\; |U| && \text{(by equation (2.21))} \\ &\le\; |E_k|_e + \varepsilon. \end{aligned}$$

Since $|E_k|_e$ is finite, we can subtract it from both sides to obtain $|U \setminus E_k|_e \le \varepsilon$. Thus E_k is measurable, and therefore $E = \cup E_k$ is measurable as well. $\square$

2.2.6 Almost Everywhere and the Essential Supremum

We introduce some terminology related to sets whose measure is zero.

Notation 2.2.24 (Almost Everywhere). A property that holds at all points of a set E except possibly for those that lie in a subset $Z \subseteq E$ whose measure is zero is said to hold *almost everywhere* on E. We often abbreviate "almost everywhere" by "a.e." ◇

Example 2.2.25. (a) The Cantor set C has measure zero. Its characteristic function χ_C satisfies $\chi_C(x) = 0$ for all $x \in \mathbb{R}$ with the exception of those points x that belong to C. Since $|C| = 0$, we therefore say that

$$\chi_C(x) \,=\, 0 \textit{ for almost every } x,$$

and we abbreviate this by writing $\chi_C = 0$ *a.e.*

(b) Define $f\colon [0,\infty) \to [0,\infty]$ by $f(x) = 1/x$ for $x > 0$ and $f(0) = \infty$. This function takes finite values at all but a single point. Thus the set

$$Z \,=\, \{x \in [0,\infty) : f(x) = \pm\infty\}$$

where f is not finite has measure zero, so we say that

$$f(x) \textit{ is finite for almost every } x \in [0,\infty),$$

or simply that f *is finite a.e.*

(c) If $f\colon E \to \mathbb{C}$ is a complex-valued function, then $f(x)$ is never $\pm\infty$. Therefore *every complex-valued function is finite at every point*, where we interpret the word "finite" in this context to mean "not $\pm\infty$." As a consequence, *every complex-valued function is finite a.e.* ◇

To motivate the next definition, let $f\colon E \to [-\infty,\infty]$ be an extended real-valued function. One way to express the supremum of f on E is by the formula

$$\sup_{x\in E} f(x) \,=\, \inf\bigl\{M \in [-\infty,\infty] \,:\, f(x) \le M \text{ for all } x \in E\bigr\}.$$

The *essential supremum* of f will be defined by a similar formula, except that we will ignore sets of measure zero. That is, instead of taking the infimum over those M such that $f(x) \le M$ for *all* $x \in E$, we take the infimum over those M for which the inequality $f(x) \le M$ holds *almost everywhere* on E. Here is the precise definition.

Definition 2.2.26 (Essential Supremum). Let E be a subset of $\mathbb{R}^d$.

(a) The *essential supremum* of a function $f\colon E \to [-\infty,\infty]$ is

$$\operatorname*{ess\,sup}_{x\in E} f(x) \,=\, \inf\bigl\{M \in [-\infty,\infty] \,:\, f(x) \le M \text{ for a.e. } x \in E\bigr\}. \tag{2.22}$$

(b) If f is either an extended real-valued or complex-valued function on E, then we say that f is *essentially bounded* if

$$\operatorname*{ess\,sup}_{x\in E} |f(x)| \;<\; \infty. \qquad \diamondsuit$$

Example 2.2.27. Consider $f(x) = x\chi_{\mathbb{Q}}(x)$ for $x \in \mathbb{R}$. This function is zero whenever x is irrational, but it takes arbitrarily large values at rational x. Hence f is unbounded and $\sup_{x\in\mathbb{R}} f(x) = \infty$. On the other hand, $f(x) = 0$ for almost every $x \in \mathbb{R}$, so

$$\operatorname*{ess\,sup}_{x\in\mathbb{R}} |f(x)| \;=\; \operatorname*{ess\,sup}_{x\in\mathbb{R}} f(x) \;=\; 0.$$

Therefore, even though f is unbounded, it is *essentially bounded.* $\diamondsuit$

Here are some properties of the essential supremum.

Lemma 2.2.28. *If $f\colon E \to [-\infty, \infty]$ and we set $m = \operatorname{ess\,sup}_{x\in E} f(x)$, then the following statements hold.*

(a) *$f(x) \le m$ for a.e. $x \in E$.*

(b) *m is the smallest extended real number M such that $f \le M$ a.e.*

Proof. (a) If $k \in \mathbb{N}$ then $m + \frac{1}{k} > m$, so, by the definition of the essential supremum, we must have $f(x) \le m + \frac{1}{k}$ for all x except those in a set Z_k of measure zero. Let $Z = \cup Z_k$. If $x \notin Z$ then $x \notin Z_k$ for any k, so $f(x) \le m + \frac{1}{k}$ for every k. Therefore $f(x) \le m$ for all $x \notin Z$.

(b) This follows from part (a) and the definition of an infimum. $\square$

By applying Lemma 2.2.28 to the absolute value of a function, we obtain the following corollary.

Corollary 2.2.29. *Let $E \subseteq \mathbb{R}^d$, and let f be a function on E that is either extended real-valued or complex-valued.*

(a) *If f is essentially bounded, then there exists a finite constant $M \ge 0$ such that $|f(x)| \le M$ for a.e. x. In particular, f is finite a.e.*

(b) $\operatorname{ess\,sup}_{x\in E} |f(x)| = 0$ *if and only if $f = 0$ a.e.*

Proof. (a) If f is essentially bounded, then $M = \operatorname{ess\,sup} |f(x)| < \infty$. Applying Lemma 2.2.28(a) to the function $|f|$, we see that $|f(x)| \le M < \infty$ for almost every $x \in E$.

(b) If $\operatorname{ess\,sup} |f(x)| = 0$, then part (a) of Lemma 2.2.28 implies that $|f(x)| \le 0$ a.e., and therefore $f = 0$ a.e. $\square$

While every essentially bounded function is finite a.e., there are functions that are finite a.e. but not essentially bounded. An example is the function $f(x) = 1/x$ considered in Example 2.2.25(b).

The essential supremum of a function is always less than or equal to its supremum. According to the following exercise, these two quantities coincide for continuous functions whose domain is an open set.

Exercise 2.2.30. Let U be a nonempty open subset of $\mathbb{R}^d$, and suppose that $f\colon U \to \mathbb{R}$ is continuous. Prove that the essential supremum of f coincides with its supremum, i.e.,

$$f \text{ is continuous on } U \quad \Longrightarrow \quad \operatorname*{ess\,sup}_{x\in U} f(x) \;=\; \sup_{x\in U} f(x). \qquad \diamondsuit$$

For dimension $d = 1$, a small extension of Exercise 2.2.30 shows that if I is any type of interval in $\mathbb{R}$ and $f\colon I \to \mathbb{R}$ is continuous, then the essential supremum of f equals its supremum on I. However, if E is a generic measurable set and $f\colon E \to \mathbb{R}$ is continuous, then the essential supremum of f need not equal its supremum on E (this is Problem 2.2.45).

Problems

2.2.31. Suppose that F and K are nonempty, disjoint subsets of $\mathbb{R}^d$ such that F is closed and K is compact. Prove that $\operatorname{dist}(F,K) > 0$. Exhibit nonempty disjoint closed sets E and F such that $\operatorname{dist}(E,F) = 0$.

2.2.32. Show that if A and B are any measurable subsets of $\mathbb{R}^d$, then

$$|A \cup B| + |A \cap B| \;=\; |A| + |B|.$$

2.2.33. Assume that $\{E_n\}_{n\in\mathbb{N}}$ is a sequence of measurable subsets of $\mathbb{R}^d$ such that $|E_m \cap E_n| = 0$ whenever $m \neq n$. Prove that $|\cup E_n| = \sum |E_n|$.

2.2.34. Let $S_r = \{x \in \mathbb{R}^d : \|x\| = r\}$ be the sphere of radius r in $\mathbb{R}^d$ centered at the origin. Prove that $|S_r| = 0$.

2.2.35. Suppose that E is a measurable subset of $\mathbb{R}$ and $|E \cap (E+t)| = 0$ for every $t \neq 0$. Prove that $|E| = 0$.

2.2.36. Let $E \subseteq \mathbb{R}^m$ and $F \subseteq \mathbb{R}^n$ be measurable sets. Assume that $\mathbf{P}(x,y)$ is a statement that is either true or false for each pair $(x,y) \in E \times F$. Suppose that

$$\text{for every } x \in E, \quad \mathbf{P}(x,y) \text{ is true for a.e. } y \in F.$$

Must it then be true that

$$\text{for a.e. } y \in F, \quad \mathbf{P}(x,y) \text{ is true for every } x \in E?$$

2.2.37. Given a set $E \subseteq \mathbb{R}^d$, prove that the following three statements are equivalent.

(a) E is Lebesgue measurable.

(b) For every $\varepsilon > 0$ there exists an open set U and a closed set F such that $F \subseteq E \subseteq U$ and $|U \backslash F| < \varepsilon$.

(c) There exists a G_δ-set G and an F_σ-set H such that $H \subseteq E \subseteq G$ and $|G \backslash H| = 0$.

2.2.38. Given a set $E \subseteq \mathbb{R}^d$ with $|E|_e < \infty$, show that the following two statements are equivalent.

(a) E is Lebesgue measurable.

(b) For each $\varepsilon > 0$ we can write $E = (S \cup A) \setminus B$ where S is a union of finitely many nonoverlapping boxes and $|A|_e$, $|B|_e < \varepsilon$.

2.2.39. Let E be a subset of $\mathbb{R}^d$ such that $0 < |E|_e < \infty$. Given $0 < \alpha < 1$, prove that there exists a cube Q such that $|E \cap Q|_e \geq \alpha\, |Q|$.

Remark: This problem will be used in the proof of Theorem 2.4.3.

2.2.40. Let E be a measurable subset of $\mathbb{R}^d$. Show that if A is any subset of $\mathbb{R}^d$ that is disjoint from E, then $|E \cup A|_e = |E| + |A|_e$.

2.2.41. Construct a two-dimensional analogue of the Cantor set C as follows. Subdivide the unit square $[0,1]^2$ into nine subsquares, and keep only the four closed corner squares. Repeat this process forever, and let S be the intersection of all of these sets. Prove that S has measure zero, equals its own boundary, has empty interior, and equals $C \times C$.

2.2.42. This problem will show that there exist closed sets with *positive measure* that have *empty interior.*

The Cantor set construction given in Example 2.1.23 removes 2^{n-1} intervals from F_n, each of length 3^{-n}, to obtain F_{n+1}. Modify this construction by removing 2^{n-1} intervals from F_n that each have length a_n instead of 3^{-n}, and set $P = \cap F_n$.

(a) Show that P is closed, P contains no open intervals, $P^\circ = \varnothing$, $P = \partial P$, and $U = [0,1] \backslash P$ is dense in $[0,1]$.

(b) Show that if $a_n \to 0$ quickly enough, then $|P| > 0$. In fact, given $0 < \varepsilon < 1$, exhibit a_n such that $|P| = 1 - \varepsilon$.

Remark: P is called a *Smith–Volterra–Cantor set* or a *fat Cantor set.*

2.2.43. Define the *inner Lebesgue measure* of a set $A \subseteq \mathbb{R}^d$ to be

$$|A|_i \;=\; \sup\bigl\{|F| \,:\, F \text{ is closed and } F \subseteq A\bigr\}.$$

Prove the following statements.

(a) If A is Lebesgue measurable, then $|A|_e = |A|_i$.

(b) If $|A|_e < \infty$ and $|A|_e = |A|_i$, then A is Lebesgue measurable.

(c) There exists a nonmeasurable set A that satisfies $|A|_e = |A|_i = \infty$ (assume that nonmeasurable sets exist; this will be proved in Section 2.4).

(d) If $E \subseteq \mathbb{R}^d$ is Lebesgue measurable and $A \subseteq E$, then

$$|E| \;=\; |A|_i \;+\; |E \backslash A|_e.$$

2.2.44. Let E be a measurable subset of $\mathbb{R}^d$ such that $|E| < \infty$. Suppose that A and B are disjoint subsets of E such that $E = A \cup B$. Show that

$$A \text{ and } B \text{ are measurable} \quad \Longleftrightarrow \quad |E| \;=\; |A|_e + |B|_e.$$

2.2.45. Exhibit a set E and a function $f\colon E \to \mathbb{R}$ that is continuous on E, yet $\operatorname{ess\,sup}_{x\in E} |f(x)| \neq \sup_{x\in E} |f(x)|$.

2.2.46. (a) Show that the complement of a G_δ-set is an F_σ-set, and the complement of an F_σ-set is a G_δ-set.

(b) Show that every countable set is an F_σ-set.

(c) Is any countable set a G_δ-set? Is every countable set a G_δ-set? Is $\{1/n\}_{n\in\mathbb{N}}$ a G_δ-set?

(d) Exhibit a subset of $\mathbb{R}$ that belongs to one of the classes $G_{\delta\sigma}$, $F_{\sigma\delta}$, $G_{\delta\sigma\delta}$ $F_{\sigma\delta\sigma}$, etc., but is not a G_δ-set or an F_σ-set.

2.2.47. Given a function $f\colon \mathbb{R}^d \to \mathbb{C}$, the *oscillation of f at the point x* is

$$\operatorname{osc}_f(x) \;=\; \inf_{\delta>0} \sup\bigl\{|f(y) - f(z)| : y, z \in B_\delta(x)\bigr\}.$$

Prove the following statements.

(a) f is continuous at x if and only if $\operatorname{osc}_f(x) = 0$.

(b) For each $\varepsilon > 0$, the set $\{x \in \mathbb{R}^d : \operatorname{osc}_f(x) \ge \varepsilon\}$ is closed.

(c) $D = \{x \in \mathbb{R}^d : f \text{ is discontinuous at } x\}$ is an F_σ-set, and therefore the set of continuities of f is a G_δ-set.

2.2.48. Given $A \subseteq \mathbb{R}^d$, prove the following statements.

(a) There exists a measurable set $H \supseteq A$ that satisfies $|A \cap E|_e = |H \cap E|$ for every measurable set $E \subseteq \mathbb{R}^d$.

(b) We can choose the set H in part (a) to be a G_δ-set.

(c) If $\{E_k\}_{k\in\mathbb{N}}$ is any collection of disjoint measurable subsets of $\mathbb{R}^d$, then

$$\left|A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)\right|_e \;=\; \sum_{k=1}^{\infty} |A \cap E_k|_e.$$

2.2.49. (a) Let A be any subset of $\mathbb{R}^d$, and let $\mathcal{L}(A) = \bigl\{E \cap A : E \in \mathcal{L}(\mathbb{R}^d)\bigr\}$ be the restriction of all Lebesgue measurable sets to A. Show that $\mathcal{L}(A)$ is a σ-algebra on A.

(b) Prove that if A is Lebesgue measurable, then $\mathcal{L}(A)$ consists of all subsets of A that are Lebesgue measurable, i.e.,

$$\mathcal{L}(A) \;=\; \bigl\{E \subseteq A : E \in \mathcal{L}(\mathbb{R}^d)\bigr\}.$$

2.2.50. Let X be a set, and let Σ be the collection of all $E \subseteq X$ such that at least one of E or $X \backslash E$ is countable. Prove that Σ is a σ-algebra on X.

2.2.51. (a) Given a set X and σ-algebras Σ_1 and Σ_2 on X, prove that

$$\Sigma_1 \cap \Sigma_2 = \bigl\{A \subseteq X : A \in \Sigma_1 \text{ and } A \in \Sigma_2\bigr\}$$

is a σ-algebra on X.

(b) Prove that the intersection of an arbitrary collection of σ-algebras on X is a σ-algebra on X.

(c) Let $\mathcal{E}$ be a collection of subsets of X. Show that

$$\Sigma(\mathcal{E}) = \bigcap \bigl\{\Sigma : \Sigma \text{ is a } \sigma\text{-algebra on } X \text{ and } \mathcal{E} \subseteq \Sigma\bigr\}$$

is a σ-algebra on X. (We say that $\Sigma(\mathcal{E})$ is the *σ-algebra generated by $\mathcal{E}$*.)

2.3 More Properties of Lebesgue Measure

We will prove several important properties of Lebesgue measurable sets in this section. In particular, we will show in Section 2.3.1 that if $E_1 \subseteq E_2 \subseteq \cdots$ is an increasing sequence of nested measurable sets and $E = \cup E_k$, then the measure of E_k converges to the measure of E as $k \to \infty$ (but there is an interesting twist for nested *decreasing* sequences of sets; see Example 2.3.3). In Section 2.3.2 we will prove that the measure of a Cartesian product $E \times F$ is the product of the measures of E and F. Finally, in Section 2.3.3 we will prove that Lebesgue measure is invariant under rotations, and more generally we will determine the relationship between the measure of a measurable set E and the measure of its image $L(E)$ under a linear transformation $L\colon \mathbb{R}^d \to \mathbb{R}^d$.

2.3.1 Continuity from Above and Below

Suppose that A is a measurable set that is contained in another measurable set B. Monotonicity tells us that $|A| \le |B|$, but we can refine this a little further. The sets A and $B \backslash A$ are measurable and disjoint and their union is B, so by countable additivity we know that

$$|B| = |A| + |B \backslash A|. \tag{2.23}$$

If $|A| = \infty$ then both sides of equation (2.23) are infinity. If $|A| < \infty$ then we can take one more step and subtract $|A|$ from both sides of the equation to obtain $|B \backslash A| = |B| - |A|$. As long as $|A|$ is finite, this equality holds in the extended real sense, even if $|B|$ is infinite. We formalize this as follows.

Lemma 2.3.1. *If $A \subseteq B$ are Lebesgue measurable sets and $|A| < \infty$ then*

$$|B \setminus A| = |B| - |A|, \tag{2.24}$$

in the sense that if $|B| < \infty$ then both sides of equation (2.24) *are finite and equal, while if $|B| = \infty$ then both sides of equation* (2.24) *are ∞.* $\diamondsuit$

We will use Lemma 2.3.1 to determine the behavior of the measures of a sequence of nested increasing measurable sets $E_1 \subseteq E_2 \subseteq \cdots$. Let $E = \cup E_k$, and write E as the following countable union of disjoint measurable sets:

$$E = E_1 \cup (E_2 \setminus E_1) \cup (E_3 \setminus E_2) \cup \cdots.$$

Applying countable additivity gives

$$|E| = |E_1| + |E_2 \setminus E_1| + |E_3 \setminus E_2| + \cdots. \tag{2.25}$$

By Lemma 2.3.1, if E_k has finite measure, then $|E_k \setminus E_{k-1}| = |E_k| - |E_{k-1}|$. This suggests that we can turn equation (2.25) into a telescoping sum, at least if every set E_k has finite measure. In fact, in this case we see that

$$\begin{aligned}
|E| &= |E_1| + \sum_{k=2}^{\infty} |E_k \setminus E_{k-1}| \\
&= |E_1| + \lim_{N \to \infty} \sum_{k=2}^{N} \Bigl(|E_k| - |E_{k-1}|\Bigr) \\
&= |E_1| + \Bigl(\lim_{N \to \infty} |E_N|\Bigr) - |E_1| \\
&= \lim_{N \to \infty} |E_N|.
\end{aligned}$$

On the other hand, if any one of the sets E_k has infinite measure, then monotonicity implies that $|E| = \infty = \lim |E_k|$. In any case, we have shown that the measure of E_k increases to the measure of E. We call this property *continuity from below*, and state it precisely as the following theorem.

Theorem 2.3.2 (Continuity from Below). *If $E_1, E_2, \dots$ are measurable subsets of $\mathbb{R}^d$ such that $E_1 \subseteq E_2 \subseteq \cdots$, then $|E_1| \le |E_2| \le \cdots$ and*

$$\left| \bigcup_{k=1}^{\infty} E_k \right| = \lim_{k \to \infty} |E_k|. \qquad \diamondsuit$$

In contrast, the following example demonstrates that the measure of nested *decreasing* sets $E_1 \supseteq E_2 \supseteq \cdots$ need not converge to the measure of $\cap E_k$.

Example 2.3.3. Let $B_k(0)$ be the open ball of radius k centered at the origin, and let E_k be its complement:

$$E_k = \mathbb{R}^d \setminus B_k(0) = \{x \in \mathbb{R}^d : \|x\| \ge k\}.$$

Each E_k is measurable, and $E_1 \supseteq E_2 \supseteq \cdots$. Furthermore, the intersection of all of these sets is $\cap E_k = \varnothing$. Therefore

$$\left|\bigcap_{k=1}^{\infty} E_k\right| = 0 \qquad \text{yet} \qquad \lim_{k\to\infty} |E_k| = \infty. \qquad \diamondsuit$$

Although "continuity from above" does not *always* hold, the next theorem shows that if all of the sets E_k have finite measure (or finite measure from some point onward), then continuity from above applies to that sequence.

Theorem 2.3.4 (Continuity from Above). *If $E_1 \supseteq E_2 \supseteq \cdots$ are measurable subsets of $\mathbb{R}^d$ and $|E_k| < \infty$ for some k, then $|E_1| \ge |E_2| \ge \cdots$ and*

$$\left|\bigcap_{k=1}^{\infty} E_k\right| = \lim_{k\to\infty} |E_k|.$$

Proof. Suppose that $E_1 \supseteq E_2 \supseteq \cdots$ are measurable and $|E_k| < \infty$ for some k. Since our sets are nested decreasing, by ignoring $E_1, \dots, E_{k-1}$ and reindexing, we may assume that $|E_1| < \infty$.

Set $F_j = E_1 \setminus E_j$. Then $F_1 \subseteq F_2 \subseteq \cdots$. Further, since $|E_1| < \infty$, we have $|F_j| = |E_1| - |E_j|$. Also

$$E_1 \setminus \left(\bigcap_{k=1}^{\infty} E_k\right) = \bigcup_{j=1}^{\infty} F_j,$$

so we compute that

$$\begin{aligned}
|E_1| - \left|\bigcap_{k=1}^{\infty} E_k\right| &= \left|\bigcup_{j=1}^{\infty} F_j\right| && \text{(by Lemma 2.3.1)}\\
&= \lim_{j\to\infty} |F_j| && \text{(by continuity from below)}\\
&= \lim_{j\to\infty} \bigl(|E_1| - |E_j|\bigr) && \text{(by Lemma 2.3.1)}\\
&= |E_1| - \lim_{j\to\infty} |E_j|.
\end{aligned}$$

All of the above quantities are finite, so we can rearrange and obtain the desired result. $\square$

Combining continuity from above with Lemma 2.2.20 gives us the following corollary.

Corollary 2.3.5. *If $E \subseteq \mathbb{R}^d$ is measurable and $|E| < \infty$, then there exist open sets $V_1 \supseteq V_2 \supseteq \cdots \supseteq E$ such that $\lim_{k\to\infty} |V_k| = |E|$.*

Proof. By Lemma 2.2.20, there exists a G_δ-set H that contains E and has exactly the same measure as E. Furthermore, that lemma tells us that we can find a sequence of nested decreasing open sets $U_1 \supseteq U_2 \supseteq \cdots$ whose intersection is H. By Theorem 2.1.27, there exists an open set $U \supseteq H$ such that $|U| \le |H| + \varepsilon < \infty$. Therefore, if we set $V_k = U \cap U_k$ then we obtain a decreasing sequence of open sets V_k, each with finite measure, whose intersection is H. Consequently, continuity from above implies that

$$\lim_{k\to\infty} |V_k| = |H| = |E|. \qquad \square$$

2.3.2 Cartesian Products

Now we will establish the seemingly "obvious" fact that the measure of a Cartesian product

$$E \times F = \{(x,y) : x \in E, y \in F\}$$

of measurable sets E and F equals the product of the measures of the two sets. This is certainly true if E and F are *boxes.* For general measurable sets E and F, we can easily obtain an *inequality* that relates $|E \times F|$ to $|E|\,|F|$, for if $\{Q_k\}_k$ is a covering of E by boxes and $\{R_\ell\}_\ell$ is a covering of F by boxes then $\{Q_k \times R_\ell\}_{k,\ell}$ is a covering of $E \times F$ by boxes, and therefore

$$|E \times F| \le \sum_{k,\ell} \mathrm{vol}(Q_k \times R_\ell) = \Bigl(\sum_k \mathrm{vol}(Q_k)\Bigr)\Bigl(\sum_\ell \mathrm{vol}(R_\ell)\Bigr).$$

If E and F have finite measure, then by taking the infimum over all such coverings of E and F we obtain $|E \times F| \le |E|\,|F|$ (and, with a bit more care, we can likewise show that $|E \times F| \le |E|\,|F|$ holds if either $|E| = \infty$ or $|F| = \infty$, the difficult cases being where the measure of one set is zero and the other is infinite).

However, it is not so easy to prove that $|E \times F|$ must *equal* $|E|\,|F|$. We present the proof as an extended exercise that proceeds through cases to ultimately show that equality holds for arbitrary measurable sets. This exercise applies many of the techniques and properties of Lebesgue measure that we have established so far, including countable additivity, continuity from above, and the equivalent characterizations of measurability that appear in Lemma 2.2.21. As declared in the Preliminaries, we use the convention that $0 \cdot \infty = 0$. Indeed, the next exercise is a good illustration of why this is the "correct" way to define $0 \cdot \infty$, at least in the context of measure theory.

Exercise 2.3.6. (a) Observe that if $Q \subseteq \mathbb{R}^m$ and $R \subseteq \mathbb{R}^n$ are boxes, then $Q \times R$ is a box in $\mathbb{R}^{m+n}$ and $|Q \times R| = |Q|\,|R|$ (easy).

(b) Suppose that $U \subseteq \mathbb{R}^m$ and $V \subseteq \mathbb{R}^n$ are nonempty open sets. Show that $U \times V$ is open, and $|U \times V| = |U|\,|V|$.

(c) Suppose that $G \subseteq \mathbb{R}^m$ and $H \subseteq \mathbb{R}^n$ are bounded G_δ-sets. Show that $G \times H$ is a G_δ-set, and use Lemma 2.2.20(b) to prove that $|G \times H| = |G|\,|H|$.

(d) Suppose that $E \subseteq \mathbb{R}^m$ is a measurable set and $Z \subseteq \mathbb{R}^n$ satisfies $|Z| = 0$. Prove that $|E \times Z| = 0 = |E|\,|Z|$.

(e) Suppose that $E \subseteq \mathbb{R}^m$ and $F \subseteq \mathbb{R}^n$ are any measurable sets. Prove that $E \times F$ is measurable and $|E \times F| = |E|\,|F|$. $\diamondsuit$

We formalize the conclusion of Exercise 2.3.6 as a theorem.

Theorem 2.3.7 (Cartesian Products). *If $E \subseteq \mathbb{R}^m$ and $F \subseteq \mathbb{R}^n$ are Lebesgue measurable sets, then $E \times F \subseteq \mathbb{R}^{m+n}$ is a Lebesgue measurable subset of $\mathbb{R}^{m+n}$, and*

$$|E \times F| \;=\; |E|\,|F|. \qquad \diamondsuit$$

2.3.3 Linear Changes of Variable

We have already seen that Lebesgue measure is invariant under translations, and Problem 2.1.38 considered the behavior of Lebesgue measure under certain types of dilations. Now we want to consider the relation between the measure of a set $E \subseteq \mathbb{R}^d$ and the measure of its image under an arbitrary linear transformation $L\colon \mathbb{R}^d \to \mathbb{R}^d$. We will show that if E is measurable, then the measure of $L(E)$ equals the measure of E multiplied by the absolute value of the determinant of the transformation L. In particular, it follows that Lebesgue measure is invariant under rotations. This seems like another "obvious" property that should be trivial to establish, but the proof is not as straightforward as it might appear at first glance (try to prove this directly from the definition).

Before we can determine the measure of $L(E)$, we must first establish that $L(E)$ is measurable. Contrary to what we might expect, it is not true that the image of a measurable set under a generic continuous function need be measurable! In fact, the following example shows that if $n > m$ then we can even find a *linear* function $L\colon \mathbb{R}^n \to \mathbb{R}^m$ that maps some measurable sets to nonmeasurable sets.

Example 2.3.8. (a) Let N be any nonmeasurable subset of $\mathbb{R}$ (we will prove that such sets exist in Section 2.4). As a subset of $\mathbb{R}^2$, $E = N \times \{0\}$ has measure zero and therefore is a measurable subset of $\mathbb{R}^2$. However, if we define $L\colon \mathbb{R}^2 \to \mathbb{R}$ by $L(x_1, x_2) = x_1$, then L is linear and E is measurable, yet $L(E) = N$ is not measurable. The same idea can be used to prove that whenever $m < n$, there exists a linear function $L\colon \mathbb{R}^n \to \mathbb{R}^m$ that maps some measurable subset of $\mathbb{R}^n$ to a nonmeasurable set in $\mathbb{R}^m$.

(b) The situation is quite different when $n < m$. If $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is linear then range(L) is a subspace of $\mathbb{R}^m$ with dimension at most n. Consequently range(L) is a *proper* subspace of $\mathbb{R}^m$, and therefore it has measure zero (see Problem 2.1.37). Thus if $n < m$ then a linear function $L\colon \mathbb{R}^n \to \mathbb{R}^m$ maps *every* subset of $\mathbb{R}^n$ to a set of measure zero. $\diamondsuit$

The following lemma shows that the question of whether a continuous function maps measurable sets to measurable sets can be reduced to the question of whether the function maps sets with measure zero to sets with measure zero.

Lemma 2.3.9. *Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be a continuous function. Suppose that f maps sets with measure zero to sets with measure zero, i.e.,*

$$Z \subseteq \mathbb{R}^n \text{ and } |Z| = 0 \implies |f(Z)| = 0. \tag{2.26}$$

Then f maps measurable sets to measurable sets, i.e.,

$$E \subseteq \mathbb{R}^n \text{ is measurable} \quad \implies \quad f(E) \subseteq \mathbb{R}^m \text{ is measurable.}$$

Proof. Assume that f is continuous and equation (2.26) holds. If E is an arbitrary measurable subset of $\mathbb{R}^n$ then Lemma 2.2.21 tells us that $E = H \cup Z$ where H is an F_σ-set and $|Z| = 0$. Therefore

$$f(E) \;=\; f(H \cup Z) \;=\; f(H) \cup f(Z).$$

Since f is continuous, Exercise 2.2.22 implies that f maps F_σ-sets to F_σ-sets. Therefore $f(H)$ is an F_σ-set. On the other hand, equation (2.26) implies that $f(Z)$ has measure zero. Therefore $f(H)$ and $f(Z)$ are both measurable, so $f(E)$ is measurable as well. $\square$

This issue of whether a function maps sets with measure zero to sets with measure zero is quite important. In particular, we will encounter this condition again when we consider *absolutely continuous functions* in Chapter 6, especially in connection with the *Banach–Zaretsky Theorem* (Theorem 6.3.1), which gives several equivalent characterizations of absolutely continuous functions.

In light of Lemma 2.3.9, we would like to find sufficient criteria that ensure that a function maps sets with measure zero to sets with measure zero. The next definition, which extends the notion of *Lipschitz continuity* introduced for functions $f\colon \mathbb{R} \to \mathbb{R}$ in Definition 1.4.1 to functions $f\colon \mathbb{R}^n \to \mathbb{R}^m$, will be instrumental in this regard.

Definition 2.3.10 (Lipschitz Function). A function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is *Lipschitz* if there exists a constant $K \geq 0$ such that

$$\|f(x) - f(y)\| \;\leq\; K\,\|x-y\|, \qquad \text{for all } x, y \in \mathbb{R}^n.$$

The number K is called a *Lipschitz constant* for f. $\diamondsuit$

Thus, for a Lipschitz function there is some control over how far apart $f(x)$ and $f(y)$ can be in comparison to the distance between the points x and y. Every Lipschitz function is continuous, but not every continuous function is Lipschitz. The following lemma shows that all linear functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ are Lipschitz.

Lemma 2.3.11. *Every linear function $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz.*

Proof. Let $\{e_1, \dots, e_n\}$ be the standard basis for $\mathbb{R}^n$. Then $L(e_1), \dots, L(e_n)$ are finitely many vectors in $\mathbb{R}^m$, so $M = \max\{\|L(e_1)\|, \dots, \|L(e_n)\|\}$ is a finite number. Given a vector $x = (x_1, \dots, x_n) = x_1e_1 + \cdots + x_ne_n \in \mathbb{R}^n$, we have

$$\begin{aligned}
\|L(x)\| &= \|x_1L(e_1) + \cdots + x_nL(e_n)\| && \text{(linearity)}\\
&\le |x_1|\,\|L(e_1)\| + \cdots + |x_n|\,\|L(e_n)\| && \text{(Triangle Inequality)}\\
&\le M\sum_{k=1}^{n} |x_k| && \text{(definition of } M\text{)}\\
&\le Mn^{1/2}\left(\sum_{k=1}^{n} |x_k|^2\right)^{1/2} && \text{(exercise)}\\
&= Mn^{1/2}\,\|x\|.
\end{aligned}$$

Therefore, if $x, y \in \mathbb{R}^n$, then by using the linearity of L we see that

$$\|L(x) - L(y)\| = \|L(x-y)\| \le Mn^{1/2}\,\|x-y\|.$$

Hence L is Lipschitz, with Lipschitz constant $K = Mn^{1/2}$. □

For the rest of this section we will focus on the case $m = n = d$. We will prove below that any Lipschitz function that maps $\mathbb{R}^d$ *into itself* must map sets with measure zero to sets with measure zero. The key is the following exercise, which bounds the measure of the image of a cube under a Lipschitz map. Recall that continuous functions map compact sets to compact sets, so $f(Q)$ is actually a compact set in this exercise, and hence is measurable.

Exercise 2.3.12. Assume $f\colon \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz. Show that there exists a constant $C \ge 0$ such that $|f(Q)| \le C\,|Q|$ for every *cube* $Q \subseteq \mathbb{R}^d$. ♢

Now we prove that a Lipschitz function $f\colon \mathbb{R}^d \to \mathbb{R}^d$ maps measurable sets to measurable sets (it is important here that the domain and codomain have the same dimension).

Theorem 2.3.13. *If $f\colon \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz, then f maps sets with measure zero to sets with measure zero, and f maps measurable sets to measurable sets.*

Proof. Let C be the constant given by Exercise 2.3.12, and let Z be any subset of $\mathbb{R}^d$ such that $|Z| = 0$. If we fix $\varepsilon > 0$, then there exists an open set $U \supseteq Z$ such that $|U| < \varepsilon$. We can write U as the union of countably many nonoverlapping cubes Q_k. Applying countable subadditivity and Exercise 2.3.12, we obtain

$$|f(Z)|_e \;\le\; |f(U)| \;\le\; \sum_{k=1}^{\infty} |f(Q_k)| \;\le\; \sum_{k=1}^{\infty} C\,|Q_k| \;=\; C\,|U| \;<\; C\varepsilon.$$

Since ε is arbitrary, it follows that $|f(Z)| = 0$. Thus f maps sets of measure zero to sets of measure zero. Lemma 2.3.9 therefore implies that f maps measurable sets to measurable sets. □

Combining Lemma 2.3.11 with Theorem 2.3.13 yields the following result. In contrast, in Section 5.1 we will construct a continuous (but nonlinear and non-Lipschitz) function φ that maps a measurable set E to a nonmeasurable set $\varphi(E)$.

Corollary 2.3.14. *Every linear function $L\colon \mathbb{R}^d \to \mathbb{R}^d$ maps sets with measure zero to sets with measure zero, and it maps measurable sets to measurable sets.* ♢

If $L\colon \mathbb{R}^d \to \mathbb{R}^d$ is linear, then there is a $d \times d$ matrix with real entries, which we also call L, such that $L(x)$ is simply the product of the matrix L with the vector x. We identify the linear transformation L with the matrix L, and use the two objects interchangeably. In particular, the determinant of the transformation L is the determinant of the matrix L, and we say that L is *nonsingular* or *invertible* if its determinant is nonzero. Using this notation, the following theorem states that the measure of $L(E)$ is $|\det(L)|$ times the measure of E.

Theorem 2.3.15 (Linear Change of Variables). *If $L\colon \mathbb{R}^d \to \mathbb{R}^d$ is linear and $E \subseteq \mathbb{R}^d$ is Lebesgue measurable, then $L(E)$ is a measurable subset of $\mathbb{R}^d$ and*

$$|L(E)| \;=\; |\det(L)|\,|E|. \qquad \diamondsuit$$

We will present the proof of Theorem 2.3.15 in the form of an extended exercise. Before doing so, we recall an important fact about linear transformations on Euclidean space. Among the many factorization theorems for matrices, the *singular value decomposition*, or SVD, states that a $d \times d$ matrix L with real entries can be written in the form

$$L \;=\; W\Delta V^{\mathrm{T}},$$

where V and W are $d \times d$ orthogonal matrices and Δ is a nonnegative $d \times d$ real diagonal matrix. An *orthogonal* matrix V is a square matrix with real entries whose columns are orthonormal vectors (equivalently, a real square matrix V is orthogonal if and only if $V^{\mathrm{T}}V = I$). As a linear transformation, an orthogonal matrix preserves both lengths and angles, and hence is a composition of rotations and flips. In particular, an orthogonal matrix V maps the unit ball $B_1(0)$ in $\mathbb{R}^d$ bijectively onto itself, and the determinant of V is ± 1.

Consequently, if $L = W\Delta V^{\mathrm{T}}$ is the SVD of L and $s_1, \dots, s_d$ are the diagonal entries of Δ, then

$$|\det(L)| \;=\; \det(\Delta) \;=\; s_1 \cdots s_d.$$

We call $s_1, \dots, s_d$ the *singular numbers* of L. In particular, L is invertible if and only if each of its singular numbers is nonzero. The SVD of L is closely related to the diagonalization of the symmetric matrix $L^{\mathrm{T}}L$. For more details on the singular value decomposition of arbitrary real or complex $m \times n$ matrices, we refer to [Str06, Sec. 6.3], [HJ90, Sec. 7.3], or [Heil18, Sec. 7.10].

The following exercise gives a proof of Theorem 2.3.15.

Exercise 2.3.16. Let $Q_0 = [0,1]^d$ be the unit cube in $\mathbb{R}^d$. For each linear transformation $L\colon \mathbb{R}^d \to \mathbb{R}^d$, set

$$d_L \;=\; |L(Q_0)|.$$

Since L is linear, $L(Q_0)$ is a parallelepiped in $\mathbb{R}^d$ (though not necessarily a rectangular parallelepiped). Eventually we will prove that the measure of $L(Q_0)$ is precisely $|\det(L)|$, but we do not know this yet. Prove the following statements.

(a) $|L(Q)| = d_L\,|Q|$ for every cube $Q \subseteq \mathbb{R}^d$.

(b) If L is nonsingular, then $|L(U)| = d_L\,|U|$ for every open set $U \subseteq \mathbb{R}^d$.

(c) If L is nonsingular, then $|L(E)| = d_L\,|E|$ for every measurable set $E \subseteq \mathbb{R}^d$.

(d) If Δ is a diagonal matrix, then $d_\Delta = |\det(\Delta)|$.

(e) If V is an orthogonal matrix, then $d_V = 1$.

(f) If A and B are two nonsingular $d \times d$ matrices, then $d_{AB} = d_A\, d_B$.

(g) Combine the previous steps and use the SVD to show that $d_L = |\det(L)|$ for every nonsingular $d \times d$ matrix L.

Finally, determine what modifications to the proof are necessary to show that $d_L = 0$ when L is singular (alternatively, find a different approach to the singular case). $\diamondsuit$

Problems

2.3.17. Assume that $E \subseteq \mathbb{R}^d$ is measurable, $0 < |E| < \infty$, and $A_n \subseteq E$ are measurable sets such that $|A_n| \to |E|$ as $n \to \infty$. Prove that there exists a subsequence $\{A_{n_k}\}_{k\in\mathbb{N}}$ such that $\left|\bigcap A_{n_k}\right| > 0$. Show by example that this can fail if $|E| = \infty$.

2.3.18. Prove that $E \subseteq \mathbb{R}^d$ is measurable if and only if for every box Q we have $|Q| = |Q \cap E|_e + |Q\backslash E|_e$.

2.3.19. Let E be a measurable subset of $\mathbb{R}^d$, and set $f(t) = |E \cap B_t(0)|$ for $t > 0$. Prove the following statements (Problem 1.1.23 may be useful).

(a) f is monotone increasing and continuous on $(0, \infty)$.

(b) $\lim_{t\to 0^+} f(t) = 0$.

(c) $\lim_{t\to\infty} f(t) = |E|$.

(d) If $|E| < \infty$, then f is uniformly continuous on $(0, \infty)$.

2.3.20. Given a measurable set $E \subseteq \mathbb{R}^d$ such that $0 < |E| \le \infty$, prove the following statements.

(a) There exists a measurable set $A \subseteq E$ such that $|A| > 0$ and $|E\backslash A| > 0$.

(b) There exist infinitely many disjoint measurable sets $E_1, E_2, \dots$ contained in E such that $|E_k| > 0$ for every k.

(c) If $|E| < \infty$, then we can choose the sets E_k in part (b) so that

$$|E_k| \;=\; 2^{-k}\,|E|, \qquad \text{for } k \in \mathbb{N}.$$

(d) There exist compact sets $K_n \subseteq E$ such that $\lim_{n\to\infty} |K_n| = |E|$.

(e) If $|E| = \infty$, then there exist disjoint measurable sets $A_1, A_2, \dots \subseteq E$ such that $|A_k| = 1$ for every $k \in \mathbb{N}$.

2.3.21. Let E be a measurable subset of $\mathbb{R}^d$ such that $|E| > 0$. Prove that there exists a point $x \in E$ such that for every $\delta > 0$ we have $|E \cap B_\delta(x)| > 0$.

2.3.22. Suppose that $m > n$ and $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is a Lipschitz, but not necessarily linear, function. Prove that $|\mathrm{range}(f)| = 0$.

2.3.23. Prove that if E is a measurable subset of $\mathbb{R}$, then $\{(x,y) \in \mathbb{R}^2 : x - y \in E\}$ is a measurable subset of $\mathbb{R}^2$.

2.3.24. Let E be a subset of $\mathbb{R}^d$. Set

$$d_E(x) \;=\; \mathrm{dist}(x,E) \;=\; \inf\{\|x-y\| : y \in E\}, \qquad \text{for } x \in \mathbb{R}^d,$$

and for each $r > 0$ let $E_r = \{x \in \mathbb{R}^d : \mathrm{dist}(x,E) < r\}$. Prove that the following statements hold.

(a) d_E is continuous on $\mathbb{R}^d$.

(b) E_r is open for each $r > 0$.

(c) If $E \subseteq \mathbb{R}^d$ is closed, then $d_E(x) = 0$ if and only if $x \in E$.

(d) Every closed set in $\mathbb{R}^d$ is a G_δ-set.

(e) Every open set in $\mathbb{R}^d$ is an F_δ-set.

(f) If E is compact, then $|E| = \lim_{r\to 0^+} |E_r|$. However, this can fail if E is a noncompact closed set, or if E is an open set (even if E is bounded).

2.3.25. Let $\mathcal{U} = \{U \subseteq \mathbb{R}^d : U \text{ is open}\}$ be the collection of all open subsets of $\mathbb{R}^d$ (i.e., $\mathcal{U}$ is the *topology* of $\mathbb{R}^d$). Let $\mathcal{B} = \Sigma(\mathcal{U})$ be the σ-algebra generated by $\mathcal{U}$ (see Problem 2.2.51). Prove the following statements.

(a) $\mathcal{B}$ contains every open set, closed set, G_δ-set, F_σ-set, $G_{\delta\sigma}$-set, $F_{\sigma\delta}$-set, and so forth.

(b) $\mathcal{B} \subseteq \mathcal{L}$, i.e., every element of $\mathcal{B}$ is a Lebesgue measurable set.

(c) If $E \subseteq \mathbb{R}^d$ is Lebesgue measurable, then $E = B \backslash Z$ where $B \in \mathcal{B}$ and $|Z| = 0$.

Remark: The elements of $\mathcal{B}$ are called the *Borel subsets* of $\mathbb{R}^d$, and $\mathcal{B}$ is the *Borel σ-algebra on* $\mathbb{R}^d$. Part (b) shows that every Borel set is Lebesgue measurable, and part (c) shows that every Lebesgue measurable set differs from a Borel set by at most a set of measure zero. There do exist Lebesgue measurable sets that are not Borel sets (see the remark following Problem 5.1.7, or the argument based on cardinality given in [Fol99, Sec. 1.6]).

2.4 Nonmeasurable Sets

We have not yet shown that nonmeasurable sets exist. For simplicity of presentation we will restrict our discussion to one dimension, but the same techniques can be applied in higher dimensions.

2.4.1 The Axiom of Choice

We will use the *Axiom of Choice* to prove the existence of a nonmeasurable set. The Axiom of Choice is one of the axioms of the standard form of set theory most commonly accepted in mathematics (Zermelo–Fraenkel set theory with the Axiom of Choice, or ZFC). Here is the formal statement of this axiom.

Axiom 2.4.1 (Axiom of Choice). Let S be a nonempty set, and let $\mathcal{P}$ be the family of all nonempty subsets of S. Then there exists a function $f\colon \mathcal{P} \to S$ such that $f(A) \in A$ for each set $A \in \mathcal{P}$. $\diamondsuit$

There are many statements that are equivalent to the Axiom of Choice. For example, Axiom 2.4.1 is equivalent to the statement that *every vector space has a Hamel basis*. Here is another equivalent statement (for the meaning of a Cartesian product of an arbitrary collection of sets, and for a proof of that Axioms 2.4.1 and 2.4.2 are equivalent, we refer to [Rot02, App. A]).

Axiom 2.4.2. The Cartesian product $\prod_{i \in I} A_i$ of any collection $\{A_i\}_{i \in I}$ of nonempty sets is nonempty. $\diamondsuit$

Axiom 2.4.2 implies that if $\{A_i\}_{i \in I}$ is a collection of disjoint, nonempty sets, then there exists a set $N \subseteq \cup A_i$ such that $N \cap A_i$ contains exactly one element for each $i \in I$. In other words, the set N contains precisely one element of each set A_i.

2.4.2 Existence of a Nonmeasurable Set

We define an equivalence relation $\sim$ on the real line $\mathbb{R}$ by declaring that two points x and y in $\mathbb{R}$ are related if the distance between them is rational. That is,

$$x \sim y \quad \iff \quad x - y \in \mathbb{Q}. \tag{2.27}$$

The *equivalence class* of a point $x \in \mathbb{R}$ is the set of all points that are related to x. We denote this equivalence class by $[x]$. For the relation $\sim$ defined in equation (2.27), the equivalence class of x is the set of rationals translated by x:

$$[x] \;=\; \bigl\{y \in \mathbb{R} : x - y \in \mathbb{Q}\bigr\} \;=\; \bigl\{r + x : r \in \mathbb{Q}\bigr\} \;=\; \mathbb{Q} + x.$$

As for any equivalence relation, any two equivalence classes are either identical or disjoint (for example, $[\pi] = [\pi + 2]$, while $[\pi]$ and $[\sqrt{2}]$ are disjoint). Therefore the set of distinct equivalence classes partitions the real line $\mathbb{R}$. Each equivalence class $[x] = \mathbb{Q} + x$ is a countable set, so there are uncountably many distinct equivalence classes. The Axiom of Choice implies that there exists a set $N \subseteq \mathbb{R}$ that contains exactly one element of each of the distinct equivalence classes of $\sim$. We will show that this set N is not Lebesgue measurable. To do this, we will need the following fact about measurable sets (which may seem surprising at first glance).

Theorem 2.4.3 (Steinhaus Theorem). *If $E \subseteq \mathbb{R}$ is Lebesgue measurable and $|E| > 0$, then the set of differences*

$$E - E \;=\; \bigl\{x - y : x, y \in E\bigr\}$$

contains an interval centered at 0.

Proof. By Problem 2.2.39, there exists a closed interval $I = [a, b]$ such that the measure of the set $F = E \cap I$ satisfies

$$|F| \;=\; |E \cap I| \;>\; \frac{3}{4}\,|I|. \tag{2.28}$$

If $t \geq 0$ then $I \cup (I + t) \subseteq [a, b + t]$, while if $t \leq 0$ then $I \cup (I + t) \subseteq [a - |t|, b]$. In any case, we see that

$$|I \cup (I+t)| \;\le\; |I| + |t|. \tag{2.29}$$

If F and $F+t$ are disjoint, then we must have

$$\begin{aligned}
2\,|I| \;&<\; 2 \cdot \frac{4}{3}\,|F| && \text{(by equation (2.28))}\\
&=\; \frac{4}{3}\,|F \cup (F+t)| && \text{(since F and $F+t$ are disjoint)}\\
&\le\; \frac{4}{3}\,|I \cup (I+t)| && \text{(by monotonicity)}\\
&\le\; \frac{4}{3}\,\bigl(|I| + |t|\bigr) && \text{(by equation (2.29)).}
\end{aligned}$$

This equation cannot hold when $|t|$ is small, so F and $F+t$ must intersect for all small enough $|t|$. Specifically,

$$|t| \;<\; \frac{1}{2}\,|I| \qquad \Longrightarrow \qquad F \cap (F+t) \,\neq\, \varnothing.$$

Hence $F-F$ contains the interval $\bigl(-\frac{|I|}{2}, \frac{|I|}{2}\bigr)$, and therefore $E-E$ must contain this interval as well. □

Problem 4.6.29 gives an appealing alternative proof of Theorem 2.4.3 based on Lebesgue integration and the operation of *convolution*.

Theorem 2.4.4. *The set N is not Lebesgue measurable.*

Proof. Recall that N contains exactly one element of each distinct equivalence class of the relation $\sim$. The distinct equivalence classes partition the real line, so their union is $\mathbb{R}$. Therefore

$$\mathbb{R} \;=\; \bigcup_{x\in N} (\mathbb{Q}+x) \;=\; \bigcup_{x\in N}\,\bigcup_{r\in\mathbb{Q}} \{r+x\} \;=\; \bigcup_{r\in\mathbb{Q}} (N+r). \tag{2.30}$$

Since exterior Lebesgue measure is translation-invariant, the exterior measure of $N+r$ is exactly the same as the exterior measure of N. Combining this fact with countable subadditivity, we see that

$$\infty \;=\; |\mathbb{R}|_e \;=\; \biggl|\bigcup_{r\in\mathbb{Q}} (N+r)\biggr|_e \;\le\; \sum_{r\in\mathbb{Q}} |N+r|_e \;=\; \sum_{r\in\mathbb{Q}} |N|_e.$$

Consequently, we must have $|N|_e > 0$. However, any two distinct points $x \neq y$ in N belong to distinct equivalence classes of the relation $\sim$, so x and y must differ by an irrational amount. Therefore $N-N$ contains no intervals, so Theorem 2.4.3 implies that N cannot be Lebesgue measurable. □

2.4.3 Further Results

In the very first paragraphs of this chapter we claimed that there is no nonzero function that is defined on *every* subset of $\mathbb{R}$, is *nonnegative*, and is both *countably additive* and *translation-invariant*. We will prove this claim now. As a corollary we obtain another proof, similar in spirit to the proof of Theorem 2.4.4 but without needing an appeal to Theorem 2.4.3, that there exist subsets of $\mathbb{R}$ that are not Lebesgue measurable.

Theorem 2.4.5. *There does not exist a function* $\mu\colon \mathcal{P}(\mathbb{R}) \to [0,\infty]$ *that satisfies all of the following properties:*
(a) $\mu\bigl([0,1)\bigr) = 1$,
(b) *if* $E_1, E_2, \dots$ *are disjoint subsets of* $\mathbb{R}$, *then* $\mu\bigl(\cup E_k\bigr) = \sum \mu(E_k)$, *and*
(c) $\mu(E+h) = \mu(E)$ *for all* $E \subseteq \mathbb{R}$ *and* $h \in \mathbb{R}$.

Proof. For this proof we use the same equivalence relation that was introduced in equation (2.27), but restricted to elements of $[0,1)$. That is, given points $x, y \in [0,1)$, we declare that $x \sim y$ if and only if x and y differ by a rational (note that this rational will lie between -1 and 1). The equivalence class of $x \in [0,1)$ is

$$[x] \;=\; \bigl\{y \in [0,1) : x - y \in \mathbb{Q}\bigr\}.$$

By the Axiom of Choice, there exists a set M that contains one element of each distinct equivalence class of this relation. Let $\{r_k\}_{k\in\mathbb{N}}$ be an enumeration of $\mathbb{Q} \cap [-1,1]$. The sets $M_k = M + r_k$ are disjoint, and

$$[0,1) \;\subseteq\; \bigcup_{k=1}^{\infty} M_k \;\subseteq\; [-1,2). \tag{2.31}$$

Suppose that there did exist a function $\mu\colon \mathcal{P}(\mathbb{R}) \to [0,\infty]$ that satisfies the properties (a)–(c) listed in the statement of the theorem. Then, by applying the countable additivity and translation-invariance properties of μ, we see that

$$\mu\bigl([-1,2)\bigr) \;=\; \mu\bigl([-1,0)\bigr) \,+\, \mu\bigl([0,1)\bigr) \,+\, \mu\bigl([1,2)\bigr) \;=\; 3. \tag{2.32}$$

Further, if we choose any sets $A \subseteq B \subseteq \mathbb{R}$ then, since μ is nonnegative and countably additive,

$$\mu(B) \;=\; \mu\bigl(A \cup (B\backslash A)\bigr) \;=\; \mu(A) \,+\, \mu(B\backslash A) \;\geq\; \mu(A).$$

Therefore μ is monotonic. Combining this with equations (2.31) and (2.32), we obtain

$$1 \;=\; \mu\bigl([0,1)\bigr) \;\leq\; \mu\Biggl(\bigcup_{k=1}^{\infty} M_k\Biggr) \;\leq\; \mu\bigl([-1,2)\bigr) \;=\; 3. \tag{2.33}$$

On the other hand, the countable additivity and translation-invariance properties of μ imply that

$$\mu\biggl(\bigcup_{k=1}^{\infty} M_k\biggr) \;=\; \sum_{k=1}^{\infty} \mu(M_k) \;=\; \sum_{k=1}^{\infty} \mu(M).$$

However, since $\mu(M) \ge 0$, the only possible values for the sum $\sum_{k=1}^{\infty} \mu(M)$ are zero (if $\mu(M) = 0$), or infinity (if $\mu(M) > 0$). This contradicts equation (2.33), so no such function μ can exist. $\square$

Corollary 2.4.6. *There exist subsets of* $\mathbb{R}$ *that are not Lebesgue measurable. In particular, the set* M *constructed in the proof of Theorem 2.4.5 is a subset of* $[0,1]$ *that is not Lebesgue measurable.*

Proof. If every subset of $\mathbb{R}$ were Lebesgue measurable, then $\mu(E) = |E|$ would define a nonnegative function on $\mathcal{P}(\mathbb{R})$ that satisfies statements (a), (b), and (c) of Theorem 2.4.5. Since no such function can exist, this is a contradiction.

This does not imply that the specific set M is nonmeasurable. However, if M were measurable, then the argument used in the proof of Theorem 2.4.5 would imply that $1 \le \sum_{k=1}^{\infty} |M| \le 3$, which is impossible. $\square$

At the beginning of Section 2.2, we motivated the definition of measurable sets by saying that it can be shown that exterior Lebesgue measure is not countably additive. Now we explain why that claim is a consequence of the existence of nonmeasurable sets.

Example 2.4.7. Since M is not measurable, by definition there must exist some $\varepsilon > 0$ such that for *every* open set $V \supseteq M$ we have

$$|V \backslash M|_e \;>\; \varepsilon.$$

On the other hand, because M has finite exterior measure, Theorem 2.1.27 implies that there is *some* open set $U \supseteq M$ such that

$$|M|_e \;\le\; |U| \;\le\; |M|_e + \varepsilon.$$

The sets M and $U \backslash M$ are disjoint, yet $|U \backslash M|_e > \varepsilon$, so

$$|M \cup (U \backslash M)|_e \;=\; |U|_e \;\le\; |M|_e + \varepsilon \;<\; |M|_e + |U \backslash M|_e. \qquad \diamondsuit$$

Problems

2.4.8. (a) Prove that continuity from below holds for *exterior* Lebesgue measure. That is, if $E_1 \subseteq E_2 \subseteq \cdots$ is any nested increasing sequence of subsets of $\mathbb{R}^d$ (even nonmeasurable sets), then $|\cup E_k|_e = \lim_{k\to\infty} |E_k|_e$.

Remark: This problem will be used in the proof of Lemma 6.2.1.

(b) Show that there exist sets $E_1 \supseteq E_2 \supseteq \cdots$ in $\mathbb{R}$ such that $|E_k|_e < \infty$ for every k and

$$\left|\bigcap_{k=1}^{\infty} E_k\right|_e < \lim_{k\to\infty} |E_k|_e.$$

Hence continuity from above does not hold for exterior Lebesgue measure.

2.4.9. Show that every subset of $\mathbb{R}$ that has positive exterior Lebesgue measure contains a nonmeasurable subset.

2.4.10. Given any integer $d > 0$, show that there exists a set $N \subseteq \mathbb{R}^d$ that is not Lebesgue measurable.

2.4.11. Assume that $E \subseteq \mathbb{R}^m$, $F \subseteq \mathbb{R}^n$, and $A \subseteq \mathbb{R}^{m+n}$ are all measurable sets. If we fix $x \in E$ and define

$$A_x = \bigl\{y \in F : (x,y) \in A\bigr\},$$

must it be true that A_x is a measurable subset of $\mathbb{R}^n$?

2.4.12. If X is a finite set, let $\#X$ denote the number of elements of X. Define $\mu\colon \mathcal{P}(\mathbb{R}) \to [0,\infty]$ by

$$\mu(E) = \begin{cases} \#E, & \text{if } E \text{ is finite,} \\ \infty, & \text{if } E \text{ is infinite.} \end{cases}$$

Determine which of the properties (a), (b), and (c) stated in Theorem 2.4.5 hold for μ and which fail.

Remark: This function μ is called *counting measure* on $\mathbb{R}$.

2.4.13. Define $\delta\colon \mathcal{P}(\mathbb{R}) \to [0,\infty]$ by

$$\delta(E) = \begin{cases} 1, & \text{if } 0 \in E, \\ 0, & \text{if } 0 \notin E. \end{cases}$$

Determine which of the properties (a), (b), and (c) stated in Theorem 2.4.5 hold for δ and which fail.

Remark: This function δ is called the *δ measure* or *Dirac measure* on $\mathbb{R}$.

2.4.14.* Assume that E is a bounded, measurable subset of $\mathbb{R}$.

(a) Let $E - x = \{y - x : y \in E\}$, and define

$$f(x) = |E \cap (E - x)|, \qquad \text{for } x \in \mathbb{R}.$$

Prove that f is continuous at $x = 0$.

Remark: This is easy to do using the techniques that we will develop in Chapter 4, but challenging to prove using only the results that we have developed so far.

(b) Use part (a) to give another proof of the Steinhaus Theorem.

Chapter 3

Measurable Functions

In this chapter we lay the groundwork for the definition of the Lebesgue integral of functions on $\mathbb{R}^d$, which will be presented in Chapter 4. We will not be able to integrate every function. In particular, the functions that we can integrate must be *measurable* in a sense that we will introduce in Section 3.1. After discussing measurability of functions in Sections 3.1–3.3, we consider some issues related to the convergence of sequences of measurable functions in Sections 3.4–3.5.

3.1 Definition and Properties of Measurable Functions

We will deal with real-valued, extended real-valued, and complex-valued functions. Since our domain is the real Euclidean space $\mathbb{R}^d$, it may seem odd at first to consider functions that take complex values. However, such functions are regularly encountered in practical settings. For example, given a fixed number $\xi \in \mathbb{R}$, the *complex exponential function with frequency* ξ is the function $e_\xi \colon \mathbb{R} \to \mathbb{C}$ defined by

$$e_\xi(x) \;=\; e^{2\pi i \xi x}, \qquad x \in \mathbb{R}.$$

These functions play key roles in many areas of mathematics, physics, and engineering, including harmonic analysis, quantum mechanics, and signal processing (for example, see [DM72], [Dau92], [Ben97], [Grö01], [SS03], [Kat04]), [Heil11]). We will see some of the importance of the complex exponential functions when we discuss the Fourier transform in Section 9.2.

By definition, a complex-valued function must take values in $\mathbb{C}$; it cannot take the values $\pm\infty$. An extended real-valued function takes values in $\mathbb{R} \cup \{\pm\infty\} = [-\infty, \infty]$. Every real-valued function is both an extended real-valued and a complex-valued function. However, an extended real-valued function need not be complex-valued, and a complex-valued function need

C. Heil, *Introduction to Real Analysis*, Graduate Texts in Mathematics 280,
https://doi.org/10.1007/978-3-030-26903-6_3

not be extended real-valued. Consequently, we end up needing to define measurability for two types of functions: Extended real-valued functions and complex-valued functions (each of which include the real-valued functions as a special case). We will consider extended real-valued functions first, and then consider complex-valued functions. Once we have finished defining measurability for both cases, it will be convenient to have a means of addressing both possibilities simultaneously, so that we do not have to state every result separately for extended real-valued and complex-valued functions. We introduced some terminology for this purpose in the Preliminaries; for ease of reference we restate that notation here.

Notation 3.1.1 (Scalars and the Symbol $\overline{\mathbf{F}}$). We let the symbol $\overline{\mathbf{F}}$ denote a choice of either the extended real line $[-\infty, \infty]$ or the complex plane $\mathbb{C}$. Associated with this choice, we make the following declarations.

- If $\overline{\mathbf{F}} = [-\infty, \infty]$, then the word *scalar* means a *real number* $c \in \mathbb{R}$.
- If $\overline{\mathbf{F}} = \mathbb{C}$, then the word *scalar* means a *complex number* $c \in \mathbb{C}$.

In particular, $\pm\infty$ are not scalars. $\diamondsuit$

Thus, when we write "$f\colon E \to \overline{\mathbf{F}}$," we mean that f is a function on the domain E and f is either extended real-valued or complex-valued.

Remark 3.1.2. Most of the extended real-valued functions that we encounter only take the values $\pm\infty$ on a set of measure zero. Such a function is said to be *finite almost everywhere*. Interpreting "finite" as meaning "not $\pm\infty$," a complex-valued function is finite at *every* point, and therefore is automatically finite a.e. Combining these two possibilities, we see that the phrase

- $f\colon E \to \overline{\mathbf{F}}$ *is finite a.e.*

is equivalent to the phrase

- *f is a function on E that is either complex-valued or is extended real-valued but finite at almost every point.*

The first phrase is more concise, but sometimes for emphasis we will write out the second phrase in full. $\diamondsuit$

3.1.1 Extended Real-Valued Functions

According to the next definition, an extended real-valued function f is measurable if the inverse image of each extended interval $(a, \infty]$ is a measurable set. To simplify the notation, it will be convenient to use some of the abbreviations that were laid out in the Preliminaries. These include shorthands such as

$$\{f > a\} \;=\; \bigl\{x \in E : f(x) > a\bigr\} \;=\; f^{-1}(a, \infty],$$

and

$$\{f \le g\} = \{x \in E : f(x) \le g(x)\}.$$

Definition 3.1.3 (Extended Real-Valued Measurable Functions). Let $E \subseteq \mathbb{R}^d$ and $f\colon E \to [-\infty, \infty]$ be given. We say that f is a *Lebesgue measurable function* on E, or simply a *measurable function* for short, if

$$\{f > a\} = f^{-1}(a, \infty]$$

is a measurable subset of $\mathbb{R}^d$ for each number $a \in \mathbb{R}$. $\diamondsuit$

Example 3.1.4. Let E be a subset of $\mathbb{R}^d$, and consider the characteristic function χ_E. If a is a real number, then

$$\{\chi_E > a\} = \begin{cases} \varnothing, & \text{if } a \ge 1, \\ E, & \text{if } 0 \le a < 1, \\ \mathbb{R}^d, & \text{if } a < 0. \end{cases}$$

Hence χ_E is a Lebesgue measurable function on $\mathbb{R}^d$ if and only if E is a Lebesgue measurable subset of $\mathbb{R}^d$. $\diamondsuit$

We do not explicitly require the domain E in the definition of a measurable function to be measurable, but in most circumstances this will be the case. In general, measurability of f "almost" implies measurability of the domain E. This statement is made precise in Problem 3.1.16, which shows that if $f\colon E \to [-\infty, \infty]$ is a measurable function and $\{f = -\infty\}$ is a measurable set, then E is measurable.

Sometimes it is useful to replace the intervals $(a, \infty]$ that appear in Definition 3.1.3 with other sets. The next lemma shows that the definition of measurability is unchanged if we replace the intervals $(a, \infty]$ by $[a, \infty]$, $[-\infty, a)$, or $[-\infty, a]$. The proof follows from the fact that any one of these types of intervals is a complement, countable union, or countable intersection of the other types of intervals.

Lemma 3.1.5. *Let E be a subset of $\mathbb{R}^d$. If $f\colon E \to [-\infty, \infty]$, then the following four statements are equivalent.*

(a) *f is a measurable function, i.e., $\{f > a\}$ is measurable for each $a \in \mathbb{R}$.*

(b) *$\{f \ge a\}$ is measurable for each $a \in \mathbb{R}$.*

(c) *$\{f < a\}$ is measurable for each $a \in \mathbb{R}$.*

(d) *$\{f \le a\}$ is measurable for each $a \in \mathbb{R}$.*

Proof. We will only prove two of the implications, as the others are similar.

(a) $\Rightarrow$ (b). Assume that $\{f > a\}$ is measurable for each $a \in \mathbb{R}$. Writing

$$\{f \ge a\} = \bigcap_{k=1}^{\infty} \{f > a - \tfrac{1}{k}\},$$

we see that $\{f \geq a\}$ is a countable intersection of measurable sets and hence is measurable.

(b) $\Rightarrow$ (c). If $\{f \geq a\}$ is measurable then so is its complement, which is $\{f < a\}$. ☐

We stated Definition 3.1.3 without motivation. To explain why it is reasonable, consider the inverse image $f^{-1}(U)$ of an open set $U \subseteq \mathbb{R}$. We can write U as a union of at most countably many (not necessarily disjoint) bounded open intervals (a_k, b_k), so the inverse image of U under f is

$$f^{-1}(U) \;=\; \bigcup_k f^{-1}(a_k, b_k) \;=\; \bigcup_k \{a_k < f < b_k\} \;=\; \bigcup_k \Big(\{a_k < f\} \cap \{f < b_k\}\Big).$$

If f is a measurable function, then $\{f > a_k\}$ and $\{f < b_k\}$ are both measurable sets, so $f^{-1}(U)$ is measurable as well. Hence, if f is measurable then

the inverse image of every open set is measurable.

Contrast this with the fact that a function is continuous if and only if *the inverse image of every open set is open.* In this sense measurability is a generalization of continuity. In particular, we have the following fact.

Lemma 3.1.6. *Every continuous real-valued function $f\colon \mathbb{R}^d \to \mathbb{R}$ is Lebesgue measurable.*

Proof. Since f is finite at each point, the inverse image of $(a, \infty]$ equals the inverse image of (a, ∞):

$$\{f > a\} \;=\; f^{-1}(a, \infty] \;=\; f^{-1}(a, \infty).$$

But f is continuous and (a, ∞) is an open set, so $\{f > a\}$ is an open set in $\mathbb{R}^d$. Open sets are measurable, so we conclude that f is a measurable function. ☐

In many circumstances, sets that have measure zero "don't matter." The next lemma shows that this philosophy holds for measurability of functions, in the sense that changing the values of a function on a set of measure zero does not affect the measurability of the function.

Lemma 3.1.7. *Let $E \subseteq \mathbb{R}^d$ be a measurable set, and let $f\colon E \to [-\infty, \infty]$ be a measurable function. If $g\colon E \to [-\infty, \infty]$ satisfies $g = f$ a.e., then g is measurable.*

Proof. Assume that f is measurable and $g = f$ a.e. Then $Z = \{f \neq g\}$ has measure zero, so it is measurable. Given $a \in \mathbb{R}$, let $Z_a = \{x \in Z : g(x) > a\}$. Then

$$\{g > a\} \;=\; \big(\{f > a\} \setminus Z\big) \cup Z_a.$$

Since $\{f > a\}$ is measurable and Z and Z_a both have measure zero, we conclude that $\{g > a\}$ is measurable. ☐

Combining Lemma 3.1.7 with the fact that continuous functions are measurable gives us the following result.

Corollary 3.1.8. *If $f\colon \mathbb{R}^d \to [-\infty,\infty]$ and there exists a continuous function $g\colon \mathbb{R}^d \to \mathbb{R}$ that equals f almost everywhere, then f is measurable.* $\diamondsuit$

It is important to note that equaling a continuous function almost everywhere is not the same as being continuous almost everywhere. The *Heaviside function* $H = \chi_{[0,\infty)}$ is continuous at all but one point, and therefore is continuous a.e., but there is no continuous function g such that $H = g$ a.e. In contrast, the characteristic function of the rationals, $\chi_{\mathbb{Q}}$, is not continuous at any point, yet $\chi_{\mathbb{Q}} = 0$ a.e., and the zero function is continuous at every point. While Corollary 3.1.8 shows that a function that equals a continuous function a.e. is measurable, we have not yet developed enough machinery to prove that a function that is continuous a.e. is measurable (we will do this in Exercise 3.2.9).

Remark 3.1.9. In addition to changing a function on a set of measure zero, it is sometimes convenient to allow f to actually be undefined on a set of measure zero. If Z is a subset of E that has measure zero, then a function f whose domain is $E\backslash Z$ is said to be *defined almost everywhere on E*. We say that such a function is measurable if it is measurable when we assign values to $f(x)$ for $x \in Z$. Since Z has measure zero, the measurability of f is unaffected by the choice of values that we assign to f on Z. $\diamondsuit$

Whenever we deal with an extended real-valued function f, the following related functions often appear.

Definition 3.1.10 (Positive and Negative Parts). Given an extended real-valued function $f\colon X \to [-\infty,\infty]$, the *positive part* of f is

$$f^+(x) \;=\; \max\{f(x),\, 0\},$$

and the *negative part* of f is

$$f^-(x) \;=\; \max\{-f(x),\, 0\}. \qquad \diamondsuit$$

By construction, f^+ and f^- are *nonnegative* extended real-valued functions, and we have the relations

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

We will show in Lemma 3.2.5 that f^+ and f^- are measurable whenever f is measurable.

3.1.2 Complex-Valued Functions

Every complex-valued function f can be written in the form $f = f_r + if_i$ where f_r and f_i are real-valued. We declare that a complex-valued function f is measurable if and only if its real part f_r and its imaginary part f_i are each measurable in the sense of Definition 3.1.3.

Definition 3.1.11 (Complex-Valued Measurable Functions). Let E be a subset of $\mathbb{R}^d$. Given a function $f\colon E \to \mathbb{C}$, write f in real and imaginary parts as $f = f_r + if_i$. Then we say that f is *Lebesgue measurable* on E, or simply *measurable* for short, if both f_r and f_i are measurable real-valued functions. $\diamondsuit$

A function $f\colon \mathbb{R}^d \to \mathbb{C}$ is continuous if and only if f_r and f_i are both continuous, so we have the following complex analogue of Lemma 3.1.6.

Lemma 3.1.12. *Every continuous function $f\colon \mathbb{R}^d \to \mathbb{C}$ is measurable.* $\diamondsuit$

The complex-valued analogue of Lemma 3.1.7 takes the following form and is proved in exactly the same manner.

Lemma 3.1.13. *Let $E \subseteq \mathbb{R}^d$ be a Lebesgue measurable set. If $f\colon E \to \mathbb{C}$ is measurable and $g = f$ a.e., then g is measurable.* $\diamondsuit$

Problems

3.1.14. Show that if $E \subseteq \mathbb{R}$ is measurable and $f\colon E \to \mathbb{R}$ is monotone increasing on E, then f is measurable.

3.1.15. Given $E \subseteq \mathbb{R}^d$, prove that $f\colon E \to [-\infty,\infty]$ is measurable if and only if $\{f > r\}$ is measurable for every rational number r.

3.1.16. Let E be a subset of $\mathbb{R}^d$. Prove that if $f\colon E \to [-\infty,\infty]$ is a measurable function and $\{f = -\infty\}$ is a measurable set, then E is measurable.

3.1.17. (a) Prove that if f is a measurable function, then $\{f = a\}$ is a measurable set for every $a \in \mathbb{R}$.

(b) Exhibit a *nonmeasurable* function $f\colon \mathbb{R} \to \mathbb{R}$ such that $\{f = a\}$ is measurable for every $a \in \mathbb{R}$.

3.1.18. (a) Prove that $f\colon \mathbb{R}^d \to \mathbb{R}$ is a measurable function if and only if $f^{-1}(U)$ is a measurable set for every open set $U \subseteq \mathbb{R}$.

(b) Prove that $f\colon \mathbb{R}^d \to \mathbb{C}$ is a measurable function if and only if $f^{-1}(U)$ is a measurable set for every open set $U \subseteq \mathbb{C}$.

3.1.19. Let $E \subseteq \mathbb{R}^d$ be a measurable set with $|E| > 0$, and assume that $f\colon E \to \overline{\mathbf{F}}$ is measurable.

(a) Show that if f is finite a.e., then there exists a measurable set $A \subseteq E$ such that $|A| > 0$ and f is bounded on A.

(b) Suppose that it is not the case that $f = 0$ a.e. (that is, f is nonzero on a set of positive measure). Prove that there exists a measurable set $A \subseteq E$ and a number $\delta > 0$ such that $|A| > 0$ and $|f| \geq \delta$ on A.

3.2 Operations on Functions

Now we investigate whether measurability is preserved under operations such as addition, multiplication, limits, and compositions. We will see that measurability is preserved in many cases, but there are situations where we need to be careful.

3.2.1 Sums and Products

We begin with addition of functions. This is an operation where there is a potential difficulty, because if we attempt to add two extended real-valued functions f and g then there may be points x where $f(x) + g(x)$ takes the indeterminate form $\infty - \infty$ or $-\infty + \infty$. The function $f + g$ is not defined at any such point. The following lemma shows that if $f(x) + g(x)$ never takes an indeterminate form, then $f + g$ will be measurable (assuming f and g are themselves measurable).

Lemma 3.2.1. *Let $E \subseteq \mathbb{R}^d$ be a Lebesgue measurable set, and assume that $f, g\colon E \to [-\infty, \infty]$ are measurable functions such that $f(x) + g(x)$ never takes the form $\infty - \infty$ or $-\infty + \infty$. Then the following statements hold.*

(a) *$\{f < g\}$ is a measurable set.*

(b) *$g + b$ and $-g + b$ are measurable for each number $b \in \mathbb{R}$.*

(c) *$f + g$ is measurable.*

Proof. (a) Since $\{f < r\}$ and $\{r < g\}$ are measurable and since a countable union of measurable sets is measurable, it follows that

$$\{f < g\} \;=\; \bigcup_{r \in \mathbb{Q}} \{f < r < g\} \;=\; \bigcup_{r \in \mathbb{Q}} \Big(\{f < r\} \cap \{r < g\}\Big)$$

is measurable.

(b) If we fix $b \in \mathbb{R}$, then for every $a \in \mathbb{R}$ we have

$$\{g + b > a\} \;=\; \{g > a - b\}.$$

This is measurable for every a, so $g+b$ is a measurable function. The function $-g$ is measurable since $\{-g > a\} = \{g < -a\}$ is measurable for every a. Consequently, $-g + b$ is measurable as well.

(c) Fix a number $a \in \mathbb{R}$. Part (b) implies that $a - g$ is measurable, so it follows from part (a) that

$$\{f + g > a\} \;=\; \{f > a - g\}$$

is measurable. This is true for every a, so $f + g$ is measurable. $\square$

Even if a function does take extended real values, in practice the set of points where $f(x)$ is $\pm\infty$ is typically a set of measure zero (such a function is said to be *finite almost everywhere*; see Remark 3.1.2). If f and g are both finite a.e., then $f(x)+g(x)$ will only be undefined on a set Z of measure zero. By Lemma 3.1.7, we can assign to $f(x) + g(x)$ any values we like for $x \in Z$ without affecting the measurability of $f + g$, or we can simply view $f + g$ as being undefined on Z. The following lemma proves that $f + g$ is measurable in this case (also compare Problem 3.2.16).

Lemma 3.2.2. *Let $E \subseteq \mathbb{R}^d$ be a Lebesgue measurable set and assume that $f, g\colon E \to [-\infty, \infty]$ are measurable functions that are finite a.e. Then $f + g$ and $f - g$ are measurable functions.*

Proof. Let Z be the set of measure zero where $f + g$ is not defined. Let $f_1(x) = f(x)$ for $x \notin Z$ and set $f_1(x) = 0$ for $x \in Z$, and define g_1 similarly. Then $f_1 = f$ a.e. and $g_1 = g$ a.e., so both f_1 and g_1 are measurable by Lemma 3.1.7. Further, Lemma 3.2.1 implies that $f_1 + g_1$ is measurable. Since $f + g = f_1 + g_1$ a.e., it follows that $f + g$ is measurable no matter how we define $f(x)+g(x)$ for $x \in Z$. Finally, since $-g$ is also measurable, we see that $f - g = f + (-g)$ is measurable as well. $\square$

Because of our convention that $0 \cdot \infty = 0$, the product of any two extended real-valued functions is defined at all points in their domain. The following lemma shows that the product of any two measurable functions that are finite a.e. is measurable (also compare Problem 3.2.17).

Lemma 3.2.3. *Let $E \subseteq \mathbb{R}^d$ be a measurable set. If $f, g\colon E \to [-\infty, \infty]$ are measurable and finite a.e., then fg is measurable as well.*

Proof. If $a \geq 0$ then

$$\{f^2 > a\} \;=\; \{f > a^{1/2}\} \cup \{f < -a^{1/2}\}$$

is measurable, so f^2 is a measurable function.

By Lemma 3.2.2, both $f+g$ and $f-g$ are measurable, so the preceding reasoning implies that $(f+g)^2$ and $(f-g)^2$ are measurable. Since these functions are finite a.e., we can apply Lemma 3.2.2 again and conclude that

$$fg = \frac{(f+g)^2-(f-g)^2}{4}$$

is measurable. □

Next we observe that measurability is preserved under quotients as long as we avoid division by zero and the indeterminate forms $\pm\infty/\infty$.

Lemma 3.2.4. *Let $E \subseteq \mathbb{R}^d$ be a measurable set. If $f, g\colon E \to [-\infty,\infty]$ are measurable, f is finite a.e., and $g \neq 0$ a.e., then f/g is measurable.*

Proof. Suppose first that g is nonzero at every point. In this case, if $a>0$ then $\bigl\{1/g > a\bigr\} = \bigl\{0 < g < 1/a\bigr\}$, which is measurable. Likewise, $\{1/g > a\}$ is measurable if $a = 0$ or $a < 0$, so we conclude that $1/g$ is measurable.

Now assume that g is nonzero almost everywhere. Define $h(x) = g(x)$ when $g(x) \neq 0$, and $h(x) = 1$ otherwise. Then $h = g$ a.e., so h is measurable and everywhere nonzero. Hence $1/h$ is measurable by our prior reasoning, and therefore $1/g$ is measurable since it equals $1/h$ a.e.

Since we have shown that $1/g$ is measurable, Lemma 3.2.3 implies that the product $f\cdot(1/g)$ is measurable. But f is finite a.e. so $f\cdot(1/g) = f/g$ a.e., and therefore f/g is measurable. □

3.2.2 Compositions

Now we consider compositions. We will show that if we compose a measurable function with a continuous function in the correct order, then the result will be measurable. As a consequence, the positive and negative parts f^+ and f^- of an extended real-valued function f are measurable, as is $|f|$ and positive powers of $|f|$.

Lemma 3.2.5. *Let $E \subseteq \mathbb{R}^d$ be a measurable set, and let $f\colon E \to [-\infty,\infty]$ be a measurable function that is finite a.e.*

(a) *If $\varphi\colon \mathbb{R} \to \mathbb{R}$ is continuous, then $\varphi \circ f$ is measurable.*

(b) *$|f|$, f^2, f^+, f^-, and $|f|^p$ for $p > 0$ are all measurable functions.*

Proof. (a) *Case 1.* Assume first that f is finite at all points, and fix $a \in \mathbb{R}$. Since φ is continuous and (a,∞) is an open set, the inverse image $\varphi^{-1}(a,\infty)$ is an open subset of $\mathbb{R}$. Since f is measurable and the inverse image of an open set under a measurable function is measurable (see Problem 3.1.18), we conclude that

$$\{\varphi \circ f > a\} \;=\; (\varphi \circ f)^{-1}(a,\infty) \;=\; f^{-1}(\varphi^{-1}(a,\infty))$$

is a measurable subset of $\mathbb{R}^d$. Hence $\varphi \circ f$ is measurable.

Case 2. Now suppose that f is finite at *almost every* point. Then we can create a function g that is finite at all points and equals f almost everywhere (for example, set $g(x) = 0$ at any point where $f(x) = \pm\infty$). Since f is measurable and $g = f$ a.e., the function g is measurable. Since g is also finite everywhere, Case 1 implies that $\varphi \circ g$ is measurable. Therefore $\varphi \circ f$ is measurable since it equals $\varphi \circ g$ almost everywhere.

(b) If $p > 0$, then $\varphi(x) = |x|^p$ is continuous on $\mathbb{R}$. It therefore follows from part (a) that $|f|^p = \varphi \circ f$ is measurable. Similarly, $\psi(x) = \max\{x, 0\}$ is continuous, so $f^+ = \psi \circ f$ is measurable. □

Although the composition $\varphi \circ f$ of a continuous function φ with a measurable function f must be measurable, it is not true that the composition $f \circ \varphi$ need be measurable, even if φ is continuous (a counterexample is given in Problem 5.1.7). Consequently, it is possible for the composition of two measurable functions to be nonmeasurable. On the other hand, the following lemma states that $f \circ L$ is measurable if f is measurable and $L\colon \mathbb{R}^d \to \mathbb{R}^d$ is a linear bijection.

Lemma 3.2.6. *Let E be a measurable subset of $\mathbb{R}^d$. If $f\colon E \to [-\infty,\infty]$ is a measurable function and $L\colon \mathbb{R}^d \to \mathbb{R}^d$ is an invertible linear transformation, then $f \circ L\colon L^{-1}(E) \to [-\infty,\infty]$ is measurable.*

Proof. Since L^{-1} is a linear mapping of $\mathbb{R}^d$ into itself, Corollary 2.3.14 tells us that L^{-1} maps measurable sets to measurable sets. Therefore the domain $L^{-1}(E)$ of the composition $f \circ L$ is a measurable set. If we choose any $a \in \mathbb{R}$, then

$$\{f \circ L > a\} \;=\; (f \circ L)^{-1}(a,\infty] \;=\; L^{-1}\big(f^{-1}(a,\infty]\big) \;=\; L^{-1}\big(\{f > a\}\big).$$

Since f is measurable and L^{-1} maps measurable sets to measurable sets, we conclude that $\{f \circ L > a\}$ is measurable. □

3.2.3 Suprema and Limits

Next we turn to suprema, infima, limsups, liminfs, and limits.

Lemma 3.2.7. *Assume $E \subseteq \mathbb{R}^d$ is measurable. If $f_n\colon E \to [-\infty,\infty]$ is measurable and finite a.e. for each $n \in \mathbb{N}$, then the following statements hold.*

(a) *Each of*

$$\sup_{n\in\mathbb{N}} f_n, \qquad \inf_{n\in\mathbb{N}} f_n, \qquad \limsup_{n\to\infty} f_n, \qquad \liminf_{n\to\infty} f_n,$$

is a measurable function on E.

(b) *If* $f(x) = \lim_{n\to\infty} f_n(x)$ *exists for a.e.* $x \in E$, *then* f *is measurable.*

(c) *If* $f(x) = \sum_{n=1}^{\infty} f_n(x)$ *exists for a.e.* $x \in E$, *then* f *is measurable.*

Proof. (a) Let $f(x) = \sup f_n(x)$. Then

$$\{f > a\} \;=\; \bigcup_{n=1}^{\infty} \{f_n > a\},$$

which is a measurable set. Therefore f is measurable. Since $-f_n$ is measurable, it follows that

$$\inf_{n\in\mathbb{N}} f_n \;=\; -\sup_{n\in\mathbb{N}} (-f_n)$$

is also measurable. Finally,

$$\limsup_{n\to\infty} f_n(x) \;=\; \inf_{m\in\mathbb{N}} \biggl(\sup_{n\geq m} f_n(x)\biggr),$$

so $\limsup f_n$ is measurable, and likewise $\liminf f_n$ is measurable.

(b) We know from part (a) that $\limsup f_n$ is measurable. Consequently, if $f(x) = \lim f_n(x)$ exists for a.e. x then f is equal almost everywhere to the measurable function $\limsup f_n$, so f is measurable.

(c) By Lemma 3.2.2, the partial sums $s_N(x) = \sum_{n=1}^{N} f_n(x)$ are measurable for each $N \in \mathbb{N}$. If these partial sums converge at almost every point, then part (b) implies that

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

is measurable. $\square$

We use the following notation to describe the type of situation that appears in part (b) of Lemma 3.2.7.

Notation 3.2.8. We say that functions f_n *converge pointwise a.e. to* f if

$$f(x) \;=\; \lim_{n\to\infty} f_n(x) \quad \text{for a.e. } x.$$

In this case we write $f_n \to f$ *pointwise a.e.*, or simply $f_n \to f$ *a.e.* $\diamondsuit$

Using this notation, Lemma 3.2.7(b) says that the pointwise a.e. limit of measurable functions is measurable.

As an application, we give an exercise that shows that any function that is continuous a.e. is measurable.

Exercise 3.2.9. Fix any function $f\colon \mathbb{R} \to \mathbb{R}$.

(a) For each $n \in \mathbb{N}$ set

$$\phi_n = \sum_{k \in \mathbb{Z}} f\left(\tfrac{k}{n}\right) \chi_{\left[\frac{k}{n}, \frac{k+1}{n}\right)}.$$

Prove that ϕ_n is measurable (even if f is not), and show that if f is continuous at a particular point x then $\phi_n(x) \to f(x)$ as $n \to \infty$.

(b) Show that if f is continuous at almost every point $x \in \mathbb{R}$, then f is Lebesgue measurable.

(c) By replacing intervals with boxes, extend part (a) to functions on $\mathbb{R}^d$, and prove that any function $f\colon \mathbb{R}^d \to \mathbb{R}$ that is continuous a.e. is Lebesgue measurable. $\diamondsuit$

Now we turn to complex-valued functions. In some ways, these are easier to deal with than extended real-valued functions because $f(x)$ must be a complex scalar for every x (hence every complex-valued function $f\colon E \to \mathbb{C}$ is finite at every point, and therefore is finite a.e.). On the other hand, we usually cannot take the sup, inf, limsup, or liminf of a sequence of complex-valued functions (although we can apply those operations to the real and imaginary parts separately). The proofs for the complex case mostly follow by breaking a function into its real and imaginary parts.

Exercise 3.2.10. Let $E \subseteq \mathbb{R}^d$ be a measurable set, and let $f, g, f_n\colon E \to \mathbb{C}$ be complex-valued measurable functions. Prove the following statements.

(a) $f + g$ is measurable.

(b) fg is measurable.

(c) If $g(x) \neq 0$ a.e., then f/g is measurable.

(d) If $h(x) = \lim_{n\to\infty} f_n(x)$ exists for a.e. $x \in E$, then h is measurable.

(e) If $s(x) = \sum_{n=1}^{\infty} f_n(x)$ exists for a.e. $x \in E$, then s is measurable.

(f) If $\varphi\colon \mathbb{C} \to \mathbb{C}$ is continuous, then $\varphi \circ f$ is measurable.

(g) $|f|^p$ is measurable for each $p > 0$.

(h) If $L\colon \mathbb{R}^d \to \mathbb{R}^d$ is an invertible linear transformation, then the composition $f \circ L\colon L^{-1}(E) \to \mathbb{C}$ is measurable. $\diamondsuit$

3.2.4 Simple Functions

In order to define the Lebesgue integral in Chapter 4, we will need to have a class of functions for which it is clear what the integral should be. For this purpose, the "simplest" functions to deal with are those that take only finitely many distinct scalar values. For example, the characteristic function χ_A of a measurable set A takes only the values 0 and 1, so it is "simple" in this sense. We consider some of the basic properties of these *simple functions* now.

Definition 3.2.11 (Simple Function). Let $E \subseteq \mathbb{R}^d$ be a Lebesgue measurable set. A *simple function* on E is a measurable function $\phi\colon E \to \mathbb{C}$ that takes only finitely many distinct values. $\diamondsuit$

A simple function can be real-valued, but it cannot take the values $\pm\infty$. In order for ϕ to be called a simple function, ϕ must be measurable, $\phi(x)$ must be a real or complex scalar for each $x \in E$, and the set of all values taken by ϕ must be a finite set. The set of all values of ϕ is just another name for the range of ϕ, so a simple function is precisely a measurable function whose range is a finite subset of $\mathbb{C}$. A simple function is nonnegative if its range is a finite subset of $[0, \infty)$.

Every characteristic function of a measurable set is a simple function. Furthermore, any finite linear combination of measurable characteristic functions is measurable and takes only finitely many scalar values, so is also simple. Hence if $E_1, \dots, E_N$ are measurable subsets of E and $c_1, \dots, c_N$ are complex scalars, then $\phi = \sum_{k=1}^N c_k \chi_{E_k}$ is a simple function. The next lemma (whose proof essentially follows "from inspection") states that every simple function has this form.

Lemma 3.2.12. *Let ϕ be a simple function whose domain is a measurable set $E \subseteq \mathbb{R}^d$. If $c_1, \dots, c_N$ are the distinct values taken by ϕ and we define*

$$E_k \;=\; \phi^{-1}\{c_k\} \;=\; \bigl\{\phi = c_k\bigr\}, \qquad \textit{for } k = 1, \dots, N, \tag{3.1}$$

then

$$\phi \;=\; \sum_{k=1}^{N} c_k \chi_{E_k}.$$

Moreover, the sets $E_1, \dots, E_N$ given in equation (3.1) *partition E into disjoint measurable sets.* $\diamondsuit$

There may be many ways to write a given simple function as a linear combination of characteristic functions, but the form given in Lemma 3.2.12 is particularly useful, so we give it the following special name.

Definition 3.2.13 (Standard Representation). The *standard representation* of a simple function ϕ is the representation given by Lemma 3.2.12, i.e., $\phi = \sum_{k=1}^N c_k \chi_{E_k}$ where $c_1, \dots, c_N$ are the distinct values taken by ϕ and $E_k = \{\phi = c_k\}$ for $k = 1, \dots, N$. $\diamondsuit$

For example, $\phi = \chi_{[0,2]} + \chi_{[1,3]}$ is a simple function on $\mathbb{R}$ because it takes only three distinct values. Its standard representation is

$$\phi \;=\; 0\chi_{E_1} + 1\chi_{E_2} + 2\chi_{E_3},$$

where $E_1 = (-\infty, 0) \cup (3, \infty)$, $E_2 = [0,1) \cup (2,3]$, and $E_3 = [1,2]$. Of course, we can also write ϕ in the form

$$\phi \;=\; 1\chi_{E_2} + 2\chi_{E_3},$$

but while the sets E_2, E_3 are disjoint, they do not partition the domain $\mathbb{R}$. In general, one of the scalars c_k in the standard representation of a simple function ϕ might be zero.

If $\phi = \sum_{j=1}^{M} c_j \chi_{E_j}$ and $\psi = \sum_{k=1}^{N} d_k \chi_{F_k}$ are the standard representations of simple functions ϕ and ψ, then $\phi + \psi$ is a linear combination of the characteristic functions of the sets $E_j \cap F_k$, because

$$\phi + \psi \;=\; \sum_{j=1}^{M} \sum_{k=1}^{N} (c_j + d_k)\, \chi_{E_j \cap F_k}. \tag{3.2}$$

This need not be the standard representation of $\phi + \psi$, since the scalars $c_j + d_k$ may coincide for different values of j and k. However, equation (3.2) does show that the sum of two simple functions is simple, and a similar computation shows that the product of two simple functions is simple.

Much of the utility of simple functions comes from our next theorem, which states that every nonnegative measurable function (including those that take the value ∞) can be written as the pointwise limit of a sequence of simple functions ϕ_n. In fact, we will be able to construct simple functions ϕ_n that increase monotonically to f at each point, and the convergence is uniform on any subset where f is bounded.

Theorem 3.2.14. *Let $E \subseteq \mathbb{R}^d$ be a measurable set, and let $f\colon E \to [0,\infty]$ be a nonnegative, measurable function on E.*

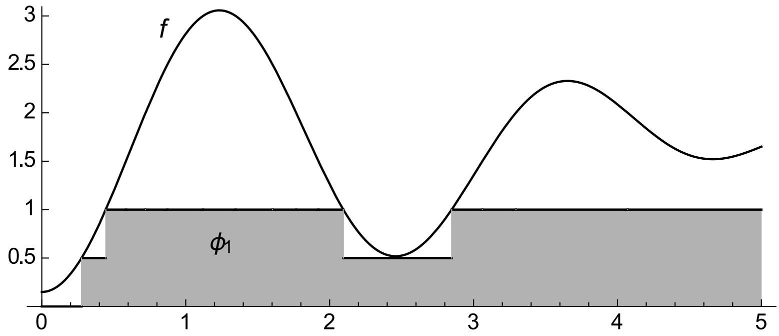
(a) *There exist nonnegative simple functions ϕ_n such that $\phi_n \nearrow f$. That is, $0 \le \phi_1 \le \phi_2 \le \cdots$, and $\lim_{n\to\infty} \phi_n(x) = f(x)$ for each $x \in E$.*

(b) *If f is bounded on some set $A \subseteq E$, then we can construct the functions ϕ_n in statement* (a) *so that they converge uniformly to f on A, i.e.,*

$$\lim_{n\to\infty} \|(f - \phi_n)\cdot \chi_A\|_{\mathrm{u}} \;=\; \lim_{n\to\infty} \Bigl(\sup_{x\in A} |f(x) - \phi_n(x)|\Bigr) \;=\; 0.$$

Proof. The idea is that we construct ϕ_n by simply rounding f down to the nearest integer multiple of 2^{-n}. However, if f is unbounded then this would give ϕ_n infinitely many values, yet a simple function can only take finitely many values. Therefore we stop the rounding-down process at the finite height n, which means that we define ϕ_n by

$$\phi_n(x) \;=\; \begin{cases} \dfrac{j-1}{2^n}, & \text{if } \dfrac{j-1}{2^n} \le f(x) < \dfrac{j}{2^n}, \quad j = 1, \dots, n2^n, \\[2ex] n, & \text{if } f(x) \ge n. \end{cases} \tag{3.3}$$



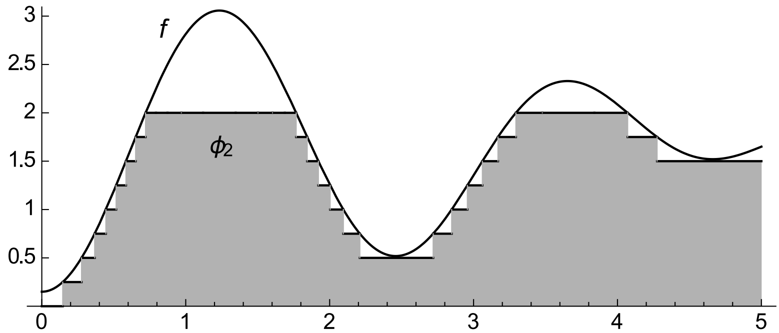


Fig. 3.1 Illustration of a function f and the approximating simple functions ϕ_1 and ϕ_2 constructed in the proof of Theorem 3.2.14 (the region under the graphs of ϕ_1 and ϕ_2 is shaded).

Illustrations for $n = 1$ and $n = 2$ appear in Figure 3.1.

By construction, ϕ_n is measurable, $\phi_n(x) \le \phi_{n+1}(x)$ for every x, and

$$f(x) \le n \quad \Longrightarrow \quad |f(x) - \phi_n(x)| \le 2^{-n}. \tag{3.4}$$

If $f(x) = \infty$ then $\phi_n(x) = n$ for every n, so $\phi_n(x) \to f(x)$ in this case. If $f(x)$ is finite, then n will eventually exceed $f(x)$, so equation (3.4) implies that $\phi_n(x) \to f(x)$. In fact, if $f(x) \le M < \infty$ for all x in some set A, then for each $n \ge M$ we simultaneously have $|f(x) - \phi_n(x)| \le 2^{-n}$ for every $x \in A$. This implies that ϕ_n converges uniformly to f on A. $\square$

Theorem 3.2.14 shows us how to write a nonnegative measurable function as a pointwise limit of simple functions. We will use this to prove that an arbitrary measurable function is a pointwise limit of simple functions. To do this, we follow a standard approach that we will see many times in the coming pages: We write an arbitrary function as a linear combination of nonnegative functions. Specifically, if a measurable function f takes extended real values then we write f as a difference of two nonnegative functions, and if f takes complex values then we write f as a linear combination of its real and imaginary parts, each of which is real-valued and can therefore be written as a difference of nonnegative functions. By applying Theorem 3.2.14 to the nonnegative functions that result from this splitting and then putting the pieces together, we create a sequence of simple functions that converge pointwise to f (although the convergence need not be monotone, as it is for nonnegative functions).

Corollary 3.2.15. *Let $E \subseteq \mathbb{R}^d$ be a measurable set. If $f\colon E \to \overline{\mathbf{F}}$ is a measurable function on E, then there exist simple functions ϕ_n on E such that:*

(a) $\lim_{n\to\infty} \phi_n(x) = f(x)$ *for each* $x \in E$,

(b) $|\phi_n(x)| \le |f(x)|$ *for every* $n \in \mathbb{N}$ *and* $x \in E$, *and*

(c) *the convergence is uniform on every set on which f is bounded.*

Proof. Suppose first that f is extended real-valued, and let f^+ and f^- be the positive and negative parts of f introduced in Definition 3.1.10. Since f^+ and f^- are nonnegative, there exist simple functions ϕ_n^+ and ϕ_n^-, such that $0 \le \phi_n^+ \nearrow f^+$ and $0 \le \phi_n^- \nearrow f^-$, and the convergence is uniform on any set on which f^+ and f^- are bounded. The result then follows by setting $\phi_n = \phi_n^+ - \phi_n^-$.

Exercise: Extend the proof to complex-valued functions by writing $f = f_r + i f_i$. $\square$

Problems

3.2.16. Let $E \subseteq \mathbb{R}^d$ be measurable, and assume that $f, g\colon E \to [-\infty, \infty]$ are measurable (but not necessarily finite a.e.). Given $c \in [-\infty, \infty]$, define

$$h(x) \;=\; \begin{cases} c, & \text{if } f(x) + g(x) \text{ is an indeterminate form,} \\ f(x) + g(x), & \text{otherwise.} \end{cases}$$

Prove that h is measurable.

3.2.17. Assume that $E \subseteq \mathbb{R}^d$ is measurable, and $f, g\colon E \to [-\infty, \infty]$ are any two measurable functions on E. Prove that fg is measurable.

3.2.18. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of measurable functions, either extended real-valued or complex-valued, whose domain is a measurable set $E \subseteq \mathbb{R}^d$. Show that

$$L \;=\; \Big\{x \in E : \lim_{n\to\infty} f_n(x) \text{ exists}\Big\} \quad \text{and} \quad S \;=\; \Big\{x \in E : \sum_{n=1}^{\infty} |f_n(x)| < \infty\Big\}$$

are measurable subsets of E.

3.2.19. Let $E \subseteq \mathbb{R}$ be a measurable set that is contained in an interval I, and assume that $f\colon I \to \mathbb{C}$ is a measurable function that is differentiable at each point in E, i.e.,

$$f'(x) \;=\; \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \quad \text{exists and is a scalar for all } x \in E.$$

Show that f' is a measurable function on E.

Remark: This problem will be used in the proof of Lemma 6.2.4.

3.2.20. Suppose that $f\colon \mathbb{R}^d \to \mathbb{R}$ is measurable, and $\varphi\colon \mathbb{R}^d \to \mathbb{R}^d$ is a bijection such that φ^{-1} is Lipschitz. Prove that $f \circ \varphi$ is measurable.

3.2.21. Assume that E is a measurable subset of $\mathbb{R}^d$ such that $|E| < \infty$.

(a) Suppose that $f\colon E \to \mathbb{R}$ is measurable. Prove that for each $\varepsilon > 0$, there is a closed set $F \subseteq E$ such that $|E\backslash F| < \varepsilon$ and f is bounded on F.

(b) Let f_n be a measurable function on E for each $n \in \mathbb{N}$. Suppose that for all $x \in E$ we have

$$M_x \;=\; \sup_{n\in\mathbb{N}} |f_n(x)| \;<\; \infty.$$

Prove that for each $\varepsilon > 0$, there exists a closed set $F \subseteq E$ and a finite constant M such that $|E\backslash F| < \varepsilon$ and $|f_n(x)| \le M$ for all $x \in F$ and $n \in \mathbb{N}$.

3.2.22. This problem is a continuation of Problem 2.3.25. Assume that $f\colon \mathbb{R}^d \to \mathbb{R}$ is a measurable function, and define

$$\Sigma \;=\; \bigl\{B \subseteq \mathbb{R} : B \text{ is measurable and } f^{-1}(B) \text{ is measurable}\bigr\}.$$

Prove the following statements.

(a) Σ is a σ-algebra of subsets of $\mathbb{R}^d$.

(b) $\mathcal{B} \subseteq \Sigma$, where $\mathcal{B}$ is the Borel σ-algebra.

(c) If B is a Borel set (i.e., $B \in \mathcal{B}$), then $f^{-1}(B)$ is a measurable set.

3.3 The Lebesgue Space $L^\infty(E)$

We will study several different spaces of measurable functions as we progress further through the text. The first of these is $L^\infty(E)$, which consists of all measurable, essentially bounded functions on E. By Definition 2.2.26, *essentially bounded* means that $\operatorname{ess\,sup}_{x\in E} |f(x)|$ is finite. For convenience, given a measurable function f on E, we define

$$\|f\|_\infty \;=\; \operatorname*{ess\,sup}_{x\in E} |f(x)|.$$

We call $\|f\|_\infty$ the *L^∞-norm of f* (although, as we will see, it is technically not a norm but rather is a *seminorm*).

Remark 3.3.1. For comparison, recall that the *uniform norm* of a function f on E is

$$\|f\|_{\mathrm{u}} \;=\; \sup_{x\in E} |f(x)|.$$

By Exercise 2.2.30, if f is a continuous function whose domain is an open set $U \subseteq \mathbb{R}^d$, then $\|f\|_\infty = \|f\|_{\mathrm{u}}$. However, in general we only have the inequality $\|f\|_\infty \le \|f\|_{\mathrm{u}}$. $\diamondsuit$

An essentially bounded function need not be bounded, but we do have the following, as an immediate consequence of Lemma 2.2.28.

Lemma 3.3.2. *Let E be a measurable subset of $\mathbb{R}^d$. If $f \in L^\infty(E)$, then*

$$|f(x)| \;\le\; \|f\|_\infty \quad \textit{for a.e. } x \in E. \qquad \diamondsuit$$

Every extended real-valued or complex-valued measurable function f on a measurable set $E \subseteq \mathbb{R}^d$ has a well-defined L^∞-norm, although $\|f\|_\infty$ could be infinite. A function is essentially bounded if and only if $\|f\|_\infty < \infty$. By Lemma 3.3.2, every essentially bounded function is finite a.e. (but not conversely—consider $f(x) = 1/x$).

We collect the essentially bounded functions to form the space $L^\infty(E)$. Technically, there are two versions of $L^\infty(E)$, one consisting of complex-valued functions and one consisting of extended real-valued functions (which must be finite a.e., since they are essentially bounded). Both cases are important in applications, and in any particular circumstance it is usually clear from context whether our functions are extended real-valued or complex-valued. Following Notation 3.1.1, we combine these two possibilities into a single definition by letting the symbol $\overline{\mathbf{F}}$ denote a choice of either $[-\infty,\infty]$ or $\mathbb{C}$. In conjunction with this (and as specified in Notation 3.1.1), the word *scalar* means a (finite) real number when $\overline{\mathbf{F}} = [-\infty,\infty]$, and it means a complex number when $\overline{\mathbf{F}} = \mathbb{C}$. Using these conventions, here is the precise definition of $L^\infty(E)$.

Definition 3.3.3 (The Lebesgue Space $L^\infty(E)$). If E is a measurable subset of $\mathbb{R}^d$, then the *Lebesgue space of essentially bounded functions on E* is the set of all essentially bounded measurable functions $f\colon E \to \overline{\mathbf{F}}$. That is,

$$L^\infty(E) \;=\; \Bigl\{ f\colon E \to \overline{\mathbf{F}} \;:\; f \text{ is measurable and } \|f\|_\infty < \infty \Bigr\}. \qquad \diamondsuit$$

The following exercise gives some properties of $L^\infty(E)$ and the L^∞-norm.

Exercise 3.3.4. Assume that $E \subseteq \mathbb{R}^d$ is measurable. Show that if f and g are any two functions in $L^\infty(E)$, then $af+bg \in L^\infty(E)$ for all scalars a and b.

Conclude that $L^\infty(E)$ is a vector space. Also prove that the following four statements hold for all functions $f, g \in L^\infty(E)$ and all scalars c.

(a) Nonnegativity: $0 \le \|f\|_\infty < \infty$.

(b) Homogeneity: $\|cf\|_\infty = |c|\,\|f\|_\infty$.

(c) The Triangle Inequality: $\|f+g\|_\infty \le \|f\|_\infty + \|g\|_\infty$.

(d) Almost Everywhere Uniqueness: $\|f\|_\infty = 0$ if and only if $f = 0$ a.e. $\diamondsuit$

Exercise 3.3.4 tells us that the "L^∞-norm" $\|\cdot\|_\infty$ is *almost* a norm on $L^\infty(E)$. Specifically, parts (a)–(c) of Exercise 3.3.4 say that $\|\cdot\|_\infty$ is a *seminorm* in the sense of Definition 1.2.3. In order to be called a *norm*, it would have to be the case that $\|f\|_\infty = 0$ if and only if f is the zero function (the function that is *identically zero*). However, part (d) of Exercise 3.3.4 implies that there exist nonzero functions that satisfy $\|f\|_\infty = 0$; in fact, this is true for any function f that is zero almost everywhere. For example, taking $E = \mathbb{R}$ we have $\|\chi_{\mathbb{Q}}\|_\infty = 0$ even though $\chi_{\mathbb{Q}}$ is not identically zero. Still, although the uniqueness requirement of a norm is not strictly satisfied, the "L^∞-norm" does satisfy "almost everywhere uniqueness" in the sense that $\|f\|_\infty = 0$ if and only if $f = 0$ a.e.

3.3.1 Convergence and Completeness in $L^\infty(E)$

A norm (or a seminorm) provides us with a way to measure the distance between vectors. Measured with respect to the L^∞-norm, the distance between two functions f and g is $\|f-g\|_\infty$, which is the essential supremum of $|f(x)-g(x)|$. As spelled out in Definition 1.1.2, once we have a distance function we can define a corresponding notion of convergence. For convenience we state this formally for the L^∞-norm. We will see many other norms and other types of convergence criteria later in the volume (and Chapter 1 contains a review of convergence in generic metric and normed spaces).

Definition 3.3.5 (Convergence in L^∞-Norm). Let E be a measurable subset of $\mathbb{R}^d$. A sequence of essentially bounded functions $\{f_n\}_{n\in\mathbb{N}}$ on E (either extended real-valued or complex-valued) is said to *converge to a function f in L^∞-norm* if

$$\lim_{n\to\infty} \|f-f_n\|_\infty \;=\; \lim_{n\to\infty}\Bigl(\operatorname*{ess\,sup}_{x\in E} |f(x)-f_n(x)|\Bigr) \;=\; 0.$$

In this case we write $f_n \to f$ *in L^∞-norm.* $\diamondsuit$

Remark 3.3.6. Because $\|\cdot\|_\infty$ is only a seminorm, the L^∞-norm limit of a sequence is unique only up to sets of measure zero. That is, if $f_n \to f$ and $f_n \to g$ in L^∞-norm, then f and g need not be identical, but they will satisfy $f = g$ a.e. $\diamondsuit$

A sequence $\{f_n\}_{n\in\mathbb{N}}$ is *Cauchy in L^∞-norm* if for each $\epsilon > 0$ there exists some $N > 0$ such that $\|f_m - f_n\|_\infty < \varepsilon$ for all $m, n \geq N$ (compare Definition 1.1.2). A space in which every Cauchy sequence converges to an element of the space is said to be *complete*. We prove next that $L^\infty(E)$ is complete. Our proof is very similar to the proof of Theorem 1.3.3, except that we need to keep track of certain sets of measure zero.

Lemma 3.3.7. *If $E \subseteq \mathbb{R}^d$ is measurable and $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^\infty(E)$, then there exists some function $f \in L^\infty(E)$ such that $f_n \to f$ in L^∞-norm as $n \to \infty$.*

Proof. Given positive integers m and n, let

$$Z_{mn} \;=\; \bigl\{|f_m - f_n| > \|f_m - f_n\|_\infty\bigr\}.$$

Lemma 3.3.2 tells us that Z_{mn} has measure zero, so $Z = \cup_{m,n} Z_{mn}$ has measure zero as well.

Given $\varepsilon > 0$, there is some N such that $\|f_m - f_n\|_\infty < \varepsilon$ for all $m, n \geq N$. Therefore, if $x \notin Z$ then $|f_m(x) - f_n(x)| \leq \|f_m - f_n\|_\infty < \varepsilon$ for all $m, n \geq N$. Hence $\{f_n(x)\}_{n\in\mathbb{N}}$ is a Cauchy sequence of scalars when $x \notin Z$, so it must converge, say to $f(x)$. This gives us a function f that is defined at almost every point of E.

If $n \geq N$, then for every $x \notin Z$ we have

$$|f(x) - f_n(x)| \;=\; \lim_{m\to\infty} |f_m(x) - f_n(x)| \;\leq\; \|f_m - f_n\|_\infty \;<\; \varepsilon.$$

Since Z has measure zero, this implies that $f_n \to f$ a.e., so f is measurable. Further, the computation above shows that $\|f - f_n\|_\infty \leq \varepsilon$ whenever $n \geq N$, so $f_n \to f$ in L^∞-norm. □

A normed space that is complete is called a *Banach space* (see Definition 1.2.5). Technically, the fact that the L^∞-norm is only a seminorm means that $L^\infty(E)$ is not a Banach space with respect to $\|\cdot\|_\infty$. However, we will see in Section 7.2.2 that *if we identify functions that are equal a.e.* then $\|\cdot\|_\infty$ becomes a norm and, with this identification, $L^\infty(E)$ is a Banach space.

Problems

3.3.8. Let $E \subseteq \mathbb{R}^d$ be measurable. Given functions $f_n, f \in L^\infty(E)$, prove that $f_n \to f$ in L^∞-norm if and only if there exists a set $Z \subseteq E$ with $|Z| = 0$ such that $f_n \to f$ uniformly on $E \backslash Z$.

3.3.9. For each $a \in \mathbb{R}$, let $f_a = \chi_{[a,a+1]}$. Prove that $\{f_a\}_{a\in\mathbb{R}}$ is an uncountable set of functions in $L^\infty(\mathbb{R})$ such that $\|f_a - f_b\|_\infty = 1$ for all real numbers $a \neq b$.

3.3.10. Let E be a measurable subset of $\mathbb{R}^d$ such that $|E| > 0$. Prove that there exist countably many disjoint measurable subsets $E_1, E_2, \dots$ of E such that $|E_k| > 0$ for every k. Use this to show that there exist *uncountably many* functions $f_i \in L^\infty(E)$ such that $\|f_i - f_j\|_\infty = 1$ for all $i \neq j$.

3.4 Egorov's Theorem

Suppose that we have a sequence of functions $\{f_n\}_{n\in\mathbb{N}}$ defined on a domain E. There are many different ways in which the functions f_n might "converge" to a limit function f. For example, f_n *converges pointwise to* f if

$$f(x) = \lim_{n\to\infty} f_n(x) \quad \text{for every } x \in E,$$

and we declared in Notation 3.2.8 that f_n *converges pointwise a.e. to* f (denoted $f_n \to f$ a.e.) if

$$f(x) = \lim_{n\to\infty} f_n(x) \quad \text{for a.e. } x \in E.$$

Sometimes we need to know that f_n converges to f in other senses. For example, f_n *converges uniformly to* f if the uniform norm of the difference between f and f_n converges to zero, i.e., if

$$\lim_{n\to\infty} \|f - f_n\|_{\mathrm{u}} = \lim_{n\to\infty} \left(\sup_{x\in E} |f(x) - f_n(x)| \right) = 0.$$

Convergence in L^∞-norm, which was introduced in Definition 3.3.5, is essentially an "almost everywhere" version of uniform convergence. Specifically, f_n *converges to* f *in* L^∞*-norm* if

$$\lim_{n\to\infty} \|f - f_n\|_\infty = \lim_{n\to\infty} \left(\operatorname*{ess\,sup}_{x\in E} |f(x) - f_n(x)| \right) = 0.$$

For the moment we will focus on pointwise and uniform convergence. Uniform convergence implies pointwise convergence, but the next example shows that pointwise convergence does not imply uniform convergence in general.

Example 3.4.1 (Shrinking Triangles). Set $E = [0,1]$. For each $n \in \mathbb{N}$, let f_n be the continuous function on $[0,1]$ defined by

$$f_n(x) = \begin{cases} 0, & \text{if } x = 0, \\ \text{linear}, & \text{if } 0 < x < \frac{1}{2n}, \\ 1, & \text{if } x = \frac{1}{2n}, \\ \text{linear}, & \text{if } \frac{1}{2n} < x < \frac{1}{n}, \\ 0, & \text{if } \frac{1}{n} \le x \le 1. \end{cases}$$

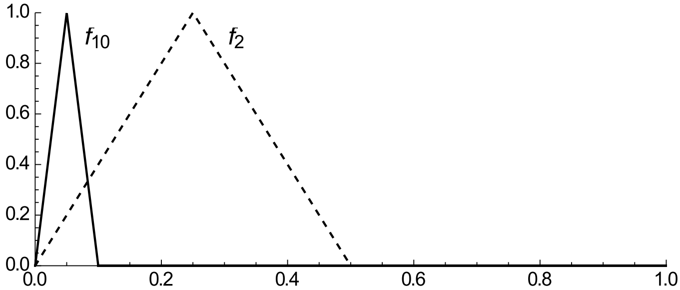


Fig. 3.2 Graphs of the functions f_2 (dashed) and f_{10} (solid) from Example 3.4.1.

For each fixed point $x \in [0,1]$ we have $f_n(x) \to 0$ as $n \to \infty$ (see the illustration in Figure 3.2). Hence f_n converges to zero pointwise. However, f_n does not converge uniformly to the zero function because for every positive integer n we have

$$\|f - f_n\|_{\mathrm{u}} \;=\; \sup_{x\in[0,1]} |0 - f_n(x)| \;=\; 1. \qquad \diamondsuit$$

Even though the Shrinking Triangles of Example 3.4.1 do not converge uniformly (or in L^∞-norm) on the domain $[0,1]$, we can find a *subset* of $[0,1]$ on which we have uniform convergence. For example, if $0 < \delta < 1$, then for all n large enough the restriction of f_n to the interval $[\delta,1]$ is zero. Hence f_n converges uniformly to the zero function *on the interval* $[\delta,1]$. We obtain uniform convergence on $[\delta,1]$, no matter how small we take δ. *Egorov's Theorem*, which we prove next, shows that this example is typical: If a sequence of measurable functions converges pointwise a.e. on a set that has finite measure, then there is a "large" subset of the domain on which the functions converge uniformly. In the proof, we use the notion of the limsup of a sequence of sets that was introduced in Definition 2.1.14.

Theorem 3.4.2 (Egorov's Theorem). *Let E be a measurable subset of $\mathbb{R}^d$ with $|E| < \infty$. Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of measurable functions on E (either complex-valued or extended real-valued) such that $f_n \to f$ a.e., where f is finite a.e. Then for each $\varepsilon > 0$ there exists a measurable set $A \subseteq E$ such that:*

(a) $|A| < \varepsilon$, *and*

(b) f_n *converges uniformly to f on $E\backslash A$, i.e.,*

$$\lim_{n\to\infty} \|(f - f_n)\cdot\chi_{A^{\mathrm{C}}}\|_{\mathrm{u}} \;=\; \lim_{n\to\infty}\left(\sup_{x\notin A}|f(x) - f_n(x)|\right) \;=\; 0.$$

Proof. Case 1: Complex-Valued Functions. Assume that the f_n are complex-valued. Since the pointwise a.e. limit of measurable functions is measurable, we know that f is measurable.

Let Z be the set of points where $f_n(x)$ does not converge to $f(x)$. By hypothesis, Z has measure zero. In order to quantify more precisely the points where $f_n(x)$ is far from $f(x)$, for each $k \in \mathbb{N}$ we let

$$Z_k = \Bigl\{x \in E : |f(x) - f_n(x)| \geq \tfrac{1}{k} \text{ for } \textit{infinitely many } n\Bigr\}.$$

Since $Z_k \subseteq Z$, we have $|Z_k| = 0$. By Exercise 2.1.15,

$$Z_k = \limsup_{n\to\infty} \Bigl\{|f - f_n| \geq \tfrac{1}{k}\Bigr\} = \bigcap_{n=1}^{\infty} A_n(k),$$

where for $k, n \in \mathbb{N}$ we take

$$A_n(k) = \bigcup_{m=n}^{\infty} \Bigl\{|f - f_m| \geq \tfrac{1}{k}\Bigr\}.$$

Each set $A_n(k)$ is measurable. By construction,

$$A_1(k) \supseteq A_2(k) \supseteq \cdots \qquad \text{and} \qquad \bigcap_{n=1}^{\infty} A_n(k) = Z_k.$$

Since $|E|$ has finite measure we can therefore apply continuity from above to obtain

$$\lim_{n\to\infty} |A_n(k)| = |Z_k| = 0. \tag{3.5}$$

Fix any $\varepsilon > 0$. By equation (3.5), for each integer $k \in \mathbb{N}$ we can find an integer $n_k \in \mathbb{N}$ such that

$$\bigl|A_{n_k}(k)\bigr| < \frac{\varepsilon}{2^k}.$$

By subadditivity, the set

$$A = \bigcup_{k=1}^{\infty} A_{n_k}(k)$$

has measure $|A| < \varepsilon$. Moreover, if $x \notin A$ then $x \notin A_{n_k}(k)$ for any k, so $|f(x) - f_m(x)| < \frac{1}{k}$ for all $m \geq n_k$.

In summary, we have found a set A with measure $|A| < \varepsilon$ such that for each integer k there exists an integer n_k such that

$$m \geq n_k \quad \Longrightarrow \quad \sup_{x \notin A} |f(x) - f_m(x)| \leq \frac{1}{k}.$$

This says that f_n converges uniformly to f on $E \backslash A$.

Case 2: Extended Real-Valued Functions. Now assume that f_n and f are extended real-valued functions with f finite a.e. Let $Y = \{f = \pm\infty\}$ be the set of measure zero consisting of all points where $f(x) = \pm\infty$. Then $F = E \backslash Y$ is measurable, f is finite on F, and $f_n \to f$ a.e. on F. Now repeat the proof of Case 1 with E replaced by F. Although $f_n(x)$ can be $\pm\infty$, if $x \in F$ then

$f(x) - f_n(x)$ never takes an indeterminate form, and the proof proceeds just as before to construct a measurable set $A \subseteq F$ such that $|A| < \varepsilon$ and $f_n \to f$ uniformly on $F \backslash A$. Consequently, $B = A \cup Y$ is a measurable subset of E that satisfies $|B| = |A| < \varepsilon$, and $f_n \to f$ uniformly on $E \backslash B$. □

The hypothesis in Egorov's Theorem that E has finite measure is necessary, as is the hypothesis that f is finite a.e. (see Problem 3.4.5).

The type of convergence that appears in the conclusion of Egorov's Theorem is sometimes called "almost uniform convergence." Here is the precise definition.

Definition 3.4.3 (Almost Uniform Convergence). Let E be a measurable subset of $\mathbb{R}^d$. We say that functions $f_n \colon E \to \overline{\mathbf{F}}$ converge *almost uniformly* to f on the set E, and write $f_n \to f$ *almost uniformly*, if for each $\varepsilon > 0$ there exists a measurable set $A \subseteq E$ such that:

(a) $|A| < \varepsilon$, and

(b) f_n converges uniformly to f on $E \backslash A$. ◇

The following exercise gives relations between L^∞-norm convergence, almost uniform convergence, and pointwise a.e. convergence.

Exercise 3.4.4. Let E be a measurable subset of $\mathbb{R}^d$, and let $f_n, f \colon E \to \overline{\mathbf{F}}$ be measurable functions on E. Prove the following statements.

(a) If $f_n \to f$ in L^∞-norm, then $f_n \to f$ almost uniformly.

(b) If $f_n \to f$ almost uniformly, then $f_n \to f$ pointwise a.e. ◇

The converse of the implications in Exercise 3.4.4 fail in general; see Problem 3.4.6. On the other hand, Egorov's Theorem tells us that if $|E| < \infty$, then pointwise a.e. convergence implies almost uniform convergence. These and other implications among various types of convergence criteria are summarized later in Figure 3.3 (also see Figures 4.3 and 7.5).

Problems

3.4.5. (a) Show by example that the assumption in Egorov's Theorem that $|E| < \infty$ is necessary.

(b) Show by example that, even if we assume $|E| < \infty$, the assumption in Egorov's Theorem that f is finite a.e. is necessary.

3.4.6. (a) Exhibit a sequence of functions that converges almost uniformly but does not converge in L^∞-norm.

(b) Exhibit a sequence of functions that converges pointwise a.e. but does not converge almost uniformly.

3.4.7. Let $E \subseteq \mathbb{R}^d$ be a measurable set such that $|E| < \infty$, and assume that f_n and f are measurable functions that are finite a.e. and satisfy $f_n \to f$ a.e. on E. Prove that there exist measurable sets $E_k \subseteq E$ such that $E \setminus \bigl(\cup_{k=1}^\infty E_k\bigr)$ has measure zero and for each individual k we have that $f_n \to f$ uniformly on E_k. Even so, show by example that f_n need not converge uniformly to f on E.

3.5 Convergence in Measure

In the preceding section we saw several ways to quantify the meaning of the convergence of a sequence of functions. We introduce another important type of convergence criterion in this section.

Definition 3.5.1 (Convergence in Measure). Let $E \subseteq \mathbb{R}^d$ be measurable, and assume that functions $f_n, f \colon E \to \overline{\mathbf{F}}$ are measurable and finite a.e. We say that f_n *converges in measure to* f *on* E, and write $f_n \xrightarrow{\mathrm{m}} f$, if for every $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \bigl|\{|f - f_n| > \varepsilon\}\bigr| \;=\; 0. \qquad \diamondsuit \tag{3.6}$$

Writing out equation (3.6) explicitly, we see that $f_n \xrightarrow{\mathrm{m}} f$ if and only if for every $\varepsilon > 0$ and every $\eta > 0$, there is an $N > 0$ such that

$$n \ge N \quad \Longrightarrow \quad \bigl|\{|f - f_n| > \varepsilon\}\bigr| \;<\; \eta.$$

Problem 3.5.17 gives some other equivalent formulations of convergence in measure.

To summarize Definition 3.5.1, convergence in measure requires that if we fix any $\varepsilon > 0$, then the measure of the set where f and f_n differ by more than ε must decrease to zero as $n \to \infty$. Here is an example.

Example 3.5.2 (Shrinking Boxes I). The domain for this example is $E = [0,1]$. Let $f = 0$, and set $f_n = \chi_{[0,\frac{1}{n}]}$. If we fix $0 < \varepsilon < 1$, then the set of points where f_n differs from 0 by more than ε is precisely the interval $[0, \frac{1}{n}]$, which has measure $\frac{1}{n}$. Therefore $f_n \xrightarrow{\mathrm{m}} 0$. $\diamondsuit$

The following example shows that pointwise a.e. convergence does not imply convergence in measure in general.

Example 3.5.3 (Boxes Marching to Infinity). For this example we take $E = \mathbb{R}$. The functions $f_n = \chi_{[n,n+1]}$ converge pointwise to the zero function. However, if we fix $0 < \varepsilon < 1$ then

$$\bigl\{|0 - f_n| > \varepsilon\bigr\} \;=\; [n, n+1],$$

which has measure 1. Therefore f_n does not converge in measure to the zero function. In fact, there is no function f such that $f_n \xrightarrow{\mathrm{m}} f$ (why?). $\diamondsuit$

Remark 3.5.4. We will see in Corollary 3.5.8 that *if E has finite measure* then pointwise a.e. convergence on E does imply convergence in measure. $\diamondsuit$

The following example shows that convergence in measure does not imply pointwise a.e. convergence in general (*even if* the domain has finite measure).

Example 3.5.5 (Boxes Marching in Circles). Set $E = [0,1]$, and define

$$\begin{aligned}
&f_1 = \chi_{[0,1]}, \\
&f_2 = \chi_{[0,\frac{1}{2}]}, \quad f_3 = \chi_{[\frac{1}{2},1]}, \\
&f_4 = \chi_{[0,\frac{1}{3}]}, \quad f_5 = \chi_{[\frac{1}{3},\frac{2}{3}]}, \quad f_6 = \chi_{[\frac{2}{3},1]}, \\
&f_7 = \chi_{[0,\frac{1}{4}]}, \quad f_8 = \chi_{[\frac{1}{4},\frac{1}{2}]}, \quad f_9 = \chi_{[\frac{1}{2},\frac{3}{4}]}, \quad f_{10} = \chi_{[\frac{3}{4},1]},
\end{aligned}$$

and so forth. Picturing the graphs of these functions as boxes, the boxes march from left to right across the interval $[0,1]$, then shrink in size and march across the interval again, and do this over and over.

Fix $0 < \varepsilon < 1$. For the indices $n = 1, \dots, 10$, the Lebesgue measure of $\{|f_n| > \varepsilon\}$ has the values

$$1, \ \frac{1}{2}, \ \frac{1}{2}, \ \frac{1}{3}, \ \frac{1}{3}, \ \frac{1}{3}, \ \frac{1}{4}, \ \frac{1}{4}, \ \frac{1}{4}, \ \frac{1}{4}.$$

We see that

$$\lim_{n\to\infty} \bigl|\{|0 - f_n| > \varepsilon\}\bigr| \;=\; 0,$$

so $f_n \xrightarrow{\mathrm{m}} 0$, i.e., f_n converges in measure to the zero function.

We do not have pointwise a.e. convergence in this example, because no matter what point $x \in [0,1]$ that we choose, there are infinitely many different values of n such that $f_n(x) = 0$, *and* infinitely many n such that $f_n(x) = 1$. Hence $f_n(x)$ does not converge at any point x in $[0,1]$. This sequence of functions $\{f_n\}_{n\in\mathbb{N}}$ *does not* converge pointwise a.e. to any function f. $\diamondsuit$

Even though the Marching Boxes in Example 3.5.5 do not converge pointwise a.e., there is a *subsequence* of these boxes that converges pointwise a.e. For example, the subsequence $f_1, f_2, f_4, f_7, \dots$ converges pointwise a.e. to the zero function. The next lemma shows that every sequence of functions that converges in measure contains a *subsequence* that converges pointwise almost everywhere.

Lemma 3.5.6. *Let E be a measurable subset of $\mathbb{R}^d$, and assume that functions $f_n, f\colon E \to \overline{\mathbf{F}}$ are measurable and finite a.e. If $f_n \xrightarrow{\mathrm{m}} f$, then there exists a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ such that $f_{n_k} \to f$ a.e.*

Proof. Since $f_n \xrightarrow{\mathrm{m}} f$, we can find indices $n_1 < n_2 < \cdots$ such that for each $n \ge n_k$ we have

$$\left|\{|f - f_n| > \tfrac{1}{k}\}\right| \le 2^{-k}.$$

Define $E_k = \left\{|f - f_{n_k}| > \frac{1}{k}\right\}$, and set

$$Z = \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} E_k = \limsup_{k\to\infty} E_k.$$

Since $\sum |E_k| < \infty$, the Borel–Cantelli Lemma (Exercise 2.1.16) implies that $|Z| = 0$. Also, since

$$Z^{\mathrm{C}} = \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} E_k^{\mathrm{C}} = \liminf_{k\to\infty} E_k^{\mathrm{C}},$$

Exercise 2.1.15 implies that if $x \notin Z$ then there exists some m such that $x \notin E_k$ for all $k \ge m$. Thus $|f(x) - f_{n_k}(x)| \le \frac{1}{k}$ for all $k \ge m$, so we conclude that $f_{n_k}(x) \to f(x)$ for all $x \notin Z$. $\square$

Although pointwise a.e. convergence does not imply convergence in measure, the following exercise shows that almost uniform convergence does imply convergence in measure.

Exercise 3.5.7. Assume $E \subseteq \mathbb{R}^d$ is measurable, and functions $f_n, f \colon E \to \overline{\mathbf{F}}$ are measurable and finite a.e. Prove that if f_n converges to f almost uniformly, then $f_n \xrightarrow{\mathrm{m}} f$. $\diamondsuit$

However, convergence in measure does not imply almost uniform convergence in general (consider the "Boxes Marching in Circles" in Example 3.5.5).

Combining Exercise 3.5.7 with Egorov's Theorem, we obtain the following result.

Corollary 3.5.8. *Let E be a measurable subset of $\mathbb{R}^d$, and assume that functions $f_n, f \colon E \to \overline{\mathbf{F}}$ are measurable and finite a.e. If $|E| < \infty$ and $f_n \to f$ a.e., then $f_n \xrightarrow{\mathrm{m}} f$.*

Proof. Since E has finite measure, Egorov's Theorem tells us that pointwise almost everywhere convergence on E implies almost uniform convergence on E. By Exercise 3.5.7, this implies convergence in measure. $\square$

We summarize in Figure 3.3 some of the relationships between the types of convergence criteria that we have studied so far in this chapter (these implications follow from Exercise 3.4.4, Lemma 3.5.6, Exercise 3.5.7, and Corollary 3.5.8). We will introduce other convergence criteria in later chapters, and we update Figure 3.3 accordingly in Figures 4.3 and 7.5.

Most types of convergence criteria have a corresponding Cauchy criterion. Here is the Cauchy criterion for convergence in measure.

$$
\begin{tikzcd}
 & \text{pointwise a.e. convergence} \arrow[d, Rightarrow, "\text{(if } |E| < \infty)"] & & \\
L^\infty\text{-norm convergence} \arrow[r, Rightarrow] & \text{almost uniform convergence} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \text{convergence in measure} \arrow[r, Rightarrow] & \text{pointwise a.e. convergence of a subsequence} \\
 & \text{pointwise a.e. convergence} & &
\end{tikzcd}
$$

Fig. 3.3 Relations among certain convergence criteria (valid for sequences of functions that are either complex-valued or extended real-valued but finite a.e.).

Definition 3.5.9 (Cauchy in Measure). Let E be a measurable subset of $\mathbb{R}^d$, and assume functions $f_n\colon E \to \overline{\mathbf{F}}$ are measurable and finite a.e. We say that the sequence $\{f_n\}_{n\in\mathbb{N}}$ is *Cauchy in measure on E* if for every $\varepsilon > 0$, there exists an $N > 0$ such that

$$m, n \ge N \implies \bigl|\{|f_m - f_n| > \varepsilon\}\bigr| < \varepsilon. \qquad \diamondsuit$$

The following theorem shows that every sequence that is Cauchy in measure must converge in measure to some measurable function (see Problem 3.5.17 for some further equivalent reformulations of convergence in measure).

Theorem 3.5.10. *Let $E \subseteq \mathbb{R}^d$ be a measurable set. If $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of measurable functions that is Cauchy in measure on E, then there exists a measurable function f such that $f_n \xrightarrow{\mathrm{m}} f$.*

Proof. If $\{f_n\}_{n\in\mathbb{N}}$ is Cauchy in measure then, just as in Problem 1.1.21, we can find indices $n_1 < n_2 < \cdots$ such that

$$\bigl|\{|f_{n_{k+1}} - f_{n_k}| > 2^{-k}\}\bigr| \le 2^{-k} \quad \text{for all } k \in \mathbb{N}.$$

For simplicity of notation, let

$$g_k = f_{n_k}, \qquad E_k = \bigl\{|g_{k+1} - g_k| > 2^{-k}\bigr\}, \qquad H_m = \bigcup_{k=m}^{\infty} E_k.$$

Since $\sum |E_k| < \infty$, the Borel–Cantelli Lemma implies that

$$Z = \bigcap_{m=1}^{\infty} H_m = \bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} E_k = \limsup_{k\to\infty} E_k$$

has measure zero. Since $Z^{\mathrm{C}} = \liminf E_k^{\mathrm{C}}$ is the set of points that belong to all but finitely many E_k^{C}, if $x \notin Z$ then there exists some $N > 0$ such that $x \notin E_k$ for all $k \ge N$. That is, $|g_{k+1}(x) - g_k(x)| \le 2^{-k}$ for all $k \ge N$, so

$\{g_k(x)\}_{k\in\mathbb{N}}$ is a Cauchy sequence of scalars, and must therefore converge. Setting

$$f(x) \;=\; \begin{cases} \lim\limits_{k\to\infty} g_k(x), & \text{if } x \notin Z, \\ 0, & \text{if } x \in Z, \end{cases}$$

we see that f is measurable and $g_k \to f$ pointwise a.e.

Now we will show that g_k converges in measure to f. Fix $\varepsilon > 0$, and choose m large enough that $2^{-m} \le \varepsilon$. If $x \notin H_m$, then for all $n > k > m$ we have

$$|g_n(x) - g_k(x)| \;\le\; \sum_{j=k}^{n-1} |g_{j+1}(x) - g_j(x)| \;\le\; \sum_{j=k}^{n-1} 2^{-j} \;\le\; 2^{-k+1} \;\le\; 2^{-m} \;\le\; \varepsilon.$$

Taking the limit as $n \to \infty$, this implies that $|f(x) - g_k(x)| \le \varepsilon$ for all $x \notin H_m$ and $k > m$. Hence $\bigl\{|f - g_k| > \varepsilon\bigr\} \subseteq H_m$ for $k > m$, and therefore

$$\limsup_{k\to\infty} \bigl|\{|f - g_k| > \varepsilon\}\bigr| \;\le\; |H_m| \;\le\; 2^{-m+1}.$$

This is true for every m, so we conclude that $\lim_{k\to\infty} |\{|f - g_k| > \varepsilon\}| = 0$, and therefore $g_k \xrightarrow{\text{m}} f$.

So, we have shown that $\{f_n\}_{n\in\mathbb{N}}$ has a subsequence $\{g_k\}_{k\in\mathbb{N}}$ that converges in measure. This, combined with the fact that $\{f_n\}_{n\in\mathbb{N}}$ is Cauchy in measure, implies that $f_n \xrightarrow{\text{m}} f$ (see Problem 3.5.16). $\square$

Problems

3.5.11. Let $f_n(x) = x/n$ for $x \in \mathbb{R}$. Prove that f_n converges pointwise to the zero function, but f_n does not converge in measure to 0 (or any other function).

3.5.12. For each $n \in \mathbb{N}$, define

$$f_n(x) \;=\; \frac{1 - |x|^n}{1 + |x|^n}, \qquad \text{for } x \in \mathbb{R}.$$

Show that there exists a measurable function f such that $f_n \to f$ pointwise and $f_n \xrightarrow{\text{m}} f$, but f_n does not converge to f uniformly.

3.5.13. Let $E \subseteq \mathbb{R}^d$ be a measurable set, and assume $f_n, f, g_n, g \colon E \to \overline{\mathbf{F}}$ are measurable and finite a.e. Prove the following statements.

(a) If $f_n \xrightarrow{\text{m}} f$ and $f_n \xrightarrow{\text{m}} g$, then $f = g$ a.e.

(b) If $f_n \xrightarrow{\text{m}} f$ and $g_n \xrightarrow{\text{m}} g$, then $f_n + g_n \xrightarrow{\text{m}} f + g$.

(c) If $|E| < \infty$, $f_n \xrightarrow{\text{m}} f$, and $g_n \xrightarrow{\text{m}} g$, then $f_n g_n \xrightarrow{\text{m}} fg$.

(d) The conclusion of part (c) can fail if $|E| = \infty$.

(e) If $f_n \xrightarrow{\text{m}} f$ and there is some $\delta > 0$ such that $|f_n| \geq \delta$ a.e. for every n, then $\frac{1}{f_n} \xrightarrow{\text{m}} \frac{1}{f}$.

3.5.14. Let E be a measurable subset of $\mathbb{R}^d$, and assume that $f_n, f \colon E \to \overline{\mathbf{F}}$ are measurable and finite a.e. Prove that the following two statements are equivalent.

(a) $f_n \xrightarrow{\text{m}} f$.

(b) If $\{g_n\}_{n\in\mathbb{N}}$ is any subsequence of $\{f_n\}_{n\in\mathbb{N}}$, then there exists a subsequence $\{h_n\}_{n\in\mathbb{N}}$ of $\{g_n\}_{n\in\mathbb{N}}$ such that $h_n \xrightarrow{\text{m}} f$.

3.5.15. Let $E \subseteq \mathbb{R}^d$ be measurable, and let $f_n, f \colon E \to \overline{\mathbf{F}}$ be measurable functions that are finite a.e. Assume that $\varphi \colon \mathbb{R} \to \mathbb{R}$ (if $\overline{\mathbf{F}} = [-\infty, \infty]$) or $\varphi \colon \mathbb{C} \to \mathbb{C}$ (if $\overline{\mathbf{F}} = \mathbb{C}$) is continuous.

(a) Suppose that $f_n \xrightarrow{\text{m}} f$ and φ is *uniformly* continuous, and prove that $\varphi \circ f_n \xrightarrow{\text{m}} \varphi \circ f$. Show by example that this can fail if φ is continuous but not uniformly continuous.

(b) Prove that if $f_n \xrightarrow{\text{m}} f$ and $|E| < \infty$, then $\varphi \circ f_n \xrightarrow{\text{m}} \varphi \circ f$. Show that this can fail when $|E| = \infty$.

3.5.16. Let $E \subseteq \mathbb{R}^d$ be measurable, and let f_n and f be measurable functions on E, either complex-valued or extended real-valued but finite a.e. Prove that if $\{f_n\}_{n\in\mathbb{N}}$ is Cauchy in measure and there exists a subsequence such that $f_{n_k} \xrightarrow{\text{m}} f$, then $f_n \xrightarrow{\text{m}} f$.

3.5.17. Let $E \subseteq \mathbb{R}^d$ be a measurable set. Let f_n and f be measurable functions on E, either complex-valued or extended real-valued but finite a.e. Prove that the following four statements are equivalent.

(a) There exists a measurable function f such that $f_n \xrightarrow{\text{m}} f$. That is, for each $\varepsilon, \eta > 0$ there exists an $N > 0$ such that

$$n \geq N \implies \bigl|\{|f - f_n| > \varepsilon\}\bigr| < \eta.$$

(b) There exists a measurable function f such that for every $\varepsilon > 0$ there exists an $N > 0$ such that

$$n \geq N \implies \bigl|\{|f - f_n| > \varepsilon\}\bigr| < \varepsilon.$$

(c) For each $\varepsilon, \eta > 0$ there exists an $N > 0$ such that

$$m, n \geq N \implies \bigl|\{|f_m - f_n| > \varepsilon\}\bigr| < \eta.$$

(d) $\{f_n\}_{n\in\mathbb{N}}$ is Cauchy in measure, i.e., for each $\varepsilon > 0$ there exists an $N > 0$ such that

$$m, n \geq N \implies \bigl|\{|f_m - f_n| > \varepsilon\}\bigr| < \varepsilon.$$

3.6 Luzin's Theorem

In this section we will use Egorov's Theorem and facts about approximation by simple functions to prove *Luzin's Theorem*, which, in essence, states that every measurable function is "nearly continuous." Precisely, if f is a measurable function, then there is a closed subset F such that f is continuous on F and the complement of F has measure ε.

Given a function $f\colon E \to \mathbb{C}$ and a set $F \subseteq E$, recall that the restriction of f to F is the function $f|_F\colon F \to \mathbb{C}$ defined by $f|_F(x) = f(x)$ for $x \in F$. We say that f *is continuous on* F if $f|_F$ is a continuous function. There are various equivalent ways to define continuity, but for the purposes of this result it will be most convenient to use the formulation, given in Exercise 1.1.15, that *a function g is continuous on F if and only if*

$$\forall\, x_n, x \in F, \qquad x_n \to x \implies g(x_n) \to g(x).$$

Using this notation, we can state Luzin's Theorem as follows.

Theorem 3.6.1 (Luzin's Theorem). *Let E be a bounded, measurable subset of $\mathbb{R}^d$, and let $f\colon E \to \overline{\mathbf{F}}$ be measurable and finite a.e. Then for each $\varepsilon > 0$, there exists a closed set $F \subseteq E$ such that $|E \backslash F| < \varepsilon$ and $f|_F$ is continuous.*

Proof. Step 1. Let $\phi = \sum_{k=1}^N c_k \chi_{E_k}$ be the standard representation of a simple function ϕ on E, and fix $\varepsilon > 0$. Since each subset E_k is measurable, Lemma 2.2.15 implies that there exist closed sets $F_k \subseteq E_k$ such that

$$|E_k \backslash F_k| < \frac{\varepsilon}{N}, \qquad \text{for } k = 1, \dots, N.$$

The set $F = F_1 \cup \cdots \cup F_N$ is closed, and since $E_1, \dots, E_N$ partition E we have $|E \backslash F| < \varepsilon$. Since E is bounded, the sets $F_1, \dots, F_N$ are compact and disjoint. Consequently, F_j is separated from F_k by a positive distance when $j \neq k$ (see Problem 2.2.31). Since ϕ is constant on each individual set F_k, it follows that the restriction of ϕ to F is continuous.

Step 2. Now let f be an arbitrary measurable function on E, and fix $\varepsilon > 0$. By Corollary 3.2.15, there exist simple functions ϕ_n that converge pointwise to f on E. Applying Step 1, for each integer $n > 0$ we can find a closed set $F_n \subseteq E$ such that

$$|E \backslash F_n| < \frac{\varepsilon}{2^{n+1}} \qquad \text{and} \qquad \phi_n|_{F_n} \text{ is continuous.}$$

By Egorov's Theorem, there exists a measurable set $A \subseteq E$ with measure $|A| < \varepsilon/4$ such that ϕ_n converges to f uniformly on $E \backslash A$. By Lemma 2.2.15, there exists a closed set $F_0 \subseteq E \backslash A$ such that

$$|(E \backslash A) \backslash F_0| < \frac{\varepsilon}{4}.$$

Writing $E\backslash F_0 = (E\backslash A)\backslash F_0 \cup A$, we see that

$$|E\backslash F_0| \leq |(E\backslash A)\backslash F_0| + |A| < \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2}.$$

Further, ϕ_n converges to f uniformly on F_0 since F_0 is contained in $E\backslash A$.

Next, let

$$F = \bigcap_{n=0}^{\infty} F_n.$$

Since F is closed and bounded, it is compact. Further,

$$|E\backslash F| = \left|\bigcup_{n=0}^{\infty} (E\backslash F_n)\right| \leq \sum_{n=0}^{\infty} |E\backslash F_n| < \sum_{n=0}^{\infty} \frac{\varepsilon}{2^{n+1}} = \varepsilon.$$

Since ϕ_n is continuous on F_n, it is continuous on the smaller set F. Thus $\{\phi_n|_F\}_{n\in\mathbb{N}}$ is a sequence of continuous functions that converges uniformly on F to $f|_F$. Therefore $f|_F$ is continuous, because the uniform limit of a sequence of continuous functions is continuous (see Theorem 1.3.3). □

Luzin's Theorem tells us that a measurable function f on a bounded set E is continuous on a closed subset F that is "nearly all" of E. Because F is closed and $\mathbb{R}^d$ is a metric space, the *Tietze Extension Theorem* implies that there exists a continuous function $g\colon \mathbb{R}^d \to \mathbb{C}$ such that $g = f$ on the set F (for one proof, see [Heil18, Thm. 4.8.2]). Hence $g|_E$ is a continuous function on E that equals f on the subset F. Problem 3.6.2 incorporates this conclusion into the statement of Luzin's Theorem, and additionally removes the hypothesis in Theorem 3.6.1 that the set E is bounded.

Problems

3.6.2. Let E be a measurable subset of $\mathbb{R}^d$, and assume that $f\colon E \to \overline{\mathbf{F}}$ is finite a.e. Prove that the following three statements are equivalent.

(a) f is measurable.

(b) For each $\varepsilon > 0$, there exists a closed set $F \subseteq E$ such that $|E\backslash F| < \varepsilon$ and $f|_F$ is continuous.

(c) For each $\varepsilon > 0$, there exist a closed set $F \subseteq E$ and a continuous function $g\colon E \to \mathbb{C}$ such that $|E\backslash F| < \varepsilon$ and $g(x) = f(x)$ for all $x \in F$.

Chapter 4
The Lebesgue Integral

In this chapter we define and study the Lebesgue integral of functions on $\mathbb{R}^d$ (or on subsets of $\mathbb{R}^d$). We first define the Lebesgue integral for nonnegative functions in Section 4.1, and in Section 4.2 prove two fundamental results on convergence of integrals: *Fatou's Lemma* and the *Monotone Convergence Theorem*. We define the integral of extended real-valued and complex-valued functions in Section 4.3. *Integrable functions* (those functions for which the integral of $|f|$ is finite) are introduced in Section 4.4, as is the *Lebesgue space* $L^1(E)$, which is the set of all integrable functions on E. In Section 4.5 we prove the *Dominated Convergence Theorem*, or *DCT*, which is one of the most useful theorems in analysis. In particular, we use the DCT to show that integrable functions can be well-approximated by a wide variety of functions that have special properties, including simple functions, continuous functions, and step functions. Among other applications, this allows us to characterize Riemann integrable functions and to establish the relationship between Lebesgue and Riemann integrals. Finally, Section 4.6 covers the important theorems of Fubini and Tonelli, which tell us when we can exchange the order of iterated integrals.

4.1 The Lebesgue Integral of Nonnegative Functions

We will define the Lebesgue integral of a measurable function in this chapter. There are some functions whose integral is undefined, but we will be able to define the integral of "most" measurable functions. If a function happens to be Riemann integrable, then we will see that its Lebesgue integral coincides with its Riemann integral. The Riemann integral is quite restrictive in the sense that only a "few" functions are Riemann integrable. For example, the *Dirichlet function* $\chi_{\mathbb{Q}}$, which is discontinuous at every point, is not Riemann integrable, but it is Lebesgue integrable. In fact, since $\chi_{\mathbb{Q}} = 0$ a.e., we will see that $\int_E \chi_{\mathbb{Q}} = \int_E 0 = 0$ for every measurable set $E \subseteq \mathbb{R}$.

C. Heil, *Introduction to Real Analysis*, Graduate Texts in Mathematics 280,
https://doi.org/10.1007/978-3-030-26903-6_4

In this section and the next we will focus on the definition and properties of the Lebesgue integral of nonnegative measurable functions, and in Section 4.3 we will consider how to extend the definition of the integral to measurable functions that are extended real-valued or complex-valued. An important difference between nonnegative functions and generic functions is that we will be able to assign a value (in the extended real sense) to the integral of *every* nonnegative measurable function. When we consider arbitrary functions in Section 4.3, we will see that we can encounter indeterminate forms when attempting to define the integral, and in such cases the integral is undefined.

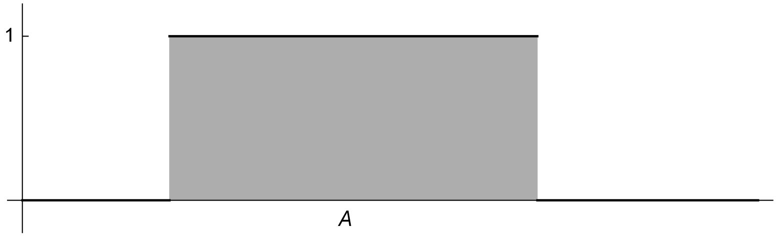


Fig. 4.1 The shaded region is $A \times [0,1]$, which is the region under the graph of χ_A.

It is not obvious how we should define the Lebesgue integral of an arbitrary nonnegative function, so we begin with a class of functions for which we know how we want the integration to come out, namely, characteristic functions. If we fix a measurable set $A \subseteq \mathbb{R}^d$, then χ_A is 0 outside of A and is identically 1 on A. The "region under the graph" of χ_A is the set $A \times [0,1]$ (see Figure 4.1). At least intuitively, the integral of a nonnegative function should be the "area of the region under its graph." Therefore it is reasonable to define the integral of χ_A to be the measure of $A \times [0,1]$. Exercise 2.3.6 showed that the Lebesgue measure of $A \times [0,1]$ (which is a measurable subset of $\mathbb{R}^{d+1}$) is the product of the measures of A and $[0,1]$, and so we define the integral of χ_A to be $\int \chi_A = |A|$.

This gets us started. In the remainder of this section we will define the integral of finite linear combinations of characteristic functions, which are precisely the *simple functions* defined in Section 3.2.4, and then see how to use simple functions to define the integral of an arbitrary nonnegative measurable function. Along the way we will need to consider convergence issues—for example, if functions f_n converge to a function f in some sense, will it be true that the integral of f_n converges to the integral of f? Unfortunately, this does not always happen. In particular, we will see examples of functions f_n that converge pointwise to some function f, yet $\int f_n$ does not converge to $\int f$ (Example 4.2.6). On the other hand, if we impose stricter hypotheses on the f_n than just pointwise convergence, then we can sometimes infer convergence of the integrals. For example, the Monotone Convergence Theorem

(Theorem 4.2.1) will show that if nonnegative functions $f_n(x)$ increase monotonically to $f(x)$ at each point x, then $\int f_n$ converges to $\int f$.

4.1.1 Integration of Nonnegative Simple Functions

Recall from Definition 3.2.11 that a *simple function* is a measurable function ϕ, defined on a set E, that takes only finitely many distinct scalar values. If these distinct values are $c_1, \dots, c_N$, then the *standard representation* of ϕ is

$$\phi = \sum_{k=1}^{N} c_k \chi_{E_k},$$

where

$$E_k = \phi^{-1}\{c_k\} = \bigl\{\phi = c_k\bigr\}, \qquad \text{for } k = 1, \dots, N.$$

The sets E_k are disjoint and measurable, and they partition the set E.

To define the integral of a nonnegative simple function we simply linearly extend the idea that the integral of a characteristic function χ_A is the measure of the set A. In considering this definition, recall our convention that $0 \cdot \infty = 0$.

Definition 4.1.1 (Integral of a Nonnegative Simple Function). Let ϕ be a nonnegative simple function on a measurable set $E \subseteq \mathbb{R}^d$, and let $\phi = \sum_{k=1}^{N} c_k \chi_{E_k}$ be its standard representation. The *Lebesgue integral of* ϕ *over* E is

$$\int_E \phi = \int_E \phi(x)\,dx = \sum_{k=1}^{N} c_k |E_k|. \qquad \diamondsuit$$

The integral of any nonnegative simple function is a uniquely defined extended real number that lies in the range $0 \le \int_E \phi \le \infty$. Some of the basic properties of the Lebesgue integral of nonnegative simple functions are given in the next lemma.

Lemma 4.1.2. *If ϕ and ψ are nonnegative simple functions defined on a measurable set $E \subseteq \mathbb{R}^d$ and $c \ge 0$, then the following statements hold.*

(a) $\displaystyle\int_E (\phi + \psi) = \int_E \phi + \int_E \psi$ *and* $\displaystyle\int_E c\phi = c\int_E \phi.$

(b) *If $E_1, \dots, E_N$ are any measurable subsets of E and $c_1, \dots, c_N$ are any nonnegative scalars, then*

$$\int_E \sum_{k=1}^{N} c_k \chi_{E_k} = \sum_{k=1}^{N} c_k |E_k|. \tag{4.1}$$

Proof. (a) The equality $\int_E c\phi = c\int_E \phi$, where c is a nonnegative real scalar, follows directly from the definition of the integral of a simple function. To address the integral of a sum, let

$$\phi \;=\; \sum_{j=1}^{M} a_j\,\chi_{E_j} \qquad \text{and} \qquad \psi \;=\; \sum_{k=1}^{N} b_k\,\chi_{F_k}$$

be the standard representations of ϕ and ψ. Then, by definition, $\{E_j\}_{j=1}^M$ and $\{F_k\}_{k=1}^N$ are each partitions of E. Therefore, for each set E_j and F_k we have

$$E_j \;=\; \bigcup_{k=1}^{N} (E_j \cap F_k) \qquad \text{and} \qquad F_k \;=\; \bigcup_{j=1}^{M} (E_j \cap F_k),$$

where these are unions of disjoint sets. Therefore, by the definition of the integral and the fact that Lebesgue measure is countably additive,

$$\int_E \phi \;=\; \sum_{j=1}^{M} a_j\,|E_j| \;=\; \sum_{j=1}^{M}\sum_{k=1}^{N} a_j\,|E_j \cap F_k| \tag{4.2}$$

and

$$\int_E \psi \;=\; \sum_{k=1}^{N} b_k\,|F_k| \;=\; \sum_{k=1}^{N}\sum_{j=1}^{M} b_k\,|E_j \cap F_k|. \tag{4.3}$$

Summing, we obtain

$$\int_E \phi \;+\; \int_E \psi \;=\; \sum_{j=1}^{M}\sum_{k=1}^{N} (a_j + b_k)\,|E_j \cap F_k|. \tag{4.4}$$

On the other hand, as we observed in equation (3.2),

$$\phi + \psi \;=\; \sum_{j=1}^{M}\sum_{k=1}^{N} (a_j + b_k)\,\chi_{E_j \cap F_k}. \tag{4.5}$$

If this were the standard representation of $\phi+\psi$, then Definition 4.1.1 would immediately tell us that

$$\int_E (\phi + \psi) \;=\; \sum_{j=1}^{M}\sum_{k=1}^{N} (a_j + b_k)\,|E_j \cap F_k|. \tag{4.6}$$

Unfortunately, equation (4.5) need not be the standard representation of $\phi + \psi$, since some values of $a_j + b_k$ may coincide. However, because the sets $E_j \cap F_k$ are disjoint, the standard representation of $\phi + \psi$ is obtained by collecting together those sets $E_j \cap F_k$ that correspond to equal values of

$a_j + b_k$. After writing out the integral of $\phi + \psi$ defined by this standard representation and applying the countable additivity of Lebesgue measure, we precisely obtain equation (4.6). Comparing equations (4.4) and (4.6), we see that $\int_E \phi + \int_E \psi$ and $\int_E (\phi + \psi)$ are equal.

(b) Set $\varphi = \sum_{k=1}^N c_k \chi_{E_k}$. If this were the standard representation of φ, then equation (4.1) would follow from the definition of the integral of a simple function. The point of this part of the theorem is that equation (4.1) holds even if $\varphi = \sum_{k=1}^N c_k \chi_{E_k}$ is not the standard representation of φ. The proof of this follows by applying part (a) and an argument by induction. □

We assign the proof of the following further properties of the integral to the reader.

Exercise 4.1.3. Let ϕ and ψ be nonnegative simple functions defined on a measurable set $E \subseteq \mathbb{R}^d$. Prove the following statements.

(a) If $\phi \le \psi$, then $\int_E \phi \le \int_E \psi$.

(b) $\int_E \phi = 0$ if and only if $\phi = 0$ a.e.

(c) If $A \subseteq E$ is measurable, then $\phi\chi_A$ is a simple function and

$$\int_A \phi \;=\; \int_E \phi\chi_A.$$

(d) If $A_1, A_2, \dots$ are disjoint measurable subsets of E and $A = \cup A_n$, then

$$\int_A \phi \;=\; \sum_{n=1}^{\infty} \int_{A_n} \phi.$$

(e) If $A_1 \subseteq A_2 \subseteq \cdots$ are nested measurable subsets of E and $A = \cup A_n$, then

$$\int_A \phi \;=\; \lim_{n\to\infty} \int_{A_n} \phi. \qquad \diamondsuit \tag{4.7}$$

Remark 4.1.4. Part (d) of Exercise 4.1.3 says that the integral satisfies "countable additivity," while part (e) is a form of "continuity from below" for the integral. ♢

4.1.2 Integration of Nonnegative Functions

So far we have only defined the integral of nonnegative simple functions. We will define the integral of an arbitrary nonnegative measurable function $f\colon E \to [0,\infty]$ in terms of approximations to f by simple functions. To motivate this, suppose that ϕ is a simple function such that $0 \le \phi \le f$. In this case, the region under the graph of ϕ is a subset of the corresponding region under the graph of f (consider Figure 3.1). Whatever the integral of f turns out to be, we should have $\int_E \phi \le \int_E f$. Each simple function ϕ gives us an approximation from below to the integral of f. We declare that $\int_E f$ is the supremum of $\int_E \phi$ over all approximations from below by simple functions.

Definition 4.1.5 (Lebesgue Integral of a Nonnegative Function). Let $E \subseteq \mathbb{R}^d$ be a measurable set. If $f\colon E \to [0,\infty]$ is a measurable function, then the *Lebesgue integral of f over E* is

$$\int_E f \;=\; \int_E f(x)\,dx \;=\; \sup\biggl\{\int_E \phi \;:\; 0 \le \phi \le f,\ \phi \text{ simple}\biggr\}. \qquad \diamondsuit$$

Notation 4.1.6. When E is an interval (a,b), we usually write the integral of f over (a,b) as $\int_a^b f$ or $\int_a^b f(x)\,dx$. Because a singleton has measure zero, the integral of f over (a,b) turns out to equal the integral of f over $(a,b]$, $[a,b)$, or $[a,b]$. $\diamondsuit$

If f is a simple function, then Definitions 4.1.1 and 4.1.5 each assign a meaning to the symbols $\int_E f$. The next lemma shows that there is no conflict between these two meanings.

Lemma 4.1.7. *If ϕ is a simple function, then the integral of ϕ given in Definition 4.1.1 coincides with the integral of ϕ given in Definition 4.1.5.*

Proof. Let $\int_E \phi$ denote the integral of ϕ given by Definition 4.1.1, and let

$$I \;=\; \sup\biggl\{\int_E \psi \;:\; 0 \le \psi \le \phi,\ \psi \text{ simple}\biggr\}. \tag{4.8}$$

If ψ is any simple function such that $0 \le \psi \le \phi$, then $0 \le \int_E \psi \le \int_E \phi$ by Exercise 4.1.3. Taking the supremum over all such ψ, we see that $I \le \int_E \phi$. On the other hand, ϕ is a simple function and $\phi \le \phi$, so ϕ is one of the functions ψ that we are taking the supremum over on the right side of equation (4.8). Therefore we also have $\int_E \phi \le I$. $\square$

Next we derive some of the basic properties of the integral of a nonnegative measurable function.

Lemma 4.1.8. *Let $E \subseteq \mathbb{R}^d$ be a measurable set, and let $f, g\colon E \to [0,\infty]$ be nonnegative measurable functions.*

(a) *If A is a measurable subset of E, then $\int_A f = \int_E f\chi_A$ and $\int_A f \le \int_E f$.*

(b) *If $f \le g$, then $\int_E f \le \int_E g$.*

(c) *If $c \ge 0$, then $\int_E cf = c\int_E f$.*

(d) *If $\int_E f < \infty$, then $f(x) < \infty$ for a.e. $x \in E$.*

Proof. (a) By Definition 4.1.5,

$$\int_A f \;=\; \sup\biggl\{\int_A \phi \,:\, 0 \le \phi \le f,\ \phi \text{ simple on } A\biggr\}.$$

Let ϕ be any simple function on A such that $\phi \le f$ on A, and let ψ be the simple function on E that equals ϕ on A and is zero on $E \backslash A$. Then

$$\begin{aligned} \int_A \phi \;=\; \int_A \psi\chi_A \;&=\; \int_E \psi\chi_A && \text{(by Exercise 4.1.3(c))} \\ &\le\; \int_E f\chi_A && \text{(since } \psi\chi_A \text{ is simple and } \psi\chi_A \le f\chi_A). \end{aligned}$$

Taking the supremum over all such simple functions ϕ, we conclude that $\int_A f \;\le\; \int_E f\chi_A$. The converse inequality, and the inequality $\int_A f \;\le\; \int_E f$, follow similarly.

(b), (c) Exercise: Prove these parts.

(d) If $f = \infty$ on a set A that has positive measure, then for each $n \in \mathbb{N}$ we have

$$\int_E f \;\ge\; \int_E f\chi_A \;\ge\; \int_A n \;=\; n\,|A|.$$

Since n is arbitrary, we conclude that $\int_E f = \infty$. $\square$

Now we prove an inequality that relates the measure of the set where f exceeds a number α to the integral of f. Although the proof of this inequality is simple, it is a surprisingly useful result.

Theorem 4.1.9 (Tchebyshev's Inequality). *Let E be a measurable subset of $\mathbb{R}^d$, and let $f\colon E \to [0,\infty]$ be a measurable nonnegative function. Then for each real number $\alpha > 0$ we have*

$$|\{f > \alpha\}| \;\le\; \frac{1}{\alpha}\int_{\{f>\alpha\}} f \;\le\; \frac{1}{\alpha}\int_E f.$$

Proof. By definition, if x belongs to the set $\{f > \alpha\}$, then $f(x) > \alpha$. Moreover, $\{f > \alpha\}$ is a subset of E, so by combining this with monotonicity we obtain

$$\int_E f(x)\,dx \;\ge\; \int_{\{f>\alpha\}} f(x)\,dx \;\ge\; \int_{\{f>\alpha\}} \alpha\,dx \;=\; \alpha\,|\{f > \alpha\}|. \qquad \square$$

The following exercise shows that sets of measure zero "don't matter" when it comes to the value of an integral. The hint for the proof is to apply Theorem 4.1.9 with $\alpha = 1/n$.

Exercise 4.1.10. Let $f\colon E \to [0,\infty]$ be a measurable, nonnegative function defined on a measurable set $E \subseteq \mathbb{R}^d$. Prove that

$$\int_E f = 0 \quad \Longleftrightarrow \quad f = 0 \text{ a.e.} \qquad \diamondsuit$$

Problems

4.1.11. Exhibit a set E and a nonnegative measurable function f such that $\int_E f = \infty$ yet $f(x) < \infty$ for every $x \in E$.

4.1.12. Let E be a measurable subset of $\mathbb{R}^d$. Suppose that f and g are measurable functions on E such that $0 \le f \le g$ and $\int_E f < \infty$. Prove that $g - f$ is measurable, $0 \le \int_E (g - f) \le \infty$, and, as extended real numbers,

$$\int_E (g - f) = \int_E g - \int_E f.$$

4.2 The Monotone Convergence Theorem and Fatou's Lemma

Given measurable nonnegative functions f and g on E, intuition suggests that $\int_E (f + g)$ and $\int_E f + \int_E g$ should be equal—but are they? Suppose that ϕ is any simple function that satisfies $0 \le \phi \le f$, and ψ is any simple function that satisfies $0 \le \psi \le g$. Then $\phi + \psi$ is a simple function and $0 \le \phi + \psi \le f + g$, so

$$\int_E \phi + \int_E \psi = \int_E (\phi + \psi) \le \int_E (f + g).$$

Keeping ψ fixed and taking the supremum over all such simple functions ϕ, it follows that

$$\int_E f + \int_E \psi \le \int_E (f + g).$$

Taking the supremum next over all such simple functions ψ, we obtain

$$\int_E f + \int_E g \le \int_E (f + g).$$

But this gives us only an *inequality*, not an equality. It is not at all clear whether we can derive the opposite inequality by similar reasoning, for if we start with an arbitrary simple function $\theta \le f + g$, then it is not obvious how to relate θ to simple functions that are bounded by f and g individually.

The difficulty here is that we have defined the integral to be a supremum of approximations by simple functions, but in general the supremum of a sum need not equal the sum of the suprema. Proving linearity of the integral would be much easier if we could employ *limits* instead of suprema. This raises the important question of how limits interact with integrals. We will explore this issue (which is a ubiquitous problem in analysis) and then consider the integral of a sum.

4.2.1 The Monotone Convergence Theorem

The following result (also known as the *Beppo Levi Theorem*) shows that if nonnegative measurable functions f_n increase monotonically to a function f, then the integrals of the f_n converge to the integral of f. The shorthand notation $f_n \nearrow f$ means that $\{f_n(x)\}_{n\in\mathbb{N}}$ is monotone increasing at each point x and $f_n(x) \to f(x)$ pointwise as $n \to \infty$.

Theorem 4.2.1 (Monotone Convergence Theorem). *Let $E \subseteq \mathbb{R}^d$ be a measurable set, and let $f_n \colon E \to [0,\infty]$ be measurable functions on E such that $f_n \nearrow f$. Then*

$$\lim_{n\to\infty} \int_E f_n \;=\; \int_E f.$$

Proof. By hypothesis, for each $x \in E$ we have (in the extended real sense) that

$$f_1(x) \;\le\; f_2(x) \;\le\; \cdots \qquad \text{and} \qquad f(x) \;=\; \lim_{n\to\infty} f_n(x).$$

Consequently, Lemma 4.1.8(b) implies that we at least have the inequalities

$$0 \;\le\; \int_E f_1 \;\le\; \int_E f_2 \;\le\; \cdots \;\le\; \int_E f \;\le\; \infty. \tag{4.9}$$

Note that we have not assumed that any of the integrals on the preceding line are finite. However, an increasing sequence of nonnegative extended real numbers must converge to a nonnegative extended real number, so

$$I \;=\; \lim_{n\to\infty} \int_E f_n \tag{4.10}$$

exists in the extended real sense. Further, it follows from equation (4.9) that $0 \le I \le \int_E f \le \infty$. We must prove that $I \ge \int_E f$.

Let ϕ be any simple function such that $0 \le \phi \le f$, and fix $0 < \alpha < 1$. Set $E_n = \bigl\{f_n \ge \alpha\phi\bigr\}$, and observe that

$$E_1 \;\subseteq\; E_2 \;\subseteq\; \cdots\,.$$

Further, $\cup E_n = E$ (this is where we use the assumption $\alpha < 1$). The continuity from below property of the integral given in equation (4.7) therefore implies that $\int_{E_n} \phi \to \int_E \phi$. Consequently,

$$\begin{aligned}
I &= \lim_{n\to\infty} \int_E f_n && \text{(definition of } I\text{)} \\
&= \limsup_{n\to\infty} \int_E f_n && \text{(lim = limsup when the limit exists)} \\
&\geq \limsup_{n\to\infty} \int_{E_n} f_n && \text{(since } E_n \subseteq E\text{)} \\
&\geq \limsup_{n\to\infty} \int_{E_n} \alpha\phi && \text{(by definition of } E_n\text{)} \\
&= \alpha \int_E \phi && \text{(by equation (4.7)).}
\end{aligned}$$

Letting $\alpha \to 1$, we see that $I \geq \int_E \phi$. Finally, by taking the supremum over all such simple functions ϕ we obtain the inequality $I \geq \int_E f$. $\square$

We often use the acronym MCT as an abbreviation for "Monotone Convergence Theorem." Note that equation (4.9) implies that the integrals $\int_E f_n$ in the conclusion of the MCT increase monotonically to $\int_E f$.

Remark 4.2.2. We cannot replace Lebesgue integrals by Riemann integrals in the MCT. For example, the characteristic function of the rationals, $f = \chi_{\mathbb{Q}}$, is not Riemann integrable on the domain $E = [0,1]$. However, we can create a sequence of Riemann integrable functions that increase monotonically to f. To do this, let $\{r_n\}_{n\in\mathbb{N}}$ be an enumeration of $\mathbb{Q} \cap [0,1]$, and let f_n be the function that takes the value 1 at the points $r_1, \dots, r_n$ and is zero elsewhere, i.e., $f_n = \chi_{\{r_1,\dots,r_n\}}$. The Riemann integral of f_n on $[0,1]$ exists and is zero for every n. Yet the Riemann integral of f does not exist, even though $0 \leq f_n \nearrow f$ on $[0,1]$. $\diamondsuit$

Given a measurable function $f\colon E \to [0,\infty]$, Theorem 3.2.14 showed us how to construct simple functions ϕ_n that increase pointwise to f. Applying the MCT to this sequence of functions, it follows that $\int_E \phi_n \to \int_E f$ as $n \to \infty$. We will use this to prove that the integral of nonnegative functions is finitely additive.

Theorem 4.2.3. *Let $E \subseteq \mathbb{R}^d$ be a measurable set. If $f, g\colon E \to [0,\infty]$ are nonnegative measurable functions on E, then*

$$\int_E (f+g) \;=\; \int_E f \;+\; \int_E g.$$

Proof. Let ϕ_n and ψ_n be nonnegative simple functions such that $\phi_n \nearrow f$ and $\psi_n \nearrow g$. Then $\phi_n + \psi_n$ is simple and $\phi_n + \psi_n \nearrow f + g$. Using Lemma 4.1.2 and the Monotone Convergence Theorem, we therefore obtain

$$\begin{aligned}
\int_E (f+g) &= \lim_{n\to\infty} \int_E (\phi_n + \psi_n) && \text{(MCT)} \\
&= \lim_{n\to\infty} \biggl(\int_E \phi_n + \int_E \psi_n \biggr) && \text{(Lemma 4.1.2)} \\
&= \int_E f + \int_E g && \text{(MCT).} \qquad \square
\end{aligned}$$

Combining Theorem 4.2.3 with the Monotone Convergence Theorem gives us the following corollary for infinite series of nonnegative functions.

Corollary 4.2.4. *If $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of measurable, nonnegative functions on a measurable set $E \subseteq \mathbb{R}^d$, then*

$$\int_E \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int_E f_n.$$

Proof. Since each f_n is nonnegative, the series $f(x) = \sum_{n=1}^{\infty} f_n(x)$ converges in the extended real sense at each point $x \in E$. In fact, the partial sums $s_N = \sum_{n=1}^{N} f_n$ increase pointwise to f as $N \to \infty$. Hence, by the MCT, $\int_E s_N$ converges to $\int_E f$. On the other hand, Theorem 4.2.3 tells us that $\int_E s_N = \sum_{n=1}^{N} \int_E f_n$. Therefore

$$\int_E f = \lim_{N\to\infty} \int_E s_N = \lim_{N\to\infty} \sum_{n=1}^{N} \int_E f_n = \sum_{n=1}^{\infty} \int_E f_n. \qquad \square$$

We assign the proof of the following "countable additivity" and "continuity from below" properties of the integral to the reader.

Exercise 4.2.5. Let $E \subseteq \mathbb{R}^d$ be a measurable set. Given a nonnegative measurable function $f\colon E \to [0,\infty]$, prove the following statements.

(a) If $A_1, A_2, \dots$ are disjoint measurable subsets of E and $A = \cup A_n$, then

$$\int_A f = \sum_{n=1}^{\infty} \int_{A_n} f.$$

(b) If $A_1 \subseteq A_2 \subseteq \cdots$ are nested measurable subsets of E and $A = \cup A_n$, then

$$\int_A f = \lim_{n\to\infty} \int_{A_n} f. \qquad \diamondsuit$$

4.2.2 Fatou's Lemma

Suppose that $f_n \colon E \to [0,\infty]$ is a measurable function for each $n \in \mathbb{N}$, and $f_n \to f$ pointwise on E. Must $\int_E f_n$ converge to $\int_E f$? The Monotone Convergence Theorem says that *if f_n increases pointwise to f*, then this is the case. Unfortunately, the following example shows that convergence of the integrals can fail if our sequence is not monotonically increasing.

Example 4.2.6 (Shrinking Boxes II). Let $E = [0,1]$ and set $f_n = n\,\chi_{(0,\frac{1}{n}]}$. Then $f_n(x) \to 0$ for every $x \in \mathbb{R}$, yet $\int_0^1 f_n = 1$ for every n. Hence

$$\int_0^1 \Bigl(\lim_{n\to\infty} f_n\Bigr) \;=\; 0 \;<\; 1 \;=\; \lim_{n\to\infty} \int_0^1 f_n.$$

Thus, for these functions the integral of the limit is not the limit of the integrals. It is true that the functions in this example are discontinuous, but that is not the issue. For example, if we replace the "boxes" $f_n = n\,\chi_{(0,\frac{1}{n}]}$ with "triangles" that have height n and base $[0,\frac{1}{n}]$ (similar to the Shrinking Triangles of Example 3.4.1 except with height n instead of height 1), then f_n converges pointwise to the zero function yet $\int_0^1 f_n = \frac{1}{2}$ for every n. $\diamondsuit$

Although Example 4.2.6 shows that pointwise convergence of functions need not imply convergence of the corresponding integrals, the next theorem gives a weaker but still very useful *inequality* that relates $\lim_{n\to\infty} \int_E f_n$ to $\int_E f$ when each function f_n is nonnegative. In fact, for this result we do not even need to assume that the functions f_n converge pointwise or that their integrals converge. Even without convergence, we obtain an inequality stated in terms of liminfs instead of limits.

Theorem 4.2.7 (Fatou's Lemma). *If $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of nonnegative measurable functions on a measurable set $E \subseteq \mathbb{R}^d$, then*

$$\int_E \Bigl(\liminf_{n\to\infty} f_n\Bigr) \;\le\; \liminf_{n\to\infty} \int_E f_n. \tag{4.11}$$

In particular, if $f_n(x) \to f(x)$ for each $x \in E$, then

$$\int_E f \;\le\; \liminf_{n\to\infty} \int_E f_n. \tag{4.12}$$

Proof. Define

$$f(x) \;=\; \liminf_{n\to\infty} f_n(x) \;=\; \lim_{k\to\infty} \inf_{n\ge k} f_n(x) \;=\; \lim_{k\to\infty} g_k(x)$$

where

$$g_k(x) \;=\; \inf_{n\ge k} f_n(x).$$

The functions g_k increase monotonically to f, i.e., $g_k \nearrow f$. The Monotone Convergence Theorem therefore implies that

$$\int_E f \;=\; \lim_{k\to\infty} \int_E g_k.$$

However, $g_k \le f_k$ and therefore $\int g_k \le \int f_k$ for every k. Consequently,

$$\int_E f \;=\; \lim_{k\to\infty} \int_E g_k \;=\; \liminf_{k\to\infty} \int_E g_k \;\le\; \liminf_{k\to\infty} \int_E f_k.$$

This proves equation (4.11). Equation (4.12) follows by recalling that if the limit of a sequence exists, then it equals the liminf of the sequence. □

Problems

4.2.8. Assume Fatou's Lemma and deduce the Monotone Convergence Theorem from it.

4.2.9. Let $f_n\colon E \to [0,\infty]$ be measurable functions defined on a measurable set $E \subseteq \mathbb{R}^d$. Suppose that $f_n \to f$ pointwise and $f_n \le f$ for each $n \in \mathbb{N}$. Show that $\int_E f_n \to \int_E f$ as $n \to \infty$ (note that $\int_E f$ might be ∞).

4.2.10. Assume $E \subseteq \mathbb{R}^d$ and $f\colon E \to [0,\infty]$ are measurable, and $\int_E f < \infty$. Prove that $\sum_{n=1}^\infty \bigl|\{f \ge n\}\bigr| < \infty$.

4.2.11. Assume $E \subseteq \mathbb{R}^d$ and $f\colon E \to [0,\infty]$ are measurable, and $\int_E f < \infty$. Given $\varepsilon > 0$, prove that there exists a measurable set $A \subseteq E$ such that $|A| < \infty$ and $\int_A f \ge \int_E f - \varepsilon$.

4.2.12. Let $E \subseteq \mathbb{R}^d$ and $f\colon E \to [0,\infty]$ be measurable and suppose that $\int_E f(x)^n\,dx = \int_E f(x)\,dx < \infty$ for every positive integer n. Prove that there is a measurable set $A \subseteq E$ such that $f = \chi_A$ a.e.

4.2.13. Let E be a measurable subset of $\mathbb{R}^d$, and let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of nonnegative measurable functions on E such that $f_n \to f$ a.e. Suppose that $\lim_{n\to\infty} \int_E f_n = \int_E f$ and $\int_E f < \infty$. Prove that $\lim_{n\to\infty} \int_A f_n = \int_A f$ for every measurable set $A \subseteq E$. Show by example that this can fail if $\int_E f = \infty$.

4.2.14. Let f be a continuous, nonnegative function on the interval $[a,b]$. Prove that the Riemann integral of f on $[a,b]$ coincides with its Lebesgue integral $\int_a^b f(x)\,dx$.

4.2.15. Let E be a measurable subset of $\mathbb{R}^d$, and suppose that f_n and f are nonnegative measurable functions on E such that $f_n \searrow f$ pointwise. Prove that if $\int_E f_k < \infty$ for some k, then $\int_E f_n \to \int_E f$ as $n \to \infty$. Show by example that the assumption that some f_k has finite integral is necessary.

4.2.16. Let E be a measurable subset of $\mathbb{R}^d$ such that $|E| < \infty$, and let f be a nonnegative, bounded function on E. Prove that f is measurable if and only if

$$\sup\biggl\{\int_E \phi \,:\, 0 \le \phi \le f,\ \phi \text{ simple}\biggr\} \;=\; \inf\biggl\{\int_E \psi \,:\, f \le \psi,\ \psi \text{ simple}\biggr\}.$$

4.2.17. Let $f\colon E \to [0,\infty]$ be a nonnegative, measurable function defined on a measurable set $E \subseteq \mathbb{R}^d$. This problem will quantify the idea that the integral of f equals "the area of the region under its graph."

(a) The *graph of f* is

$$\Gamma_f \;=\; \bigl\{(x, f(x)) : x \in E,\ f(x) < \infty\bigr\}.$$

Show that $|\Gamma_f| = 0$.

(b) The *region under the graph of f* is the set R_f that consists of all points $(x,y) \in \mathbb{R}^{d+1} = \mathbb{R}^d \times \mathbb{R}$ such that $x \in E$ and y satisfies

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

Show that R_f is a measurable subset of $\mathbb{R}^{d+1}$, and its Lebesgue measure is

$$|R_f| \;=\; \int_E f(x)\, dx.$$

4.2.18. (a) Prove *Fatou's Lemma for series*: If $a_{kn} \ge 0$ for every $k, n \in \mathbb{N}$, then

$$\sum_{k=1}^{\infty} \liminf_{n\to\infty} a_{kn} \;\le\; \liminf_{n\to\infty} \sum_{k=1}^{\infty} a_{kn}.$$

Show by example that strict inequality can hold.

(b) Formulate and prove a Monotone Convergence Theorem for series.

4.3 The Lebesgue Integral of Measurable Functions

In the preceding section we defined the integral of nonnegative measurable functions. Now we will consider functions that can take extended real values or complex values.

4.3.1 Extended Real-Valued Functions

We begin with extended real-valued functions. A generic measurable, extended real-valued function f can take both positive and negative values, so to define its integral we split f into its positive and negative parts $f^+(x) = \max\{f(x), 0\}$ and $f^-(x) = \max\{-f(x), 0\}$. Since f^+ and f^- are nonnegative and measurable, they each have well-defined Lebesgue integrals. Furthermore, $f = f^+ - f^-$, so we will declare the integral of f to be the difference of $\int_E f^+$ and $\int_E f^-$. However, we must be careful to exclude any cases that would assign an indeterminate form to the integral.

Definition 4.3.1 (Lebesgue Integral of an Extended Real-Valued Function). Let $f\colon E \to [-\infty, \infty]$ be a measurable extended real-valued function defined on a measurable set $E \subseteq \mathbb{R}^d$. The *Lebesgue integral of f over E* is

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

as long as this does not have the form $\infty - \infty$ (in that case, the integral is undefined). $\diamondsuit$

Here is an example of a function whose Lebesgue integral does not exist.

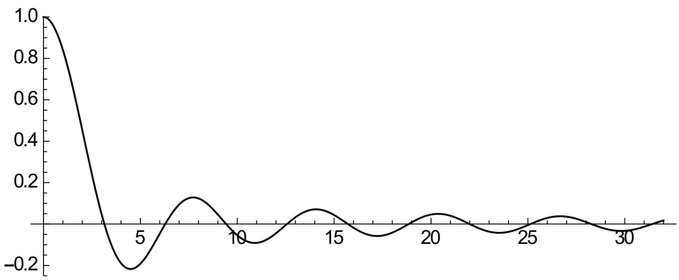


Fig. 4.2 Graph of $\operatorname{sinc}(x) = \frac{\sin x}{x}$ for $x \geq 0$.

Exercise 4.3.2. The (unnormalized) *sinc function* is

$$\operatorname{sinc}(x) = \frac{\sin x}{x}, \qquad x \neq 0.$$

This function is continuous on $\mathbb{R}$ if we set $\operatorname{sinc}(0) = 1$ (see the illustration in Figure 4.2). Prove that the Lebesgue integrals of the positive and negative parts of the sinc function over $[0, \infty)$ are both infinite, i.e.,

$$\int_0^\infty \operatorname{sinc}^+(x)\,dx = \infty = \int_0^\infty \operatorname{sinc}^-(x)\,dx.$$

Conclude that the Lebesgue integral of sinc on $E = [0, \infty)$ does not exist. Even so, Problem 4.6.19 will show that the *improper Riemann integral of the sinc function over* $[0, \infty)$ does exist, and it has the value

$$\lim_{a \to \infty} \int_0^a \frac{\sin x}{x}\, dx \;=\; \frac{\pi}{2}. \qquad \diamondsuit$$

The next lemma gives a simple but useful inequality that relates the integral of f to the integral of $|f|$. Note that since $|f|$ is nonnegative and measurable, its Lebesgue integral always exists (in the extended real sense), even if the integral of f is undefined.

Lemma 4.3.3. *Let $f \colon E \to [-\infty, \infty]$ be a measurable function defined on a measurable set $E \subseteq \mathbb{R}^d$.*

(a) *If $\int_E f$ exists, then*

$$0 \;\le\; \left| \int_E f \right| \;\le\; \int_E |f| \;\le\; \infty.$$

(b) *$\int_E f$ exists and is finite if and only if $\int_E |f| < \infty$.*

Proof. Each of the three functions f^+, f^-, and $|f| = f^+ + f^-$ are nonnegative and measurable, so their integrals are well-defined nonnegative extended real numbers. Further, $0 \le f^+, f^- \le |f|$, so

$$0 \;\le\; \int_E f^- \;\le\; \int_E |f| \;\le\; \infty \qquad \text{and} \qquad 0 \;\le\; \int_E f^+ \;\le\; \int_E |f| \;\le\; \infty.$$

(a) Assume that the integral of f exists. Then, by definition, one or both of $\int_E f^+$ and $\int_E f^-$ must be finite. Therefore

$$0 \;\le\; \left| \int_E f \right| \;=\; \left| \int_E f^+ - \int_E f^- \right| \;\le\; \int_E f^+ + \int_E f^- \;=\; \int_E |f| \;\le\; \infty.$$

(b) Since $\int_E f^+$ and $\int_E f^-$ are nonnegative,

$$\begin{aligned} \int_E f \text{ exists and is finite} \quad &\Longleftrightarrow \quad \int_E f^+,\ \int_E f^- \;<\; \infty \\ &\Longleftrightarrow \quad \int_E f^+ + \int_E f^- \;<\; \infty \\ &\Longleftrightarrow \quad \int_E |f| \;<\; \infty. \qquad \square \end{aligned}$$

Looking ahead to Definition 4.4.1, a function that satisfies $\int_E |f| < \infty$ is said to be *integrable* on E.

4.3.2 Complex-Valued Functions

Now we turn to the complex-valued setting. We define the integral of a complex-valued function by breaking it into real and imaginary parts.

Definition 4.3.4 (Lebesgue Integral of a Complex-Valued Function). Let $f\colon E \to \mathbb{C}$ be a measurable complex-valued function defined on a measurable set $E \subseteq \mathbb{R}^d$. Write f in real and imaginary parts as $f = f_r + if_i$, where f_r and f_i are real-valued. If $\int_E f_r$ and $\int_E f_i$ *both exist and are finite*, then the *Lebesgue integral of f over E* is

$$\int_E f \;=\; \int_E f_r \;+\; i\int_E f_i.$$

Otherwise, the integral is undefined. $\diamondsuit$

While the integral of an extended real-valued function can be $\pm\infty$, the integral of a complex-valued function is always a complex scalar (if it exists).

Now we derive an analogue of Lemma 4.3.3 for complex-valued functions.

Lemma 4.3.5. *Let $f\colon E \to \mathbb{C}$ be a measurable function defined on a measurable set $E \subseteq \mathbb{R}^d$. Then*

$$\int_E f \text{ exists} \quad\Longleftrightarrow\quad \int_E |f| \;<\; \infty. \tag{4.13}$$

Further, in this case we have

$$0 \;\le\; \left|\int_E f\right| \;\le\; \int_E |f| \;<\; \infty. \tag{4.14}$$

Proof. First note that since $|f|$ is nonnegative, $\int_E |f|$ exists as a nonnegative, extended real number (although it could be ∞). Write $f = f_r + if_i$, where f_r and f_i are real-valued.

Suppose that $\int_E f$ exists. Then Definition 4.3.4 requires that $\int_E f_r$ and $\int_E f_i$ both be finite real numbers. Consequently, Lemma 4.3.3 implies that $\int_E |f_r|$ and $\int_E |f_i|$ are finite. Therefore

$$\int_E |f| \;=\; \int_E |f_r + if_i| \;\le\; \int_E \bigl(|f_r| + |f_i|\bigr) \;=\; \int_E |f_r| \;+\; \int_E |f_i| \;<\; \infty.$$

Conversely, if $\int_E |f|$ is finite, then both $\int_E |f_r|$ and $\int_E |f_i|$ must be finite, and therefore $\int_E f$ is defined. This establishes equation (4.13).

To prove equation (4.14), assume that $\int_E |f| < \infty$. Then $z = \int_E f$ exists and is a complex number. Let α be a complex number with $|\alpha| = 1$ such that $\alpha z = |z|$ (if $z \neq 0$ then α is uniquely determined, while otherwise α can be any complex number with unit modulus). That is, $|\alpha| = 1$ and

$$\left|\int_E f\right| = \alpha \int_E f.$$

Now write αf (not f!) in real and imaginary parts, i.e., $\alpha f = g + ih$ where g and h are real-valued. Assuming that $\int_E \alpha f = \alpha \int_E f$ (the formal justification is assigned below as part of Exercise 4.3.6), we compute that

$$\left|\int_E f\right| = \alpha \int_E f = \int_E \alpha f = \int_E g + i \int_E h.$$

Since $\left|\int_E f\right|$ is a real number, we must have $\int_E h = 0$ (though we cannot infer from this that h is zero). As g is real-valued, we apply Lemma 4.3.3 to obtain

$$\left|\int_E f\right| = \int_E g \le \int_E |g| \le \int_E |f|,$$

the final inequality following from the fact that g is the real part of αf, and therefore $|g| \le |\alpha f| = |f|$. $\square$

4.3.3 Properties of the Integral

The following exercise gives some properties of the integrals of extended real-valued or complex-valued functions. In the statement of this exercise, when we write a condition like "$f \le g$ a.e." we implicitly assume that f and g are extended real-valued functions. However, a hypothesis such as "$f = 0$ a.e." can be satisfied by either an extended real-valued or a complex-valued function.

Exercise 4.3.6. Let $E \subseteq \mathbb{R}^d$ be measurable, and assume that $f, g \colon E \to \overline{\mathbf{F}}$ are measurable. Prove the following statements.

(a) If $\int_E f$ and $\int_E g$ both exist and $f \le g$ a.e., then $\int_E f \le \int_E g$.

(b) If $\int_E f$ and $\int_E g$ both exist and $f = g$ a.e., then $\int_E f = \int_E g$.

(c) If $\int_E f$ exists and A is a measurable subset of E, then $\int_A f$ exists.

(d) If $f = 0$ a.e. on E, then $\int_E f$ exists and $\int_E f = 0$.

(e) If $\int_E f$ exists and c is a scalar, then $\int_E cf$ exists and $\int_E cf = c \int_E f$.

(f) If $\int_E f$ exists and $A_1, A_2, \dots$ are disjoint measurable subsets of E, then

$$\int_{\cup A_n} f = \sum_{n=1}^{\infty} \int_{A_n} f.$$

(g) If $\int_E f$ exists and $A_1 \subseteq A_2 \subseteq \cdots$ are nested measurable subsets of E, then

$$\int_{\cup A_n} f \;=\; \lim_{n\to\infty} \int_{A_n} f. \qquad \diamondsuit$$

In particular, statement (b) of the preceding exercise shows that changing the value of a function on a set of zero measure does not change the value of its integral. Consequently, many of our earlier theorems that required hypotheses to hold at all points are still valid if we assume only that the hypotheses hold almost everywhere. Here is such a version of the Monotone Convergence Theorem.

Theorem 4.3.7 (Monotone Convergence Theorem). *Assume that E is a measurable subset of $\mathbb{R}^d$. If functions $f_n\colon E \to [-\infty,\infty]$ are measurable, $f_n \ge 0$ a.e., and $f_n(x) \nearrow f(x)$ for a.e. $x \in E$, then*

$$\lim_{n\to\infty} \int_E f_n \;=\; \int_E f.$$

Proof. Let Z be the set of all points x where either some $f_n(x)$ is negative or $f_n(x)$ does not converge to $f(x)$. For $x \notin Z$ set $g_n(x) = f_n(x)$ and $g(x) = f(x)$, and let $g_n(x) = g(x) = 0$ for all $x \in Z$. Then the set Z has measure zero, $g_n \ge 0$ everywhere, and $g_n \nearrow g$, so the Monotone Convergence Theorem implies that

$$\int_E f_n \;=\; \int_E g_n \;\nearrow\; \int_E g \;=\; \int_E f. \qquad \square$$

An entirely similar approach establishes the following extension of Fatou's Lemma.

Theorem 4.3.8 (Fatou's Lemma). *Assume that $E \subseteq \mathbb{R}^d$ is measurable. If functions $f_n\colon E \to [-\infty,\infty]$ are measurable with $f_n \ge 0$ a.e., then*

$$\int_E \Bigl(\liminf_{n\to\infty} f_n\Bigr) \;\le\; \liminf_{n\to\infty} \int_E f_n. \qquad \diamondsuit$$

Problems

4.3.9. Assume that $f\colon \mathbb{R}^d \to \overline{\mathbf{F}}$ is measurable. Show that if $\int_{\mathbb{R}^d} f$ exists, then for each point $a \in \mathbb{R}^d$ we have

$$\int_{\mathbb{R}^d} f(x-a)\,dx \;=\; \int_{\mathbb{R}^d} f(x)\,dx \;=\; \int_{\mathbb{R}^d} f(a-x)\,dx.$$

4.3.10. Let $L\colon \mathbb{R}^d \to \mathbb{R}^d$ be an invertible linear transformation, let $E \subseteq \mathbb{R}^d$ be a measurable set, and let $f\colon E \to \overline{\mathbf{F}}$ be a measurable function such that $\int_E f$ exists. Show that

$$\int_E f(x)\,dx \;=\; |\det(L)| \int_{L^{-1}(E)} f(Lx)\,dx.$$

4.4 Integrable Functions and $L^1(E)$

We regularly encounter the quantity $\int_E |f|$ and the condition $\int_E |f| < \infty$, so we introduce the following terminology.

Definition 4.4.1 (L^1-Norm and Integrable Functions). Let $E \subseteq \mathbb{R}^d$ be a measurable set, and let $f\colon E \to \overline{\mathbf{F}}$ be a measurable function on E.

(a) The extended real number

$$\|f\|_1 \;=\; \int_E |f|$$

is called the *L^1-norm of f on E* (it could be infinite).

(b) We say that f is *integrable* on E if $\|f\|_1 = \int_E |f| < \infty$. $\diamondsuit$

Although we refer to $\|\cdot\|_1$ as a "norm," it is actually only a seminorm on the space of integrable functions because $\|f\|_1 = 0$ if and only if $f = 0$ a.e. (see Exercise 4.4.5).

We will study integrable functions and the L^1-norm in this section. First, we give some examples.

Example 4.4.2. (a) If $f = 0$ a.e., then $\|f\|_1 = \int_E |f| = 0$ by Exercise 4.3.6(d).

(b) If $|E| < \infty$ and f is bounded on E, then f is integrable. However, if $|E| = \infty$, then the function that is identically 1 on E is bounded yet not integrable.

(c) An unbounded function can be integrable, e.g., consider $f(x) = x^{-1/2}$ on the interval $[0,1]$.

(d) An integrable function must be finite at almost every point of E (why?). However, there are functions that are finite a.e. but not integrable (for example, consider $g(x) = x^{-1}$ on the interval $[0,1]$).

(e) An integrable function need not decay to zero at $\pm\infty$. In fact, there exist *unbounded, continuous functions* $f\colon \mathbb{R} \to \mathbb{R}$ that are integrable (see Problem 4.4.16). $\diamondsuit$

4.4.1 The Lebesgue Space $L^1(E)$

The Lebesgue space $L^\infty(E)$ introduced in Definition 3.3.3 consists of the essentially bounded functions on E. We similarly collect the integrable functions to form a space that we call $L^1(E)$. Technically, there are two versions of $L^1(E)$, one consisting of complex-valued functions and one consisting of extended real-valued functions (which must be finite a.e., since they are integrable). Both cases are important, and in practice it is usually clear from context whether we are working with extended real-valued functions or complex-valued functions. As usual, we combine the two possibilities into a single definition by letting $\overline{\mathbf{F}}$ denote either $[-\infty, \infty]$ or $\mathbb{C}$. Implicitly, the word *scalar* denotes a real number $c \in \mathbb{R}$ if $\overline{\mathbf{F}} = [-\infty, \infty]$, and a complex number $c \in \mathbb{C}$ if $\overline{\mathbf{F}} = \mathbb{C}$.

Definition 4.4.3 (The Lebesgue Space $L^1(E)$). If E is a measurable subset of $\mathbb{R}^d$, then the *Lebesgue space of integrable functions on* E is

$$L^1(E) = \left\{ f : E \to \overline{\mathbf{F}} \ : \ f \text{ is measurable and } \|f\|_1 = \int_E |f| < \infty \right\}. \qquad \diamondsuit$$

Suppose that f and g are integrable functions on E and a and b are scalars. Regardless of whether we are considering extended real-valued or complex-valued functions, $|af + bg|$ is an extended real-valued function. Therefore we can apply Theorem 4.2.3 and compute that

$$\int_E |af + bg| \le \int_E \bigl(|a||f| + |b||g|\bigr) = |a| \int_E |f| + |b| \int_E |g| < \infty. \quad (4.15)$$

This shows that $af + bg$ is integrable. Consequently $L^1(E)$ is closed under the operations of addition of functions and multiplication of a function by a scalar, so it is a vector space with respect to these operations.

Remark 4.4.4. In contrast, $L^1(E)$ need not be closed under products. For example, if $E = [0, 1]$ then $f(x) = x^{-1/2} \in L^1[0, 1]$, but the product of f with itself is

$$f^2(x) = f(x)\, f(x) = \frac{1}{x} \notin L^1[0, 1].$$

More generally, Problem 4.4.21 asks for a proof that $L^1(E)$ is *never* closed under products (except in the trivial case that $|E| = 0$). On the other hand, in Section 4.6.3 we will introduce a "multiplication-like" operation known as *convolution* that is defined for functions on the domain $\mathbb{R}^d$, and we will prove that $L^1(\mathbb{R}^d)$ is closed with respect to convolution. $\diamondsuit$

The following exercise shows that the L^1-norm has properties similar to those of the L^∞-norm (see Exercise 3.3.4).

Exercise 4.4.5. Assume that $E \subseteq \mathbb{R}^d$ is measurable. Prove that the following statements hold for all functions $f, g \in L^1(E)$ and all scalars c.

(a) Nonnegativity: $0 \le \|f\|_1 < \infty$.

(b) Homogeneity: $\|cf\|_1 = |c|\, \|f\|_1$.

(c) The Triangle Inequality: $\|f+g\|_1 \le \|f\|_1 + \|g\|_1$.

(d) Almost Everywhere Uniqueness: $\|f\|_1 = 0$ if and only if $f = 0$ a.e. $\diamondsuit$

Considering the definition of *norms* and *seminorms* from Section 1.2.2, parts (a)–(c) of Exercise 4.4.5 tell us that $\|\cdot\|_1$ is a *seminorm* on $L^1(E)$. However, $\|\cdot\|_1$ is not a norm because $\|f\|_1 = 0$ does not imply that f is identically zero. Instead, $\|f\|_1 = 0$ implies that f is zero *almost everywhere.* We will explore this issue in more depth in Chapter 7, where we discuss both $L^1(E)$ and related spaces $L^p(E)$ in detail.

4.4.2 Convergence in L^1-Norm

The *distance* between two functions f and g with respect to the L^1-norm is $\|f-g\|_1$. Once we have a notion of distance, we also have a corresponding notion of convergence, made precise in the following definition.

Definition 4.4.6 (Convergence in L^1-Norm). Let E be a measurable subset of $\mathbb{R}^d$. A sequence of integrable functions $\{f_n\}_{n\in\mathbb{N}}$ on E (either extended real-valued or complex-valued) is said to *converge to f in L^1-norm* if

$$\lim_{n\to\infty} \|f-f_n\|_1 \;=\; \lim_{n\to\infty} \int_E |f-f_n| \;=\; 0.$$

In this case we write $f_n \to f$ *in L^1-norm.* $\diamondsuit$

The following examples compare L^1-norm convergence to pointwise a.e. convergence.

Example 4.4.7. The domain for this example is $E = [0,1]$.

(a) The *Shrinking Boxes* $f_n = \chi_{[0,\frac{1}{n}]}$ from Example 3.5.2 converge pointwise a.e. to the zero function, and they also converge to the zero function in L^1-norm, because

$$\|0-f_n\|_1 \;=\; \|f_n\|_1 \;=\; \int_0^1 \chi_{[0,\frac{1}{n}]} \;=\; \frac{1}{n} \;\to\; 0.$$

(b) The *Shrinking Boxes* $f_n = n\,\chi_{[0,\frac{1}{n}]}$ from Example 4.2.6 converge pointwise a.e. to the zero function, but they do not converge in L^1-norm to the zero function because for every n we have

$$\|0-f_n\|_1 \;=\; \|f_n\|_1 \;=\; n\int_0^1 \chi_{[0,\frac{1}{n}]} \;=\; 1.$$

Hence pointwise a.e. convergence does not imply L^1-norm convergence in general.

(c) Let $\{f_n\}_{n\in\mathbb{N}}$ be the sequence of *Boxes Marching in Circles* defined in Example 3.5.5. The values of $\|f_n\|_1$ for $n = 1, \dots, 10$ are

$$1,\ \frac{1}{2},\ \frac{1}{2},\ \frac{1}{3},\ \frac{1}{3},\ \frac{1}{3},\ \frac{1}{4},\ \frac{1}{4},\ \frac{1}{4},\ \frac{1}{4}.$$

Continuing this sequence, we see that the functions f_n converge in L^1-norm to the zero function (slowly, to be sure, but they do converge). However, f_n does not converge pointwise a.e., so convergence in L^1-norm does not imply pointwise a.e. convergence. $\diamondsuit$

Although L^1-norm convergence does not imply pointwise a.e. convergence, we will use Tchebyshev's Inequality to prove that L^1-norm convergence implies *convergence in measure*, and consequently there must exist a *subsequence* that converges pointwise a.e.

Lemma 4.4.8. *Let $E \subseteq \mathbb{R}^d$ be a measurable set, and let f_n and f be integrable functions on E. If $f_n \to f$ in L^1-norm, then:*

(a) $f_n \xrightarrow{\mathrm{m}} f$, *and*

(b) *there exists a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ such that $f_{n_k} \to f$ pointwise a.e.*

Proof. If we fix any $\varepsilon > 0$, then Tchebyshev's Inequality (Theorem 4.1.9) implies that

$$\lim_{n\to\infty} \bigl|\{|f - f_n| > \varepsilon\}\bigr| \;\le\; \lim_{n\to\infty} \frac{1}{\varepsilon} \int_E |f - f_n| \;=\; \frac{1}{\varepsilon} \lim_{n\to\infty} \|f - f_n\|_1 \;=\; 0.$$

This shows that f_n converges in measure to f. Consequently we can apply Lemma 3.5.6, which states that any sequence that converges in measure has a subsequence that converges pointwise a.e. $\square$

In Figure 3.3, we showed some implications that hold between certain types of convergence criteria. Figure 4.3 shows the implications that hold when we also include the results of Lemma 4.4.8.

Sometimes we need to deal with families indexed by a real parameter. In particular, if $f \in L^1(E)$ and we are given functions $f_t \in L^1(E)$ for each t in some interval $(0, c)$, then we declare that $f_t \to f$ in L^1-norm as $t \to 0^+$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $\|f - f_t\|_1 < \varepsilon$ whenever $0 < t < \delta$. The following lemma (essentially a restatement of Problem 1.1.23) deals with L^1-norm convergence in this context, and shows that convergence as $t \to 0^+$ can be reduced to consideration of sequences indexed by the natural numbers.

Lemma 4.4.9. *Let $E \subseteq \mathbb{R}^d$ be measurable, and let $f_t, f \in L^1(E)$ be given for t in some interval $(0, c)$, where $c > 0$. Then $f_t \to f$ in L^1-norm as $t \to 0^+$ if and only if $\|f - f_{t_k}\|_1 \to 0$ for every sequence of real numbers $\{t_k\}_{k\in\mathbb{N}}$ in $(0, c)$ that satisfy $t_k \to 0$.* $\diamondsuit$

$$
\begin{tikzcd}
& \text{pointwise a.e. convergence} \arrow[d, Rightarrow, "\text{(if } |E|<\infty)"] & L^1\text{-norm convergence} \arrow[d, Rightarrow] & \\
L^\infty\text{-norm convergence} \arrow[r, Rightarrow] & \text{almost uniform convergence} \arrow[r, Rightarrow] \arrow[d, Rightarrow] & \text{convergence in measure} \arrow[r, Rightarrow] & \text{pointwise a.e. convergence of a subsequence} \\
& \text{pointwise a.e. convergence} & &
\end{tikzcd}
$$

Fig. 4.3 Relations among certain convergence criteria (valid for sequences of functions that are either complex-valued or extended real-valued but finite a.e.).

4.4.3 Linearity of the Integral for Integrable Functions

By Theorem 4.2.3, $\int_E (f+g) = \int_E f + \int_E g$ for all nonnegative functions f and g. We will enlarge the class of functions for which this conclusion holds, but we must impose *some* restrictions in order to exclude indeterminate forms. The following result achieves this by focusing on integrable functions.

Theorem 4.4.10 (Linearity of the Integral). *Let $E \subseteq \mathbb{R}^d$ be a measurable set. If $f, g \colon E \to \overline{\mathbf{F}}$ are integrable functions and a and b are scalars, then*

$$\int_E (af+bg) \;=\; a\int_E f \;+\; b\int_E g. \tag{4.16}$$

Proof. Case 1: $\overline{\mathbf{F}} = [-\infty,\infty]$. Assume that $f, g \colon E \to [-\infty,\infty]$ are integrable functions on E. By equation (4.15), their sum $f+g$ is also integrable. Define the measurable sets

$$
\begin{aligned}
E_1 &= \{f \ge 0,\, g \ge 0\}, & E_4 &= \{f < 0,\, g \ge 0,\, f+g \ge 0\},\\
E_2 &= \{f \ge 0,\, g < 0,\, f+g \ge 0\}, & E_5 &= \{f < 0,\, g \ge 0,\, f+g < 0\},\\
E_3 &= \{f \ge 0,\, g < 0,\, f+g < 0\}, & E_6 &= \{f < 0,\, g < 0\}.
\end{aligned}
$$

Consider the integral of $f+g$ on the set E_2. Since $f+g$ and $-g$ are each nonnegative on E_2, we compute that

$$
\begin{aligned}
\int_{E_2}(f+g) \;-\; \int_{E_2} g &= \int_{E_2}(f+g) \;+\; \int_{E_2}(-g) && \text{(by Exercise 4.3.6(e))}\\
&= \int_{E_2}(f+g) \;+\; (-g) && \text{(by Theorem 4.2.3)}\\
&= \int_{E_2} f.
\end{aligned}
$$

Since each integral is finite, we can rearrange to obtain

$$\int_{E_2} (f+g) \;=\; \int_{E_2} f \;+\; \int_{E_2} g.$$

A similar argument shows that equality holds for each of the other sets E_k. Consequently, since $E_1, \dots, E_6$ partition E, we can use Exercise 4.3.6(f) to compute that

$$\int_E (f+g) \;=\; \sum_{k=1}^{6} \int_{E_k} (f+g) \;=\; \sum_{k=1}^{6} \biggl(\int_{E_k} f \;+\; \int_{E_k} g \biggr) \;=\; \int_E f \;+\; \int_E g.$$

Equation (4.16) therefore follows by combining this equality with the homogeneity property of the integral given in Exercise 4.3.6(e).

Case 2: $\overline{\mathbf{F}} = \mathbb{C}$. This follows by splitting into real and imaginary parts and applying Case 1. $\square$

We will use the linearity of the integral to prove that if a sequence of functions $\{f_n\}_{n\in\mathbb{N}}$ converges in L^1-norm, then the integrals of the f_n converge. That is, if $\int_E |f - f_n| \to 0,$ then we must have $\int_E f_n \to \int_E f$ as well.

Lemma 4.4.11. *Let E be a measurable subset of $\mathbb{R}^d$. If $f_n, f\colon E \to \overline{\mathbf{F}}$ are integrable functions on E and $f_n \to f$ in L^1-norm, then*

$$\lim_{n\to\infty} \int_E f_n \;=\; \int_E f.$$

Proof. Applying linearity and either Lemma 4.3.3 (for extended real-valued functions) or Lemma 4.3.5 (for complex-valued functions), we see that

$$\biggl| \int_E f \;-\; \int_E f_n \biggr| \;=\; \biggl| \int_E (f - f_n) \biggr| \;\le\; \int_E |f - f_n| \;=\; \|f - f_n\|_1 \;\to\; 0. \qquad \square$$

4.4.4 Inclusions between $L^1(E)$ and $L^\infty(E)$

The L^1-norm and the L^∞-norm measure the distance between functions in different ways. For example, consider the two functions f and g shown in Figure 4.4. There is a set of positive measure (in fact, an interval centered at $x=1$) where $|f(x) - g(x)| \ge 3$. Consequently, $\|f-g\|_\infty \ge 3$, so as measured by the L^∞-norm, the distance between these two functions is large. However, the *integral* of $|f(x) - g(x)|$ is small (numerically, $\|f-g\|_1 \approx 0.3$ for these two functions). Hence f and g are close together, at least as measured by the L^1-norm. We take a closer look now at the relationship between $\|\cdot\|_1$ and $\|\cdot\|_\infty$.

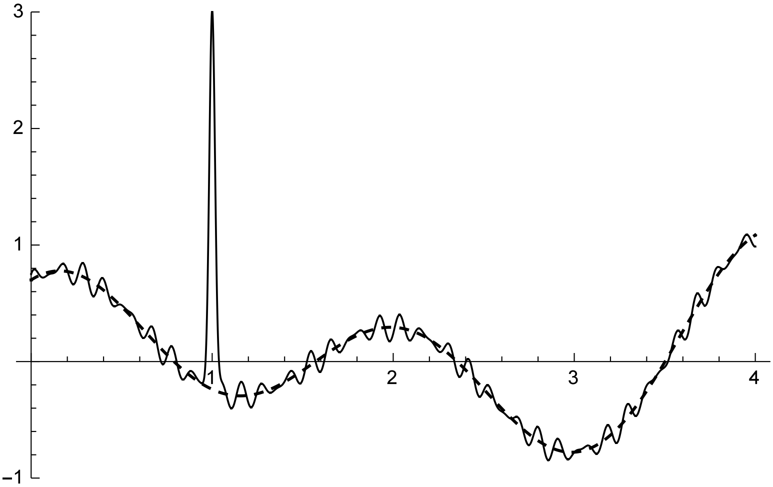


Fig. 4.4 The distance between the function f (solid curve) and g (dashed curve) is small when measured by the L^1-norm, but large when measured by the L^∞-norm.

An integrable function need not be essentially bounded. For example, $f(x) = x^{-1/2}$ is integrable even though it is unbounded on the interval $[0,1]$. In fact, we will show that there exist unbounded integrable functions on any domain that has positive measure.

Lemma 4.4.12. *If E is a measurable subset of $\mathbb{R}^d$ and $|E| > 0$, then there exists a function $f \in L^1(E) \setminus L^\infty(E)$.*

Proof. By Problem 2.3.20(a), there exists a measurable set $A \subseteq E$ with measure $0 < |A| < \infty$. By part (c) of that same problem, there exist disjoint, measurable subsets A_k of A such that $|A_k| = 2^{-k}\,|A|$ for each $k \in \mathbb{N}$. The function

$$f \;=\; \sum_{k=1}^{\infty} 2^{k/2} \chi_{A_k}$$

is integrable on E, but it is not essentially bounded. □

In the converse direction, $L^\infty(E)$ is not contained in $L^1(E)$ if E has infinite measure, because the constant function 1 is bounded but not integrable when $|E| = \infty$. On the other hand, the following lemma shows that $L^\infty(E)$ is contained in $L^1(E)$ whenever $|E| < \infty$. Moreover, convergence in L^∞-norm implies convergence in L^1-norm in this case.

Lemma 4.4.13. *If E is a measurable subset of $\mathbb{R}^d$ such that $|E| < \infty$, then the following statements hold.*

(a) *If $f \colon E \to \overline{\mathbf{F}}$ is measurable, then $\|f\|_1 \le |E|\,\|f\|_\infty$.*

(b) $L^\infty(E) \subseteq L^1(E)$, *and if* $|E| > 0$ *then* $L^\infty(E) \neq L^1(E)$.

(c) *If* $f_n, f \in L^\infty(E)$ *and* $f_n \to f$ *in* L^∞*-norm, then* $f_n \to f$ *in* L^1*-norm.*

Proof. (a) By definition of the essential supremum, we have $|f| \le \|f\|_\infty$ a.e. It therefore follows from Exercise 4.3.6(a) that

$$\|f\|_1 \;=\; \int_E |f| \;\le\; \int_E \|f\|_\infty \;=\; |E|\,\|f\|_\infty.$$

(b) If $f \in L^\infty(E)$ then $\|f\|_\infty < \infty$, and therefore $\|f\|_1 < \infty$ by part (a). This shows that $L^\infty(E)$ is contained in $L^1(E)$, and Lemma 4.4.12 implies that the inclusion is proper if E has positive measure.

(c) If $\|f - f_n\|_\infty \to 0$, then $\|f - f_n\|_1 \to 0$ by part (a). $\square$

The following corollary of Lemma 4.4.13 follows immediately.

Corollary 4.4.14 (Uniform Convergence Theorem). *Let E be a measurable subset of $\mathbb{R}^d$ such that $|E| < \infty$. If $f_n, f \colon E \to \overline{\mathbf{F}}$ are bounded, measurable functions and $f_n \to f$ uniformly, then $f_n \to f$ in L^1-norm, and consequently $\int_E f_n \to \int_E f$.* $\diamondsuit$

Problems

4.4.15. Determine all values of $\alpha, \beta \in \mathbb{R}$ for which $f_\alpha(x) = x^\alpha\, \chi_{[0,1]}(x)$ or $g_\beta(x) = x^\beta\, \chi_{[1,\infty)}(x)$ belong to $L^1(\mathbb{R})$.

4.4.16. Prove the following statements.

(a) There exists a function $f \in C_0(\mathbb{R})$ that is not integrable on $\mathbb{R}$.

(b) There exists an unbounded continuous function that is integrable on $\mathbb{R}$ (such a function cannot be monotonically increasing!).

(c) If f is uniformly continuous and integrable on $\mathbb{R}$, then $\lim_{x\to\infty} f(x)$ exists and equals zero.

(d) If f is integrable on $\mathbb{R}$ and $a = \lim_{x\to\infty} f(x)$ exists, then $a = 0$.

4.4.17. (a) Suppose that $f, g \colon E \to [-\infty, \infty]$ are measurable functions, where E is a measurable subset of $\mathbb{R}^d$. Prove that if f is integrable and $f \le g$ a.e., then $g - f$ is measurable and $\int_E (g - f) = \int_E g \;-\; \int_E f$.

(b) Show that the Monotone Convergence Theorem and Fatou's Lemma remain valid if we replace the assumption $f_n \ge 0$ with $f_n \ge g$ a.e., where g is an integrable function on E. However, this can fail if g is not integrable.

4.4.18. Show by example that the hypothesis $|E| < \infty$ in Corollary 4.4.14 is necessary, even if we explicitly require each f_n to be integrable on E.

4.4.19. Prove that if $f \in L^1(\mathbb{R})$ is differentiable at $x = 0$ and $f(0) = 0$, then $\int_{-\infty}^{\infty} \frac{f(x)}{x}\,dx$ exists.

4.4.20. Prove that $L^1(\mathbb{R}^d)$ is closed under invertible linear changes of variable. That is, show that if $L\colon \mathbb{R}^d \to \mathbb{R}^d$ is an invertible linear transformation and $f \in L^1(\mathbb{R}^d)$, then $f \circ L \in L^1(\mathbb{R}^d)$.

4.4.21. Given a measurable set $E \subseteq \mathbb{R}^d$, prove the following statements.

(a) If $f \in L^1(E)$ and $g \in L^\infty(E)$, then $fg \in L^1(E)$.

(b) If $|E| > 0$, then $L^1(E)$ is not closed under products, i.e., there exist functions $f, g \in L^1(E)$ such that $fg \notin L^1(E)$.

(c) If f, g are measurable functions on E such that $|f|^2$ and $|g|^2$ each belong to $L^1(E)$, then $fg \in L^1(E)$.

4.4.22. Suppose that $f \in L^1[a,b]$ satisfies $\int_a^x f(t)\,dt = 0$ for all $x \in [a,b]$. Prove that $f = 0$ a.e.

Remark: If we are allowed to appeal to later results, this follows easily from the Lebesgue Differentiation Theorem (Theorem 5.5.7). The challenge is to find a solution that uses only the tools that have been developed so far.

4.4.23. (a) Let E be a measurable subset of $\mathbb{R}^d$, and assume that $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of integrable functions on E such that $\sup \|f_n\|_1 < \infty$ and $f_n \to f$ pointwise a.e. Prove that $f \in L^1(E)$ and

$$\lim_{n\to\infty} \left(\int_E |f_n| \;-\; \int_E |f - f_n| \right) \;=\; \int_E |f|. \tag{4.17}$$

Remark: This is sometimes referred to as the "missing term in Fatou's Lemma" [LL01] or "Lieb's version of Fatou's Lemma" [Str11].

(b) Exhibit integrable functions f_n such that $\sup \|f_n\|_1 = \infty$ and $f_n \to f$ pointwise a.e., but equation (4.17) fails.

4.4.24. Let E be a measurable subset of $\mathbb{R}^d$, and assume that f_n and f are integrable functions on E such that $f_n \to f$ pointwise a.e. Prove that

$$\lim_{n\to\infty} \|f - f_n\|_1 \;=\; 0 \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} \|f_n\|_1 \;=\; \|f\|_1.$$

4.5 The Dominated Convergence Theorem

Example 4.2.6 showed that pointwise convergence of functions need not imply convergence of the integrals of those functions. The Monotone Convergence Theorem tells us that if we have nonnegative functions f_n that increase pointwise to a function f, then the integral of f_n will converge to the integral of f.

However, this is a rather strong hypothesis that is not often satisfied in practice. In this section we will prove the *Dominated Convergence Theorem*, or *DCT*, which gives a different sufficient condition that implies convergence of the integrals of the f_n. We will use the Dominated Convergence Theorem to prove several results regarding approximation of integrable functions by functions that have various special properties.

4.5.1 The Dominated Convergence Theorem

The Dominated Convergence Theorem states that if f_n converges pointwise almost everywhere to f and we can find a *single, integrable function* g that simultaneously dominates *every* $|f_n|$, then $f_n \to f$ in L^1-norm, and therefore $\int_E f_n$ converges to $\int_E f$.

Theorem 4.5.1 (Dominated Convergence Theorem). *Let* $\{f_n\}_{n\in\mathbb{N}}$ *be a sequence of measurable functions (either extended real-valued or complex-valued) defined on a measurable set* $E \subseteq \mathbb{R}^d$. *If*

(a) $f(x) = \lim_{n\to\infty} f_n(x)$ *exists for a.e.* $x \in E$, *and*

(b) *there exists a single integrable function* g *such that for each* $n \in \mathbb{N}$ *we have* $|f_n(x)| \le g(x)$ *a.e.,*

then f_n *converges to* f *in* L^1*-norm, i.e.,*

$$\lim_{n\to\infty} \|f - f_n\|_1 \;=\; \lim_{n\to\infty} \int_E |f - f_n| \;=\; 0. \tag{4.18}$$

As a consequence,

$$\lim_{n\to\infty} \int_E f_n \;=\; \int_E f. \tag{4.19}$$

Proof. The hypotheses imply that g is integrable and nonnegative almost everywhere. Therefore

$$0 \;\le\; \int_E g \;=\; \int_E |g| \;<\; \infty.$$

Step 1. Suppose first that $f_n \ge 0$ a.e. for each n. In this case we can apply Fatou's Lemma to obtain

$$0 \;\le\; \int_E f \;=\; \int_E \liminf_{n\to\infty} f_n \;\le\; \liminf_{n\to\infty} \int_E f_n \;\le\; \int_E g \;<\; \infty. \tag{4.20}$$

We also have $g - f_n \ge 0$ a.e., so we can apply Fatou's Lemma to the functions $g - f_n$. Doing this, we obtain

$$\begin{aligned}
\int_E g - \int_E f &= \int_E (g-f) && (f \text{ and } g \text{ are integrable})\\
&= \int_E \liminf_{n\to\infty} (g-f_n) && (\text{since } f_n \to f \text{ a.e.})\\
&\le \liminf_{n\to\infty} \int_E (g-f_n) && (\text{Fatou's Lemma})\\
&= \liminf_{n\to\infty} \biggl(\int_E g - \int_E f_n\biggr) && (f_n \text{ and } g \text{ are integrable})\\
&= \int_E g - \limsup_{n\to\infty} \int_E f_n && (\text{properties of liminf}).
\end{aligned}$$

All of the integrals that appear in the preceding calculation are finite, so by rearranging we see that $\limsup_{n\to\infty} \int_E f_n \le \int_E f$. Combining this with equation (4.20) yields

$$\int_E f \le \liminf_{n\to\infty} \int_E f_n \le \limsup_{n\to\infty} \int_E f_n \le \int_E f.$$

Hence $\lim_{n\to\infty} \int_E f_n$ exists and equals $\int_E f$. This does not show that f_n converges to f in L^1-norm, but we will establish that in Step 2.

Step 2. Now assume that the f_n are arbitrary functions (either extended real-valued or complex-valued) that satisfy hypotheses (a) and (b). In this case, the functions $|f-f_n|$ are nonnegative a.e., converge pointwise a.e. to the zero function, and satisfy

$$|f-f_n| \le |f| + |f_n| \le 2g \text{ a.e.}$$

Since $2g$ is integrable, we can apply Step 1 to $|f-f_n|$, which gives us

$$\lim_{n\to\infty} \|f-f_n\|_1 = \lim_{n\to\infty} \int_E |f-f_n| = \int_E 0 = 0.$$

This proves that f_n converges to f in L^1-norm, so equation (4.18) holds. Applying Lemma 4.4.11, it follows that the integral of f_n converges to the integral of f, so equation (4.19) holds as well. □

The reader should consider why the Shrinking Boxes of Example 4.2.6 do not satisfy the hypotheses of the DCT, and contrast this with the Shrinking Triangles of Example 3.4.1, which do.

The following special case of the DCT for domains with finite measure is encountered often enough that it has its own name.

Corollary 4.5.2 (Bounded Convergence Theorem). *Let E be a measurable subset of $\mathbb{R}^d$ such that $|E| < \infty$. If $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of measurable functions on E such that $f_n \to f$ a.e. and there exists a single finite constant M such that $|f_n| \le M$ a.e. for every n, then $f_n \to f$ in L^1-norm.*

Proof. Since $|E| < \infty$, the constant function M is integrable. The result therefore follows by applying the DCT with $g(x) = M$. $\square$

Here is a sketch of an alternative proof of the Dominated Convergence Theorem. The spirit of this proof is quite similar to that of the proof we gave previously, but it is more concise and well worth working out.

Exercise 4.5.3. Assume that the hypotheses of the Dominated Convergence Theorem are satisfied. Observe that $2g - |f - f_n| \ge 0$ a.e. Write

$$2\int_E g \;=\; \int_E \liminf_{n\to\infty} \bigl(2g - |f - f_n|\bigr),$$

and apply Fatou's Lemma. $\diamondsuit$

4.5.2 First Applications of the DCT

To illustrate the use of the DCT, we prove a simple but important fact about approximation of integrable functions by functions that are zero outside of a bounded set.

Lemma 4.5.4. *Assume that $E \subseteq \mathbb{R}^d$ is measurable and $f\colon E \to \overline{\mathbf{F}}$ is integrable. For each $n \in \mathbb{N}$, set*

$$f_n(x) \;=\; f(x)\,\chi_{B_n(0)}(x) \;=\; \begin{cases} f(x), & \text{if } x \in E \text{ and } \|x\| < n,\\ 0, & \text{if } x \in E \text{ and } \|x\| \ge n. \end{cases}$$

Then $f_n \to f$ in L^1-norm.

Proof. Note that $f_n \to f$ pointwise and $|f_n| \le |f|$ for every n. Since $|f|$ is integrable, the DCT implies that $\|f - f_n\|_1 \to 0$. $\square$

Part (a) of the next exercise applies the DCT in a similar but slightly different way to show that every integrable function can be well-approximated in L^1-norm by bounded functions. The result contained in part (b) of this exercise is much more important than it may appear at first glance. In particular, we will make use of part (b) in the proofs of Theorem 6.3.1 and Lemma 6.4.1.

Exercise 4.5.5. Let $E \subseteq \mathbb{R}^d$ be measurable, and assume that $f\colon E \to \overline{\mathbf{F}}$ is integrable.

(a) Set $E_n = \bigl\{|f| \le n\bigr\}$, and show that $f \cdot \chi_{E_n}$ converges to f in L^1-norm, i.e., $\|f - f \cdot \chi_{E_n}\|_1 \to 0$ as $n \to \infty$.

(b) Given $\varepsilon > 0$, show that there exists a constant $\delta > 0$ such that for every measurable set $A \subseteq E$ we have

$$|A| < \delta \quad \Longrightarrow \quad \int_A |f| < \varepsilon. \qquad \diamondsuit \tag{4.21}$$

4.5.3 Approximation by Continuous Functions

Now we focus on functions whose domain is all of $\mathbb{R}^d$. How well can we approximate an arbitrary integrable function on $\mathbb{R}^d$ by a *continuous function*, or perhaps by a *compactly supported continuous function*? That is, given an integrable function f on $\mathbb{R}^d$, can we find an element of

$$C_c(\mathbb{R}^d) \;=\; \bigl\{f \in C(\mathbb{R}^d) : \operatorname{supp}(f) \text{ is compact}\bigr\}$$

that lies as close as we like to f, or is there a limit to how closely we can approximate f? We measure "closeness" here in terms of the L^1-norm, i.e., we wish to know if it is true that for every $\varepsilon > 0$ there exists a function $\theta \in C_c(\mathbb{R}^d)$ such that $\|f - \theta\|_1 < \varepsilon$.

We will show that we can approximate an integrable function as closely as we like by an element of $C_c(\mathbb{R}^d)$ when we measure error by the L^1-norm. A key tool in the proof is *Urysohn's Lemma*, which gives us a way of constructing a continuous function that "separates" disjoint closed sets. We give an exercise regarding the distance from a point to a set in a metric space, and then prove Urysohn's Lemma.

Exercise 4.5.6. Let X be a metric space. Define the *distance* from a point $x \in X$ to a nonempty set $A \subseteq X$ to be $\operatorname{dist}(x, A) = \inf\bigl\{\mathrm{d}(x, y) : y \in A\bigr\}$, where $\mathrm{d}(\cdot, \cdot)$ is the metric on X. Prove the following statements.

(a) If A is closed, then $x \in A$ if and only if $\operatorname{dist}(x, A) = 0$.

(b) $\operatorname{dist}(x, A) \le \mathrm{d}(x, y) + \operatorname{dist}(y, A)$ for all $x, y \in X$.

(c) $|\operatorname{dist}(x, A) - \operatorname{dist}(y, A)| \le \mathrm{d}(x, y)$ for all $x, y \in X$.

(d) The function $f(x) = \operatorname{dist}(x, A)$ is uniformly continuous on X. $\diamondsuit$

Theorem 4.5.7 (Urysohn's Lemma). *If E and F are disjoint closed subsets of a metric space X, then there exists a continuous function $\theta\colon X \to \mathbb{R}$ such that $0 \le \theta \le 1$ on X, $\theta = 0$ on E, and $\theta = 1$ on F.*

Proof. If $E = \varnothing$ then we just take $\theta = 1$, and likewise if $F = \varnothing$ then we can take $\theta = 0$. Therefore we assume that E and F are both nonempty. Applying Exercise 4.5.6, it follows that the function

$$\theta(x) \;=\; \frac{\operatorname{dist}(x, E)}{\operatorname{dist}(x, E) + \operatorname{dist}(x, F)}, \qquad \text{for } x \in X,$$

has the required properties. $\square$

Now we prove that we can approximate any integrable function by a continuous function that has compact support.

Theorem 4.5.8. *If $f \in L^1(\mathbb{R}^d)$ and $\varepsilon > 0$, then there exists a function $\theta \in C_c(\mathbb{R}^d)$ such that $\|f-\theta\|_1 < \varepsilon$.*

Proof. Step 1. First we consider a characteristic function $f = \chi_E$, where E is a bounded subset of $\mathbb{R}^d$ (we assume that E is bounded so that χ_E is integrable). If we fix $\varepsilon > 0$, then Theorem 2.1.27 implies that there exists a bounded open set $U \supseteq E$ such that $|U \backslash E| < \varepsilon$. By Problem 2.2.43, there also exists a compact set $K \subseteq E$ such that $|E \backslash K| < \varepsilon$. Applying Urysohn's Lemma to the disjoint closed sets K and $\mathbb{R}^d \backslash U$, we can find a continuous function $\theta \colon \mathbb{R}^d \to \mathbb{R}$ that satisfies

- $0 \le \theta \le 1$ everywhere on $\mathbb{R}^d$,
- $\theta = 1$ on K, and
- $\theta = 0$ on $\mathbb{R}^d \backslash U$.

This function θ belongs to $C_c(\mathbb{R}^d)$, and

$$\|\chi_E - \theta\|_1 \;=\; \int_{\mathbb{R}^d} |\chi_E - \theta| \;=\; \int_{U \backslash K} |\chi_E - \theta| \;\le\; |U \backslash K| \;<\; 2\varepsilon.$$

Hence χ_E can be approximated as closely as we like in L^1-norm by an element of $C_c(\mathbb{R}^d)$.

Step 2. Let ϕ be a simple function of the form

$$\phi \;=\; \sum_{k=1}^{N} a_k \chi_{E_k},$$

where each set E_k is bounded and each scalar a_k is nonzero. By Step 1, there exist functions $\theta_k \in C_c(\mathbb{R}^d)$ such that

$$\big\|\chi_{E_k} - \theta_k\big\|_1 \;<\; \frac{\varepsilon}{|a_k|\,N}, \qquad \text{for } k = 1, \dots, N.$$

Then the function $\theta = \sum_{k=1}^N a_k \theta_k$ belongs to $C_c(\mathbb{R}^d)$, and by applying the Triangle Inequality we see that

$$\|\phi - \theta\|_1 \;=\; \left\|\sum_{k=1}^{N} a_k \chi_{E_k} \;-\; \sum_{k=1}^{N} a_k \theta_k\right\|_1 \;\le\; \sum_{k=1}^{N} |a_k|\, \|\chi_{E_k} - \theta_k\|_1 \;<\; \varepsilon.$$

Step 3. Let f be an arbitrary element of $L^1(\mathbb{R}^d)$. By Lemma 4.5.4, there exists a function g that is zero outside of some bounded set and satisfies

$$\|f - g\|_1 \;<\; \varepsilon.$$

By Corollary 3.2.15, there exist simple functions ϕ_n that converge pointwise to g and satisfy $|\phi_n| \le |g|$ a.e. Since g is integrable, the Dominated Convergence Theorem implies that $\|g-\phi_n\|_1 \to 0$ as $n \to \infty$. Therefore, if we choose n large enough then we will have

$$\|g-\phi_n\|_1 \; < \; \varepsilon.$$

Applying Step 2, there exists a function $\theta \in C_c(\mathbb{R}^d)$ such that

$$\|\phi_n-\theta\|_1 \; < \; \varepsilon.$$

Therefore, by the Triangle Inequality,

$$\|f-\theta\|_1 \; \le \; \|f-g\|_1 \, + \, \|g-\phi_n\|_1 \, + \, \|\phi_n-\theta\|_1 \; < \; 3\varepsilon. \qquad \square$$

By taking $\varepsilon = 1/n$ in Theorem 4.5.8, we see that if f is any integrable function on $\mathbb{R}^d$, then there exist functions $\theta_n \in C_c(\mathbb{R}^d)$ such that

$$\lim_{n\to\infty} \|f-\theta_n\|_1 \; = \; 0.$$

That is, every function in $L^1(\mathbb{R}^d)$ is an L^1-norm limit of functions from $C_c(\mathbb{R}^d)$. Using the terminology introduced in Section 1.1.2, this says that $C_c(\mathbb{R}^d)$ is a *dense subset* of $L^1(\mathbb{R}^d)$. This also shows that $C_c(\mathbb{R}^d)$ is not a closed subset of $L^1(\mathbb{R}^d)$ with respect to the L^1-norm, because a sequence of elements of $C_c(\mathbb{R}^d)$ can converge in L^1-norm to a function that does not belong to $C_c(\mathbb{R}^d)$.

An analogous situation is the set of rationals $\mathbb{Q}$ in the real line $\mathbb{R}$. Every real number can be written as a limit of rational numbers, so $\mathbb{Q}$ is a dense subset of $\mathbb{R}$, but $\mathbb{Q}$ is not closed because a limit of rational numbers can be irrational. However, there is an interesting difference between $\mathbb{Q}$ and $C_c(\mathbb{R}^d)$. While $\mathbb{Q}$ is a proper dense *subset* of $\mathbb{R}$, it is not a dense *subspace* (because it is not closed under multiplication by arbitrary real scalars). In contrast, $C_c(\mathbb{R}^d)$ is a dense *subspace* of $L^1(\mathbb{R}^d)$. Only an infinite-dimensional normed space can contain a proper dense subspace, because proper subspaces of finite-dimensional normed spaces are closed (for one proof of this, see [Heil11, Thm. 1.22]).

The following important exercise is an application of Theorem 4.5.8. The "easy" way to solve this is to first prove that equation (4.22) holds for functions $\theta \in C_c(\mathbb{R}^d)$, and then extend to arbitrary functions $f \in L^1(\mathbb{R}^d)$ by approximating by continuous functions (keep in mind that every function in $C_c(\mathbb{R}^d)$ is compactly supported and therefore is uniformly continuous).

Exercise 4.5.9 (Strong Continuity of Translation). Given $f \in L^1(\mathbb{R}^d)$, let $T_af(x) = f(x-a)$ denote the translation of f by $a \in \mathbb{R}^d$. Prove that $T_af \to f$ in L^1-norm as $a \to 0$, i.e.,

$$\lim_{a\to 0} \|T_af-f\|_1 \; = \; 0. \qquad \diamondsuit \tag{4.22}$$

We often summarize equation (4.22) by saying that *translation is strongly continuous on* $L^1(\mathbb{R}^d)$. In contrast, translation is not strongly continuous on $L^\infty(\mathbb{R}^d)$. For example, if we set $\chi = \chi_{[0,1]}$, then for every $a \neq 0$ we have $\|T_a\chi - \chi\|_\infty = 1$ (see the illustration in Figure 7.6).

Remark 4.5.10. Exercise 4.5.9 does *not* imply that $T_a f \to f$ pointwise or even pointwise a.e. as $a \to 0$. For example, if $f = \chi_E$ where $E = [0,1] \setminus \mathbb{Q}$ is the set of irrationals in $[0,1]$ then there is no point $x \in [0,1]$ where $T_a f(x) \to f(x)$ as $a \to 0$. $\diamondsuit$

4.5.4 Approximation by Really Simple Functions

Corollary 3.2.15 tells us that if f is a measurable function on a set E, then there exist simple functions ϕ_n that converge pointwise to f and satisfy $|\phi_n| \le |f|$ for every n. If it so happens that f is integrable, then we can apply the Dominated Convergence Theorem and conclude that ϕ_n converges to f in L^1-norm as well as pointwise. Unfortunately, although a simple function takes only finitely many values, the sets on which those values are taken can be arbitrary measurable sets. Sometimes we need to know that we can approximate by actual "step functions," i.e., functions that are finite linear combinations of characteristic functions of *intervals*. These functions are sometimes called the *really simple functions* on $\mathbb{R}$ (for example, see the terminology in [LL01, Sec. 1.17]). Here is the precise definition.

Definition 4.5.11 (Really Simple Function). A *really simple function* on $\mathbb{R}$ is a measurable function ϕ of the form

$$\phi = \sum_{k=1}^{N} c_k \, \chi_{[a_k, b_k)}, \tag{4.23}$$

where $N \in \mathbb{N}$, $a_k < b_k$ are real numbers, and c_k is a scalar. $\diamondsuit$

We use half-open intervals $[a_k, b_k)$ in Definition 4.5.11 for convenience. Other types of finite intervals can usually be substituted if minor adjustments are made to the proofs.

We saw in Theorem 4.5.8 that we can approximate an integrable function by a continuous function. By approximating a continuous function with a step function, we obtain the following result.

Theorem 4.5.12. *If $f \in L^1(\mathbb{R})$, then for each $\varepsilon > 0$ there exists a really simple function ϕ such that $\|f - \phi\|_1 < \varepsilon$.*

Proof. By Theorem 4.5.8, there exists some function $\theta \in C_c(\mathbb{R})$ such that $\|f - \theta\|_1 < \varepsilon/2$. Since θ is compactly supported, we can choose R large enough that $\theta(x) = 0$ for $|x| \geq R$. Then, since θ is uniformly continuous, there exists some $0 < \delta < 1$ such that

$$|x - y| < \delta \quad \Longrightarrow \quad |\theta(x) - \theta(y)| < \frac{\varepsilon}{4R+4}.$$

The really simple function

$$\phi(x) = \sum_{k \in \mathbb{Z}} \theta(k\delta)\, \chi_{[k\delta,(k+1)\delta)}$$

is identically zero outside of $[-R-1, R+1]$ and satisfies

$$|\theta(x) - \phi(x)| < \frac{\varepsilon}{4R+4}, \qquad \text{for } x \in \mathbb{R}.$$

Therefore

$$\begin{aligned} \|f - \phi\|_1 &\leq \|f - \theta\|_1 + \|\theta - \phi\|_1 \\ &= \|f - \theta\|_1 + \int_{-R-1}^{R+1} |\theta(x) - \phi(x)|\, dx \\ &\leq \frac{\varepsilon}{2} + (2R+2)\, \frac{\varepsilon}{4R+4} = \varepsilon. \qquad \square \end{aligned}$$

Using the terminology of Section 1.1.2, Theorem 4.5.12 says that the set of really simple functions is a dense subspace of $L^1(\mathbb{R})$.

4.5.5 Relation to the Riemann Integral

A measurable bounded function f on a finite interval $[a, b]$ is necessarily integrable, so its Lebesgue integral $\int_a^b f(x)\, dx$ exists and is a finite scalar. Some bounded functions on $[a, b]$ are also Riemann integrable (for example, this is true for all continuous functions). However, there are functions that are Lebesgue integrable but not Riemann integrable. One example is the Dirichlet function $\chi_{\mathbb{Q}}$, the characteristic function of the rational numbers. Even though the Riemann integral of $\chi_{\mathbb{Q}}$ does not exist, its Lebesgue integral does; in fact, $\int_a^b \chi_{\mathbb{Q}} = 0$ since $\chi_{\mathbb{Q}} = 0$ a.e.

It is important to know whether these two types of integrals coincide when they exist. For example, we need to know whether the formulas that we learned in undergraduate calculus still hold if we replace Riemann integrals by Lebesgue integrals. The following theorem shows if a bounded function is Riemann integrable on a finite interval, then it is also Lebesgue integrable on that interval and the two integrals coincide. Moreover, this theorem provides a complete characterization of the functions that are Riemann integrable—they are precisely those functions that are continuous a.e.

Theorem 4.5.13. *Let $f\colon [a,b] \to \mathbb{C}$ be a bounded function whose domain is a finite closed interval $[a,b]$.*

(a) *If f is Riemann integrable on $[a,b]$, then it is Lebesgue integrable on $[a,b]$, and its Riemann integral equals its Lebesgue integral $\int_a^b f$.*

(b) *f is Riemann integrable on $[a,b]$ if and only if f is continuous at almost every point of $[a,b]$.*

Proof. Since f is bounded, it is finite at every point. By considering the real and imaginary parts of f separately, it suffices to consider real-valued functions. Therefore we assume throughout this proof that f is real-valued.

We make some observations and lay out some notation before beginning the main part of the proof. Given a partition

$$\Gamma = \{a = x_0 < x_1 < \cdots < x_n = b\},$$

set $|\Gamma| = \max\{x_j - x_{j-1}\}$ (this is called the *mesh size* of Γ), and define

$$m_j = \inf_{x\in[x_{j-1},x_j]} f(x) \qquad \text{and} \qquad M_j = \sup_{x\in[x_{j-1},x_j]} f(x),$$

for $j = 1, \dots, n$. The numbers

$$L_\Gamma = \sum_{j=1}^n m_j\,(x_j - x_{j-1}) \qquad \text{and} \qquad U_\Gamma = \sum_{j=1}^n M_j\,(x_j - x_{j-1}),$$

are called *lower* and *upper Riemann sums* for f, respectively. Further,

$$\phi_\Gamma = \sum_{j=1}^n m_j\,\chi_{[x_{j-1},x_j)} \qquad \text{and} \qquad \psi_\Gamma = \sum_{j=1}^n M_j\,\chi_{[x_{j-1},x_j)},$$

are simple functions that satisfy

$$\phi_\Gamma \le f \le \psi_\Gamma \tag{4.24}$$

on the interval $[a,b)$. By setting $\phi_\Gamma(b) = f(b) = \psi_\Gamma(b)$, we can assume that ϕ_Γ and ψ_Γ are simple functions such that equation (4.24) holds on all of $[a,b]$. Since ϕ_Γ and ψ_Γ are simple, their *Lebesgue integrals* are precisely

$$\int_a^b \phi_\Gamma = L_\Gamma \qquad \text{and} \qquad \int_a^b \psi_\Gamma = U_\Gamma.$$

For ease of notation, given a sequence of partitions $\{\Gamma_k\}_{k\in\mathbb{N}}$, we will use the shorthands

$$L_k = L_{\Gamma_k}, \qquad U_k = U_{\Gamma_k}, \qquad \phi_k = \phi_{\Gamma_k}, \qquad \psi_k = \psi_{\Gamma_k}.$$

Now we proceed to establish the validity of statements (a) and (b) of the theorem.

(a) Assume that f is a real-valued Riemann integrable function, and let I denote the value of the *Riemann integral* of f over $[a,b]$. Let $\{\Gamma_k\}_{k\in\mathbb{N}}$ be any sequence of partitions of $[a,b]$ such that:

- Γ_{k+1} is a refinement of Γ_k for each $k \in \mathbb{N}$, and
- $|\Gamma_k| \to 0$ as $k \to \infty$, where $|\Gamma_k|$ is the mesh size of the partition Γ_k.

Then it follows from the definition of the Riemann integral that $L_k \to I$ and $U_k \to I$ as $k \to \infty$.

We have not yet shown that f is measurable, so we do not yet know whether its Lebesgue integral exists. However, since each partition Γ_{k+1} is a refinement of the preceding partition Γ_k, we do know that $\{\phi_k\}_{k\in\mathbb{N}}$ is a monotone increasing sequence of simple functions, and similarly $\{\psi_k\}_{k\in\mathbb{N}}$ is a monotone decreasing sequence of simple functions. Therefore the functions

$$\phi(x) = \lim_{k\to\infty} \phi_k(x) \qquad \text{and} \qquad \psi(x) = \lim_{k\to\infty} \psi_k(x),$$

are measurable. Further, if we set $M = \sup_{x\in[a,b]} |f(x)|$, then M is finite and $|\phi_k|, |\psi_k| \le M$ for every k. Applying the *Bounded Convergence Theorem* (Corollary 4.5.2), it follows that the *Lebesgue* integrals of ϕ and ψ satisfy

$$\int_a^b \phi = \lim_{k\to\infty} \int_a^b \phi_k = \lim_{k\to\infty} L_k = I = \lim_{k\to\infty} U_k = \lim_{k\to\infty} \int_a^b \psi_k = \int_a^b \psi.$$

Therefore, the Lebesgue integral of $\psi-\phi$ is $\int_a^b (\psi-\phi) = 0$. Since $\psi-\phi \ge 0$, it follows that $\psi-\phi = 0$ a.e. But $\phi \le f \le \psi$, so this implies that $\phi = f = \psi$ a.e. Consequently, f is measurable and its Lebesgue integral is $\int_a^b f = I$.

(b) Suppose that f is real-valued and Riemann integrable on $[a,b]$. Using the same partitions and notation from part (a), let E be the set of all points $x \in [a,b]$ such that $\phi(x) = f(x) = \psi(x)$. The proof of part (a) shows that $Z = [a,b]\backslash E$ has measure zero. Since each partition Γ_k contains finitely many partitioning points, the set S that contains every partitioning point of every Γ_k is countable and therefore also has measure zero. Suppose that f is discontinuous at a point $x \notin Z \cup S$. Then there exists some $\varepsilon > 0$ such that for every $\delta > 0$ there is a point $t \in (x-\delta, x+\delta)$ such that $|f(x) - f(t)| \ge \varepsilon$. It follows from this that

$$\psi_k(x) - \phi_k(x) \ge \varepsilon \quad \text{for every } k \in \mathbb{N}.$$

However, since $x \in E$, this implies that

$$\varepsilon \le \lim_{k\to\infty} \bigl(\psi_k(x) - \phi_k(x)\bigr) = \psi(x) - \phi(x) = 0,$$

which is a contradiction. Therefore f must be continuous at every point $x \notin Z \cup S$, so f is continuous a.e.

For the converse, suppose that f is continuous a.e. Let $\{\Gamma_k\}_{k\in\mathbb{N}}$ be *any* sequence of partitions of $[a,b]$ such that $|\Gamma_k| \to 0$. We are no longer assuming that Γ_{k+1} is a refinement of Γ_k, so the sequence $\{\phi_k\}_{k\in\mathbb{N}}$ need not be monotone increasing, and $\{\psi_k\}_{k\in\mathbb{N}}$ need not be monotone decreasing. On the other hand, the fact that f is continuous almost everywhere implies that $\phi_k(x) \to f(x)$ at each point of continuity of f (compare Exercise 3.2.9). Thus $\phi_k \to f$ a.e., and similarly $\psi_k \to f$ a.e. It therefore follows from the Bounded Convergence Theorem that

$$\lim_{k\to\infty} L_k \;=\; \lim_{k\to\infty} \int_a^b \phi_k \;=\; \int_a^b f \;=\; \lim_{k\to\infty} \int_a^b \psi_k \;=\; \lim_{k\to\infty} U_k,$$

where the integrals on the preceding line are all Lebesgue integrals. This tells us that the upper and lower Riemann sums for f converge to the number $\int_a^b f$. Since we have shown that this is true for every sequence of partitions whose mesh size converges to zero, we conclude that f is Riemann integrable and its Riemann integral is $\int_a^b f$. $\square$

As we have noted before, the two statements "f is continuous a.e." and "f equals a continuous function a.e." are distinct. The first means that $\lim_{y\to x} f(y) = f(x)$ for almost every x, while the second means that there exists a continuous function g such that $f(x) = g(x)$ for almost every x. For example, the characteristic function $\chi_{\mathbb{Q}}$ equals a continuous function a.e. but it is not continuous at any point, while $\chi_{[0,1]}$ is continuous a.e. on $\mathbb{R}$ but there is no continuous function that equals it almost everywhere.

Remark 4.5.14. Somewhat more care is required when dealing with *improper* Riemann integrals. For example, Problem 4.6.19 shows that the improper Riemann integral of $f(x) = \frac{\sin x}{x}$ over $[0,\infty)$ exists and has the value $\frac{\pi}{2}$. However, the integrals of the positive and negative parts of f are $\int_0^\infty f^+ = \infty$ and $\int_0^\infty f^- = \infty$, so f is not integrable on $[0,\infty)$ and the Lebesgue integral of f on $[0,\infty)$ does not even exist (see Exercise 4.3.2). In essence, improper Riemann integrals may exist because of "fortunate cancellations," while the existence of the Lebesgue integral requires "absolute convergence." $\diamondsuit$

Problems

4.5.15. Evaluate the following limits.

(a) $\displaystyle \lim_{n\to\infty} \int_1^2 \frac{n^2 \sin(x/n)}{1+nx^2}\, dx.$ (b) $\displaystyle \lim_{n\to\infty} \int_0^\infty \frac{\sin x^n}{x^n}\, dx.$

4.5.16. Let f be an integrable function on a measurable set $E \subseteq \mathbb{R}^d$. Prove the following statements.

(a) $f = 0$ a.e. if and only if $\int_A f = 0$ for every measurable set $A \subseteq E$.

(b) If $\varepsilon > 0$, then there is a measurable set $A \subseteq E$ such that f is bounded on A and $\int_{E \setminus A} |f| < \varepsilon$.

4.5.17. Show that if $f \in L^1(\mathbb{R})$, then its indefinite integral $F(x) = \int_0^x f(t)\,dt$ is uniformly continuous on $\mathbb{R}$.

4.5.18. Prove the *Dominated Convergence Theorem for Series*: If scalars $a_{kn} \in \mathbb{C}$ are such that $\lim_{n\to\infty} a_{kn} = b_k$ exists for each k and

$$\sum_{k=1}^{\infty} \Bigl(\sup_{n\in\mathbb{N}} |a_{kn}|\Bigr) \;<\; \infty,$$

then

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} |b_k - a_{kn}| \;=\; 0 \qquad \text{and} \qquad \lim_{n\to\infty} \sum_{k=1}^{\infty} a_{kn} \;=\; \sum_{k=1}^{\infty} b_k.$$

4.5.19. Assume that f is a nonnegative function on $[a,b]$, and f is bounded and Riemann integrable on $[a+\delta, b]$ for each $\delta > 0$. Let I_δ denote the Riemann integral of f on $[a+\delta, b]$, and suppose that $I = \lim_{\delta\to 0} I_\delta$ exists and is finite. Prove that f is integrable on $[a,b]$ and I equals the Lebesgue integral $\int_a^b f$.

4.5.20. Show by example that the hypothesis $|E| < \infty$ is necessary in the Bounded Convergence Theorem (Corollary 4.5.2), even if we explicitly require each function f_n to be integrable on E.

4.5.21. Use Egorov's Theorem to prove the Bounded Convergence Theorem.

4.5.22. Show that the conclusion of the Dominated Convergence Theorem continues to hold if we replace the hypothesis $f_n \to f$ a.e. with $f_n \xrightarrow{\mathrm{m}} f$.

4.5.23. Let $f \colon E \to [0,\infty]$ be an integrable function defined on a measurable set $E \subseteq \mathbb{R}^d$, and suppose that $I = \int_E f > 0$. Given $0 \le t \le I$, prove that there exists a measurable set $A \subseteq E$ such that $\int_A f = t$. Does anything change if $f \colon E \to [-\infty,\infty]$ is integrable?

4.5.24. Let E be a measurable subset of $\mathbb{R}^d$, and suppose that f is integrable and nonnegative on E. Prove that

$$\lim_{n\to\infty} \int_E n \ln\!\Bigl(1 + \frac{f(x)}{n}\Bigr)\,dx \;=\; \int_E f(x)\,dx.$$

4.5.25. Assume $K \subset \mathbb{R}^d$ is compact, and let $g(x) = \max\{1 - \operatorname{dist}(x,K),\, 0\}$. Prove that

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} g(x)^n\, dx \;=\; |K|.$$

4.5.26. Let E be a measurable subset of $\mathbb{R}^d$ such that $|E| < \infty$. Prove that $\lim_{h\to 0} |E \cap (E+h)| = |E|$.

4.5.27. This problem will establish a *Generalized Dominated Convergence Theorem*. Let E be a measurable subset of $\mathbb{R}^d$. Assume that:

(a) $f_n,\, g_n,\, f,\, g \in L^1(E)$,

(b) $f_n \to f$ pointwise a.e.,

(c) $g_n \to g$ pointwise a.e.,

(d) $|f_n| \le g_n$ a.e., and

(e) $\int_E g_n \to \int g$.

Prove that $\int_E f_n \to \int_E f$ and $\|f - f_n\|_1 \to 0$.

4.5.28. Compute $\displaystyle\lim_{n\to\infty} \int_0^\infty \left(1 + \tfrac{x}{n}\right)^{-n} \sin \tfrac{x}{n}\, dx.$

4.5.29. Suppose that f is a bounded, measurable function on $[0,1]$ such that $\int_0^1 x^n f(x)\, dx = 0$ for $n = 0, 1, 2, \dots$. Show that $f(x) = 0$ a.e.

4.5.30. Prove the following continuous-parameter version of the DCT. Let E be a measurable subset of $\mathbb{R}^d$, and fix $c > 0$. Assume that:

(a) f_t is a measurable function on E for each real number $t \in (0,c)$,

(b) $f_t \to f$ pointwise a.e. as $t \to 0^+$, and

(c) there exists a single function $g \in L^1(E)$ such that $|f_t| \le g$ a.e. for each $t \in (0,c)$.

Prove that $\lim_{t\to 0^+} \|f - f_t\|_1 = 0$.

4.5.31. (a) Given $f \in L^1(\mathbb{R})$, define

$$F(\omega) \;=\; \int_{-\infty}^{\infty} f(x)\, \sin \omega x\, dx, \qquad \omega \in \mathbb{R}.$$

Prove that F is continuous at $\omega = 0$, and if $\int_{-\infty}^{\infty} |x f(x)|\, dx < \infty$ then F is differentiable at $\omega = 0$.

(b) Given $f \in L^1(\mathbb{R})$, define

$$G(\omega) \;=\; \int_{-\infty}^{\infty} f(x)\, \frac{\sin \omega x}{x}\, dx, \qquad \omega \in \mathbb{R}.$$

Prove that G is differentiable at $\omega = 0$.

(c) Show that parts (a) and (b) remain valid if ω is any point in $\mathbb{R}$.

4.5.32. Assume that $f\colon [0,1]^2 \to \mathbb{C}$ satisfies the following two conditions:

(i) for each fixed $x \in [0,1]$, $f(x,y)$ is an integrable function of y, and

(ii) $\frac{\partial f}{\partial x}(x,y)$ exists at all points and is bounded on $[0,1]^2$.

Prove that $\frac{\partial f}{\partial x}(x,y)$ is a measurable function of y for each $x \in [0,1]$, and

$$\frac{d}{dx} \int_0^1 f(x,y)\,dy \;=\; \int_0^1 \frac{\partial f}{\partial x}(x,y)\,dy.$$

4.5.33. Let X be a set, and let Σ be a σ-algebra of subsets of X (see Definition 2.2.14). A function $\nu\colon \Sigma \to [-\infty,\infty]$ is a *signed measure* on (X,Σ) if: $\nu(\varnothing) = 0$, $\nu(E)$ takes at most one of the values ∞ and $-\infty$, and ν is countably additive, i.e., if $E_1, E_2, \dots$ are countably many disjoint sets in Σ, then

$$\nu\biggl(\bigcup_k E_k\biggr) \;=\; \sum_k \nu(E_k).$$

We say that ν is a *positive measure* if $\nu(E) \ge 0$ for every $E \in \Sigma$.

(a) Let $\mathcal{P}(\mathbb{R}^d)$ be the set of all subsets of $\mathbb{R}^d$. *Counting measure* on $(\mathbb{R}^d, \mathcal{P}(\mathbb{R}^d))$ is the function $\mu\colon \mathcal{P}(\mathbb{R}^d) \to [0,\infty]$ defined by

$$\mu(E) \;=\; \begin{cases} \#E, & \text{if } E \text{ is finite},\\ \infty, & \text{if } E \text{ is infinite}, \end{cases}$$

where $\#E$ is the number of elements of E. Prove that μ is a positive measure on $(\mathbb{R}^d, \mathcal{P}(\mathbb{R}^d))$.

(b) The δ *measure* or *Dirac measure* on $(\mathbb{R}^d, \mathcal{P}(\mathbb{R}^d))$ is the function $\delta\colon \mathcal{P}(\mathbb{R}^d) \to [0,\infty]$ defined by

$$\delta(E) \;=\; \begin{cases} 1, & \text{if } 0 \in E,\\ 0, & \text{if } 0 \notin E. \end{cases}$$

Prove that δ is a positive measure on $(\mathbb{R}^d, \mathcal{P}(\mathbb{R}^d))$.

(c) Let $\mathcal{L}$ be the set of all Lebesgue measurable subsets of $\mathbb{R}^d$, and let $f\colon \mathbb{R}^d \to [-\infty,\infty]$ be a measurable function such that at least one of $\int f^+$ or $\int f^-$ is finite. For each measurable set $E \subseteq \mathbb{R}^d$, define $\nu_f(E) = \int_E f(t)\,dt$. Prove that ν_f is a signed measure on $(\mathbb{R}^d, \mathcal{L})$.

(d) We say that a signed measure ν on $(\mathbb{R}^d, \mathcal{L})$ is *absolutely continuous with respect to Lebesgue measure* if for each measurable set $A \subseteq \mathbb{R}^d$ we have

$$|A| \,=\, 0 \quad\Longrightarrow\quad \nu(A) \,=\, 0.$$

Restricting μ and δ to the σ-algebra $\mathcal{L}$, determine whether the measures μ, δ, and ν_f are absolutely continuous with respect to Lebesgue measure.

4.6 Repeated Integration

Let $E \subseteq \mathbb{R}^m$ and $F \subseteq \mathbb{R}^n$ be measurable sets. If f is a measurable function on $E \times F$ then there are at least three natural integrals of f over $E \times F$ whose existence we can consider. First, there is the integral of f over the set $E \times F \subseteq \mathbb{R}^{m+n}$ with respect to Lebesgue measure on $\mathbb{R}^{m+n}$. We will formally write this as the *double integral*

$$\iint_{E\times F} f \;=\; \iint_{E\times F} f(x,y)\,(dx\,dy).$$

This double integral is simply the Lebesgue integral of f on $E \times F$. The double integral may or may not actually exist, but it is one possible way that we can attempt to integrate f.

A second possibility is to perform an *iterated integration* where for each fixed y we integrate $f(x,y)$ as a function of x, and then integrate the result in y. This gives us the *iterated integral*

$$\int_F \biggl(\int_E f(x,y)\,dx\biggr)\,dy.$$

Again, this iterated integral may or may not exist.

The third possibility is the iterated integral performed in the opposite order, which is

$$\int_E \biggl(\int_F f(x,y)\,dy\biggr)\,dx.$$

In general the three integrals given above need not be equal (for some specific examples, see Problems 4.6.12–4.6.14). Our goal in this section is to derive the theorems of Fubini and Tonelli, which give sufficient conditions under which these three integrals all exist and are equal.

4.6.1 Fubini's Theorem

We begin by giving the statement of Fubini's Theorem. According to this result, the double integral and the two iterated integrals are all equal if f is an *integrable* function on the Cartesian product $E \times F$.

Theorem 4.6.1 (Fubini's Theorem). *Let E be a measurable subset of $\mathbb{R}^m$ and let F be a measurable subset of $\mathbb{R}^n$. If $f\colon E \times F \to \overline{\mathbf{F}}$ is integrable on $E \times F$, then the following statements hold.*

(a) *$f_x(y) = f(x,y)$ is measurable and integrable on F for almost every $x \in E$.*

(b) *$f^y(x) = f(x,y)$ is measurable and integrable on E for almost every $y \in F$.*

(c) $g(x) = \int_F f_x(y)\,dy$ *is measurable and integrable on* E.

(d) $h(y) = \int_E f^y(x)\,dx$ *is measurable and integrable on* F.

(e) *The following three integrals exist and are finite (i.e., they are real or complex scalars), and they are equal as indicated:*

$$\iint_{E\times F} f(x,y)\,(dx\,dy) \;=\; \int_F \biggl(\int_E f(x,y)\,dx\biggr)\,dy$$
$$= \int_E \biggl(\int_F f(x,y)\,dy\biggr)\,dx. \qquad \diamondsuit$$

Before beginning the proof of Fubini's Theorem, we point out that statements (a) and (b) of the theorem are not trivial. If f is measurable on $E \times F$ and we fix $x \in E$, then $f_x(y) = f(x,y)$ *need not* be a measurable function on F! For example, let Z be a subset of $\mathbb{R}$ that has measure zero and let N be a nonmeasurable subset of $\mathbb{R}$. Then $Z \times N$ has measure zero as a subset of $\mathbb{R}^2$, so

$$f(x,y) \;=\; \chi_{Z\times N}(x,y) \;=\; \chi_Z(x)\,\chi_N(y), \qquad (x,y)\in\mathbb{R}^2,$$

is a measurable function on $\mathbb{R}^2$. However, if we fix a point $x \in Z$, then

$$f_x(y) \;=\; \chi_Z(x)\,\chi_N(y) \;=\; \chi_N(y), \qquad y \in \mathbb{R},$$

is not measurable on $\mathbb{R}$. To prove Fubini's Theorem, we will have to show that if f is measurable on $E \times F$ then the restriction f_x is measurable on F for almost every x, and the restriction f^y is measurable on E for almost every y. We must be careful not to try to integrate f_x or f^y before we have verified that they are measurable.

The idea of the proof of Fubini's Theorem is to proceed from characteristic functions to simple functions to arbitrary integrable functions. We will make this procedure explicit through a series of lemmas. Because we can split a complex-valued function into real and imaginary parts, it will suffice to consider extended real-valued functions. To further simplify the presentation, we will first establish Fubini's Theorem for the case $E = \mathbb{R}^m$ and $F = \mathbb{R}^n$, and afterward discuss the (easy) extension to arbitrary Cartesian product domains $E \times F$.

To begin the proof, let $\mathcal{F}$ denote the set of all extended real-valued, integrable functions on $\mathbb{R}^{m+n}$ that satisfy statements (a)–(e) in Fubini's Theorem for $E = \mathbb{R}^m$ and $F = \mathbb{R}^n$:

$$\mathcal{F} \;=\; \bigl\{f\colon\mathbb{R}^{m+n} \to [-\infty,\infty] \;:\; f \text{ is integrable and (a)–(e) hold}\bigr\}.$$

Our ultimate goal is to show that every integrable function on $\mathbb{R}^{m+n}$ belongs to $\mathcal{F}$. As a first step, we show that *certain* characteristic functions belong to $\mathcal{F}$.

Lemma 4.6.2. *If $A \subseteq \mathbb{R}^m$ and $B \subseteq \mathbb{R}^n$ are measurable and $|A|, |B| < \infty$, then $\chi_{A\times B} \in \mathcal{F}$.*

Proof. Let $f = \chi_E$ where $E = A \times B$. Fix any point $y \in \mathbb{R}^m$, and consider the function of x defined by $f^y(x) = f(x,y)$. Because E is a Cartesian product,

$$f^y(x) \;=\; \chi_{A\times B}(x,y) \;=\; \chi_A(x)\,\chi_B(y).$$

Thus, when we hold y fixed, f^y is simply the constant $\chi_B(y)$ times the characteristic function χ_A:

$$f^y \;=\; \chi_B(y)\,\chi_A.$$

Since χ_A is measurable and integrable, we conclude that f^y is measurable and integrable for every y.

Since f^y is a measurable and integrable function of x, its integral exists. In fact, if we let $h(y)$ denote this integral, then

$$h(y) \;=\; \int_{\mathbb{R}^m} f^y(x)\,dx \;=\; \chi_B(y)\int_{\mathbb{R}^m} \chi_A(x)\,dx \;=\; |A|\,\chi_B(y).$$

Thus h is simply a constant multiple of χ_B, so h is both measurable and integrable. Further, since $f = \chi_{A\times B}$ and $|A \times B| = |A|\,|B|$, we compute that

$$\begin{aligned}\int_{\mathbb{R}^n}\biggl(\int_{\mathbb{R}^m} f(x,y)\,dx\biggr)\,dy \;&=\; \int_{\mathbb{R}^n} |A|\,\chi_B(y)\,dy\\ &=\; |A|\,|B| \;=\; \iint_{\mathbb{R}^{m+n}} f(x,y)\,(dx\,dy).\end{aligned}$$

Combining this with a symmetric calculation for the other iterated integral, it follows that $f \in \mathcal{F}$. $\square$

If Q is a box in $\mathbb{R}^{m+n}$ then we can write $Q = Q_1 \times Q_2$ where Q_1 is a box in $\mathbb{R}^m$ and Q_2 is a box in $\mathbb{R}^n$. Therefore, a corollary of Lemma 4.6.2 is that $\chi_Q \in \mathcal{F}$ for every box Q contained in $\mathbb{R}^{m+n}$.

Before proceeding to characteristic functions of more general types of sets, we will consider some properties of the collection $\mathcal{F}$. One immediate fact is that $\mathcal{F}$ is closed under addition and scalar multiplication. This is because linear combinations of measurable functions are measurable, and the Lebesgue integral is linear when applied to integrable functions (see Theorem 4.4.10). We state this formally as a lemma.

Lemma 4.6.3. *$\mathcal{F}$ is closed under finite linear combinations, and hence is a subspace of $L^1(\mathbb{R}^{m+n})$.* $\diamondsuit$

Next, by applying the Monotone Convergence Theorem, we will show that $\mathcal{F}$ is closed under monotone limits of nonnegative functions.

Lemma 4.6.4. *Assume that $0 \le f_k \in \mathcal{F}$ for $k \in \mathbb{N}$, and let f be an integrable function on $\mathbb{R}^{m+n}$.*

(a) *If* $f_k \nearrow f$, *then* $f \in \mathcal{F}$.

(b) *If* $f_k \searrow f$, *then* $f \in \mathcal{F}$.

Proof. (a) Assume that $f_k \nearrow f$. By the definition of the family $\mathcal{F}$, the function f_k^y is integrable for almost every y. Further, the function

$$h_k(y) \;=\; \int_{\mathbb{R}^m} f_k^y(x)\,dx, \qquad y \in \mathbb{R}^n,$$

is defined a.e., and it is measurable and integrable.

Let Z_k be the set of y such that f_k^y is not integrable. Then $Z = \cup_{k=1}^\infty Z_k$ has measure zero, and if $y \notin Z$ then f_k^y is measurable for every k. Since $f_k^y \nearrow f^y$, it follows that f^y is measurable. Thus f^y is measurable for almost every y.

If $y \notin Z$ then f^y is both measurable and nonnegative, so its integral exists and is nonnegative (though it might be infinite). Therefore we can define

$$h(y) \;=\; \int_{\mathbb{R}^m} f^y(x)\,dx, \qquad \text{for } y \notin Z.$$

We do not yet know that h is measurable. However, if $y \notin Z$ then the measurable functions f_k^y increase to the measurable function f^y, so the Monotone Convergence Theorem implies that

$$0 \;\le\; h_k(y) \;=\; \int_{\mathbb{R}^m} f_k^y(x)\,dx \;\nearrow\; \int_{\mathbb{R}^m} f^y(x)\,dx \;=\; h(y).$$

Thus $h_k(y) \to h(y)$ for a.e. y. Since each h_k is measurable and the pointwise a.e. limit of measurable functions is measurable, we conclude that h is measurable. Further, h is nonnegative, so its integral exists in the extended real sense.

Now we apply the Monotone Convergence Theorem again, this time to the measurable functions h_k. Since $h_k \nearrow h$, we have

$$0 \;\le\; \int_{\mathbb{R}^n} h_k(y)\,dy \;\nearrow\; \int_{\mathbb{R}^n} h(y)\,dy.$$

At this point, we do not know whether the integral of h is finite. However, using the definition of $\mathcal{F}$ and applying the Monotone Convergence Theorem yet again, we see that

$$\begin{aligned}
\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^m} f(x,y)\,dx\right)dy \;&=\; \int_{\mathbb{R}^n} h(y)\,dy && \text{(definition of } h\text{)}\\
&=\; \lim_{k\to\infty}\int_{\mathbb{R}^n} h_k(y)\,dy && \text{(MCT on } \mathbb{R}^n\text{)}
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{k\to\infty} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} f_k^y(x)\,dx \right) dy && \text{(definition of } h_k\text{)} \\
&= \lim_{k\to\infty} \iint_{\mathbb{R}^{m+n}} f_k(x,y)\,(dx\,dy) && \text{(since } f_k \in \mathcal{F}\text{)} \\
&= \iint_{\mathbb{R}^{m+n}} f(x,y)\,(dx\,dy) && \text{(MCT on } \mathbb{R}^{m+n}\text{)} \\
&< \infty.
\end{aligned}
$$

Hence h is integrable. This implies that $h(y) = \int f^y$ is finite a.e., and therefore f^y is integrable for a.e. y. Finally, the calculation above shows that

$$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} f(x,y)\,dx \right) dy \;=\; \iint_{\mathbb{R}^{m+n}} f(x,y)\,(dx\,dy).$$

A symmetric argument applies to the other iterated integral, so we conclude that $f \in \mathcal{F}$.

(b) Assume that $f_k \searrow f$, and set $g_k = f_1 - f_k$ and $g = f_1 - f$. Then $g_k \in \mathcal{F}$ since $\mathcal{F}$ is closed under linear combinations. Further, g is integrable and $0 \le g_k \nearrow g$, so part (a) implies that $g \in \mathcal{F}$. Therefore $f = f_1 - g \in \mathcal{F}$ as well. $\square$

Now we return to the task of showing that $\mathcal{F}$ contains every characteristic function χ_A with $A \subseteq \mathbb{R}^{m+n}$ and $|A| < \infty$. So far, we know that $\chi_Q \in \mathcal{F}$ when Q is a box in $\mathbb{R}^{m+n}$. Since every open set is a countable union of nonoverlapping boxes, we expect that we should be able to show that $\chi_U \in \mathcal{F}$ for any bounded open set U (we assume boundedness so that χ_U is integrable). Unfortunately, although we can write $U = \cup Q_k$ where the boxes Q_k are nonoverlapping, we have that

$$\chi_U \;\ne\; \sum_{k=1}^{\infty} \chi_{Q_k}$$

because the Q_k are *not disjoint*. This means that we cannot simply combine our previous lemmas to get the conclusion that χ_U belongs to $\mathcal{F}$. We can find disjoint sets $A_k \subseteq Q_k$ such that $\chi_U = \sum \chi_{A_k}$, but the A_k are not boxes, and hence we do not yet know whether χ_{A_k} belongs to $\mathcal{F}$. These problems make the proof of our next lemma longer than we might have expected.

Lemma 4.6.5. *If U is a bounded open subset of $\mathbb{R}^{m+n}$, then $\chi_U \in \mathcal{F}$.*

Proof. Step 1. We will show that $\chi_Z \in \mathcal{F}$ for any set Z that is contained in the boundary of a box Q in $\mathbb{R}^{m+n}$.

Since Q is a box in $\mathbb{R}^{m+n}$, we can write it as

$$Q \;=\; \prod_{k=1}^{m+n} [a_k, b_k] \;=\; R \times S,$$

where R is a box in $\mathbb{R}^m$ and S is a box in $\mathbb{R}^n$. If

$$(x, y) = (x_1, \dots, x_m, x_{m+1}, \dots, x_{m+n}) \in \partial Q,$$

then there must be some k such that x_k equals either a_k or b_k. If $1 \le k \le m$, then this says that $x \in \partial R$, while if $m+1 \le k \le m+n$ then we have $y \in \partial S$.

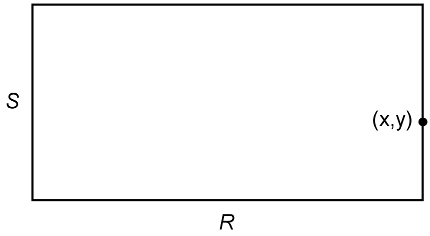


Fig. 4.5 Illustration for $d = 2$. Let $Q = R \times S$ where R, S are closed intervals. If $(x, y) \in \partial Q$ and $y \notin \partial S$, then $x \in \partial R$.

Fix any set $Z \subseteq \partial Q$. Suppose that $y \notin \partial S$ and $\chi_Z^y(x) = 1$. Then $(x, y) \in Z \subseteq \partial Q$, but since $y \notin \partial S$ we must have $x \in \partial R$ (see the illustration in Figure 4.5). Since ∂R has measure zero, we conclude that

$$y \notin \partial S \quad \Longrightarrow \quad \chi_Z^y = 0 \text{ a.e.}$$

Hence χ_Z^y is measurable and integrable except possibly for those y that belong to the measure-zero set ∂S. Further, for a.e. y (those not in ∂S) we have

$$h(y) = \int_{\mathbb{R}^m} \chi_Z^y(x)\, dx = 0.$$

Thus $h = 0$ a.e. Hence h is measurable and integrable, and

$$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} \chi_Z(x, y)\, dx \right) dy = \int_{\mathbb{R}^n} h(y)\, dy = 0 = \iint_{\mathbb{R}^{m+n}} \chi_Z(x, y)\, (dx\, dy),$$

where the last equality follows from the fact that $\chi_Z = 0$ a.e. on $\mathbb{R}^{m+n}$. Combining this with a symmetric calculation for the other iterated integral, we conclude that $\chi_Z \in \mathcal{F}$.

Step 2. Let U be any bounded open subset of $\mathbb{R}^{m+n}$. By Lemma 2.1.5, we can write U as the union of countably many nonoverlapping boxes Q_k contained in $\mathbb{R}^{m+n}$. "Disjointize" these boxes by setting

$$A_1 = Q_1, \quad A_2 = Q_2 \setminus Q_1, \quad A_3 = Q_3 \setminus (Q_1 \cup Q_2),$$

and so forth. The sets A_k are measurable and disjoint, and their union is U. Further, $Z_k = Q_k \setminus A_k \subseteq \partial Q_k$, so $\chi_{Z_k} \in \mathcal{F}$ by Step 1. Since we also have

$\chi_{Q_k} \in \mathcal{F}$, it follows that $\chi_{A_k} = \chi_{Q_k} - \chi_{Z_k} \in \mathcal{F}$. Consequently,

$$\phi_N = \sum_{k=1}^{N} \chi_{A_k} \in \mathcal{F}, \qquad \text{for all } N \in \mathbb{N}.$$

Since $0 \le \phi_N \nearrow \chi_U$ and χ_U is integrable, we can apply Lemma 4.6.4 and conclude that $\chi_U \in \mathcal{F}$. □

If H is a bounded G_δ-set, then we can write $H = \cap U_k$ where $\{U_k\}_{k\in\mathbb{N}}$ is a nested decreasing sequence of open sets. Noting that $\chi_{U_k} \searrow \chi_H$ and applying Lemma 4.6.4, it follows that $\chi_H \in \mathcal{F}$. Since every bounded measurable set $A \subseteq \mathbb{R}^{m+n}$ can be written as $A = H \backslash Z$ where $|Z| = 0$, we are near to proving that $\chi_A \in \mathcal{F}$ for arbitrary bounded measurable sets A.

Lemma 4.6.6. (a) *If $Z \subseteq \mathbb{R}^{m+n}$ and $|Z| = 0$, then $\chi_Z \in \mathcal{F}$.*

(b) *If A is any bounded measurable subset of $\mathbb{R}^{m+n}$, then $\chi_A \in \mathcal{F}$.*

Proof. (a) If $Z \subseteq \mathbb{R}^{m+n}$ has zero measure, then there exists a G_δ-set H that contains Z and has the same measure as Z. As we remarked before the statement of the lemma, the results we have established so far imply that $\chi_H \in \mathcal{F}$. Therefore

$$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} \chi_H^y(x)\, dx \right) dy \;=\; \iint_{\mathbb{R}^{m+n}} \chi_H(x,y)\,(dx\,dy) \;=\; |H| \;=\; |Z| \;=\; 0.$$

The integrands on the preceding line are nonnegative, so this implies that

$$h(y) \;=\; \int_{\mathbb{R}^m} \chi_H^y(x)\, dx \;=\; 0 \quad \text{for a.e. } y.$$

Consequently, for a.e. y we have $\chi_H^y = 0$ a.e., and since $Z \subseteq H$, it follows that

$$\text{for a.e. } y, \quad \chi_Z^y \;=\; 0 \text{ a.e.}$$

Therefore χ_Z^y is measurable and integrable for a.e. y. Further, $h = 0$ a.e., so h is measurable and integrable, and

$$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} \chi_Z(x,y)\, dx \right) dy \;=\; \int_{\mathbb{R}^n} h(y)\, dy \;=\; 0 \;=\; \iint_{\mathbb{R}^{m+n}} \chi_Z(x,y)\,(dx\,dy).$$

Combining this with a symmetric calculation for the other iterated integral, we conclude that $\chi_Z \in \mathcal{F}$.

(b) If A is bounded and measurable, then $A = H \backslash Z$ where H is a bounded G_δ-set and $|Z| = 0$. By replacing Z with $H \cap Z$, we may assume that $Z \subseteq H$. Hence $\chi_A = \chi_H - \chi_Z$. But χ_H and χ_Z both belong to $\mathcal{F}$ and we know that $\mathcal{F}$ is closed under finite linear combinations, so $\chi_A \in \mathcal{F}$. □

By combining the preceding lemmas we will obtain the proof of Fubini's Theorem for extended real-valued functions whose domain is $\mathbb{R}^{m+n}$.

Theorem 4.6.7. *If f is an integrable extended real-valued function on $\mathbb{R}^{m+n}$, then $f \in \mathcal{F}$.*

Proof. Assume first that f is nonnegative, and let ϕ_k be nonnegative simple functions such that $\phi_k \nearrow f$. Let $Q_k = [-k,k]^{m+n}$, and define

$$\psi_k = \phi_k \cdot \chi_{Q_k}.$$

Each ψ_k is a compactly supported simple function, and $\psi_k \nearrow f$. A compactly supported simple function is a finite linear combination of characteristic functions of bounded sets, so by combining Lemma 4.6.6 with the fact that $\mathcal{F}$ is closed under linear combinations, we see that $\psi_k \in \mathcal{F}$. Consequently, by applying Lemma 4.6.4 we obtain $f \in \mathcal{F}$.

Now let f be an arbitrary integrable extended real-valued function. Then we can write $f = f^+ - f^-$ where f^+ and f^- are both nonnegative. Since f^+ and f^- are integrable, they belong to $\mathcal{F}$. Hence $f \in \mathcal{F}$ since $\mathcal{F}$ is closed under finite linear combinations. $\square$

Thus, we have shown that Fubini's Theorem holds for integrable extended real-valued functions whose domain is $\mathbb{R}^{m+n}$. By splitting a complex-valued function into its real and imaginary parts, the corresponding result for complex-valued functions on $\mathbb{R}^{m+n}$ also follows.

The final step is to extend to functions whose domain is $E \times F$ instead of $\mathbb{R}^m \times \mathbb{R}^n$. This is easy, for if f is defined on $E \times F$ then we can extend the domain of f to $\mathbb{R}^{m+n}$ by setting $f = 0$ outside of $E \times F$. Applying Fubini's Theorem for functions on $\mathbb{R}^{m+n}$ and recalling that f vanishes outside of $E \times F$, we see that all of statements (a)–(e) in Fubini's Theorem hold for f on the domain $E \times F$. This completes the proof of Theorem 4.6.1.

4.6.2 Tonelli's Theorem

Our next result, which is known as *Tonelli's Theorem*, is complementary to Fubini's Theorem. It states that the interchange in the order of integration is allowed if f is a *nonnegative* function. In this case all of the integrals involved are nonnegative, although they might be infinite.

Theorem 4.6.8 (Tonelli's Theorem). *Let E be a measurable subset of $\mathbb{R}^m$ and let F be a measurable subset of $\mathbb{R}^n$. If $f\colon E \times F \to [0,\infty]$ is measurable, then the following statements hold.*

(a) *$f_x(y) = f(x,y)$ is a measurable function on F for almost every $x \in E$.*

(b) *$f^y(x) = f(x,y)$ is a measurable function on E for almost every $y \in F$.*

(c) $g(x) = \int_F f_x(y)\,dy$ *is a measurable function on* E.

(d) $h(y) = \int_E f^y(x)\,dx$ *is a measurable function on* F.

(e) *The following three integrals exist as nonnegative extended real numbers, and are equal as indicated:*

$$\iint_{E\times F} f(x,y)\,(dx\,dy) \;=\; \int_F \biggl(\int_E f(x,y)\,dx\biggr)\,dy \tag{4.25}$$

$$= \int_E \biggl(\int_F f(x,y)\,dy\biggr)\,dx. \tag{4.26}$$

Proof. The idea of the proof is that we create an integrable approximation f_k to f to which we can apply Fubini's Theorem, and then use the Monotone Convergence Theorem to move to the limit.

Let f be any nonnegative measurable function on $E \times F$. For each $k \in \mathbb{N}$, set $Q_k = [-k,k]^{m+n}$, and for $x \in E \times F$ define

$$f_k(x) \;=\; \begin{cases} k, & \text{if } x \in Q_k \text{ and } f(x) > k,\\ f(x), & \text{if } x \in Q_k \text{ and } 0 \le f(x) \le k,\\ 0, & \text{otherwise.}\end{cases}$$

Each f_k is integrable and nonnegative, and $f_k \nearrow f$.

By Fubini's Theorem, f_k^y is measurable and integrable for a.e. y. Since $f_k^y \nearrow f^y$, it follows that f^y is measurable for a.e. y. It also follows from Fubini's Theorem that the function

$$h_k(y) \;=\; \int_E f_k(x,y)\,dx$$

is measurable and integrable. Since f^y is nonnegative, its integral exists (although the integral could be infinite). Further, by the Monotone Convergence Theorem, for a.e. y we have that

$$h_k(y) \;=\; \int_E f_k(x,y)\,dx \;\nearrow\; \int_E f(x,y)\,dx \;=\; h(y).$$

Hence h is defined a.e. and is measurable. Applying the Monotone Convergence Theorem again, we see that

$$\begin{aligned}
\int_F \biggl(\int_E f(x,y)\,dx\biggr)\,dy \;&=\; \int_F h(y)\,dy && \text{(definition of } h)\\
&=\; \lim_{k\to\infty} \int_F h_k(y)\,dy && \text{(MCT on } F)\\
&=\; \lim_{k\to\infty} \int_F \biggl(\int_E f_k^y(x)\,dx\biggr)\,dy && \text{(definition of } h_k)
\end{aligned}$$

$$= \lim_{k\to\infty} \iint_{E\times F} f_k(x,y)\,(dx\,dy) \qquad \text{(Fubini)}$$

$$= \iint_{E\times F} f(x,y)\,(dx\,dy) \qquad \text{(MCT on } E\times F).$$

The quantities above may be infinite, but they are equal as indicated. This establishes the equality given in equation (4.25). The proof of the equality in equation (4.26) follows similarly, by interchanging the roles of x and y. □

One of the most common uses of Tonelli's Theorem is to determine if Fubini's Theorem is applicable. In order to apply Fubini's Theorem, we need to know that the function f is integrable on $E \times F$. To do this, we have to compute the integral of $|f|$ on $E \times F$. Since $|f|$ is nonnegative, Tonelli's Theorem tells us that we can prove that f is integrable by showing that any one of three possible integrals is finite. Hence we can choose whichever one of these three integrals is simplest to evaluate, and just verify that one integral is finite. Here is the precise formulation.

Corollary 4.6.9. *Let E be a measurable subset of $\mathbb{R}^m$ and let F be a measurable subset of $\mathbb{R}^n$. If $f\colon E\times F \to \overline{\mathbf{F}}$ is a measurable function on $E\times F$, then, as extended real numbers,*

$$\iint_{E\times F} |f(x,y)|\,(dx\,dy) = \int_F\left(\int_E |f(x,y)|\,dx\right)dy = \int_E\left(\int_F |f(x,y)|\,dy\right)dx.$$

Consequently, if any one of these three integrals is finite, then $f \in L^1(E\times F)$ and

$$\iint_{E\times F} f(x,y)\,(dx\,dy) = \int_F\left(\int_E f(x,y)\,dx\right)dy = \int_E\left(\int_F f(x,y)\,dy\right)dx. \quad \diamondsuit$$

Fubini's Theorem and Tonelli's Theorem can be adapted to domains that are not Cartesian products. Given a function f on a measurable set $A \subseteq \mathbb{R}^{m+n}$, the simplest way to apply Fubini or Tonelli is to extend f by zero. The following lemma illustrates this technique.

Lemma 4.6.10. *If F is a nonnegative or integrable function on the domain $D = \big\{(x,y)\in[0,\infty)^2 : y\le x\big\}$, then*

$$\int_0^\infty \int_0^x F(x,y)\,dy\,dx \;=\; \int_0^\infty \int_y^\infty F(x,y)\,dx\,dy.$$

Proof. Extend F to all of $[0,\infty)^2$ by setting $F(x,y) = 0$ for $(x,y)\notin D$. Applying Tonelli's Theorem or Fubini's Theorem (as appropriate), we see that

$$\int_0^\infty \int_0^x F(x,y)\,dy\,dx \;=\; \int_0^\infty \int_0^\infty F(x,y)\,\chi_D(x,y)\,dy\,dx$$

$$= \int_0^\infty \int_0^\infty F(x,y)\, \chi_D(x,y)\, dx\, dy$$

$$= \int_0^\infty \int_y^\infty F(x,y)\, dx\, dy. \qquad \square$$

4.6.3 Convolution

To give an application of Fubini's Theorem and Tonelli's Theorem, we introduce the operation of *convolution* and prove that $L^1(\mathbb{R}^d)$ is closed under this operation.

If f and g belong to $L^1(\mathbb{R}^d)$, then we formally define their convolution to be the function $f * g$ given by

$$(f * g)(x) \;=\; \int_{\mathbb{R}^d} f(y)\, g(x-y)\, dy. \tag{4.27}$$

This is a "formal" definition because at this point we do not know whether the integral in the definition of $(f * g)(x)$ exists.

It may not be obvious at this point why we would want to define $f * g$ by equation (4.27), or why this would lead to a useful operation. However, convolution is in fact a natural operation that arises in a wide variety of circumstances. To give a familiar example of a discrete version of a convolution, consider the product of two polynomials

$$p(x) = a_0 + a_1 x + \cdots + a_m x^m \qquad \text{and} \qquad q(x) = b_0 + b_1 x + \cdots + b_n x^n.$$

If we set $a_k = 0$ for $k > m$ and $k < 0$ and $b_k = 0$ for $k > n$ and $k < 0$, then the product of p and q is $p(x)q(x) = c_0 + c_1 x + \cdots + c_{m+n} x^{m+n}$, where

$$c_k \;=\; \sum_{j=0}^{k} a_j\, b_{k-j}, \qquad \text{for } k = 0, \dots, m+n.$$

The sequence of coefficients (c_k) of the polynomial pq is a discrete convolution of the sequence (a_k) with the sequence (b_k).

In this section we will give one particular sufficient condition on f and g that implies that $f * g$ exists. Specifically, we will use Fubini's Theorem to show that $(f * g)(x)$ is defined for a.e. x when f and g are both integrable. To apply Fubini's Theorem, we need a function of two variables, and this is

$$F(x,y) \;=\; f(y)\, g(x-y).$$

To see why F is measurable, first consider $G(x,y) = f(x)$ for $(x,y) \in \mathbb{R}^{2d}$. This is measurable on $\mathbb{R}^{2d}$ because

$$\{G > a\} = \{f > a\} \times \mathbb{R}^d.$$

Similarly $g(y)$ is measurable as a function of x and y, and therefore the product $H(x,y) = f(x)\,g(y)$ is measurable on $\mathbb{R}^{2d}$. Since $F = H \circ L$, where $L\colon \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ is the linear function $L(x,y) = (y, x-y)$, and since measurability is preserved under linear changes of variable, it follows that $F(x,y) = H(y, x-y)$ is measurable.

Now we show that F is integrable on $\mathbb{R}^{2d}$. To do this, we use Tonelli's Theorem, which allows us to choose the most convenient iterated integral to evaluate. We choose to compute $\iint |F|$ as follows:

$$\begin{aligned}
\iint_{\mathbb{R}^{2d}} |F(x,y)|\,(dx\,dy) &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(y)|\,|g(x-y)|\,dx \right) dy && \text{(Tonelli)} \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |g(x-y)|\,dx \right) |f(y)|\,dy \\
&= \int_{\mathbb{R}^d} \|g\|_1\,|f(y)|\,dy && \text{(by Problem 4.3.9)} \\
&= \|g\|_1\,\|f\|_1 \;<\; \infty. && (4.28)
\end{aligned}$$

Therefore F is integrable. Consequently Fubini's Theorem implies that $F_x(y) = f(y)\,g(x-y)$ is a measurable and integrable function of y for almost every x, and

$$(f * g)(x) \;=\; \int_{\mathbb{R}^d} F_x(y)\,dy$$

exists for almost every x and is an integrable function of x.

In summary, by using the theorems of Tonelli and Fubini, we have shown that if f and g are integrable on $\mathbb{R}^d$, then $f * g$ is defined at almost every point and is integrable on $\mathbb{R}^d$. Thus

$$f,\, g \in L^1(\mathbb{R}^d) \quad\Longrightarrow\quad f * g \in L^1(\mathbb{R}^d),$$

so $L^1(\mathbb{R}^d)$ *is closed under convolution.* Furthermore, by using equation (4.28) we obtain a relationship between the norms of f, g, and $f * g$:

$$\begin{aligned}
\|f * g\|_1 &= \int_{\mathbb{R}^d} |(f * g)(x)|\,dx \\
&= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(y)\,g(x-y)\,dy \right| dx \\
&\le \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(y)\,g(x-y)|\,dy\,dx \\
&= \iint_{\mathbb{R}^{2d}} |F(x,y)|\,(dx\,dy) \;=\; \|f\|_1\,\|g\|_1.
\end{aligned}$$

We state these results formally as a theorem (which is itself a special case of *Young's Inequality*, see Theorem 9.1.14).

Theorem 4.6.11. *If $f, g \in L^1(\mathbb{R}^d)$, then $f * g \in L^1(\mathbb{R}^d)$ and*

$$\|f * g\|_1 \ \le \ \|f\|_1 \, \|g\|_1. \qquad \diamondsuit \tag{4.29}$$

We often summarize equation (4.29) by saying that convolution is *sub-multiplicative* with respect to the L^1-norm. Some further properties of convolution are given in Problems 4.6.25–4.6.27, and we will return to study convolution in more detail in Section 9.1.

Problems

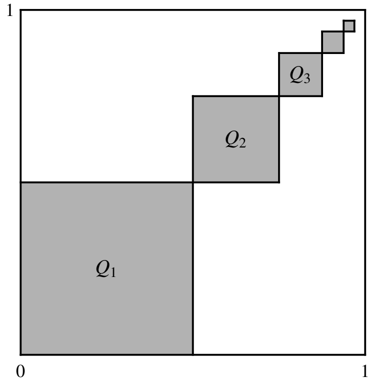


Fig. 4.6 Boxes $Q_1, Q_2, \dots$ for Problem 4.6.12.

4.6.12. Let $Q = [0,1]^2$, and let $Q_1, Q_2, \dots$ be an infinite sequence of nonoverlapping squares centered on the main diagonal of Q, as shown in Figure 4.6. Subdivide each square Q_n into four equal subsquares, and let $f = 1/|Q_n|$ on the lower left and upper right subsquares of Q_n, and $f = -1/|Q_n|$ on the lower right and upper left subsquares. Set $f = 0$ everywhere else. Prove that

$$\int_0^1 \left(\int_0^1 f(x,y)\, dx \right) dy \ = \ \int_0^1 \left(\int_0^1 f(x,y)\, dy \right) dx \ = \ 0,$$

but $\iint_Q |f(x,y)|\,(dx\,dy) = \infty$. Use this to show that $\iint_Q f(x,y)\,(dx\,dy)$, the Lebesgue integral of f on Q, is undefined.

4.6.13. Consider the two iterated integrals

$$I_1 = \int_{-1}^{1}\int_{-1}^{1} \frac{x}{1-y^2}\,dx\,dy, \qquad I_2 = \int_{-1}^{1}\int_{-1}^{1} \frac{x}{1-y^2}\,dy\,dx.$$

Prove that I_1 exists, but I_2 is undefined. Note that $\frac{x}{1-y^2}$ is continuous but unbounded on $(-1,1)^2$.

4.6.14. Use the fact that $\frac{d}{dy}\frac{y}{x^2+y^2} = \frac{x^2-y^2}{(x^2+y^2)^2}$ to prove that the following iterated integrals have the indicated values:

$$\int_1^\infty \biggl(\int_1^\infty \frac{x^2-y^2}{(x^2+y^2)^2}\,dy\biggr)\,dx = -\frac{\pi}{4},$$

$$\int_1^\infty \biggl(\int_1^\infty \frac{x^2-y^2}{(x^2+y^2)^2}\,dx\biggr)\,dy = \frac{\pi}{4},$$

$$\int_1^\infty \biggl(\int_1^\infty \biggl|\frac{x^2-y^2}{(x^2+y^2)^2}\biggr|\,dx\biggr)\,dy = \infty.$$

Note that $\frac{x^2-y^2}{(x^2+y^2)^2}$ is both continuous and bounded on $[1,\infty)^2$.

4.6.15. Given $f \in L^1(\mathbb{R})$, define $g(x) = \int_{-\infty}^x f(t)\,dt$ for $x \in \mathbb{R}$. Prove that if we fix $c \in \mathbb{R}$, then $g(x+c) - g(x)$ is an integrable function of x and

$$\int_{-\infty}^{\infty} \bigl(g(x+c) - g(x)\bigr)\,dx = c\int_{-\infty}^{\infty} f(t)\,dt.$$

4.6.16. Let $E \subseteq \mathbb{R}^m$ and $F \subseteq \mathbb{R}^n$ be measurable sets, and assume that $f\colon E \times F \to \overline{\mathbf{F}}$ is measurable. Define $f_x(y) = f(x,y)$, and prove that the following two statements are equivalent.

(a) $f = 0$ a.e. on $E \times F$.

(b) For almost every $x \in E$ we have $f_x(y) = 0$ for a.e. $y \in F$.

4.6.17. Use Tonelli's Theorem to give another solution to Problem 4.2.17.

4.6.18. Define $f\colon (0,\infty)^2 \to \mathbb{R}$ by $f(x,y) = x\,e^{-x^2(1+y^2)}$. Compute the two iterated integrals of f (one with respect to $dx\,dy$ and one with respect to $dy\,dx$), and use Fubini's Theorem to show that

$$\int_0^\infty e^{-t^2}\,dt = \frac{\sqrt{\pi}}{2}.$$

4.6.19. Use Fubini's Theorem and the substitution $\int_0^\infty e^{-tx}\,dt = \frac{1}{x}$ to evaluate the integral $\int_0^a \frac{\sin x}{x}\,dx$. Then apply the Dominated Convergence Theorem to show that $\lim_{a\to\infty} \int_0^a \frac{\sin x}{x}\,dx = \frac{\pi}{2}$.

Remark: Thus, even though $\frac{\sin x}{x}$ is not integrable on the infinite interval $[0,\infty)$, the improper Riemann integral $\int_0^\infty \frac{\sin x}{x}\,dx$ exists and equals $\frac{\pi}{2}$.

4.6.20. Given $f \in L^1[0,1]$, define

$$g(x) \;=\; \int_x^1 \frac{f(t)}{t}\,dt, \qquad 0 < x \le 1.$$

Show that g is defined a.e. on $[0,1]$, $g \in L^1[0,1]$, and $\int_0^1 g(x)\,dx = \int_0^1 f(x)\,dx$.

4.6.21. Assume that $E \subseteq \mathbb{R}^d$ is measurable. The *distribution function* of a measurable function $f\colon E \to \overline{\mathbf{F}}$ is

$$\omega(t) \;=\; \bigl|\{|f| > t\}\bigr|, \qquad t \ge 0.$$

By definition, ω is a nonnegative extended real-valued function. Prove the following facts about ω.

(a) ω is monotone decreasing on $[0,\infty)$.

(b) ω is right-continuous, i.e., $\lim_{s\to t^+} \omega(s) = \omega(t)$ for each $t \ge 0$.

(c) If f is integrable, then $\lim_{s\to t^-} \omega(s) = |\{|f| \ge t\}|$.

(d) $\int_0^\infty \omega(t)\,dt \;=\; \int_E |f(x)|\,dx$.

(e) f is integrable if and only if ω is integrable.

(f) If f is integrable, then $\lim_{n\to\infty} n\,\omega(n) = 0 = \lim_{n\to\infty} \frac{1}{n}\,\omega\bigl(\frac{1}{n}\bigr)$.

4.6.22. Prove *Fubini's Theorem for series*: If c_{mn} is a real or complex number for each $m, n \in \mathbb{N}$ and

$$\sum_{m=1}^\infty \sum_{n=1}^\infty |c_{mn}| \;<\; \infty,$$

then the following series converge and are equal as indicated:

$$\sum_{m=1}^\infty \sum_{n=1}^\infty c_{mn} \;=\; \sum_{n=1}^\infty \sum_{m=1}^\infty c_{mn}.$$

4.6.23. Prove *Tonelli's Theorem for series*: If $c_{mn} \ge 0$ for $m, n \in \mathbb{N}$, then

$$\sum_{m=1}^\infty \sum_{n=1}^\infty c_{mn} \;=\; \sum_{n=1}^\infty \sum_{m=1}^\infty c_{mn},$$

in the sense that either both sides converge and are equal, or both sides are infinite.

4.6.24. Prove the following mixed integral/series version of Fubini's Theorem: If $f_n\colon E \to \overline{\mathbf{F}}$ is measurable for each $n \in \mathbb{N}$, where $E \subseteq \mathbb{R}^d$ is measurable, and if

$$\sum_{n=1}^\infty \int_E |f_n(t)|\,dt \;<\; \infty,$$

then the series $\sum_{n=1}^{\infty} f_n(t)$ converges for a.e. t, and the following series and integrals exist and are equal as indicated:

$$\int_E \sum_{n=1}^{\infty} f_n(t)\,dt \;=\; \sum_{n=1}^{\infty} \int_E f_n(t)\,dt.$$

(For an integral/series version of Tonelli's Theorem, see Corollary 4.2.4.)

4.6.25. Let $f(x) = e^{-|x|}$, $g(x) = e^{-x^2}$, and $h(x) = xe^{-x^2}$. Compute $f*f$, $g*g$, and $h*h$.

4.6.26. Prove that the following statements hold for all $f, g, h \in L^1(\mathbb{R})$.

(a) Convolution is commutative: $f*g = g*f$ a.e.

(b) Convolution is associative: $(f*g)*h = f*(g*h)$ a.e.

(c) Convolution distributes over addition: $f*(ag+bh) = af*g+bf*h$ a.e. for all scalars a and b.

(d) Convolution commutes with translation: $f*(T_a g) = (T_a f)*g = T_a(f*g)$ a.e. for all $a \in \mathbb{R}$.

4.6.27. Given $f \in L^1(\mathbb{R})$ and $g \in L^\infty(\mathbb{R})$, prove the following statements.

(a) The integral defining $(f*g)(x)$ exists for *every* $x \in \mathbb{R}$.

(b) $f*g$ is continuous on $\mathbb{R}$.

(c) $f*g$ is bounded on $\mathbb{R}$, and $\|f*g\|_\infty \le \|f\|_1\,\|g\|_\infty$.

4.6.28. (a) Show that if $f, g \in C_c(\mathbb{R})$, then $f*g \in C_c(\mathbb{R})$ and

$$\mathrm{supp}(f*g) \;\subseteq\; \mathrm{supp}(f) + \mathrm{supp}(g) \;=\; \bigl\{x+y : x \in \mathrm{supp}(f),\, y \in \mathrm{supp}(g)\bigr\}.$$

Conclude that $C_c(\mathbb{R})$ is closed under convolution.

(b) Is $C_c^1(\mathbb{R})$ closed under convolution?

4.6.29. Let E be a measurable subset of $\mathbb{R}$ such that $0 < |E| < \infty$.

(a) Prove that the convolution $\chi_E * \chi_{-E}$ is continuous.

(b) Prove the *Steinhaus Theorem*: The set $E - E = \{x-y : x, y \in E\}$ contains an open interval centered at the origin (compare this proof to the one that appears in Theorem 2.4.3).

(c) Show that $\lim_{t\to 0} |E \cap (E+t)| = |E|$ and $\lim_{t\to\pm\infty} |E \cap (E+t)| = 0$.

4.6.30. (a) Prove that if $f \in L^1(\mathbb{R})$ and $g \in C_0(\mathbb{R})$, then $f*g \in C_0(\mathbb{R})$.

(b) Given $f \in L^1(\mathbb{R})$, evaluate $\displaystyle\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x-n)\,\frac{x}{1+|x|}\,dx.$

Chapter 5
Differentiation

In this chapter and the next, we will take a closer look at some of the fundamental properties of functions, especially those whose domain is a finite closed interval $[a, b]$. The interplay between differentiation and integration will be a recurring theme throughout Chapters 5 and 6.

An important issue that motivates much of our work is the Fundamental Theorem of Calculus (which we often refer to by the acronym *FTC*). We know from undergraduate real analysis that if a function f is differentiable at every point in a closed finite interval $[a, b]$ and if f' is continuous on $[a, b]$, then the Fundamental Theorem of Calculus holds, and it tells us that

$$\int_a^x f'(t)\,dt \;=\; f(x) - f(a), \qquad \text{for all } x \in [a, b]. \tag{5.1}$$

Since we assumed that f' is continuous, the integral on the line above exists as a Riemann integral. Does the Fundamental Theorem of Calculus hold if we assume only that f' is Lebesgue integrable? Precisely:

If $f'(x)$ exists for a.e. x and f' is integrable, must equation (5.1) *hold?*

We construct a fascinating function in Section 5.1 that shows that the answer to this question is *no* in general.

By the end of Chapter 6, we will characterize the functions for which the FTC holds. To this end, we introduce in Section 5.2 the class of functions that have *bounded variation*, and we prove that each such function is a finite linear combination of monotone increasing functions. In order to make further progress we prove two types of *covering lemmas* in Section 5.3, and use these to show in Section 5.4 that all monotone increasing functions (and hence all functions with bounded variation) are differentiable at almost every point. In Section 5.5 we prove the *Maximal Theorem*, and then use it and a covering lemma to prove the *Lebesgue Differentiation Theorem*, which is a fundamental result on the convergence of averages of a locally integrable function. All of these results will be important to us when we further study the relationship between differentiation and integration in Chapter 6, ultimately establishing

C. Heil, *Introduction to Real Analysis*, Graduate Texts in Mathematics 280,
https://doi.org/10.1007/978-3-030-26903-6_5

the connection between *absolutely continuous functions* and the Fundamental Theorem of Calculus.

Most of the functions that we encounter in this chapter will be finite at every point. Hence, we usually need only consider real-valued and complex-valued functions. Since every real-valued function is complex-valued, it therefore suffices in most of this chapter to just consider complex-valued functions.

5.1 The Cantor–Lebesgue Function

We will construct a continuous function φ that is differentiable at almost every point and whose derivative φ' is equal almost everywhere to a continuous function (the zero function!), yet the Fundamental Theorem of Calculus does not apply to φ.

The construction is closely related to the construction of the Cantor middle-thirds set presented in Example 2.1.23. We will also need to make use of the fact, proved in Theorem 1.3.3, that the space $C[0,1]$, consisting of all continuous functions $f\colon [0,1] \to \mathbb{C}$, is *complete* with respect to the uniform norm

$$\|f\|_{\mathrm{u}} = \sup_{x\in[0,1]} |f(x)|.$$

Precisely, completeness means that every sequence $\{f_n\}_{n\in\mathbb{N}}$ that is Cauchy in $C[0,1]$ with respect to the uniform norm must actually converge uniformly to some function $f \in C[0,1]$.

To construct the Cantor–Lebesgue function, first consider the functions φ_1 and φ_2 pictured in Figure 5.1. The function φ_1 takes the constant value $\frac{1}{2}$ on the interval $(\frac{1}{3}, \frac{2}{3})$ that is removed from $[0,1]$ in the first stage of the construction of the Cantor set, and it is linear on the remaining subintervals of $[0,1]$. The function φ_2 takes the same constant $\frac{1}{2}$ on the interval $(\frac{1}{3}, \frac{2}{3})$, but additionally is constant with values $\frac{1}{4}$ and $\frac{3}{4}$ on the two intervals that are removed during the second stage of the construction of the Cantor set. We continue this process and define $\varphi_3, \varphi_4, \dots$ in a similar fashion. Each function φ_k is continuous and monotone increasing on $[0,1]$, and φ_k is constant on each of the open intervals that are removed during the kth stage of the construction of the Cantor set.

Looking at Figure 5.1, we can see that $\varphi_1(x)$ and $\varphi_2(x)$ never differ by more than $\frac{1}{2}$ unit (and even that is only a gross estimate). More generally, for each $k \in \mathbb{N}$ we have

$$\|\varphi_{k+1} - \varphi_k\|_{\mathrm{u}} = \sup_{x\in[0,1]} |\varphi_{k+1}(x) - \varphi_k(x)| \le 2^{-k}.$$

Applying the Triangle Inequality, if we fix $m < n$ then we see that

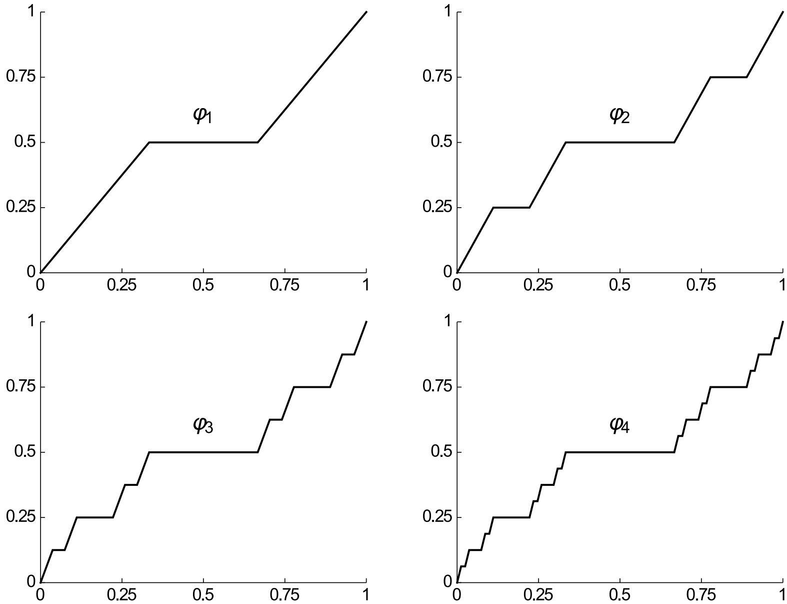


Fig. 5.1 First stages in the construction of the Cantor–Lebesgue function.

$$\begin{aligned}\|\varphi_n - \varphi_m\|_{\mathrm{u}} &= \left\|\sum_{k=m}^{n-1} (\varphi_{k+1} - \varphi_k)\right\|_{\mathrm{u}} \\ &\le \sum_{k=m}^{n-1} \|\varphi_{k+1} - \varphi_k\|_{\mathrm{u}} \\ &\le \sum_{k=m}^{n-1} 2^{-k} \le \sum_{k=m}^{\infty} 2^{-k} = 2^{-m+1}.\end{aligned}$$

Consequently, if $\varepsilon > 0$ is fixed and we choose N large enough, then we will have $\|\varphi_n - \varphi_m\|_{\mathrm{u}} < \varepsilon$ for all $m, n \ge N$. Hence $\{\varphi_n\}_{n\in\mathbb{N}}$ is a uniformly Cauchy sequence in $C[0,1]$. Since we know that every Cauchy sequence in $C[0,1]$ must converge, there is some function $\varphi \in C[0,1]$ such that φ_k converges uniformly (and therefore pointwise) to φ.

Definition 5.1.1 (Cantor–Lebesgue Function). The continuous function φ defined by

$$\varphi(x) = \lim_{k\to\infty} \varphi_k(x), \qquad \text{for } x \in [0,1],$$

is called the *Cantor–Lebesgue function.* $\diamondsuit$

More picturesquely, the Cantor–Lebesgue function is also known as the *Devil's staircase* on $[0,1]$. If we like, we can extend φ to a continuous function

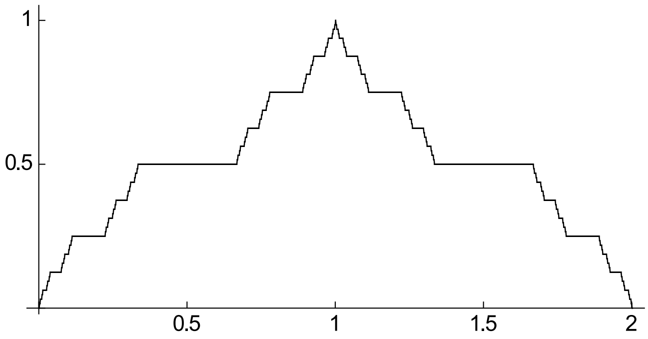


Fig. 5.2 The reflected Devil's staircase (Cantor–Lebesgue function).

on the entire real line $\mathbb{R}$ by reflecting its graph about the point $x = 1$ and declaring φ to be zero outside of $[0,2]$ (see Figure 5.2).

If x is any point in the open interval $\left(\frac{1}{3}, \frac{2}{3}\right)$, then $\varphi_k(x) = \frac{1}{2}$ for every k. Therefore

$$\varphi(x) = \lim_{k\to\infty} \varphi_k(x) = \frac{1}{2}, \qquad \text{for all } x \in \left(\tfrac{1}{3}, \tfrac{2}{3}\right).$$

Similarly,

$$\varphi = \frac{1}{4} \text{ on } \left(\tfrac{1}{9}, \tfrac{2}{9}\right) \qquad \text{and} \qquad \varphi = \frac{3}{4} \text{ on } \left(\tfrac{7}{9}, \tfrac{8}{9}\right).$$

Continuing in this way, we see that φ is differentiable on every open interval that belongs to the complement of the Cantor set C, and

$$\varphi'(x) = 0, \qquad \text{for all } x \in [0,1]\backslash C.$$

Since the Cantor set has zero measure, we have proved the following result.

Theorem 5.1.2. *The Cantor–Lebesgue function φ is differentiable at almost every point of $[0,1]$, and $\varphi' = 0$ a.e. on $[0,1]$.* $\diamondsuit$

In summary, on the interval $[0,1]$ the Cantor–Lebesgue function φ is continuous and monotone increasing, differentiable at almost every point, and $\varphi' = 0$ almost everywhere. Yet the Fundamental Theorem of Calculus does not hold for φ, because

$$\varphi(1) - \varphi(0) = 1 \neq 0 = \int_0^1 \varphi'(x)\,dx. \tag{5.2}$$

We give a name to functions that are differentiable at almost every point but whose derivative is zero a.e.

Definition 5.1.3 (Singular Function). A function f on $[a,b]$ (either extended real-valued or complex-valued) is *singular* if f is differentiable at almost every point in $[a,b]$ and $f' = 0$ a.e. on $[a,b]$. $\diamondsuit$

In particular, the Cantor–Lebesgue function is singular on $[0,1]$. There are many surprising examples of singular functions. For constructions of continuous, *strictly increasing* functions that are singular on $[0,1]$, see Problem 5.4.8 or [BC09, Ex. 4.2.5].

The existence of singular functions shows that we need more than just almost everywhere differentiability in order to conclude that the Fundamental Theorem of Calculus holds for a given function. We will give a complete characterization of the functions that satisfy the FTC in Section 6.4, and we will see there that these are precisely the functions that are *absolutely continuous* in the sense that we will introduce in Section 6.1.

The Cantor–Lebesgue function has many unusual properties. For example, Problem 5.1.5 asks for a proof that φ is *Hölder continuous* but not *Lipschitz continuous*. We show next that even though the Cantor–Lebesgue function is continuous, it does not map measurable sets to measurable sets.

Example 5.1.4. If $x \in [0,1]$ belongs to the complement of the Cantor set C, then $\varphi(x)$ has the form $k/2^n$ for some integers k and n. Hence φ maps $[0,1]\backslash C$ into the set of rational numbers $\mathbb{Q}$. Thus

$$\varphi\big([0,1]\backslash C\big) \;\subseteq\; \mathbb{Q} \cap [0,1]. \tag{5.3}$$

Since φ is a surjective mapping of $[0,1]$ onto itself, if $z \in [0,1]$ is irrational then we must have $z = \varphi(x)$ for some x. By equation (5.3), this point x must belong to C. Thus $\varphi(C)$ includes all of the irrational numbers in $[0,1]$! Therefore $|\varphi(C)| = 1$, even though $|C| = 0$.

Every set with positive measure contains a nonmeasurable subset, so there exists a set $N \subseteq [0,1]\backslash\mathbb{Q}$ that is not measurable (in fact, the nonmeasurable set constructed in Section 2.4.2 contains only one rational point, so deleting that point gives us a nonmeasurable set that contains no rationals). Since N contains no rationals, its inverse image $E = \varphi^{-1}(N)$ is entirely contained in C. Consequently,

$$|E|_e \;\le\; |C| \;=\; 0,$$

and therefore E is a Lebesgue measurable set. However, because φ maps $[0,1]$ *onto* $[0,1]$ we have $\varphi(E) = \varphi(\varphi^{-1}(N)) = N$. Thus $\varphi(E)$ is not measurable, even though E is measurable. $\diamondsuit$

Problems

5.1.5. Prove that the Cantor–Lebesgue function φ is Hölder continuous (in the sense of Definition 1.4.1) precisely for those exponents α that lie in the range $0 < \alpha \le \log_3 2 \approx 0.6309\dots$. In particular, φ is not Lipschitz.

5.1.6. Exhibit a continuous function $f\colon [0,1] \to \mathbb{R}$ that is differentiable at almost every point and satisfies $f' \geq 0$ a.e., yet f is not monotone increasing on $[0,1]$.

5.1.7. Let C be the Cantor set, let φ be the Cantor–Lebesgue function, and set $g(x) = \varphi(x) + x$ for $x \in [0,1]$.

(a) Prove that $g\colon [0,1] \to [0,2]$ is a continuous, strictly increasing bijection, and its inverse function $h = g^{-1}\colon [0,2] \to [0,1]$ is also a continuous, strictly increasing bijection.

(b) Show that $g(C)$ is a closed subset of $[0,2]$, and $|g(C)| = 1$.

(c) Since $g(C)$ has positive measure, it follows from Problem 2.4.9 that $g(C)$ contains a nonmeasurable set N. Show that $A = h(N)$ is a Lebesgue measurable subset of $[0,1]$. (Note that $N = h^{-1}(A)$ is not measurable, so this shows that the inverse image of a Lebesgue measurable set under a continuous function need not be Lebesgue measurable.)

(d) Set $f = \chi_A$. Prove that $f \circ h$ is not a Lebesgue measurable function, even though f is Lebesgue measurable and h is continuous (compare this to Lemma 3.2.5).

Remark: Since h is continuous, the inverse image under h of an open set is open. It follows from this that the inverse image of any *Borel set* under h must be a Borel set (see Problem 2.3.25 for the definition of Borel sets). Since $N = h^{-1}(A)$ is not measurable and therefore is not a Borel set, we conclude that A is not a Borel set. Hence A is an example of a Lebesgue measurable set that is not a Borel set.

5.2 Functions of Bounded Variation

The Cantor–Lebesgue function φ is "unpleasant" in the sense that it is a singular function on $[0,1]$. However, it is quite nice in other ways, e.g., it is both continuous and monotone increasing on $[0,1]$. As x increases from 0 to 1, the value of $\varphi(x)$ increases monotonically from $\varphi(0) = 0$ to $\varphi(1) = 1$. Hence the total variation in the height of $\varphi(x)$ as x moves from 0 to 1 is simply $\varphi(1) - \varphi(0) = 1$. In contrast, at least intuitively it seems that the total variation in height of the function $f(x) = \sin(1/x)$ over the interval $[0,1]$ must be infinite. Our goal in this section is to make this idea of total variation precise, and to characterize the functions that have finite total variation in height. We say that these functions have "bounded variation." We will show that a real-valued function f has bounded variation on a finite interval $[a,b]$ if and only if we can write f in the form $f = g - h$ where g and h are each monotone increasing on $[a,b]$. Consequently, the space of functions that have bounded variation on $[a,b]$ is precisely the finite linear span of the set of monotone increasing functions.

5.2.1 Definition and Examples

First we must decide exactly what we mean by the variation of a function. We could consider the arc length of the graph of f as one measure of the variation. However, here we are interested purely in the variation in height. For example, the variation in height alone of both $f(x) = x$ and $g(x) = x^2$ over the interval $[0,1]$ is 1, but the arc lengths of the graphs of these two functions are different. We also want all variations in height, both upward and downward, to be counted positively. If f is either monotone increasing or monotone decreasing on $[a,b]$, then it is clear that the total variation in the height of f over the interval $[a,b]$ is $|f(b) - f(a)|$. However, if f is more complicated, then it is not so clear how we should define the total variation. Still, we can form an approximation to the variation by examining the values of $f(x)$ at finitely many points in the interval $[a,b]$. That is, if we fix finitely many points $a = x_0 < \cdots < x_n = b$, then we can think of the quantity $\sum_{j=1}^{n} |f(x_j) - f(x_{j-1})|$ as being an approximation to how much f varies in height over the interval $[a,b]$ (note that all variations are counted positively). We declare the total variation of f to be the supremum of all such approximations. Here is the precise definition.

Definition 5.2.1 (Bounded Variation). Let $f\colon [a,b] \to \mathbb{C}$ be given. For each finite partition

$$\Gamma \;=\; \bigl\{a = x_0 < \cdots < x_n = b\bigr\}$$

of $[a,b]$, set

$$S_\Gamma \;=\; S_\Gamma[f;a,b] \;=\; \sum_{j=1}^{n} |f(x_j) - f(x_{j-1})|. \tag{5.4}$$

The *total variation* of f over $[a,b]$ (or simply the *variation* of f, for short) is

$$V[f] \;=\; V[f;a,b] \;=\; \sup\bigl\{S_\Gamma : \Gamma \text{ is a partition of } [a,b]\bigr\}. \tag{5.5}$$

We say that f has *bounded variation* on $[a,b]$ if $V[f;a,b] < \infty$. We collect the functions that have bounded variation on $[a,b]$ to form the space

$$\mathrm{BV}[a,b] \;=\; \bigl\{f\colon [a,b] \to \mathbb{C} \;:\; f \text{ has bounded variation}\bigr\}. \qquad \diamondsuit$$

Remark 5.2.2. (a) We sometimes need to consider the case $a = b$; we declare that $V[f;a,a] = 0$ for every function f.

(b) By Problem 5.2.17, a complex-valued function has bounded variation if and only if its real and imaginary parts each have bounded variation. $\diamondsuit$

Since the total variation $V[f;a,b]$ is defined in equation (5.5) to be a supremum, for each particular partition Γ of $[a,b]$ we have $S_\Gamma \le V[f;a,b]$. Applying this inequality to the smallest possible partition $\Gamma = \{a < b\}$, we obtain

$$|f(b) - f(a)| \;=\; S_\Gamma \;\le\; V[f;a,b]. \tag{5.6}$$

On the other hand, setting $\Gamma = \{a < x < b\}$, we see that

$$|f(x) - f(a)| \;\le\; |f(x) - f(a)| \;+\; |f(b) - f(x)| \;=\; S_\Gamma \;\le\; V[f;a,b].$$

Consequently,

$$\|f\|_{\mathrm{u}} \;=\; \sup_{x\in[a,b]} |f(x)| \;\le\; V[f;a,b] \;+\; |f(a)|.$$

Thus every function that has bounded variation is bounded. However, we will see in Exercise 5.2.4 that there are bounded functions that have unbounded variation, so we have the proper inclusion

$$\mathrm{BV}[a,b] \;\subsetneq\; L^\infty[a,b].$$

According to Problem 5.2.19, $\mathrm{BV}[a,b]$ is closed under function addition and scalar multiplication (and several other operations). Hence $\mathrm{BV}[a,b]$ is a subspace of $L^\infty[a,b]$. It is not complete with respect to the L^∞-norm, but Problem 5.2.26 shows how to define a norm on $\mathrm{BV}[a,b]$ with respect to which it is a Banach space.

We will give several examples. First, we observe that Definition 5.2.1 is consistent with our earlier remarks about functions that are monotone increasing or decreasing.

Example 5.2.3. If $f\colon [a,b] \to \mathbb{R}$ is monotone increasing on $[a,b]$, then equation (5.4) becomes a telescoping sum, and hence $S_\Gamma = f(b) - f(a)$ for every partition Γ. Therefore f has bounded variation, and its total variation is precisely $V[f;a,b] = f(b) - f(a) = |f(b) - f(a)|$. Similarly, if f is monotone decreasing then $V[f;a,b] = |f(b) - f(a)|$. ◇

The Dirichlet function $\chi_{\mathbb{Q}}$ does not have bounded variation on any interval $[a,b]$. While $f(x) = \sin(1/x)$ is continuous on $(0,1]$, it does not have bounded variation on the interval $[0,1]$, no matter how we define it at $x = 0$. The next exercise will show that there exist continuous (and even differentiable!) functions that do not have bounded variation. As discussed in the Preliminaries, when we say that a function is differentiable on a closed interval $[a,b]$, we mean that it is differentiable on the interior (a,b) and the appropriate one-sided limits exist at the endpoints a and b.

Exercise 5.2.4. For $x \ne 0$ define

$$f(x) = x\sin\tfrac{1}{x}, \qquad g(x) = x^2\sin\tfrac{1}{x^2}, \qquad h(x) = x^2\sin\tfrac{1}{x},$$

and for $x = 0$ set $f(0) = g(0) = h(0) = 0$ (see Figure 5.3). Prove the following statements.

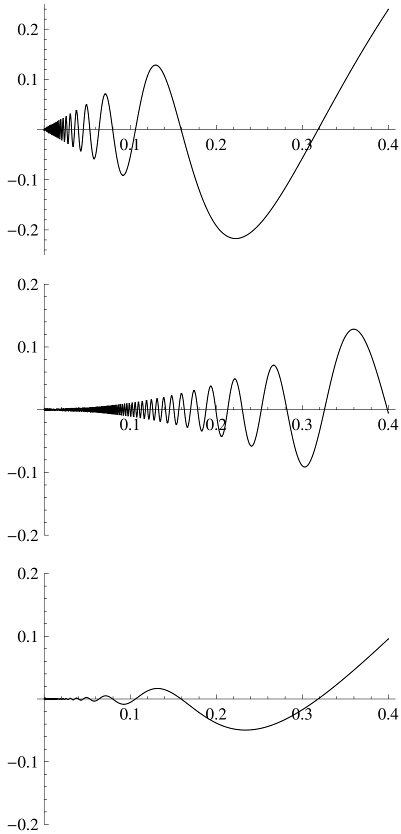


Fig. 5.3 The functions f (top), g (middle), and h (bottom) discussed in Exercise 5.2.4.

(a) f is continuous on $[-1,1]$, f is not differentiable at the point $x=0$, and $f \notin \mathrm{BV}[-1,1]$.

(b) g is differentiable everywhere on $[-1,1]$, $g \notin \mathrm{BV}[-1,1]$, g' is unbounded and therefore not continuous on $[-1,1]$, and $g' \notin L^1[-1,1]$.

(c) h is differentiable everywhere on $[-1,1]$, $h \in \mathrm{BV}[-1,1]$, h' is not continuous on $[-1,1]$, and $h' \in L^\infty[-1,1] \subseteq L^1[-1,1]$. $\diamondsuit$

Another interesting example is the function $k(x) = |x|^{3/2}\sin(1/x)$. According to Problem 6.4.19, k is differentiable on $[-1,1]$ and has bounded variation, while k' is integrable but unbounded. The properties of functions of the form $|x|^a \sin|x|^{-b}$ are studied in more detail in Problems 5.2.22, 6.3.13, and 6.4.20.

5.2.2 Lipschitz and Hölder Continuous Functions

Let I be an interval in the real line. Recall from Definition 1.4.1 that a function $f\colon I \to \mathbb{C}$ is *Hölder continuous with exponent* $\alpha > 0$ if there exists a constant $K \geq 0$ such that $|f(x) - f(y)| \leq K\,|x-y|^\alpha$ for all $x, y \in I$.

The larger that we can take α, the "smoother" that the graph of f typically appears. If we can take $\alpha = 1$ then we say that f is *Lipschitz continuous*, or simply that f is *Lipschitz*. Any number K such that

$$|f(x) - f(y)| \;\leq\; K\,|x-y|, \qquad \text{for all } x, y \in I, \tag{5.7}$$

is called a *Lipschitz constant* for f. We denote the class of Lipschitz functions on the interval I by

$$\mathrm{Lip}(I) \;=\; \bigl\{f\colon I \to \mathbb{C} \;:\; f \text{ is Lipschitz}\bigr\}.$$

By Problem 1.4.4, $f(x) = |x|^{1/2}$ is Hölder continuous but not Lipschitz on the real line, and Problem 5.1.5 shows that the Cantor–Lebesgue function φ is Hölder continuous but not Lipschitz on $[0,1]$. Here are some other examples.

- Some differentiable functions are Lipschitz, e.g., $f(x) = x$ is Lipschitz on every interval I.
- Not every differentiable function is Lipschitz, e.g., $f(x) = x^2$ is not Lipschitz on $I = \mathbb{R}$.
- A Lipschitz function need not be differentiable, e.g., $f(x) = |x|$ is Lipschitz on $I = \mathbb{R}$ but it is not differentiable at the origin.

A Lipschitz function need not be differentiable everywhere, but we will prove later that every Lipschitz function has *bounded variation* and therefore is differentiable at *almost every* point (see Lemma 5.2.7 and Corollary 5.4.3).

Suppose that we have a real-valued function $f\colon I \to \mathbb{R}$ that we know is differentiable everywhere on I. If $x \neq y \in I$, then the Mean Value Theorem implies that there is a point ξ between x and y such that $f(x) - f(y) = f'(\xi)\,(x-y)$. Therefore, *if f' is bounded* (say $|f'(t)| \leq K$ for $t \in I$), then

$$|f(x) - f(y)| \;=\; |f'(\xi)|\,|x-y| \;\leq\; K\,|x-y|.$$

Although the Mean Value Theorem only holds for real-valued functions, by applying the MVT to the real and imaginary parts of f, a similar result can be proved for complex-valued functions (this is Problem 1.4.2). This gives us the following sufficient condition for a function to be Lipschitz continuous (also compare Problem 6.4.10, which gives a *characterization* of Lipschitz continuity in terms of absolute continuity).

Lemma 5.2.5. *Let I be an interval in $\mathbb{R}$. If $f\colon I \to \mathbb{C}$ is differentiable everywhere on I and f' is bounded on I, then f is Lipschitz on I.* $\diamondsuit$

Let $C^1(I)$ be the set of all differentiable functions on I whose first derivatives are continuous, i.e.,

$$C^1(I) = \bigl\{f \in C(I) : f \text{ is differentiable on } I \text{ and } f' \in C(I)\bigr\}.$$

Specializing to the case $I = [a,b]$ (which is the setting we will mostly be working with in this chapter and the next), we obtain the following corollary.

Corollary 5.2.6. $C^1[a,b] \subsetneq \mathrm{Lip}[a,b]$.

Proof. If $f \in C^1[a,b]$, then f is differentiable and f' is continuous. Since the interval $[a,b]$ is compact, it follows that f' is bounded, so f is Lipschitz by Lemma 5.2.5. On the other hand, if we fix $a < t_0 < b$, then $g(x) = |x - t_0|$ is Lipschitz on $[a,b]$ but it does not belong to $C^1[a,b]$. $\square$

Now we prove that all Lipschitz functions have bounded variation.

Lemma 5.2.7. *If f is Lipschitz on $[a,b]$ and K is a Lipschitz constant for f, then f is uniformly continuous, f has bounded variation, and*

$$V[f;a,b] \le K\,(b-a). \tag{5.8}$$

Proof. All continuous functions on a compact domain are uniformly continuous, but we can also see this directly from equation (5.7).

If we fix any finite partition $\Gamma = \{a = x_0 < \cdots < x_n = b\}$, then

$$S_\Gamma = \sum_{j=1}^{n} |f(x_j) - f(x_{j-1})| \le \sum_{j=1}^{n} K\,(x_j - x_{j-1}) = K\,(b-a).$$

Taking the supremum over all such partitions yields equation (5.8). $\square$

Not every function that has bounded variation is Lipschitz. For example, $f(x) = |x|^{1/2}$ is not Lipschitz on $[0,1]$ (compare Problem 1.4.4), yet it is monotone increasing and therefore has bounded variation on that interval. Thus we have the proper inclusions

$$C^1[a,b] \subsetneq \mathrm{Lip}[a,b] \subsetneq \mathrm{BV}[a,b].$$

5.2.3 Indefinite Integrals and Antiderivatives

The following (easy) exercise is essentially the Fundamental Theorem of Calculus (FTC) that we learn in undergraduate calculus, stated here using our terminology.

Exercise 5.2.8 (Simple Version of the FTC). Prove that if g is a continuous function g on $[a,b]$, then its indefinite integral

$$G(x) = \int_a^x g(t)\,dt, \qquad x \in [a,b], \tag{5.9}$$

has the following properties:

(a) G is differentiable everywhere on $[a,b]$,

(b) $G'(x) = g(x)$ for every $x \in [a,b]$,

(c) $G \in C^1[a,b]$, so G is Lipschitz and has bounded variation on $[a,b]$. ♢

Thus, if g is continuous then its indefinite integral G is differentiable at every point, and it is an *antiderivative* of g because $G' = g$. What happens if we assume only that the function g is *integrable*? Here is a partial answer.

Lemma 5.2.9. *If $g \in L^1[a,b]$, then its indefinite integral*

$$G(x) = \int_a^x g(t)\,dt, \qquad x \in [a,b],$$

has the following properties:

(a) *G is continuous on $[a,b]$,*

(b) *$G \in \mathrm{BV}[a,b]$, and*

(c) *the total variation of G is bounded by the L^1-norm of g, i.e.,*

$$V[G;a,b] \le \int_a^b |g(t)|\,dt = \|g\|_1.$$

Proof. (a) Fix any point $x \in (a,b)$. If $h > 0$ is small enough that $x+h$ belongs to $[a,b]$, then

$$G(x+h) - G(x) = \int_x^{x+h} g(t)\,dt = \int_a^b g(t)\,\chi_{[x,x+h]}(t)\,dt.$$

The integrand $g \cdot \chi_{[x,x+h]}$ is bounded by the integrable function $|g|$, and it converges pointwise a.e. to zero as $h \to 0^+$. The Dominated Convergence Theorem therefore implies that $G(x+h) - G(x) \to 0$ as $h \to 0^+$. Combining this with a similar argument for $h \to 0^-$, we see that G is continuous at x. Similar one-sided arguments show that G is continuous from the right at $x = a$ and continuous from the left at $x = b$, so G is continuous on the interval $[a,b]$.

(b), (c) If $\Gamma = \{a = x_0 < \cdots < x_n = b\}$ is a partition of $[a,b]$, then

$$\begin{aligned} S_\Gamma &= \sum_{j=1}^n |G(x_j) - G(x_{j-1})| \\ &\le \sum_{j=1}^n \int_{x_{j-1}}^{x_j} |g(t)|\,dt = \int_a^b |g(t)|\,dt = \|g\|_1. \end{aligned}$$

Taking the supremum over all such partitions we see that G has bounded variation and $V[G;a,b] \le \|g\|_1$. $\square$

Remark 5.2.10. In the proof of Lemma 5.2.9 we applied the Dominated Convergence Theorem to a limit of the form $h \to 0$. Technically, we should note that the DCT stated in Theorem 4.5.1 only applies to sequences of functions indexed by the natural numbers. However, Problem 4.5.30 shows how to generalize the DCT to families indexed by a continuous parameter. $\diamondsuit$

Unfortunately, Lemma 5.2.9 is not very satisfactory when compared to Exercise 5.2.8. We are still left with the following questions.

- If $g \in L^1[a,b]$, is the indefinite integral $G(x) = \int_a^x g(t)\,dt$ a differentiable function of x?
- If the indefinite integral G is differentiable, is it the antiderivative of g? That is, is it true that $G' = g$?

The answers to these questions are **not obvious** at this point. In Chapter 6 we will see that:

- G is an *absolutely continuous* function and, as a consequence, it is differentiable at *almost every* point in $[a,b]$, and
- $G'(x) = g(x)$ for *almost every* $x \in [a,b]$.

The definition of absolute continuity will be given in Section 6.1. After we develop some machinery, we will prove that G is absolutely continuous and therefore differentiable a.e. (see Lemma 6.1.6), and $G' = g$ a.e. (Theorem 6.4.2). Furthermore, we will establish the converse fact that *every* absolutely continuous function is the indefinite integral of its derivative (Theorem 6.4.2). However, there is still work to do before we can prove these statements.

5.2.4 The Jordan Decomposition

Our next goal is to prove that every real-valued function that has bounded variation can be written as the difference of two monotone increasing functions. Before doing this, we need to develop some tools and introduce some additional terminology. We begin with an exercise that gives some of the basic properties of the variation function $V[f;a,b]$. In part (b) of this exercise, a *refinement* of a partition Γ means a partition Γ' that includes all of the points that are in Γ. Note that we are not assuming here that f has bounded variation—it is possible that $V[f;a,b]$ could be infinite.

Exercise 5.2.11. Given $f\colon [a,b] \to \mathbb{C}$, prove the following statements.

(a) $|f(b) - f(a)| \le V[f;a,b]$.

(b) If $\Gamma = \{a = x_0 < \cdots < x_n = b\}$ is a partition of $[a,b]$ and Γ' is a refinement of Γ, then $S_\Gamma \le S_{\Gamma'}$.

(c) If $[c,d] \subseteq [a,b]$, then $V[f;c,d] \le V[f;a,b]$. $\diamondsuit$

We will also need the following additivity property of the total variation.

Lemma 5.2.12. *If $f\colon [a,b] \to \mathbb{C}$ and $a < c < b$, then*

$$V[f;a,b] \;=\; V[f;a,c] \,+\, V[f;c,b].$$

Proof. Suppose that $a < c < b$. Let $\Gamma_1 = \{a = x_0 < \cdots < x_m = c\}$ and $\Gamma_2 = \{c = x_m < \cdots < x_n = b\}$ be finite partitions of $[a,c]$ and $[c,b]$, respectively. Then $\Gamma = \Gamma_1 \cup \Gamma_2$ is a partition of $[a,b]$, and

$$S_{\Gamma_1} + S_{\Gamma_2} \;=\; S_\Gamma \;\le\; V[f;a,b].$$

Holding Γ_2 fixed and taking the supremum over all partitions Γ_1 of $[a,c]$ gives us $V[f;a,c] + S_{\Gamma_2} \le V[f;a,b]$. Taking next the supremum over all partitions Γ_2 of $[c,d]$, we obtain $V[f;a,c] + V[f;c,b] \le V[f;a,b]$.

For the opposite inequality, let $\Gamma = \{a = x_0 < \cdots < x_n = b\}$ be any finite partition of $[a,b]$. There are two possibilities. If $x_j < c < x_{j+1}$ for some j, then

$$\Gamma_1 \;=\; \{a = x_0 < \cdots < x_j < c\} \quad \text{and} \quad \Gamma_2 \;=\; \{c < x_{j+1} < \cdots < x_n = b\}$$

are partitions of $[a,c]$ and $[c,b]$, respectively. Further, $\Gamma' = \Gamma_1 \cup \Gamma_2$ is a partition of $[a,b]$ and Γ' is a refinement of Γ, so

$$S_\Gamma \;\le\; S_{\Gamma'} \;=\; S_{\Gamma_1} + S_{\Gamma_2} \;\le\; V[f;a,c] + V[f;c,b].$$

On the other hand, if $c = x_j$ for some j then a similar argument shows that we also have $S_\Gamma \le V[f;a,c] + V[f;c,b]$ in this case. Taking the supremum over all partitions Γ, we conclude that $V[f;a,b] \le V[f;a,c] + V[f;c,b]$. $\square$

In order to obtain monotone increasing functions that are related to the variation of a real-valued function f, we break the total variation of f into a "positive part" and a "negative part." However, we do not accomplish this by splitting f into its positive and negative parts, but rather by splitting each term $y_j = f(x_j) - f(x_{j-1})$ into the positive and negative parts

$$y_j^+ \;=\; \max\{y_j, 0\} \qquad \text{and} \qquad y_j^- \;=\; \max\{-y_j, 0\}.$$

Note that $y_j^+ - y_j^- = y_j$ and $y_j^+ + y_j^- = |y_j|$.

Definition 5.2.13 (Positive and Negative Variation). Let $f\colon [a,b] \to \mathbb{R}$ be a real-valued function on $[a,b]$. For each finite partition $\Gamma = \{a = x_0 < \cdots < x_n = b\}$ of $[a,b]$, define

$$S_\Gamma^+ \;=\; \sum_{j=1}^n \Bigl(f(x_j)-f(x_{j-1})\Bigr)^+ \quad \text{and} \quad S_\Gamma^- \;=\; \sum_{j=1}^n \Bigl(f(x_j)-f(x_{j-1})\Bigr)^-.$$

The *positive variation* of f on $[a,b]$ is

$$V^+[f;a,b] \;=\; \sup\bigl\{S_\Gamma^+ : \Gamma \text{ is a partition of } [a,b]\bigr\},$$

and the *negative variation* is

$$V^-[f;a,b] \;=\; \sup\bigl\{S_\Gamma^- : \Gamma \text{ is a partition of } [a,b]\bigr\}. \qquad \diamondsuit$$

Comparing Definitions 5.2.1 and 5.2.13, we see that for each partition Γ we have

$$S_\Gamma^+ + S_\Gamma^- \;=\; S_\Gamma \qquad \text{and} \qquad S_\Gamma^+ - S_\Gamma^- \;=\; f(b)-f(a). \tag{5.10}$$

The next result extends these equalities from individual partitions to the variation functions. Note that we are not assuming in this lemma that f has bounded variation, so V, V^+, or V^- might be infinite.

Lemma 5.2.14. *If $f\colon [a,b]\to\mathbb{R}$, then, as extended real numbers,*

$$V^+[f;a,b] + V^-[f;a,b] \;=\; V[f;a,b].$$

Further, if any one of $V[f;a,b]$, $V^+[f;a,b]$, or $V^-[f;a,b]$ is finite then the other two are finite as well, and in this case

$$V^+[f;a,b] \;-\; V^-[f;a,b] \;=\; f(b)-f(a). \tag{5.11}$$

Proof. For every partition Γ we have $S_\Gamma^+ = S_\Gamma^- + C$, where C is the fixed, finite constant $C = f(b)-f(a)$. Hence, even if they are infinite,

$$V^+[f;a,b] \;=\; \sup_\Gamma S_\Gamma^+ \;=\; \sup_\Gamma \bigl(S_\Gamma^- + C\bigr) \;=\; V^-[f;a,b] + C.$$

In particular, $V^+[f;a,b]$ is finite if and only if $V^-[f;a,b]$ is finite.

Similarly,

$$S_\Gamma \;=\; S_\Gamma^+ + S_\Gamma^- \;=\; \bigl(S_\Gamma^- + C\bigr) + S_\Gamma^- \;=\; 2S_\Gamma^- + C,$$

so

$$V[f;a,b] \;=\; \sup_\Gamma S_\Gamma \;=\; \sup_\Gamma \bigl(2S_\Gamma^- + C\bigr) \;=\; 2V^-[f;a,b] + C.$$

Hence $V[f;a,b]$ is finite if and only if $V^-[f;a,b]$ is finite.

Finally, by combining the above equalities we see that, even if they are infinite,

$$V^+[f;a,b] + V^-[f;a,b] \;=\; 2V^-[f;a,b] + C \;=\; V[f;a,b]. \qquad \square$$

Now we prove the *Jordan decomposition*, which expresses a real-valued function with bounded variation as the difference of two monotone increasing functions. Except for an additive constant, these two monotone increasing functions are $V^+[f;a,x]$ and $V^-[f;a,x]$, the positive and negative variations of f on the interval $[a,x]$. Each of these variations increases with x, and we see from equation (5.11) that their difference is precisely $f(x) - f(a)$.

Theorem 5.2.15 (Jordan Decomposition). *If $f\colon [a,b] \to \mathbb{R}$, then the following two statements are equivalent.*

(a) $f \in \mathrm{BV}[a,b]$.

(b) *There exist monotone increasing functions g and h such that $f = g - h$.*

Proof. (a) $\Rightarrow$ (b). Assume that f has bounded variation on $[a,b]$, and set

$$g(x) \;=\; V^+[f;a,x] \;+\; f(a) \qquad \text{and} \qquad h(x) \;=\; V^-[f;a,x]$$

for $x \in [a,b]$. Exercise 5.2.11(c) implies that g and h are each monotonically increasing, and by Lemma 5.2.14 we have

$$g(x) \,-\, h(x) \;=\; V^+[f;a,x] \,+\, f(a) \,-\, V^-[f;a,x] \;=\; f(x).$$

(b) $\Rightarrow$ (a). This implication follows from Problem 5.2.19. $\square$

Applying Theorem 5.2.15 to the real and imaginary parts of a complex-valued function, we obtain the following corollary.

Corollary 5.2.16. *A function $f\colon [a,b] \to \mathbb{C}$ belongs to* $\mathrm{BV}[a,b]$ *if and only if there exist monotone increasing functions f_1, f_2, f_3, f_4 on $[a,b]$ such that*

$$f \;=\; (f_1 - f_2) \,+\, i(f_3 - f_4). \qquad \diamondsuit$$

As a consequence, the space of functions with bounded variation is precisely the finite linear span of the monotone increasing functions:

$$\mathrm{BV}[a,b] \;=\; \mathrm{span}\bigl\{f\colon[a,b] \to \mathbb{C} \;:\; f \text{ is monotone increasing on } [a,b]\bigr\}.$$

Thus, in order to make further progress understanding the properties of functions of bounded variation, we need to understand monotone increasing functions. To this end we will derive some useful tools in Section 5.3, and then in Section 5.4 we will show that a monotone increasing function can have at most *countably many* discontinuities and is differentiable at almost every point.

Problems

5.2.17. Given a function $f\colon [a,b] \to \mathbb{C}$, write $f = f_r + i f_i$ where f_r and f_i are real-valued. Prove that $f \in \mathrm{BV}[a,b]$ if and only if $f_r,\, f_i \in \mathrm{BV}[a,b]$.

5.2.18. Suppose that $f\colon [a,b] \to \mathbb{C}$. Show that there exist partitions Γ_k of $[a,b]$ such that Γ_{k+1} is a refinement of Γ_k for each k and $S_{\Gamma_k} \nearrow V[f;a,b]$ as $k \to \infty$.

5.2.19. Prove that if f and g belong to $\mathrm{BV}[a,b]$, then the following statements hold.

(a) $V^+[f;a,b] = \frac{1}{2}\big(V[f;a,b] + f(b) - f(a)\big)$.

(b) $V^-[f;a,b] = \frac{1}{2}\big(V[f;a,b] - f(b) + f(a)\big)$.

(c) $|f| \in \mathrm{BV}[a,b]$.

(d) $\alpha f + \beta g \in \mathrm{BV}[a,b]$ for all $\alpha, \beta \in \mathbb{C}$, and

$$V[\alpha f + \beta g; a,b] \;\le\; |\alpha|\, V[f;a,b] + |\beta|\, V[g;a,b].$$

(e) $fg \in \mathrm{BV}[a,b]$.

(f) If $|g(x)| \ge \delta > 0$ for all $x \in [a,b]$, then $f/g \in \mathrm{BV}[a,b]$.

5.2.20. Given functions $g\colon [a,b] \to [c,d]$ and $f\colon [c,d] \to \mathbb{C}$, prove the following statements.

(a) If f is Lipschitz and $g \in \mathrm{BV}[a,b]$, then $f \circ g \in \mathrm{BV}[a,b]$. However, this can fail if we assume only that f is continuous, even if f is continuous and has bounded variation.

(b) If $f \in \mathrm{BV}[c,d]$ and g is monotone increasing on $[a,b]$, then $f \circ g \in \mathrm{BV}[a,b]$.

Remark: This problem will be used in the proof of Corollary 6.5.5.

5.2.21. Assume that $E \subseteq \mathbb{R}$ is measurable, and suppose that $f\colon E \to \mathbb{R}$ is Lipschitz *on the set* E, i.e., there exists a constant $K \ge 0$ such that

$$|f(x) - f(y)| \;\le\; K\,|x-y|, \qquad \text{for all } x,\, y \in E.$$

Prove that $|f(A)|_e \;\le\; K\,|A|_e$ for every set $A \subseteq E$.

5.2.22. Fix $a, b > 0$, and define

$$f(x) \;=\; \begin{cases} |x|^a \sin |x|^{-b}, & \text{if } x \ne 0, \\ 0, & \text{if } x = 0. \end{cases}$$

Prove the following statements (the space $C^\alpha[-1,1]$ is defined in Problem 1.4.5).

(a) $f \in \mathrm{BV}[-1,1]$ if and only if $a > b$.

(b) If $a = b$, then $f \in C^\alpha[-1,1]$ with exponent $\alpha = \frac{b}{b+1}$, even though part (a) implies that f does not have bounded variation.

(c) $C^\alpha[-1,1]$ is not contained in $\mathrm{BV}[-1,1]$ for any $0 < \alpha < 1$.

5.2.23. (a) Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of complex-valued functions on $[a,b]$, and $f_n \to f$ pointwise on $[a,b]$. Prove that

$$V[f;a,b] \;\le\; \liminf_{n\to\infty} V[f_n;a,b].$$

(b) Exhibit functions f_n and f such that $f_n \in \mathrm{BV}[a,b]$ for each $n \in \mathbb{N}$ and $f_n \to f$ pointwise, but $f \notin \mathrm{BV}[a,b]$.

5.2.24. Fix $f \in \mathrm{BV}[a,b]$, and extend f to the real line by setting $f(x) = f(a)$ for $x < a$ and $f(x) = f(b)$ for $x > b$. Prove that there exists a constant $C > 0$ such that

$$\|T_t f - f\|_1 \;\le\; C|t|, \qquad \text{for all } t \in \mathbb{R},$$

where $T_t f(x) = f(x-t)$ denotes the translation of f by t.

5.2.25. Given functions $f_k \in \mathrm{BV}[a,b]$, suppose that $f(x) = \sum_{k=1}^{\infty} f_k(x)$ converges for each $x \in [a,b]$ and $\sum_{k=1}^{\infty} V[f_k;a,b] < \infty$. Prove that f has bounded variation, and

$$V[f;a,b] \;\le\; \sum_{k=1}^{\infty} V[f_k;a,b].$$

5.2.26. Prove the following statements.

(a) $\|f\| = V[f;a,b]$ defines a seminorm on $\mathrm{BV}[a,b]$, and

$$\|f\|_{\mathrm{BV}} \;=\; V[f;a,b] \;+\; \|f\|_{\mathrm{u}}, \quad \text{for } f \in \mathrm{BV}[a,b],$$

is a norm on $\mathrm{BV}[a,b]$.

(b) $\mathrm{BV}[a,b]$ is a Banach space with respect to $\|\cdot\|_{\mathrm{BV}}$.

(c) $|||f|||_{\mathrm{BV}} = V[f;a,b] + |f(a)|$ defines an *equivalent norm* for $\mathrm{BV}[a,b]$, i.e., it is a norm and there exist constants C_1, $C_2 > 0$ such that

$$C_1\,\|f\|_{\mathrm{BV}} \;\le\; |||f|||_{\mathrm{BV}} \;\le\; C_2\,\|f\|_{\mathrm{BV}}, \quad \text{for all } f \in \mathrm{BV}[a,b].$$

5.2.27.* Prove that if $f \in \mathrm{BV}[a,b]$ is continuous, then the following statements hold.

(a) $V[f;a,b] = \lim_{|\Gamma|\to 0} S_\Gamma$.

(b) $V(x) = V[f;a,x]$ is a continuous function on $[a,b]$.

(c) If $f \in C^1[a,b]$, then $V[f;a,b] = \int_a^b |f'|$.

5.3 Covering Lemmas

Suppose that we have a collection of open balls, or cubes, or some other type of reasonably nice sets that cover a set E. We might have infinitely many of these sets, maybe even uncountably many. Many of these sets may intersect. Is it possible to extract some subcollection of sets that are *disjoint* and *still cover* E? In general, this will not be possible, but perhaps we can weaken our goal a little and find a subcollection of disjoint sets that at least covers some prescribed fraction of E. This type of result is called a *covering lemma.* We will prove two such covering lemmas in this section.

5.3.1 The Simple Vitali Lemma

We begin with the *Simple Vitali Lemma*, which states that if we are given any collection of open balls in $\mathbb{R}^d$, then we can find *finitely many disjoint balls* from the collection that cover a fixed fraction of the measure of the union of the original balls. Up to an ε, this fraction is 3^{-d} (so in dimension $d = 1$, we can choose disjoint open intervals that cover about $1/3$ of the original collection). The proof is an example of a *greedy algorithm*—essentially we let B_1 be the largest possible ball in the original collection, then choose B_2 to be the largest possible ball that is disjoint from B_1, and so forth.

Theorem 5.3.1 (Simple Vitali Lemma). *Let $\mathcal{B}$ be any nonempty collection of open balls in $\mathbb{R}^d$. Let U be the union of all of the balls in $\mathcal{B}$, and fix $0 < c < |U|$. Then there exist finitely many* disjoint *balls $B_1, \dots, B_N \in \mathcal{B}$ such that*

$$\sum_{k=1}^{N} |B_k| \;>\; \frac{c}{3^d}.$$

Proof. Note that the number c is finite, even if $|U| = \infty$. Since $c < |U|$, Problem 2.3.20 implies that there exists a compact set $K \subseteq U$ such that

$$c \;<\; |K| \;\le\; |U|.$$

Since $\mathcal{B}$ is an open cover of K, we can find finitely many balls $A_1, \dots, A_m \in \mathcal{B}$ such that

$$K \;\subseteq\; \bigcup_{j=1}^{m} A_j.$$

Let B_1 be an A_j ball that has maximal radius.

If there are no A_j balls that are disjoint from B_1, then we set $N = 1$ and stop. Otherwise, let B_2 be an A_j ball with largest radius that is disjoint from B_1 (if there is more than one such ball, just choose one of them). We then repeat this process, which must eventually stop, to select disjoint balls $B_1, \dots, B_N$ from $A_1, \dots, A_m$. These balls need not cover K, but we hope that they will cover an appropriate portion of K.

To prove this, let B_k^* denote the open ball that has the same center as B_k, but with radius three times larger. Suppose that $1 \le j \le m$, but A_j is not one of $B_1, \ldots, B_N$. Then A_j must intersect at least one of the balls $B_1, \ldots, B_N$. Let k be the smallest index such that $A_j \cap B_k \neq \varnothing$. By construction,

$$\operatorname{radius}(A_j) \leq \operatorname{radius}(B_k).$$

It follows from this that $A_j \subseteq B_k^*$ (see the "proof by picture" in Figure 5.4).

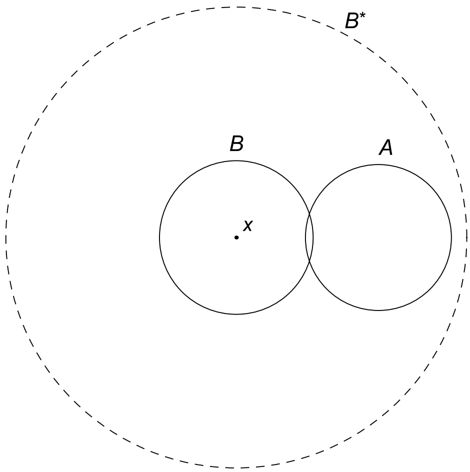


Fig. 5.4 Circle B has radius 1, circle A has radius 0.95, and circle B^* (which has the same center x as circle B) has radius 3.

The preceding paragraph tells us that every set A_j that is not one of $B_1, \ldots, B_N$ is contained in some B_k^*. Hence

$$K \subseteq \bigcup_{j=1}^{m} A_j \subseteq \bigcup_{k=1}^{N} B_k^*,$$

and therefore

$$c < |K| \leq \sum_{k=1}^{N} |B_k^*| = 3^d \sum_{k=1}^{N} |B_k|. \quad \square$$

5.3.2 The Vitali Covering Lemma

Given an arbitrary collection of open balls with union U, the Simple Vitali Lemma tells us that we can find disjoint open balls from the collection that cover a prescribed fraction of U. In general we will not be able to cover all of U with disjoint sets. Our next result, also due to Vitali, shows that if we impose more conditions on our collection, then we can draw a much stronger conclusion. We will use closed balls for this result, and we will assume that every point x of a set E is covered not just by one ball from our collection, but by infinitely many balls whose radii shrink to zero. Using these hypotheses, we will be able to prove that we can select disjoint balls that cover all of E except for a set of measure ε.

To formulate this precisely, we define the *closed ball* centered at x with radius r to be

$$\overline{B_r(x)} \;=\; \{y \in \mathbb{R}^d : \|x-y\| \le r\}.$$

We let radius(B) denote the radius r of a closed ball $B = \overline{B_r(x)}$. Here is the precise requirement that we will need to impose on our collection of balls.

Definition 5.3.2 (Vitali Cover). A collection $\mathcal{B}$ of closed balls is a *Vitali cover* of a set $E \subseteq \mathbb{R}^d$ if for each $x \in E$ and $\varepsilon > 0$ there exists some ball $B \in \mathcal{B}$ such that $x \in B$ and radius$(B) < \varepsilon$. $\diamondsuit$

We prove now that if we have a Vitali covering, then there are finitely many disjoint balls that cover all of E except possibly for a set of measure ε. Moreover, although these balls might include points outside of E, we can select them in such a way that the measure of their union is only slightly more than the measure of E.

Theorem 5.3.3 (Vitali Covering Lemma). *Let E be a subset of $\mathbb{R}^d$ with $0 < |E|_e < \infty$. If $\mathcal{B}$ is a Vitali covering of E, then for each $\varepsilon > 0$ there exist disjoint balls $B_1, \dots, B_N \in \mathcal{B}$ such that*

$$\biggl|E \setminus \bigcup_{k=1}^{N} B_k\biggr|_e < \varepsilon \qquad \textit{and} \qquad \sum_{k=1}^{N} |B_k| < |E|_e + \varepsilon. \tag{5.12}$$

Proof. Let $U \supseteq E$ be an open set such that $|U| < |E|_e + \varepsilon$. Remove all balls from $\mathcal{B}$ that are not contained in U; this still leaves us with a Vitali cover of E. We proceed to choose balls inductively from $\mathcal{B}$, using a modification of the greedy approach.

The first ball is arbitrary; we choose any ball $B_1 \in \mathcal{B}$. For the inductive step, once disjoint balls $B_1, \dots, B_n \in \mathcal{B}$ have been chosen, we proceed as follows.

If $|E \setminus (B_1 \cup \cdots \cup B_n)|_e = 0$, then we stop. The proof is complete in this case, because by additivity we have $\sum |B_k| = |\cup B_k| \le |U| < |E|_e + \varepsilon$.

Otherwise, we must keep going and somehow select a new ball B_{n+1} that is disjoint from $B_1, \dots, B_n$. We know that there exist some balls in $\mathcal{B}$ disjoint from $B_1, \dots, B_n$ because $\mathcal{B}$ is a Vitali cover. Specifically, since

$$E \setminus (B_1 \cup \cdots \cup B_n)$$

has positive measure, it contains a point x. This x belongs to the open set $U \setminus (B_1 \cup \cdots \cup B_n)$ and there are balls with arbitrarily small radius in $\mathcal{B}$ that contain x, so if we choose the radius small enough then we will have a ball that contains x and is disjoint from $B_1, \dots, B_n$. But there could be many such balls—which of them should we choose? In contrast to the proof of Theorem 5.3.1, there need not be a ball with largest radius. So, although we can define

$$s_n = \sup\bigl\{\text{radius}(B) : B \in \mathcal{B} \text{ and } B \text{ is disjoint from } B_1, \dots, B_n\bigr\},$$

this supremum need not be achieved. Therefore we settle for being "sufficiently greedy" in the sense that we choose a ball B_{n+1} that is disjoint from $B_1, \dots, B_n$ and has radius more than half of this supremum, i.e.,

$$\text{radius}(B_{n+1}) > \frac{s_n}{2}.$$

If this process stops after finitely many steps, then the proof is finished. Otherwise, we will continue forever, obtaining countably many disjoint closed balls $B_1, B_2, \dots$. These balls are contained in U, so

$$\sum_{k=1}^{\infty} |B_k| = \biggl|\bigcup_{k=1}^{\infty} B_k\biggr| \le |U| < |E|_e + \varepsilon < \infty.$$

Consequently $|B_k| \to 0$, and therefore $\text{radius}(B_k) \to 0$, as $k \to \infty$.

Fix an integer $N \in \mathbb{N}$, and suppose that x belongs to $E \setminus \bigcup_{k=1}^{N} B_k$. Then $x \in U$ but $x \notin B_1, \dots, B_N$, so x belongs to the open set

$$U_N = U \setminus (B_1 \cup \cdots \cup B_N).$$

Since $\mathcal{B}$ is a Vitali cover, there exists a ball $B \in \mathcal{B}$ that contains x and is disjoint from $B_1, \dots, B_N$.

Suppose that B was disjoint from $B_1, \dots, B_k$ for *every* $k \in \mathbb{N}$. Then, given how we constructed B_{k+1}, we must have

$$\text{radius}(B_{k+1}) \ge \frac{1}{2}\,\text{radius}(B). \tag{5.13}$$

Hence $\text{radius}(B) \le 2\,\text{radius}(B_{k+1}) \to 0$, which is a contradiction. Therefore B must intersect at least one ball B_k.

Let $n \in \mathbb{N}$ be the smallest integer such that B is disjoint from $B_1, \dots, B_{n-1}$ but $B \cap B_n \ne \varnothing$ (note that $n > N$). Just as in equation (5.13),

$$\text{radius}(B) \le 2\,\text{radius}(B_n). \tag{5.14}$$

Let B_k^* denote the closed ball that has the same center as B_k but with 5 times the radius. Since B intersects B_n and equation (5.14) holds, an argument similar to the one illustrated in Figure 5.4 shows that $B \subseteq B_n^*$. Consequently, $x \in B \subseteq B_n^*$ where $n > N$, so we have shown that

$$E \setminus \bigcup_{k=1}^{N} B_k \subseteq \bigcup_{k>N} B_k^*.$$

Therefore

$$\left| E \setminus \bigcup_{k=1}^{N} B_k \right|_e \le \sum_{k=N+1}^{\infty} |B_k^*| = 5^d \sum_{k=N+1}^{\infty} |B_k|.$$

Since $\sum_{k=1}^{\infty} |B_k| < \infty$, by choosing N large enough we will obtain

$$\left| E \setminus \bigcup_{k=1}^{N} B_k \right|_e < \varepsilon. \qquad \square$$

We could have used closed cubes instead of closed balls in Theorem 5.3.3. The proof would be identical, except that we would work with sidelengths instead of radii.

Remark 5.3.4. We can derive some further conclusions from equation (5.12) by applying Carathéodory's Criterion. Specifically, if equation (5.12) holds, then

$$\begin{aligned} |E|_e &= \left| E \cap \bigcup_{k=1}^{N} B_k \right|_e + \left| E \setminus \bigcup_{k=1}^{N} B_k \right|_e && \text{(by Carathéodory)} \\ &< \left| E \cap \bigcup_{k=1}^{N} B_k \right|_e + \varepsilon && \text{(by equation (5.12))}, \end{aligned}$$

and therefore

$$\sum_{k=1}^{N} |B_k| = \left| \bigcup_{k=1}^{N} B_k \right|_e \ge \left| E \cap \bigcup_{k=1}^{N} B_k \right|_e > |E|_e - \varepsilon. \tag{5.15}$$

These inequalities will be useful to us when we prove Theorem 5.4.2. $\diamondsuit$

Problems

5.3.5. Assume that $E \subseteq \mathbb{R}^d$ satisfies $0 < |E|_e < \infty$, and let $\mathcal{B}$ be a Vitali covering of E. Given $\varepsilon > 0$, prove that there exist countably many disjoint balls $B_k \in \mathcal{B}$ such that

$$\left| E \setminus \bigcup_{k} B_k \right|_e = 0 \qquad \text{and} \qquad \sum_{k} |B_k| < |E|_e + \varepsilon.$$

5.4 Differentiability of Monotone Functions

In this section we will prove that a monotone increasing function on $[a, b]$ is differentiable at almost every point of the interval. This fact, which may seem to be "obvious," takes a surprising amount of work to prove. We will need to use the Vitali Covering Lemma, and also make use of the following notions.

Definition 5.4.1 (Dini Numbers). Let f be a real-valued function on a set $E \subseteq \mathbb{R}$. If x is an interior point of E (so f is defined on an open interval containing x), then the four *Dini numbers* or *derivates* of f at x are

$$\begin{aligned} D^+f(x) &= \limsup_{h \to 0^+} \frac{f(x+h)-f(x)}{h}, \\ D_+f(x) &= \liminf_{h \to 0^+} \frac{f(x+h)-f(x)}{h}, \\ D^-f(x) &= \limsup_{h \to 0^-} \frac{f(x+h)-f(x)}{h}, \\ D_-f(x) &= \liminf_{h \to 0^-} \frac{f(x+h)-f(x)}{h}. \qquad \diamondsuit \end{aligned}$$

We always have $D_+f(x) \le D^+f(x)$ and $D_-f(x) \le D^-f(x)$. The function f is differentiable at $x \in E^\circ$ if and only if all four Dini numbers are equal and finite.

Now we prove that all monotone increasing functions are differentiable a.e. Further, although we know that the Fundamental Theorem of Calculus need not hold for monotone increasing functions (the Cantor–Lebesgue function is a counterexample), we prove that the integral of f' satisfies a certain *inequality* when f is monotone increasing.

Theorem 5.4.2 (Differentiability of Monotone Functions). *If a function $f\colon [a, b] \to \mathbb{R}$ is monotone increasing, then the following statements hold.*

(a) *f has at most countably many discontinuities, and they are all jump discontinuities.*

(b) *$f'(x)$ exists for almost every $x \in [a, b]$.*

(c) *f' is measurable and $f' \ge 0$ a.e.*

(d) *$f' \in L^1[a, b]$, and*

$$0 \le \int_a^b f' \le f(b) - f(a). \tag{5.16}$$

Proof. For simplicity of presentation, extend the domain of f to the entire real line by setting $f(x) = f(a)$ for $x < a$ and $f(x) = f(b)$ for $x > b$.

(a) Since f is monotone increasing and takes real values at each point of $[a,b]$, it follows that f is bounded on $[a,b]$ and the one-sided limits

$$f(x-) \;=\; \lim_{y\to x^-} f(y) \qquad \text{and} \qquad f(x+) \;=\; \lim_{y\to x^+} f(y)$$

exist at every point $x \in (a,b)$. The appropriate one-sided limits also exist at the points a and b. Consequently, each point of discontinuity of f must be a jump discontinuity. Since f is bounded, if we fix a positive integer k, then there can be at most finitely many points $x \in [a,b]$ such that

$$f(x+) \;-\; f(x-) \;\geq\; \frac{1}{k}.$$

Since every jump discontinuity must satisfy this inequality for some integer $k \in \mathbb{N}$, we conclude that there can be at most countably many discontinuities.

(b) For the proof of this part we will implicitly restrict our attention to points in the open interval (a,b). Since f is monotone increasing, Problem 5.4.9 shows that each of the four Dini numbers of f are finite a.e. on (a,b). Let

$$S \;=\; \{D^+f > D_-f\} \;=\; \bigl\{x \in (a,b) \,:\, D^+f(x) > D_-f(x)\bigr\}.$$

We will prove that S has measure zero. A similar argument works for any other pair of Dini numbers, so this will show that all four Dini numbers are equal for a.e. x.

Since f is monotone increasing, each Dini number is nonnegative. Let $0 < s < r$ be fixed rational numbers, and set

$$A \;=\; \bigl\{D_-f < s < r < D^+f\bigr\}.$$

Consider the collection of closed intervals

$$\mathcal{B} \;=\; \biggl\{[x-h,x] \subseteq (a,b) \,:\, x \in (a,b),\; h > 0, \text{ and } \frac{f(x-h)-f(x)}{-h} < s\biggr\}.$$

If $x \in A$ then $D_-f(x) < s$, so by the definition of liminf there must exist arbitrarily small values of $h > 0$ such that

$$\frac{f(x-h)-f(x)}{-h} \;<\; s. \tag{5.17}$$

This need not be true for all $h > 0$, but there must at least exist a sequence of values of h that tend to zero for which equation (5.17) holds. For each of these particular h the closed interval $[x-h,x]$ belongs to $\mathcal{B}$. This shows that $\mathcal{B}$ is a Vitali covering of the set A.

Fix $\varepsilon > 0$. By the Vitali Covering Lemma and one of the extra conclusions that appear in equation (5.15), there exist finitely many disjoint intervals in $\mathcal{B}$, say $I_n = [x_n - h_n, x_n]$ for $n = 1, \dots, N$, such that

$$\left| A \cap \bigcup_{n=1}^{N} I_n \right|_e \; > \; |A|_e - \varepsilon \qquad \text{and} \qquad \sum_{n=1}^{N} h_n \; < \; |A|_e + \varepsilon. \tag{5.18}$$

Since each interval $I_n = [x_n - h_n, x_n]$ belongs to $\mathcal{B}$, we have

$$\frac{f(x_n) - f(x_n - h_n)}{h_n} \; = \; \frac{f(x_n - h_n) - f(x_n)}{-h_n} \; < \; s.$$

Therefore

$$\sum_{n=1}^{N} \Big(f(x_n) - f(x_n - h_n) \Big) \; < \; s \sum_{n=1}^{N} h_n \; < \; s \left(|A|_e + \varepsilon \right). \tag{5.19}$$

Let

$$B \; = \; A \cap \bigcup_{n=1}^{N} I_n.$$

By equation (5.18) we have $|B|_e > |A|_e - \varepsilon$. If $y \in B$ then $y \in A$ and $y \in I_n$ for some n. We have $D^+ f(y) > r$, so by the definition of limsup there exist infinitely many values of k that tend to zero such that

$$\frac{f(y+k) - f(y)}{k} \; > \; r.$$

Proceeding similarly to before, we construct a Vitali cover of B and apply the Vitali Covering Lemma to infer the existence of disjoint intervals $J_m = [y_m, y_m + k_m]$ for $m = 1, \dots, M$ such that each J_m is contained in some I_n and

$$\sum_{m=1}^{M} k_m \; = \; \left| \bigcup_{m=1}^{M} J_m \right| \; \ge \; \left| B \cap \bigcup_{m=1}^{M} J_m \right|_e \; > \; |B|_e - \varepsilon \; > \; |A|_e - 2\varepsilon.$$

Since each interval $J_m = [y_m, y_m + k_m]$ belongs to $\mathcal{B}$, we have

$$\frac{f(y_m + k_m) - f(y_m)}{k_m} \; > \; r,$$

and therefore

$$\sum_{m=1}^{M} \Big(f(y_m + k_m) - f(y_m) \Big) \; > \; r \sum_{m=1}^{M} k_m \; > \; r \left(|A|_e - 2\varepsilon \right). \tag{5.20}$$

Now, each J_m is contained in some I_n. There may be more than one J_m in I_n, but the intervals J_m are disjoint. Since f is monotone increasing, it follows that

$$\sum_{m=1}^{M} \Big(f(y_m + k_m) - f(y_m) \Big) \;\le\; \sum_{n=1}^{N} \Big(f(x_n) - f(x_n - h_n) \Big). \tag{5.21}$$

Combining equations (5.19)–(5.21), we conclude that

$$r\,\big(|A|_e - 2\varepsilon\big) \;<\; s\,\big(|A|_e + \varepsilon\big).$$

Since ε is arbitrary, this implies that $r\,|A|_e \le s\,|A|_e$. But $r > s$, so we must have $|A|_e = 0$. Taking the union over all rational r and s with $s < r$, we see then that $S = \{D^+ f > D_- f\}$ has measure zero. A similar argument applies to any other pair of Dini numbers, so all four Dini numbers are equal for almost every $x \in (a,b)$.

(c) The functions

$$f_n(x) \;=\; \frac{f(x+\frac{1}{n}) - f(x)}{\frac{1}{n}} \;=\; n\,\big(f(x+\tfrac{1}{n}) - f(x)\big), \qquad x \in \mathbb{R},$$

converge pointwise a.e. to $f'(x)$ on $[a,b]$ as $n \to \infty$. Each f_n is measurable and nonnegative (because f is monotone increasing), so f' is measurable and $f' \ge 0$ a.e.

(d) By Fatou's Lemma,

$$\int_a^b f' \;=\; \int_a^b \liminf_{n\to\infty} f_n \;\le\; \liminf_{n\to\infty} \int_a^b f_n.$$

On the other hand, recalling how we extended the domain of f to $\mathbb{R}$, for each individual n we compute that

$$\begin{aligned}
\int_a^b f_n &= n \int_{a+\frac{1}{n}}^{b+\frac{1}{n}} f \;-\; n \int_a^b f && \text{(by the definition of } f_n\text{)}\\
&= n \int_b^{b+\frac{1}{n}} f \;-\; n \int_a^{a+\frac{1}{n}} f\\
&= n \int_b^{b+\frac{1}{n}} f(b) \;-\; n \int_a^{a+\frac{1}{n}} f && \text{(since } f \text{ is constant on } [b,\infty))\\
&\le n \int_b^{b+\frac{1}{n}} f(b) \;-\; n \int_a^{a+\frac{1}{n}} f(a) && \text{(since } f \text{ is monotone increasing)}\\
&= f(b) - f(a).
\end{aligned}$$

Therefore

$$\int_a^b f' \;\le\; \liminf_{n\to\infty} \int_a^b f_n \;\le\; f(b) - f(a) \;<\; \infty,$$

so f' is integrable. $\square$

As illustrated by the Cantor–Lebesgue function, it is possible for strict inequality to hold in equation (5.16).

Now we combine Theorem 5.4.2 with the Jordan decomposition to show that functions that have bounded variation on $[a,b]$ are differentiable a.e. and have integrable derivatives.

Corollary 5.4.3. *Choose* $f \in \mathrm{BV}[a,b]$, *and for each* $x \in [a,b]$ *let* $V(x) = V[f;a,x]$ *be the total variation of* f *over* $[a,x]$. *Then the following statements hold.*

(a) $f'(x)$ *exists for a.e.* $x \in [a,b]$.

(b) $f' \in L^1[a,b]$.

(c) $|f'| \le V'$ *a.e.*

(d) *The* L^1*-norm of* f' *is bounded by the total variation of* f, *i.e.,*

$$\|f'\|_1 = \int_a^b |f'| \le V[f;a,b]. \tag{5.22}$$

Proof. (a), (b) If f is a complex-valued function that has bounded variation, then we can write $f = (f_1 - f_2) + i(f_3 - f_4)$ where f_1, f_2, f_3, f_4 are each monotone increasing. Theorem 5.4.2 implies that each f_k is differentiable a.e. and f_k' is integrable. Since these properties are preserved by finite linear combinations, it follows that f is differentiable a.e. and f' is integrable.

(c) Exercise 5.2.11(c) implies that $V(x) = V[f;a,x]$ is monotone increasing on $[a,b]$. Therefore V is differentiable a.e. by Theorem 5.4.2.

Let Z be the set of measure zero that consists of all points $x \in [a,b]$ where either $f'(x)$ or $V'(x)$ does not exist. Fix $x \notin Z$ with $x \ne b$. If $h > 0$ is small enough that $x+h \in [a,b]$, then by applying equation (5.6) and Lemma 5.2.12 we see that

$$|f(x+h) - f(x)| \le V[f;x,x+h] = V(x+h) - V(x).$$

Since f and V are both differentiable at x, it follows that

$$|f'(x)| = \lim_{h\to 0^+} \left|\frac{f(x+h)-f(x)}{h}\right| \le \lim_{h\to 0^+} \frac{V(x+h)-V(x)}{h} = V'(x).$$

Thus $|f'| \le V'$ a.e.

(d) Using part (c) and applying equation (5.16) to the monotone increasing function V, we obtain

$$\int_a^b |f'| \le \int_a^b V' \le V(b) - V(a) = V[f;a,b]. \qquad \square$$

As an application of Theorem 5.4.2, we prove a lemma due to Fubini.

Lemma 5.4.4. *Assume f_k is monotone increasing on $[a,b]$ for each $k \in \mathbb{N}$. If the series*

$$s(x) \;=\; \sum_{k=1}^{\infty} f_k(x)$$

converges (to a finite real number) for every $x \in [a,b]$, then s is differentiable a.e. and

$$s'(x) \;=\; \sum_{k=1}^{\infty} f_k'(x) \;\; a.e. \tag{5.23}$$

Proof. For each $N \in \mathbb{N}$, let

$$s_N(x) \;=\; \sum_{k=1}^{N} f_k(x) \qquad \text{and} \qquad r_N(x) \;=\; \sum_{k=N+1}^{\infty} f_k(x).$$

By hypothesis, the series defining $r_N(x)$ converges for every x, so $s(x) = s_N(x) + r_N(x)$ for every x. Since s_N and r_N are monotone increasing on $[a,b]$, Theorem 5.4.2 implies that they are differentiable except possibly on some set Z_N that has measure zero. Further, $s_N' \ge 0$ a.e. and $r_N' \ge 0$ a.e. Consequently s is differentiable at all points $x \notin Z = \cup Z_N$, and

$$s'(x) \;=\; s_N'(x) + r_N'(x), \qquad \text{for all } x \notin Z.$$

Our goal is to show that $s_N'(x) \to s'(x)$ for a.e. x.

Now, $s_N(x) \to s(x)$ everywhere, so $r_N(x) \to 0$ for every x. For each $j \in \mathbb{N}$, choose N_j large enough that we have both $r_{N_j}(a) < 2^{-j}$ and $r_{N_j}(b) < 2^{-j}$. Then

$$0 \;\le\; \sum_{j=1}^{\infty} \bigl(r_{N_j}(b) - r_{N_j}(a)\bigr) \;<\; \infty. \tag{5.24}$$

Since $r_N' \ge 0$ a.e., the series $g(x) = \sum_{j=1}^{\infty} r_{N_j}'(x)$ converges at almost every point in the extended real sense. We compute that

$$\begin{aligned}
0 \;\le\; \int_a^b g \;&=\; \int_a^b \sum_{j=1}^{\infty} r_{N_j}' \\
&=\; \sum_{j=1}^{\infty} \int_a^b r_{N_j}' && \text{(by Corollary 4.2.4)} \\
&\le\; \sum_{j=1}^{\infty} \bigl(r_{N_j}(b) \;-\; r_{N_j}(a)\bigr) && \text{(by Theorem 5.4.2)} \\
&<\; \infty && \text{(by equation (5.24)).}
\end{aligned}$$

Thus g is integrable, so it must be finite a.e. Hence

$$0 \le g(x) = \sum_{j=1}^{\infty} r'_{N_j}(x) < \infty \quad \text{a.e.}$$

Therefore, for a.e. x,

$$\lim_{j\to\infty} \bigl(s'(x) - s'_{N_j}(x)\bigr) = \lim_{j\to\infty} r'_{N_j}(x) = 0.$$

Thus $s'_{N_j}(x) \to s'(x)$ a.e. Although this only tells us that a *subsequence* of the partial sums converges, the fact that $f'_k \ge 0$ a.e. implies that the partial sums $s'_N = \sum_{k=1}^{N} f'_k$ increase with N:

$$s'_1(x) \le s'_2(x) \le \cdots \qquad \text{for a.e. } x. \tag{5.25}$$

The reader should check that since s'_N is monotone increasing and a subsequence converges a.e. to s', we have that $s'_N \nearrow s'$ a.e. as $N \to \infty$. Hence equation (5.23) holds. $\square$

Problems

5.4.5. Let I be any interval in $\mathbb{R}$ (possibly infinite, and not necessarily closed). Prove that any monotone increasing function $f\colon I \to \mathbb{R}$ is differentiable a.e. on I (note that f need not be bounded).

5.4.6. Assume that $f\colon [a,b] \to \mathbb{R}$ is continuous and $D^+f \ge 0$ on (a,b). Prove that f is monotone increasing on $[a,b]$.

5.4.7. Let $\{r_k\}_{k\in\mathbb{N}}$ be an enumeration of the rational points in $(0,1)$. Define

$$f(x) = \sum_{n=1}^{\infty} 2^{-n}\, \chi_{[r_n,1]}(x), \qquad x \in [0,1].$$

Prove that f is monotone increasing on $[0,1]$, right-continuous at every point in $[0,1]$, discontinuous at every rational point in $(0,1)$, and continuous at every irrational point in $(0,1)$.

5.4.8. (Brown [Bro69]) Let φ be the Cantor–Lebesgue function on $[0,1]$. Extend φ to $\mathbb{R}$ by setting $\varphi(x) = \varphi(0) = 0$ for $x < 0$ and $\varphi(x) = \varphi(1) = 1$ for $x > 1$. Let $\{[a_n,b_n]\}_{n\in\mathbb{N}}$ be an enumeration of all subintervals of $[0,1]$ with rational endpoints $a_n < b_n$. For each $n \in \mathbb{N}$ set

$$f_n(x) = 2^{-n}\, \varphi\Bigl(\frac{x-a_n}{b_n-a_n}\Bigr), \qquad x \in \mathbb{R}.$$

Observe that f_n is monotone increasing on $\mathbb{R}$ and has uniform norm $\|f_n\|_u = 2^{-n}$. Prove the following statements.

(a) The series $f = \sum_{n=1}^\infty f_n$ converges uniformly on $[0,1]$.

(b) f is continuous and monotone increasing on $[0,1]$.

(c) f is *strictly increasing* on $[0,1]$, i.e., if $0 \le x < y \le 1$ then $f(x) < f(y)$.

(d) f is singular on $[0,1]$, i.e., $f'(x)$ exists for almost every $x \in [0,1]$ and $f' = 0$ a.e. (Lemma 5.4.4 is helpful here).

5.4.9. This problem will show that if $f\colon [a,b] \to \mathbb{R}$ is monotone increasing, then $D^+f < \infty$ a.e. Suppose that $A = \{D^+f = \infty\}$ had positive measure, and fix any number $M > 0$.

(a) Prove that $\mathcal{B} = \left\{[x,y] \subseteq (a,b) : x \in A,\, y \in (a,b),\, \dfrac{f(y)-f(x)}{y-x} > M\right\}$ is a Vitali cover of A.

(b) Given $0 < \varepsilon < |A|_e$, use the Vitali Covering Lemma to show that there exist disjoint intervals $[x_k, y_k] \in \mathcal{B}$, where $k = 1, \dots, N$, such that $\sum_{k=1}^N (y_k - x_k) > |A|_e - \varepsilon$.

(c) Show that $\sum_{k=1}^N \bigl(f(y_k) - f(x_k)\bigr) > M\,\bigl(|A|_e - \varepsilon\bigr)$.

(d) Derive a contradiction, and conclude that $|A|_e = 0$. Show that D^-f, D_+f, and D_-f are also finite a.e.

5.5 The Lebesgue Differentiation Theorem

Suppose that a function $f\colon \mathbb{R}^d \to \mathbb{C}$ is continuous at a point x. In this case, f "does not vary much" over a small ball $B_h(x)$ centered at x. Hence the average of f over this small ball, which we will denote by

$$\widetilde{f}_h(x) \;=\; \frac{1}{|B_h(x)|} \int_{B_h(x)} f(t)\, dt, \tag{5.26}$$

should be close to the value taken by f at the center of the ball, and we expect that this average value will converge to $f(x)$ as the radius h shrinks to zero. The next lemma makes these statements precise. Although it is true that the measure of the ball $B_h(x)$ is $C_d h^d$, where C_d is a constant that depends only on the dimension d, we will write it as $|B_h(x)|$ to emphasize the averaging operation that is being performed. The observation that

$$\frac{1}{|B_h(x)|} \int_{B_h(x)} dt \;=\; 1$$

is a trivial but surprisingly convenient fact that is employed in many proofs of this type.

Lemma 5.5.1. *If a function* $f\colon \mathbb{R}^d \to \mathbb{C}$ *is continuous at a point* $x \in \mathbb{R}^d$, *then*

$$\lim_{h\to 0} \frac{1}{|B_h(x)|} \int_{B_h(x)} |f(x)-f(t)|\,dt \;=\; 0 \tag{5.27}$$

and

$$\lim_{h\to 0} \widetilde{f}_h(x) \;=\; f(x). \tag{5.28}$$

Proof. Suppose that f is continuous at x, and fix $\varepsilon > 0$. Then there is a $\delta > 0$ such that $|f(x) - f(t)| < \varepsilon$ whenever t satisfies $\|x-t\| < \delta$. Hence for all $0 < h < \delta$ we have

$$\frac{1}{|B_h(x)|} \int_{B_h(x)} |f(x)-f(t)|\,dt \;\le\; \frac{1}{|B_h(x)|} \int_{B_h(x)} \varepsilon\,dt \;=\; \varepsilon.$$

This proves equation (5.27). Equation (5.28) follows from equation (5.27), because

$$\begin{aligned} |f(x)-\widetilde{f}_h(x)| \;&=\; \left| f(x)\,\frac{1}{|B_h(x)|}\int_{B_h(x)} dt \;-\; \frac{1}{|B_h(x)|}\int_{B_h(x)} f(t)\,dt \right| \\ &=\; \frac{1}{|B_h(x)|} \left| \int_{B_h(x)} \bigl(f(x)-f(t)\bigr)\,dt \right| \\ &\le\; \frac{1}{|B_h(x)|} \int_{B_h(x)} |f(x)-f(t)|\,dt. \qquad \square \end{aligned}$$

According to the following exercise, if f is uniformly continuous on $\mathbb{R}^d$, then the averages $\widetilde{f}_h$ converge to f *uniformly*, not just pointwise.

Exercise 5.5.2. Prove that if $f\colon \mathbb{R}^d \to \mathbb{C}$ is uniformly continuous on $\mathbb{R}^d$, then

$$\lim_{h\to 0} \|f - \widetilde{f}_h\|_{\mathrm{u}} \;=\; 0. \qquad \diamondsuit$$

5.5.1 L^1-Convergence of Averages

If f is not continuous at x, then it need not be true that averages of f over balls $B_h(x)$ will converge to $f(x)$ as $h \to 0$. Even so, we will soon prove the *Lebesgue Differentiation Theorem*, which shows that if f is an integrable function, then these averages converge *pointwise almost everywhere*. This is a nontrivial result, and it will require some work. For motivation, we first prove the easier fact that the averages $\widetilde{f}_h$ of an integrable function f converge to f *in* L^1*-norm* as $h \to 0$.

Theorem 5.5.3. *If* $f \in L^1(\mathbb{R}^d)$, *then* $\widetilde{f}_h \to f$ *in* L^1*-norm, i.e.,*

$$\lim_{h\to 0} \|f - \widetilde{f}_h\|_1 \;=\; \lim_{h\to 0} \int_{\mathbb{R}^d} |f(x) - \widetilde{f}_h(x)|\,dx \;=\; 0.$$

Proof. Let χ_h denote the characteristic function of the open ball of radius h centered at the origin, but rescaled so that $\int \chi_h = 1$. Explicitly,

$$\chi_h \;=\; \frac{1}{|B_h(0)|}\,\chi_{B_h(0)}.$$

Using this notation, we can rewrite $\widetilde{f}_h$ as

$$\widetilde{f}_h(x) \;=\; \frac{1}{|B_h(0)|}\int_{B_h(0)} f(x-t)\,dt \;=\; \int_{\mathbb{R}^d} f(x-t)\,\chi_h(t)\,dt. \tag{5.29}$$

Using Tonelli's Theorem to interchange the order of integration and noting that χ_h is only nonzero on $B_h(0)$, we can therefore estimate the L^1-norm of $f - \widetilde{f}_h$ as follows:

$$\begin{aligned}
\|f - \widetilde{f}_h\|_1 \;&=\; \int_{\mathbb{R}^d} |f(x) - \widetilde{f}_h(x)|\,dx \\
&=\; \int_{\mathbb{R}^d} \left| f(x)\int_{\mathbb{R}^d} \chi_h(t)\,dt \;-\; \int_{\mathbb{R}^d} f(x-t)\,\chi_h(t)\,dt \right| dx \\
&\le\; \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(x) - f(x-t)|\,\chi_h(t)\,dt\,dx \\
&=\; \frac{1}{|B_h(0)|}\int_{B_h(0)} \left(\int_{\mathbb{R}^d} |f(x) - f(x-t)|\,dx\right) dt \\
&=\; \frac{1}{|B_h(0)|}\int_{\|t\|<h} \|f - T_t f\|_1\,dt,
\end{aligned}$$

where $T_t f(x) = f(x-t)$ denotes the translation of f by t. The "strong continuity" property of translation on $L^1(\mathbb{R}^d)$ established in Exercise 4.5.9 tells us that

$$\lim_{t\to 0} \|f - T_t f\|_1 \;=\; 0.$$

Therefore, if we fix an $\varepsilon > 0$, then there is some $\delta > 0$ such that $\|f - T_t f\|_1 < \varepsilon$ whenever $\|t\| < \delta$. Consequently, for all $0 < h < \delta$ we have

$$\|f - \widetilde{f}_h\|_1 \;\le\; \frac{1}{|B_h(0)|}\int_{\|t\|<h} \|f - T_t f\|_1\,dt \;\le\; \frac{1}{|B_h(0)|}\int_{\|t\|<h} \varepsilon\,dt \;=\; \varepsilon. \qquad \square$$

We introduced the operation of *convolution* in Section 4.6.3. If functions f and g are defined on the domain $\mathbb{R}^d$, then their convolution is the function $f * g$ defined by

$$(f * g)(x) \;=\; \int_{\mathbb{R}^d} f(x-t)\,g(t)\,dt,$$

as long as this integral exists. Using this terminology, equation (5.29) says that the average of f over the ball $B_h(x)$ is the convolution of f with χ_h:

$$\widetilde{f}_h(x) \;=\; \int_{\mathbb{R}^d} f(x-t)\,\chi_h(t)\,dt \;=\; (f*\chi_h)(x).$$

Hence an equivalent wording of Theorem 5.5.3 is that for every $f \in L^1(\mathbb{R}^d)$, we have that

$$f*\chi_h \to f \text{ in } L^1\text{-norm as } h \to 0.$$

This is a special case of the results on *approximate identities* that we will prove when we study convolution in detail in Section 9.1. In fact, our proof of Theorem 5.5.3 is a simplified version of the proof of Theorem 9.1.11.

5.5.2 Locally Integrable Functions

When we prove the Lebesgue Differentiation Theorem, we will see that we do not need to restrict ourselves to functions that are integrable on all of $\mathbb{R}^d$. Instead, we will be able to prove the theorem for functions that are merely *locally integrable* in the following sense.

Definition 5.5.4 (Locally Integrable Functions). Let $f\colon \mathbb{R}^d \to \overline{\mathbf{F}}$ be a measurable function on $\mathbb{R}^d$. We say that f is *locally integrable* if its restriction to any compact set K is integrable. In other words, f is locally integrable if

$$\|f\cdot\chi_K\|_1 \;=\; \int_K |f| \;<\; \infty \quad \text{for every compact set } K \subseteq \mathbb{R}^d.$$

The space of *locally integrable functions* on $\mathbb{R}^d$ is

$$L^1_{\mathrm{loc}}(\mathbb{R}^d) \;=\; \bigl\{f\colon\mathbb{R}^d \to \overline{\mathbf{F}} \;:\; f \text{ is locally integrable on } \mathbb{R}^d\bigr\}. \qquad \diamondsuit$$

Since every compact set is bounded, a measurable function f is locally integrable if and only if

$$\|f\cdot\chi_{B_N(0)}\|_1 \;=\; \int_{\|x\|<N} |f(x)|\,dx \;<\; \infty, \qquad \text{for all } N \in \mathbb{N}.$$

Every continuous function, including polynomials and e^x, is locally integrable.

5.5.3 The Maximal Theorem

Our ultimate goal is to prove the Lebesgue Differentiation Theorem, which states that if f is locally integrable, then for almost every x we have

$$f(x) = \lim_{h \to 0} \widetilde{f}_h(x) = \lim_{h \to 0} \frac{1}{|B_h(x)|} \int_{B_h(x)} f(t)\, dt.$$

However, we need to develop some tools before we can do this. Specifically, before we can understand how *limits* of averages behave, we need to understand the *supremum* of these averages. In fact, to obtain a true upper estimate, we will consider the supremum of the averages of $|f|$, rather than averages of f. This leads us to the *Hardy–Littlewood maximal function*, which is defined as follows.

Definition 5.5.5 (Hardy–Littlewood Maximal Function). The *Hardy–Littlewood maximal function* of a locally integrable function f is

$$Mf(x) = \sup_{h>0} \widetilde{f}_h(x) = \sup_{h>0} \frac{1}{|B_h(x)|} \int_{B_h(x)} |f(t)|\, dt. \qquad \diamondsuit$$

Each averaged function $\widetilde{f}_h$ is measurable. In fact, $\widetilde{f}_h$ is continuous by Problem 5.5.13 (this is not so surprising, since averaging tends to be a smoothing operation). The supremum of a family of continuous functions need not be continuous, but it is *lower semicontinuous* in the sense given in Problem 1.1.24. Hence Mf is a fairly nice function in certain ways. To illustrate this, let $g = |f|$, so $Mf = \sup_{h>0} \widetilde{g}_h$. Then for any $a \in \mathbb{R}$, the set

$$\{Mf > a\} = \bigcup_{h>0} \{\widetilde{g}_h > a\}$$

is open, because $\widetilde{g}_h$ is continuous and therefore $\{\widetilde{g}_h > a\} = \widetilde{g}_h^{\,-1}(a,\infty)$ is open. Thus $\{Mf > a\}$ is an open set, not just a measurable set.

Unfortunately, Mf is *not integrable*, even if f is integrable (except in the trivial case $f = 0$ a.e.); see Problem 5.5.22. Even so, if f is integrable then Mf does possess a property that is reminiscent of integrable functions. To motivate this, recall Tchebyshev's Inequality (Theorem 4.1.9), which states that if $f \in L^1(\mathbb{R}^d)$ then we have the following inequality relating the measure of the set where $|f|$ exceeds α to the integral of $|f|$:

$$\bigl|\{|f| > \alpha\}\bigr| \le \frac{1}{\alpha} \int_{\mathbb{R}^d} |f|.$$

Hence, if Mf were integrable then we would have

$$\bigl|\{Mf > \alpha\}\bigr| \le \frac{1}{\alpha} \int_{\mathbb{R}^d} Mf. \tag{5.30}$$

Sadly, Mf is not integrable, but the following important result, known as the *Hardy–Littlewood Maximal Theorem*, or simply the *Maximal Theorem*, gives us a substitute: The equation obtained by replacing Mf with $3^d|f|$ on the right-hand side of equation (5.30) holds whenever f is integrable.

Theorem 5.5.6 (The Maximal Theorem). *If* $f \in L^1(\mathbb{R}^d)$, *then for each* $\alpha > 0$ *we have*

$$\big|\{Mf > \alpha\}\big| \;\le\; \frac{3^d}{\alpha} \int_{\mathbb{R}^d} |f| \;=\; \frac{3^d}{\alpha}\, \|f\|_1.$$

Proof. For each $\alpha > 0$, let $E_\alpha = \big\{Mf > \alpha\big\}$. If $x \in E_\alpha$, then

$$\alpha \;<\; Mf(x) \;=\; \sup_{h>0} \frac{1}{|B_h(x)|} \int_{B_h(x)} |f(t)|\, dt.$$

Hence there must exist some radius r_x such that

$$\frac{1}{|B_{r_x}(x)|} \int_{B_{r_x}(x)} |f(t)|\, dt \;>\; \alpha. \tag{5.31}$$

We trivially have

$$E_\alpha \;\subseteq\; \bigcup_{x \in E_\alpha} B_{r_x}(x).$$

Therefore, if we fix $0 < c < |E_\alpha|$, then the Simple Vitali Lemma implies that there exist finitely many points $x_1, \dots, x_N \in E_\alpha$ such that the balls $B_k = B_{r_{x_k}}(x_k)$ for $k = 1, \dots, N$ are disjoint and satisfy

$$\sum_{k=1}^N |B_k| \;>\; \frac{c}{3^d}. \tag{5.32}$$

Consequently,

$$\begin{aligned}
c \;&<\; 3^d \sum_{k=1}^N |B_k| && \text{(by equation (5.32))}\\
&\le\; 3^d \sum_{k=1}^N \frac{1}{\alpha} \int_{B_k} |f| && \text{(by equation (5.31))}\\
&\le\; 3^d \, \frac{1}{\alpha} \int_{\mathbb{R}^d} |f| && \text{(by disjointness).}
\end{aligned}$$

Since this is true for all $0 < c < |E_\alpha|$, we conclude that

$$|E_\alpha| \;\le\; \frac{3^d}{\alpha} \int_{\mathbb{R}^d} |f| \;<\; \infty. \qquad \square$$

5.5.4 The Lebesgue Differentiation Theorem

Now we prove the Lebesgue Differentiation Theorem.

Theorem 5.5.7 (Lebesgue Differentiation Theorem). *If f is locally integrable on $\mathbb{R}^d$, then for almost every $x \in \mathbb{R}^d$ we have*

$$\lim_{h\to 0} \frac{1}{|B_h(x)|} \int_{B_h(x)} |f(x)-f(t)|\,dt \;=\; 0, \tag{5.33}$$

and

$$\lim_{h\to 0} \widetilde{f}_h(x) \;=\; \lim_{h\to 0} \frac{1}{|B_h(x)|} \int_{B_h(x)} f(t)\,dt \;=\; f(x). \tag{5.34}$$

Proof. Step 1: Proof of equation (5.34) *for integrable functions.*

Assume that f is integrable. Restating equation (5.34) in an equivalent form that uses the real-parameter version of limsup, our goal is to show that

$$\limsup_{h\to 0} |f(x)-\widetilde{f}_h(x)| \;=\; 0 \quad \text{for a.e. } x \in \mathbb{R}^d. \tag{5.35}$$

Fix $\varepsilon > 0$. By Theorem 4.5.8, there exists a function $g \in C_c(\mathbb{R}^d)$ that satisfies $\|f-g\|_1 < \varepsilon$. Therefore, for every $x \in \mathbb{R}^d$ we have that

$$\begin{aligned}
&|f(x)-\widetilde{f}_h(x)| \\
&\quad\le\; |f(x)-g(x)| \;+\; |g(x)-\widetilde{g}_h(x)| \;+\; |\widetilde{g}_h(x)-\widetilde{f}_h(x)| \\
&\quad=\; |f(x)-g(x)| \;+\; |g(x)-\widetilde{g}_h(x)| \;+\; \left|\frac{1}{|B_h(x)|}\int_{B_h(x)} \bigl(g(t)-f(t)\bigr)\,dt\right| \\
&\quad\le\; |f(x)-g(x)| \;+\; \|g-\widetilde{g}_h\|_{\mathrm{u}} \;+\; M(g-f)(x).
\end{aligned}$$

Since g is uniformly continuous, Exercise 5.5.2 shows that $\widetilde{g}_h \to g$ uniformly. Therefore

$$\begin{aligned}
&\limsup_{h\to 0} |f(x)-\widetilde{f}_h(x)| \\
&\qquad\le\; |f(x)-g(x)| \;+\; \Bigl(\limsup_{h\to 0} \|g-\widetilde{g}_h\|_{\mathrm{u}}\Bigr) \;+\; M(g-f)(x) \\
&\qquad=\; |f(x)-g(x)| \;+\; 0 \;+\; M(g-f)(x).
\end{aligned} \tag{5.36}$$

Fix $\alpha > 0$, and let

$$E_\alpha \;=\; \Bigl\{\limsup_{h\to 0} |f-\widetilde{f}_h| \;>\; 2\alpha\Bigr\}.$$

By equation (5.36), if $x \in E_\alpha$ then we must have either $|f(x)-g(x)| > \alpha$ or $M(g-f)(x) > \alpha$. Therefore

$$E_\alpha \subseteq F_\alpha \cup G_\alpha$$

where

$$F_\alpha = \bigl\{|f-g| > \alpha\bigr\} \qquad \text{and} \qquad G_\alpha = \bigl\{M(g-f) > \alpha\bigr\}.$$

By Tchebyshev's Inequality,

$$|F_\alpha| = \bigl|\bigl\{|f-g| > \alpha\bigr\}\bigr| \le \frac{1}{\alpha}\int_{\mathbb{R}^d} |f-g| = \frac{1}{\alpha}\|f-g\|_1 < \frac{\varepsilon}{\alpha}.$$

On the other hand, the Maximal Theorem implies that

$$|G_\alpha| = \bigl|\{M(g-f) > \alpha\}\bigr| \le \frac{3^d}{\alpha}\int_{\mathbb{R}^d} |f-g| = \frac{3^d}{\alpha}\|f-g\|_1 < \frac{3^d\varepsilon}{\alpha}.$$

Consequently,

$$|E_\alpha| \le |F_\alpha| + |G_\alpha| < \frac{3^d+1}{\alpha}\,\varepsilon.$$

This holds for every $\varepsilon > 0$, so we conclude that $|E_\alpha| = 0$. And this is true for every $\alpha > 0$, so the set

$$Z = \Bigl\{\limsup_{h\to 0} |f - \widetilde{f}_h| > 0\Bigr\} = \bigcup_{n=1}^{\infty} E_{1/n}$$

has measure zero. Therefore equation (5.35) holds when f is integrable.

Step 2: Proof of equation (5.34) *for locally integrable functions.*

Now assume that f is locally integrable. Given an integer $N \in \mathbb{N}$, let

$$g = f \cdot \chi_{B_N(0)},$$

and observe that g is integrable since f is locally integrable. Further, if $\|x\| < N$ then $\widetilde{f}_h(x) = \widetilde{g}_h(x)$ for all small enough h. Applying Step 1 to g, it follows that

$$\lim_{h\to 0} \widetilde{f}_h(x) = \lim_{h\to 0} \widetilde{g}_h(x) = g(x) = f(x), \quad \text{for a.e. } x \in B_N(0).$$

Since the union of countably many sets with measure zero still has measure zero, this implies that $\widetilde{f}_h(x) \to f(x)$ for a.e. $x \in \mathbb{R}^d$.

Step 3: Proof of equation (5.33) *for locally integrable functions.*

Assume that f is locally integrable. Given a scalar $c \in \mathbb{C}$, set $g_c(x) = |f(x) - c|$. Then g_c is locally integrable, so by applying Step 2 to g_c we see that

$$\lim_{h\to 0} \frac{1}{|B_h(x)|}\int_{B_h(x)} |f(t) - c|\,dt = |f(x) - c| \tag{5.37}$$

for a.e. $x \in \mathbb{R}^d$. That is, for every $c \in \mathbb{C}$, equation (5.37) holds for a.e. x. However, we need to prove something different. Specifically, we need to prove that for a.e. $x \in \mathbb{R}^d$, equation (5.37) holds when we take $c = f(x)$. This does not follow from what we have established so far (consider Problem 2.2.36).

So, for each $c \in \mathbb{C}$ let Z_c denote the set of measure zero where equation (5.37) does not hold. Let $R = \mathbb{Q} + i\mathbb{Q}$ be the set of all rational complex numbers. Then R is countable, so

$$Z \;=\; \bigcup_{c \in R} Z_c$$

has measure zero.

Suppose that $x \notin Z$, and choose $\varepsilon > 0$. Since $f(x)$ is a complex scalar and since R is dense in $\mathbb{C}$, there is a point $c \in R$ such that

$$|f(x) - c| \;<\; \varepsilon.$$

Therefore

$$\begin{aligned}
&\limsup_{h \to 0} \frac{1}{|B_h(x)|} \int_{B_h(x)} |f(x) - f(t)|\, dt \\
&\qquad \le\; \limsup_{h \to 0} \frac{1}{|B_h(x)|} \int_{B_h(x)} \Big(|f(x) - c| + |c - f(t)| \Big)\, dt \\
&\qquad \le\; \limsup_{h \to 0} \frac{|f(x) - c|}{|B_h(x)|} \int_{B_h(x)} dt \;+\; \limsup_{h \to 0} \frac{1}{|B_h(x)|} \int_{B_h(x)} |c - f(t)|\, dt \\
&\qquad =\; |f(x) - c| \;+\; |f(x) - c| \qquad\qquad (\text{since } x \notin Z_c) \\
&\qquad <\; \varepsilon + \varepsilon \;=\; 2\varepsilon.
\end{aligned}$$

Since ε is arbitrary, equation (5.33) holds for this x. This is true for all $x \notin Z$, so we conclude that equation (5.33) holds for a.e. x. $\square$

Although Theorem 5.5.7 is stated for functions on $\mathbb{R}^d$, it can be applied to functions whose domain is a subset of $\mathbb{R}^d$. For example, suppose that f is integrable on some measurable set $E \subseteq \mathbb{R}^d$. Then we can extend the domain of f to all of $\mathbb{R}^d$ by declaring that $f(t) = 0$ for $t \notin E$. If x belongs to the interior of E, then the open ball $B_h(x)$ is entirely contained in E for all small enough h. Applying Theorem 5.5.7 to the extended function f, it follows that equations (5.33) and (5.34) hold for almost every $x \in E^\circ$.

5.5.5 Lebesgue Points

The points that satisfy the criterion that appears in equation (5.33) are given the following special name.

Definition 5.5.8 (Lebesgue Points and the Lebesgue Set). Let f be a locally integrable function on $\mathbb{R}^d$. If $x \in \mathbb{R}^d$ satisfies

$$\lim_{h\to 0} \frac{1}{|B_h(x)|} \int_{B_h(x)} |f(x) - f(t)|\, dt \;=\; 0,$$

then x is called a *Lebesgue point* of f. The set of all Lebesgue points is the *Lebesgue set* of f. $\diamondsuit$

Using this terminology, the Lebesgue Differentiation Theorem implies that almost every point in the domain of a locally integrable function is a Lebesgue point. In particular, we saw in Lemma 5.5.1 that every point of continuity is a Lebesgue point. However, a Lebesgue point need not be a point of continuity.

Next we give a generalization of the Lebesgue Differentiation Theorem that allows us to average over sets other than the open balls $B_h(x)$. Here are the specific types of families of sets that we will be allowed to average over.

Definition 5.5.9 (Regularly Shrinking Family). We say that a family $\{E_n\}_{n\in\mathbb{N}}$ of measurable subsets of $\mathbb{R}^d$ *shrinks regularly* to a point $x \in \mathbb{R}^d$ as $n \to \infty$ if there exists a constant $\alpha > 0$ and radii $r_n \to 0$ such that for each $n \in \mathbb{N}$ we have

$$E_n \;\subseteq\; B_{r_n}(x) \qquad \text{and} \qquad |E_n| \;\ge\; \alpha\, |B_{r_n}(x)|. \qquad \diamondsuit$$

In other words, in order for $\{E_n\}_{n\in\mathbb{N}}$ to shrink regularly to x, each set E_n must be contained in some ball centered at x and must contain some fixed fraction of the volume of that ball, although the set E_n *need not* contain x itself.

Now we prove that we can replace averages over balls with averages over sets in a regularly shrinking family.

Theorem 5.5.10. *If f is locally integrable on $\mathbb{R}^d$ and $\{E_n\}_{n\in\mathbb{N}}$ shrinks regularly to a Lebesgue point x of f, then*

$$\lim_{n\to\infty} \frac{1}{|E_n|} \int_{E_n} |f(x) - f(t)|\, dt \;=\; 0.$$

Proof. By the definition of a Lebesgue point and the properties of a regularly shrinking family, we have that

$$\begin{aligned}\frac{1}{|E_n|} \int_{E_n} |f(y) - f(x)|\, dy \;&\le\; \frac{1}{\alpha\, |B_{r_n(x)}|} \int_{B_{r_n}(x)} |f(y) - f(x)|\, dy \\ &\to\; 0 \quad \text{as } n \to \infty. \qquad \square\end{aligned}$$

An analogous result holds for families that are indexed by a real parameter. In particular, we say that a family of sets $\{E_r\}_{r>0}$ shrinks regularly to x as $r \to 0$ if there exists some constant $\alpha > 0$ such that $E_r \subseteq B_r(x)$ and $|E_r| \ge \alpha\,|B_r(x)|$ for each $r > 0$. In this case, if x is a Lebesgue point of f then

$$\lim_{r\to 0} \frac{1}{|E_r|} \int_{E_r} |f(x) - f(t)|\,dt \;=\; 0.$$

Specializing to dimension $d = 1$ gives us the following result.

Corollary 5.5.11. *If f is locally integrable on $\mathbb{R}$ and x is a Lebesgue point of f, then*

$$\lim_{h\to 0} \frac{1}{2h} \int_{x-h}^{x+h} |f(x) - f(t)|\,dt \;=\; 0 \tag{5.38}$$

and

$$\lim_{h\to 0} \frac{1}{h} \int_{x}^{x+h} |f(x) - f(t)|\,dt \;=\; 0. \tag{5.39}$$

Proof. In one dimension, the open ball of radius h centered at x is the open interval $B_h(x) = (x-h, x+h)$. Therefore equation (5.38) is just a restatement of equation (5.33). Equation (5.39) is a consequence of Theorem 5.5.10 and the fact that the family $\{[x, x+h]\}_{h>0}$ shrinks regularly to x as $h \to 0$. $\square$

Problems

5.5.12. Give another solution to Problem 4.4.22.

5.5.13. Show that if f is locally integrable on $\mathbb{R}^d$, then $\widetilde{f_h}$ is continuous. Also show that if f is integrable, then $\|\widetilde{f_h}\|_1 \le \|f\|_1$.

5.5.14. This problem gives a generalization of Theorem 5.5.3. Let g be an integrable function on $\mathbb{R}^d$ that is identically zero outside of some ball of finite radius and whose integral over $\mathbb{R}^d$ is $\int g = 1$. For each $h > 0$, define $g_h(x) = h^{-d} g(x/h)$. Prove that

$$\lim_{h\to 0} \|f - f * g_h\|_1 \;=\; 0, \qquad \text{for all } f \in L^1(\mathbb{R}^d).$$

5.5.15. Prove that the maximal function is *sublinear* in the sense that if f and g are any locally integrable functions on $\mathbb{R}^d$ and c is any scalar, then

$$M(f+g) \;\le\; Mf \,+\, Mg \qquad \text{and} \qquad M(cf) \;=\; |c|\,Mf.$$

5.5.16. Suppose that f_n and f are nonnegative locally integrable functions on $\mathbb{R}^d$, and $f_n(x) \nearrow f(x)$ for a.e. x. Prove that $Mf_n(x) \nearrow Mf(x)$ for every x.

5.5.17. Given a locally integrable function f on $\mathbb{R}^d$, define a non-centered maximal function by

$$M^*f(x) \;=\; \sup\biggl\{\frac{1}{|B|}\int_B |f| \;:\; B \text{ is any open ball that contains } x\biggr\}.$$

Prove that $Mf \le M^*f \le 2^d\, Mf$.

5.5.18. A useful space that sometimes substitutes for L^1 in theorems where L^1 is not appropriate is the set *Weak*-$L^1(\mathbb{R}^d)$ that consists of all measurable functions f on $\mathbb{R}^d$ for which there exists a constant $C > 0$ such that

$$\bigl|\{|f| > \alpha\}\bigr| \;\le\; \frac{C}{\alpha} \quad \text{for every } \alpha > 0.$$

Prove the following statements.

(a) $L^1(\mathbb{R}^d) \subsetneq \text{Weak-}L^1(\mathbb{R}^d)$.

(b) If $f \in L^1(\mathbb{R}^d)$ then $Mf \in \text{Weak-}L^1(\mathbb{R}^d)$.

5.5.19. Let A be any subset of $\mathbb{R}^d$ with $|A|_e > 0$. Define the *density of A at a point $x \in \mathbb{R}^d$* to be

$$D_A(x) \;=\; \lim_{r\to 0} \frac{|A \cap B_r(x)|_e}{|B_r(x)|},$$

whenever this limit exists. Prove the following statements.

(a) $D_A(x) = 1$ for a.e. $x \in A$.

(b) A is measurable if and only if $D_A(x) = 0$ for a.e. $x \notin A$.

Additionally, exhibit a measurable set E and a point x such that $D_E(x)$ does not exist, and given $0 < \alpha < 1$ exhibit a measurable set E and a point x such that $D_E(x) = \alpha$.

5.5.20. Suppose that $E \subseteq [0,1]$ is measurable and there exists some $\delta > 0$ such that $\bigl|E \cap [a,b]\bigr|_e \ge \delta\,(b-a)$ for all $0 \le a < b \le 1$. Prove that $|E| = 1$.

5.5.21. Fix $0 < \lambda < 1$, and suppose that $f \in L^1[0,1]$ satisfies $\int_E f = 0$ for every measurable set $E \subseteq [0,1]$ such that $|E| = \lambda$. Prove that $f = 0$ a.e.

5.5.22. Assume that f is locally integrable, and f is not zero almost everywhere. Prove the following statements.

(a) There exist C, $R > 0$ such that $Mf(x) \ge C\,|x|^{-d}$ for all $|x| > R$.

(b) Mf is not integrable on $\mathbb{R}^d$.

(c) There exist C', $\alpha_0 > 0$ such that

$$\bigl|\{Mf > \alpha\}\bigr| \;\ge\; \frac{C'}{\alpha}, \qquad \text{all } 0 < \alpha < \alpha_0.$$

Compare this estimate to the Maximal Theorem.

Chapter 6
Absolute Continuity and the Fundamental Theorem of Calculus

Every continuous function $f\colon [a,b] \to \mathbb{C}$ is measurable, but there are many ways in which a continuous function can be "badly behaved." For example, even though the Cantor–Lebesgue function φ is continuous, is differentiable almost everywhere, is monotone increasing, and maps $[0,1]$ onto itself, it also has the following properties:

- it maps a set with measure zero to a set that has positive measure;
- it maps a measurable set to a nonmeasurable set;
- the Fundamental Theorem of Calculus (FTC) does not apply to φ;
- φ is singular but not constant.

What extra condition must a continuous function satisfy in order that it not have these unpleasant properties? We will prove in this chapter that the *absolutely continuous* functions are precisely those continuous functions that do not have the types of drawbacks listed above.

We define absolute continuity in Section 6.1. Section 6.2 derives two *growth lemmas*, which we use in Section 6.3 to prove the *Banach–Zaretsky Theorem*. This key theorem shows that absolute continuity is closely related to the issue of whether a function maps sets with measure zero to sets with measure zero. In Section 6.4 we use the Lebesgue Differentiation Theorem to characterize the absolutely continuous functions as those functions that satisfy the FTC. This completes the main goals of the chapter, but two optional sections provide some additional material. In Section 6.5 we study the relationship between absolute continuity, the Chain Rule, and changes of variable, while Section 6.6 introduces *convex functions* and proves *Jensen's Inequality*.

In this chapter the functions we consider will almost exclusively be finite at every point (in fact, they will usually be bounded). Therefore we will not need to deal with extended real-valued functions in this chapter; rather we will focus on real-valued and complex-valued functions.

C. Heil, *Introduction to Real Analysis*, Graduate Texts in Mathematics 280,
https://doi.org/10.1007/978-3-030-26903-6_6

6.1 Absolutely Continuous Functions

To motivate the definition of absolute continuity, recall that a function $f\colon [a,b] \to \mathbb{C}$ is *uniformly continuous* on $[a,b]$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|x-y| < \delta \quad \Longrightarrow \quad |f(x)-f(y)| < \varepsilon.$$

Absolutely continuous functions satisfy a similar but more stringent requirement.

Definition 6.1.1 (Absolutely Continuous Function). We say that a function $f\colon [a,b] \to \mathbb{C}$ is *absolutely continuous on* $[a,b]$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for any finite or countably infinite collection of nonoverlapping subintervals $\{[a_j,b_j]\}$ of $[a,b]$, we have

$$\sum_j (b_j - a_j) < \delta \quad \Longrightarrow \quad \sum_j |f(b_j)-f(a_j)| < \varepsilon. \tag{6.1}$$

We denote the class of absolutely continuous functions on $[a,b]$ by

$$\mathrm{AC}[a,b] = \bigl\{f\colon[a,b]\to\mathbb{C} \;:\; f \text{ is absolutely continuous on } [a,b]\bigr\}. \qquad \diamondsuit$$

Problem 6.1.7 asks for a proof that a complex-valued function is absolutely continuous if and only if its real and imaginary parts are each absolutely continuous.

The Cantor–Lebesgue function φ is uniformly continuous and has bounded variation on $[0,1]$, but we will show that it is not absolutely continuous. The point is that we can find intervals $[a_j,b_j]$ with small total length such that the sum of $|\varphi(b_j)-\varphi(a_j)|$ is large.

Example 6.1.2. Let φ be the Cantor–Lebesgue function, and set

$$[a_1,b_1] = [0,\tfrac{1}{3}] \qquad \text{and} \qquad [a_2,b_2] = [\tfrac{2}{3},1].$$

Then

$$\sum_{j=1}^{2} (b_j-a_j) = \frac{2}{3} \qquad \text{and} \qquad \sum_{j=1}^{2} |\varphi(b_j)-\varphi(a_j)| = 1.$$

Using a similar idea, for each n we can find 2^n nonoverlapping intervals $[a_j,b_j]$, each of length 3^{-n}, such that $\varphi(b_j)-\varphi(a_j) = 2^{-n}$. Therefore, for this collection $\bigl\{[a_j,b_j]\bigr\}_{j=1,\dots,2^n}$ we have

$$\sum_{j=1}^{2^n} (b_j-a_j) = \left(\frac{2}{3}\right)^n \qquad \text{and} \qquad \sum_{j=1}^{2^n} |\varphi(b_j)-\varphi(a_j)| = 1.$$

Since we can do this for every $n \in \mathbb{N}$, it follows that φ is not absolutely continuous on $[0,1]$. $\diamondsuit$

By considering a collection $\{[c,d]\}$ that contains only a single subinterval of $[a,b]$, equation (6.1) implies that all absolutely continuous functions are uniformly continuous. The next lemma gives implications between Lipschitz continuity, absolute continuity, and bounded variation.

Lemma 6.1.3. (a) *Every Lipschitz function on* $[a,b]$ *is absolutely continuous on* $[a,b]$.

(b) *Every absolutely continuous function on* $[a,b]$ *has bounded variation on* $[a,b]$.

Proof. (a) Suppose that f is Lipschitz on $[a,b]$, and let K be a Lipschitz constant. Given $\varepsilon > 0$, let $\delta = \varepsilon/K$. If $\{[a_j,b_j]\}_j$ is any countable collection of nonoverlapping intervals in $[a,b]$ such that $\sum (b_j - a_j) < \delta$, then

$$\sum_j |f(b_j) - f(a_j)| \;\le\; K \sum_j (b_j - a_j) \;\le\; K\delta \;=\; \varepsilon.$$

(b) Suppose that f is absolutely continuous on $[a,b]$. Set $\varepsilon = 1$, and let δ be the corresponding number whose existence is given in the definition of absolute continuity. Let $[c,d]$ be any subinterval of $[a,b]$ with length $d-c < \delta$. If $\Gamma = \{c = x_0 < \cdots < x_n = d\}$ is a finite partition of $[c,d]$, then equation (6.1) implies that

$$S_\Gamma \;=\; \sum_{j=1}^{n} |f(x_j) - f(x_{j-1})| \;<\; \varepsilon \;=\; 1.$$

Taking the supremum over all such partitions of $[c,d]$, we obtain $V[f;c,d] \le 1$. Write $[a,b]$ as a union of N nonoverlapping intervals $[c_k,d_k]$ that each have length less than δ. Then by applying Lemma 5.2.12 we see that

$$V[f;a,b] \;=\; \sum_{k=1}^{N} V[f;c_k,d_k] \;\le\; N \;<\; \infty. \qquad \square$$

Example 6.1.2 shows that the implication in part (b) of Lemma 6.1.3 is not reversible, and the following example shows that the converse of part (a) does not hold either.

Example 6.1.4. We saw in Lemma 5.2.5 that every function that is differentiable everywhere on $[a,b]$ and has a bounded derivative is Lipschitz. We cannot prove it yet, but we will see in Corollary 6.3.3 that a function that is differentiable everywhere on $[a,b]$ and has an *integrable* derivative is absolutely continuous (this is a consequence of the *Banach–Zaretsky Theorem*, see Theorem 6.3.1). Therefore, any differentiable function whose derivative is integrable but unbounded will be absolutely continuous but not Lipschitz. Problem 6.4.19 shows that one specific example is $|x|^{3/2} \sin \frac{1}{x}$ on the interval $[-1,1]$. $\diamondsuit$

Combining these facts with other inclusions that we obtained in earlier chapters, we see that

$$C^1[a,b] \subsetneq \mathrm{Lip}[a,b] \subsetneq \mathrm{AC}[a,b] \subsetneq \mathrm{BV}[a,b] \subsetneq L^\infty[a,b] \subsetneq L^1[a,b].$$

6.1.1 Differentiability of Absolutely Continuous Functions

According to Corollary 5.4.3, all functions that have bounded variation are differentiable a.e. and have integrable derivatives. Since absolutely continuous functions have bounded variation, we immediately obtain the following result.

Corollary 6.1.5. *If $f \in \mathrm{AC}[a,b]$, then $f'(x)$ exists for almost every x, and $f' \in L^1[a,b]$.* $\diamondsuit$

The next lemma answers one of the questions that we posed immediately after Lemma 5.2.9.

Lemma 6.1.6. *If $g \in L^1[a,b]$, then its indefinite integral*

$$G(x) \;=\; \int_a^x g(t)\,dt, \qquad x \in [a,b],$$

has the following properties:

(a) *G is absolutely continuous on $[a,b]$,*

(b) *G is differentiable at almost every point of $[a,b]$, and*

(c) *$G' \in L^1[a,b]$.*

Proof. Fix any $\varepsilon > 0$. Since g is integrable, Exercise 4.5.5 implies that there exists a constant $\delta > 0$ such that $\int_E |g| < \varepsilon$ for every measurable set E with measure $|E| < \delta$. Let $\{[a_j,b_j]\}$ be a countable collection of nonoverlapping subintervals of $[a,b]$ that satisfies $\sum (b_j - a_j) < \delta$, and set $E = \cup (a_j,b_j)$. Then $|E| < \delta$, so

$$\sum_j |G(b_j) - G(a_j)| \;=\; \sum_j \left| \int_{a_j}^{b_j} g \right| \;\le\; \sum_j \int_{a_j}^{b_j} |g| \;=\; \int_E |g| \;<\; \varepsilon.$$

Thus $G \in \mathrm{AC}[a,b]$. Finally, the fact that G' exists a.e. and is integrable is a consequence of Corollary 6.1.5. $\square$

However, we still cannot say whether G' equals g! We will address this issue in Section 6.4 (see Theorem 6.4.2 in particular).

Problems

6.1.7. Given $f\colon [a,b] \to \mathbb{C}$, write $f = f_r + i f_i$ where f_r and f_i are real-valued. Prove that $f \in \mathrm{AC}[a,b]$ if and only if $f_r,\ f_i \in \mathrm{AC}[a,b]$.

6.1.8. Prove that if $f,\ g \in \mathrm{AC}[a,b]$, then the following statements hold.

(a) $|f| \in \mathrm{AC}[a,b]$.

(b) $\alpha f + \beta g \in \mathrm{AC}[a,b]$ for all $\alpha,\ \beta \in \mathbb{C}$.

(c) $fg \in \mathrm{AC}[a,b]$.

(d) If $|g(x)| \ge \delta > 0$ for all $x \in [a,b]$, then $f/g \in \mathrm{AC}[a,b]$.

6.1.9. Prove that $f \in \mathrm{AC}[a,b]$ if and only if for every $\varepsilon > 0$ there exists some $\delta > 0$ such that for every *finite* collection of nonoverlapping subintervals $\{[a_j, b_j]\}_{j=1,\dots,N}$ of $[a,b]$, we have

$$\sum_{j=1}^{N} (b_j - a_j) \,<\, \delta \quad \implies \quad \sum_{j=1}^{N} |f(b_j) - f(a_j)| \,<\, \varepsilon.$$

6.1.10. (a) Prove that $\mathrm{AC}[a,b]$ is a *closed* subspace of $\mathrm{BV}[a,b]$ with respect to the norm $\|f\|_{\mathrm{BV}}$ defined in Problem 5.2.26. That is, show that if $f_n \in \mathrm{AC}[a,b]$, $f \in \mathrm{BV}[a,b]$, and $\|f - f_n\|_{\mathrm{BV}} \to 0$, then $f \in \mathrm{AC}[a,b]$.

(b) Exhibit functions f_n and f such that $f_n \in \mathrm{AC}[a,b]$ and f_n converges uniformly to $f \in \mathrm{BV}[a,b]$, but $f \notin \mathrm{AC}[a,b]$. Thus the uniform limit of absolutely continuous functions need not be absolutely continuous.

6.1.11. Let E be a measurable subset of $\mathbb{R}^d$ with $0 < |E| < \infty$ and assume that $f\colon E \to [-\infty, \infty]$ is integrable. Define $g(x) = \int_E |f(t) - x|\, dt$ for $x \in \mathbb{R}$.

(a) Prove that g is absolutely continuous on every finite interval $[a,b]$, and $g(x) \to \infty$ as $x \to \pm\infty$.

(b) Find g', and prove that $g(x) = \inf_{y \in \mathbb{R}} g(y)$ if and only if $\bigl|\{f > x\}\bigr| = \bigl|\{f < x\}\bigr|$.

6.2 Growth Lemmas

In Section 6.3 we will prove the Banach–Zaretsky Theorem, which gives a reformulation of absolute continuity that is related to the issue of whether a function maps sets with measure zero to sets with measure zero. To prove Banach–Zaretsky we need two lemmas for real-valued functions, which are quite striking in their own right. These are “growth lemmas” in the sense that they give an upper bound to the measure of the direct image $f(E)$ in terms of the function f and the set E. A forerunner of our first lemma appeared as Problem 5.2.21, which states that if f is Lipschitz on the entire interval $[a,b]$ and K is a Lipschitz constant for f, then $|f(E)|_e \le K\,|E|_e$ for every set $E \subseteq [a,b]$. In particular, if f is differentiable on $[a,b]$ and f' is bounded on $[a,b]$, then f is Lipschitz and $K = \|f'\|_\infty$ is a Lipschitz constant. However, in order to prove the Banach–Zaretsky Theorem we will need to show that if f' is bounded *on a single subset* E then the estimate $|f(E)|_e \le K\,|E|_e$ holds *for that set* E (with $K = \sup_{x\in E} |f'(x)|$). We need to obtain this estimate without assuming that f' is bounded on all of $[a,b]$. We cannot assume that f is Lipschitz on $[a,b]$, so Problem 5.2.21 is not applicable. Instead, we have to be more sophisticated in order to obtain the desired estimate. (The first published proof of Lemma 6.2.1 of which we are aware is the comparatively "recent" paper of Varberg [Var65], though he comments that this result is "an elegant inequality which the author discovered lying buried as an innocent problem in Natanson's book [Nat55].")

Lemma 6.2.1 (Growth Lemma I). *Let E be any subset of $[a,b]$. If $f\colon [a,b] \to \mathbb{R}$ is differentiable at* every *point of E and*

$$M_E \;=\; \sup_{x\in E} |f'(x)| \;<\; \infty,$$

then

$$|f(E)|_e \;\le\; M_E\,|E|_e.$$

Proof. Choose any $\varepsilon > 0$. If $x \in E$, then

$$\lim_{\substack{y\to x,\\ y\in[a,b]}} \frac{|f(x)-f(y)|}{|x-y|} \;=\; |f'(x)| \;\le\; M_E.$$

Therefore, there exists an integer $n_x \in \mathbb{N}$ such that

$$y \in [a,b],\ |x-y| < \frac{1}{n_x} \quad\Longrightarrow\quad |f(x)-f(y)| \le (M_E+\varepsilon)\,|x-y|. \tag{6.2}$$

For each $n \in \mathbb{N}$, let

$$E_n \;=\; \bigl\{x \in E : n_x \le n\bigr\}.$$

The sets E_n are nested increasing ($E_1 \subseteq E_2 \subseteq \cdots$), and their union is E. We do not know whether E_n is a measurable set, but fortunately Problem 2.4.8 tells us that continuity from below holds for *exterior* Lebesgue measure. Therefore

$$|E|_e \;=\; \lim_{n\to\infty} |E_n|_e. \tag{6.3}$$

The images $f(E_n)$ are also nested increasing and their union is $f(E)$, so we likewise have

$$|f(E)|_e \;=\; \lim_{n\to\infty} |f(E_n)|_e. \tag{6.4}$$

Fix any particular integer n. By the definition of exterior Lebesgue measure, there exists a collection of countably many boxes $\{I_n^k\}_k$ such that

$$E_n \subseteq \bigcup_k I_n^k \qquad \text{and} \qquad \sum_k |I_n^k| \le |E_n|_e + \varepsilon. \tag{6.5}$$

Since the boxes I_n^k are subsets of the real line, they are simply closed intervals. By replacing I_n^k with $I_n^k \cap [a,b]$, we may assume that $I_n^k \subseteq [a,b]$ for each n and k. Further, by subdividing if necessary, we may assume that each interval I_n^k has length less than $1/n$.

Suppose that x and y are any two points in $E_n \cap I_n^k$. Then, since $x \in E_n$, we have $n_x \le n$. Also, since x and y belong to I_n^k, whose length is less than $1/n$,

$$|x-y| < \frac{1}{n} \le \frac{1}{n_x}.$$

It therefore follows from equation (6.2) that

$$|f(x)-f(y)| \le (M_E+\varepsilon)\,|x-y| \le (M_E+\varepsilon)\,|I_n^k|.$$

Since this is true for all $x, y \in E_n \cap I_n^k$, we conclude that

$$\operatorname{diam}\bigl(f(E_n \cap I_n^k)\bigr) = \sup\bigl\{|f(x)-f(y)| : x, y \in E_n \cap I_n^k\bigr\} \le (M_E+\varepsilon)\,|I_n^k|.$$

This implies that $f(E_n \cap I_n^k)$ is contained in an interval of length at most $(M_E+\varepsilon)\,|I_n^k|$. Hence

$$|f(E_n \cap I_n^k)|_e \le (M_E+\varepsilon)\,|I_n^k|. \tag{6.6}$$

Consequently,

$$\begin{aligned}
|f(E_n)|_e &= \Bigl|\bigcup_k f(E_n \cap I_n^k)\Bigr|_e && \text{(by equation (6.5))}\\
&\le \sum_k |f(E_n \cap I_n^k)|_e && \text{(by subadditivity)}\\
&\le (M_E+\varepsilon) \sum_k |I_n^k| && \text{(by equation (6.6))}\\
&\le (M_E+\varepsilon)\bigl(|E_n|_e+\varepsilon\bigr) && \text{(by equation (6.5)).}
\end{aligned}$$

Therefore, by applying equations (6.3) and (6.4), we see that

$$\begin{aligned}
|f(E)|_e &= \lim_{n\to\infty} |f(E_n)|_e\\
&\le (M_E+\varepsilon) \lim_{n\to\infty} \bigl(|E_n|_e+\varepsilon\bigr)\\
&= (M_E+\varepsilon)\bigl(|E|_e+\varepsilon\bigr).
\end{aligned}$$

Since ε is arbitrary, the result follows. $\square$

One immediate consequence of Lemma 6.2.1 is that if f is differentiable everywhere on E and $f' = 0$ on E, then $|f(E)| = 0$. The following lemma extends this to functions whose derivative is zero almost everywhere on E, and also proves that the converse statement holds (compare the original proof of the "$\Leftarrow$" direction that appears in [SV69]).

Corollary 6.2.2. *Let $f\colon [a,b] \to \mathbb{R}$ and $E \subseteq [a,b]$ be given. If f is differentiable at* every point *of E, then*

$$f' = 0 \text{ a.e. on } E \qquad \Longleftrightarrow \qquad |f(E)| = 0. \tag{6.7}$$

Proof. $\Rightarrow$. Suppose that $f' = 0$ a.e. on E, and let $E_0 = \{x \in E : f'(x) = 0\}$. Then, by Lemma 6.2.1,

$$|f(E_0)|_e \;\le\; 0 \cdot |E_0|_e \;=\; 0.$$

On the other hand, if $k > 0$ then $E_k = \{x \in E : 0 < |f'(x)| \le k\}$ has measure zero, so Lemma 6.2.1 implies that

$$|f(E_k)|_e \;\le\; k\,|E_k|_e \;=\; 0.$$

Since $E = \cup_{k=0}^{\infty} E_k$, it follows that $|f(E)| = |\cup_{k=0}^{\infty} f(E_k)| = 0$.

$\Leftarrow$. Assume that $|f(E)| = 0$. Our goal is to show that

$$D \;=\; \{x \in E : |f'(x)| > 0\}$$

has measure zero. For each $n \in \mathbb{N}$, let

$$D_n \;=\; \left\{x \in D \;:\; \left|\frac{f(y) - f(x)}{y - x}\right| \;\ge\; \frac{1}{n} \;\text{ for all } y \text{ with } 0 < |y - x| < \frac{1}{n}\right\}.$$

If $x \in D$, then $f'(x)$ exists and is strictly positive. It follows from this that $x \in D_n$ for some n. Therefore $D = \cup D_n$, so it suffices to show that $|D_n| = 0$ for every n.

Let n be a fixed positive integer, and let J be any closed subinterval of $[a,b]$ whose length is less than $1/n$. We will show that $|D_n \cap J| = 0$. To do this, fix any $\varepsilon > 0$ (and note for later reference that ε is chosen independently of n). Since $|f(E)| = 0$, there exist countably many boxes (closed finite intervals) Q_k such that

$$f(E) \subseteq \bigcup_k Q_k \qquad \text{and} \qquad \sum_k |Q_k| \;<\; \varepsilon.$$

If we set

$$A_k \;=\; f^{-1}(Q_k) \,\cap\, D_n \,\cap\, J,$$

then $D_n \cap J = \cup_k A_k$.

Suppose that x and y are two distinct points in A_k. Then x and y belong to J, so $0 < |y-x| < 1/n$. But we also have $x, y \in D_n$, so this implies that

$$|y-x| \ \le \ n\,|f(y)-f(x)|. \tag{6.8}$$

The preceding equation also holds if $x = y$. Assuming that A_k is nonempty, we can therefore estimate its measure as follows:

$$\begin{aligned}
|A_k|_e \ &\le \ \operatorname{diam}(A_k) \\
&= \ \sup\bigl\{|y-x| : x, y \in A_k\bigr\} && \text{(definition of diameter)} \\
&\le \ \sup\bigl\{n\,|f(y)-f(x)| : x, y \in A_k\bigr\} && \text{(by equation (6.8))} \\
&\le \ n \sup\bigl\{|w-z| : w, z \in Q_k\bigr\} && \text{(since } f(A_k) \subseteq Q_k\text{)} \\
&= \ n\,\operatorname{diam}(Q_k) && \text{(definition of diameter)} \\
&= \ n\,|Q_k| && \text{(since } Q_k \text{ is an interval).}
\end{aligned}$$

The estimate $|A_k|_e \le n\,|Q_k|$ also holds if A_k is empty, so we obtain

$$|D_n \cap J|_e \ \le \ \sum_k |A_k|_e \ \le \ n \sum_k |Q_k| \ < \ n\varepsilon.$$

Since ε is arbitrary (and independent of n), we conclude that $D_n \cap J$ has measure zero.

Finally, since $[a,b]$ is a finite interval, we can cover it with finitely many subintervals $J_1, \dots, J_m$ that each have length at most $1/n$. Our work above shows that $|D_n \cap J_k| = 0$ for each k, so finite subadditivity implies that $|D_n| = 0$. □

If we let φ be the Cantor–Lebesgue function, then $\varphi' = 0$ a.e. on the Cantor set C, simply because $|C| = 0$. However, we saw in Example 5.1.4 that $|\varphi(C)| = 1$. Therefore, we cannot relax the hypotheses of Corollary 6.2.2 from "f is differentiable at every point of E" to "f is differentiable at almost every point of E," at least for the "$\Rightarrow$" direction of equation (6.7). On the other hand, the following corollary shows that we can allow this generalization in the "$\Leftarrow$" direction.

Corollary 6.2.3. *Fix $E \subseteq [a,b]$. If $f\colon [a,b] \to \mathbb{R}$ is differentiable a.e. on E and $|f(E)| = 0$, then $f' = 0$ a.e. on E.*

Proof. Let $A = \{x \in E : f'(x) \text{ exists}\}$. Then $Z = E \setminus A$ has measure zero, and $|f(A)| \le |f(E)| = 0$. Since f is differentiable at every point of A, Corollary 6.2.2 implies that $f' = 0$ a.e. on A. Since $|Z| = 0$, it follows that $f' = 0$ a.e. on $E = A \cup Z$. □

Corollary 6.2.3 will be useful to us in Section 6.5, when we consider the Chain Rule in connection with absolutely continuous functions.

Our second growth lemma (which also appears to have been first proved in [Var65]) relates the exterior measure of $f(E)$ to the integral of $|f'|$ on E. As we have observed before, a measurable function need not map measurable sets to measurable sets. Therefore, even though we assume in this lemma that the set E and the function f are measurable, the image $f(E)$ might not be measurable.

Lemma 6.2.4 (Growth Lemma II). *Assume that $f\colon [a,b] \to \mathbb{R}$ is measurable. If E is a measurable subset of $[a,b]$ and f is differentiable at* every *point of E, then*

$$|f(E)|_e \;\le\; \int_E |f'|.$$

Proof. By Problem 3.2.19, the derivative $f'\colon E \to \mathbb{R}$ is a measurable function on E. Hence $\int_E |f'|$ exists as a nonnegative, extended real number.

Fix any $\varepsilon > 0$, and for each $k \in \mathbb{N}$ define

$$E_k \;=\; \bigl\{x \in E : (k-1)\varepsilon \;\le\; |f'(x)| \;<\; k\varepsilon\bigr\}.$$

The sets E_k are measurable and disjoint, and since f is differentiable everywhere on E we have $E = \cup E_k$. Since Lebesgue measure is countably additive, it follows that

$$|E| \;=\; \sum_{k=1}^{\infty} |E_k|.$$

Lemma 6.2.1 implies that $|f(E_k)|_e \le k\varepsilon\,|E_k|$, so we see that

$$\begin{aligned}
|f(E)|_e \;=\; \left|\bigcup_{k=1}^{\infty} f(E_k)\right|_e \;&\le\; \sum_{k=1}^{\infty} |f(E_k)|_e \\
&\le\; \sum_{k=1}^{\infty} k\varepsilon\,|E_k| \\
&=\; \sum_{k=1}^{\infty} (k-1)\varepsilon\,|E_k| \;+\; \sum_{k=1}^{\infty} \varepsilon\,|E_k| \\
&\le\; \sum_{k=1}^{\infty} \int_{E_k} |f'| \;+\; \varepsilon\,|E| \\
&=\; \int_E |f'| \;+\; \varepsilon\,|E|.
\end{aligned}$$

Since ε is arbitrary and $|E| < \infty$, the result follows. $\square$

Problems

6.2.5. Suppose that $f\colon [a,b] \to \mathbb{C}$ is differentiable at every point of $E \subseteq [a,b]$. Prove that $f' = 0$ a.e. on any subset of E where f is constant.

6.2.6. Suppose that $f\colon [a,b] \to \mathbb{R}$ is differentiable a.e. on a measurable set $E \subseteq [a,b]$. Prove that if $f \in \mathrm{AC}[a,b]$, then

$$|f(E)|_e \;\le\; \int_E |f'|.$$

Show by example that the assumption of absolute continuity is necessary.

6.3 The Banach–Zaretsky Theorem

In this section we will prove the Banach–Zaretsky Theorem, which tells us what properties that a function $f\colon [a,b] \to \mathbb{R}$ needs to possess in addition to continuity in order to be absolutely continuous. Specifically, f must map sets with measure zero to sets with measure zero, and we must also know either that f has bounded variation, or that f is differentiable almost everywhere and f' is integrable. The result is similar for complex-valued functions, except that both the real and imaginary parts of f must map sets of measure zero to sets of measure zero (compare Problem 6.3.5).

Theorem 6.3.1 (Banach–Zaretsky Theorem). *If $f\colon [a,b] \to \mathbb{R}$ is a real-valued function on $[a,b]$, then the following three statements are equivalent.*

(a) $f \in \mathrm{AC}[a,b]$.

(b) *f is continuous, $f \in \mathrm{BV}[a,b]$, and*

$$A \subseteq [a,b],\ |A| = 0 \implies |f(A)| = 0.$$

(c) *f is continuous, f is differentiable a.e., $f' \in L^1[a,b]$, and*

$$A \subseteq [a,b],\ |A| = 0 \implies |f(A)| = 0.$$

If $f\colon [a,b] \to \mathbb{C}$ is a complex-valued function and we write $f = f_r + i f_i$ where f_r and f_i are real-valued, then the same three statements are equivalent if we replace "$|f(A)| = 0$" by "$|f_r(A)| = |f_i(A)| = 0$."

Proof. Since we can split a complex-valued function into real and imaginary parts, it suffices to prove the result for real-valued functions.

(a) $\Rightarrow$ (b). Every absolutely continuous function is continuous and has bounded variation, so our task is to show that f maps sets with measure zero to sets with measure zero.

Suppose that A is a subset of $[a,b]$ that has measure zero. Since the two-element set $\{a,b\}$ has measure zero and its image $\{f(a), f(b)\}$ also has measure zero, it suffices to assume that A is contained within the open interval (a,b). Fix $\varepsilon > 0$. By the definition of absolute continuity, there exists some $\delta > 0$ such that if $\{[a_j, b_j]\}$ is any countable collection of nonoverlapping subintervals of $[a,b]$ that satisfy $\sum (b_j - a_j) < \delta$, then $\sum |f(b_j) - f(a_j)| < \varepsilon$.

By Theorem 2.1.27, there is an open set $U \supseteq A$ whose measure satisfies

$$|U| \;<\; |A| + \delta \;=\; \delta.$$

By replacing U with the open set $U \cap (a,b)$, we may assume that $U \subseteq (a,b)$. Since U is open, we can write it as a union of countably many disjoint open intervals contained in (a,b), say

$$U \;=\; \bigcup_j (a_j, b_j).$$

Fix any particular j. Since f is continuous on the closed interval $[a_j, b_j]$, there is a point in $[a_j, b_j]$ where f attains its minimum value on $[a_j, b_j]$, and another point where f attains its maximum. Let c_j and d_j be points in $[a_j, b_j]$ such that f has a max at one point and a min at the other. By interchanging their roles if necessary, we may assume that $c_j \le d_j$. Because f is continuous, the Intermediate Value Theorem implies that the image of $[a_j, b_j]$ under f is the set of all points between $f(c_j)$ and $f(d_j)$. Hence the exterior Lebesgue measure of this image is

$$\bigl|f([a_j, b_j])\bigr|_e \;=\; |f(d_j) - f(c_j)|.$$

Now, $[c_j, d_j] \subseteq [a_j, b_j]$, so $\{[c_j, d_j]\}$ is a collection of nonoverlapping subintervals of $[a,b]$. Moreover,

$$\sum_j |d_j - c_j| \;\le\; \sum_j (b_j - a_j) \;=\; |U| \;<\; \delta.$$

Therefore $\sum |f(d_j) - f(c_j)| < \varepsilon$, so

$$|f(A)|_e \;\le\; |f(U)|_e \;\le\; \sum_j \bigl|f([a_j, b_j])\bigr|_e \;=\; \sum_j \bigl|f(d_j) - f(c_j)\bigr| \;<\; \varepsilon.$$

Since ε is arbitrary, we conclude that $|f(A)| = 0$.

(b) $\Rightarrow$ (c). This follows from Corollary 5.4.3.

(c) $\Rightarrow$ (a). Assume that f is real-valued and statement (c) holds. Let D be the set of points where f is differentiable. By hypothesis, $Z = [a,b] \setminus D$ has measure zero, so $D = [a,b] \setminus Z$ is a measurable set.

Let $[c,d]$ be an arbitrary subinterval of $[a,b]$. Since f is continuous, the Intermediate Value Theorem implies that f must take every value between

$f(c)$ and $f(d)$. Therefore $f([c,d])$, the image of $[c,d]$ under f, must contain an interval of length $|f(d)-f(c)|$. Define

$$B = [c,d] \cap D \qquad \text{and} \qquad A = [c,d] \backslash D.$$

The set A has measure zero, so $|f(A)| = 0$ by hypothesis. Since f is differentiable at every point of B, we therefore compute that

$$\begin{aligned}
|f(d)-f(c)| &\le \bigl|f([c,d])\bigr|_e \\
&= \bigl|f(B) \cup f(A)\bigr|_e && \text{(since } [c,d] = B \cup A\text{)} \\
&\le |f(B)|_e + |f(A)|_e && \text{(by subadditivity)} \\
&\le \int_B |f'| + 0 && \text{(by Lemma 6.2.4)} \\
&\le \int_c^d |f'| && \text{(since } B \subseteq [c,d]\text{).} \qquad (6.9)
\end{aligned}$$

This calculation holds for every subinterval $[c,d]$ of $[a,b]$.

Now fix $\varepsilon > 0$. Because f' is integrable, Exercise 4.5.5 implies that there is some constant $\delta > 0$ such that for every measurable set $E \subseteq [a,b]$ we have

$$|E| < \delta \quad \Longrightarrow \quad \int_E |f'| < \varepsilon.$$

Let $\bigl\{[a_j,b_j]\bigr\}$ be any countable collection of nonoverlapping subintervals of $[a,b]$ such that $\sum (b_j - a_j) < \delta$. Then $E = \cup\, [a_j,b_j]$ is a measurable subset of $[a,b]$ and $|E| < \delta$, so $\int_E |f'| < \varepsilon$. Applying equation (6.9) to each subinterval $[a_j,b_j]$, it follows that

$$\sum_j \bigl|f(b_j) - f(a_j)\bigr| \le \sum_j \int_{a_j}^{b_j} |f'| = \int_E |f'| < \varepsilon.$$

Hence f is absolutely continuous on $[a,b]$. □

We will give several implications of the Banach–Zaretsky Theorem. Our first corollary shows that absolutely continuous functions preserve measurability.

Corollary 6.3.2. *Absolutely continuous functions map sets of measure zero to sets of measure zero, and they map measurable sets to measurable sets.*

Proof. Assume that f is absolutely continuous. If f is real-valued, then the Banach–Zaretsky Theorem directly implies that f maps sets of measure zero to sets of measure zero. On the other hand, if f is complex-valued then the

Banach–Zaretsky Theorem tells us that both the real and imaginary parts of f map sets of measure zero to sets of measure zero. Applying Problem 6.3.5, it follows that f maps sets of measure zero to sets of measure zero. In either case, we can apply Lemma 2.3.9 and conclude that f also maps measurable sets to measurable sets. □

To motivate our second implication, recall from Lemma 5.2.5 that if f is differentiable everywhere on $[a,b]$ and f' is bounded, then f is Lipschitz and therefore absolutely continuous. What happens if f is differentiable everywhere on $[a,b]$ but we only know that f' is *integrable*? Although such a function need not be Lipschitz, the next corollary shows that f is absolutely continuous.

Corollary 6.3.3. *If $f\colon [a,b] \to \mathbb{C}$ is differentiable everywhere on $[a,b]$ and $f' \in L^1[a,b]$, then $f \in \mathrm{AC}[a,b]$.*

Proof. We may assume that f is real-valued. Let A be any subset of $[a,b]$ that has measure zero. Since f is differentiable everywhere, it is continuous and hence measurable. Becuase A is a measurable set, we can therefore apply Lemma 6.2.4 to obtain the estimate

$$|f(A)|_e \;\le\; \int_A |f'| \;=\; 0.$$

Consequently, the Banach–Zaretsky Theorem implies that f is absolutely continuous. □

Problem 6.3.8 gives a generalization of Corollary 6.3.3: If f is differentiable at all but *countably many* points and $f' \in L^1[a,b]$, then $f \in \mathrm{AC}[a,b]$. As shown by the Cantor–Lebesgue function, we cannot weaken this hypothesis further to just differentiability *almost everywhere.*

We also cannot remove the hypothesis in Corollary 6.3.3 that f' is integrable. For example, Problem 6.3.12 shows that

$$g(x) \;=\; x^2 \sin \tfrac{1}{x^2}$$

is differentiable everywhere on $[-1,1]$, but g' is not integrable and g does not even have bounded variation on $[-1,1]$, so g is certainly not absolutely continuous.

Our final implication uses the Banach–Zaretsky Theorem to show that the only functions that are both absolutely continuous and singular are constant functions.

Corollary 6.3.4 (AC + Singular Implies Constant). *If $f\colon [a,b] \to \mathbb{C}$ is both absolutely continuous and singular, then f is constant.*

Proof. It suffices to assume that f is real-valued. Suppose that $f \in \mathrm{AC}[a,b]$ and $f' = 0$ a.e., and define

$$E = \{f' = 0\} \qquad \text{and} \qquad Z = [a,b]\backslash E.$$

Since $|Z| = 0$, the Banach–Zaretsky Theorem implies that $|f(Z)| = 0$. Since E is measurable and f is differentiable on E, Lemma 6.2.4 implies that

$$|f(E)|_e \leq \int_E |f'| = 0.$$

Therefore the range of f has measure zero, because

$$|\mathrm{range}(f)|_e = \bigl|f([a,b])\bigr|_e = |f(E) \cup f(Z)|_e \leq |f(E)|_e + |f(Z)|_e = 0.$$

However, f is continuous and $[a,b]$ is compact, so the Intermediate Value Theorem implies that the range of f is a either single point or a closed interval $[c,d]$. Since range(f) has measure zero, we conclude that it is a single point, and therefore f is constant. $\square$

Problems

6.3.5. Define Lebesgue measure on the complex plane by identifying $\mathbb{C}$ with $\mathbb{R}^2$ in the natural way. Given $f\colon X \to \mathbb{C}$, write $f = f_r + i f_i$ where f_r and f_i are real-valued. Prove that if $|f_r(X)| = |f_i(X)| = 0$, then $|f(X)| = 0$, but show by example that the converse implication can fail.

6.3.6. Assume that $g\colon [a,b] \to [c,d]$ and $f\colon [c,d] \to \mathbb{C}$ are continuous. Prove the following statements (compare Problems 5.2.20 and 6.3.7).

(a) If f is Lipschitz and $g \in \mathrm{AC}[a,b]$, then $f \circ g \in \mathrm{AC}[a,b]$.

(b) If $f \in \mathrm{AC}[c,d]$, $g \in \mathrm{AC}[a,b]$, and g is monotone increasing on $[a,b]$, then $f \circ g \in \mathrm{AC}[a,b]$.

(c) If $f \in \mathrm{AC}[c,d]$ and $g \in \mathrm{AC}[a,b]$, then

$$f \circ g \in \mathrm{AC}[a,b] \qquad \Longleftrightarrow \qquad f \circ g \in \mathrm{BV}[a,b].$$

Remark: This problem will be used in the proof of Corollary 6.5.8.

6.3.7. Prove the following statements (compare Problem 6.3.6).

(a) $f(x) = x^{1/2}$ is monotone increasing and absolutely continuous on $[0,1]$ and $g(t) = t^2 \sin^2 \frac{1}{t}$ is Lipschitz on $[0,1]$, yet $f \circ g$ is not absolutely continuous.

(b) $f(x) = x^2$ is monotone increasing and absolutely continuous on $[0,1]$ and $g(t) = t \sin \frac{1}{t}$ is not absolutely continuous on $[0,1]$, yet $f \circ g$ is absolutely continuous on $[0,1]$.

6.3.8. Suppose that $f\colon [a,b] \to \mathbb{C}$ is continuous, f is differentiable at all but countably many points of $[a,b]$, and $f' \in L^1[a,b]$. Prove that $f \in \mathrm{AC}[a,b]$.

6.3.9. Assume that $f \in \mathrm{AC}[a,b]$ and there is a continuous function g such that $f' = g$ a.e. Prove that f is differentiable everywhere on $[a,b]$ and $f'(x) = g(x)$ for every $x \in [a,b]$. Show by example that the hypothesis of absolute continuity is necessary.

6.3.10. Suppose that $f\colon [a,b] \to \mathbb{C}$ is differentiable everywhere on $[a,b]$. Prove the following statements.

(a) $f \in \mathrm{AC}[a,b]$ if and only if $f \in \mathrm{BV}[a,b]$.

(b) $f' = 0$ a.e. if and only if f is constant on $[a,b]$.

6.3.11. (a) Suppose that $f \in \mathrm{BV}[a,b]$, f is continuous from the right at $x = a$, and $f \in \mathrm{AC}[a+\delta,b]$ for each $\delta > 0$. Prove that $f \in \mathrm{AC}[a,b]$.

(b) Show by example that the assumption in part (a) that f has bounded variation is necessary.

6.3.12. Define $g(x) = x^2 \sin(1/x^2)$ for $x \neq 0$, and set $g(0) = 0$. Show that $g \in L^1[-1,1]$, g is differentiable everywhere on $[-1,1]$, $g' \notin L^1[-1,1]$, $g \notin \mathrm{BV}[-1,1]$, and $g \notin \mathrm{AC}[-1,1]$.

Remark: This is a special case of Problem 6.3.13, but it may be instructive to work it first.

6.3.13. Fix $a,\ b > 0$ and define $f(x) = |x|^a \sin|x|^{-b}$ for $x \neq 0$ and $f(0) = 0$. According to Problem 5.2.22, f belongs to $\mathrm{BV}[-1,1]$ if and only if $a > b$. Prove that $f \in \mathrm{AC}[-1,1]$ if and only if $a > b$.

6.4 The Fundamental Theorem of Calculus

Following Lemma 5.2.9, we asked two questions: First, is the indefinite integral G of an integrable function g differentiable? Second, if G is differentiable, does $G' = g$? The first question was answered affirmatively in Lemma 6.1.6, and the next lemma will show that $G' = g$ a.e.

Lemma 6.4.1. *If $g \in L^1[a,b]$, then its indefinite integral*

$$G(x) \;=\; \int_a^x g(t)\,dt, \qquad x \in [a,b],$$

is absolutely continuous and satisfies $G' = g$ a.e.

Proof. Because G is the indefinite integral of an integrable function, Lemma 6.1.6 implies that G is absolutely continuous. Applying Corollary 5.5.11 (extend g by zero outside of $[a,b]$, so that it is locally integrable on $\mathbb{R}$), we also see that if $x \in [a,b]$ is a Lebesgue point of g then

$$\frac{G(x+h)-G(x)}{h} \;=\; \frac{1}{h}\int_x^{x+h} g(t)\,dt \;\to\; g(x) \quad \text{as } h \to 0.$$

Therefore G is differentiable at x and $G'(x) = g(x)$. Since almost every point is a Lebesgue point, we conclude that $G' = g$ a.e. $\square$

Now we tie everything together and prove that the absolutely continuous functions are precisely those for which the Fundamental Theorem of Calculus holds.

Theorem 6.4.2 (Fundamental Theorem of Calculus). *If $f\colon [a,b] \to \mathbb{C}$, then the following three statements are equivalent.*

(a) $f \in \mathrm{AC}[a,b]$.

(b) *There exists a function $g \in L^1[a,b]$ such that*

$$f(x) - f(a) \;=\; \int_a^x g(t)\,dt, \qquad \textit{for all } x \in [a,b].$$

(c) *f is differentiable almost everywhere on $[a,b]$, $f' \in L^1[a,b]$, and*

$$f(x) - f(a) \;=\; \int_a^x f'(t)\,dt, \qquad \textit{for all } x \in [a,b].$$

Proof. (a) $\Rightarrow$ (c). Suppose that f is absolutely continuous on $[a,b]$. Corollary 6.1.5 implies that f' exists a.e. and is integrable. It therefore follows from Lemma 6.4.1 that the indefinite integral

$$F(x) \;=\; \int_a^x f'(t)\,dt$$

is absolutely continuous and satisfies $F' = f'$ a.e. Hence $(F-f)' = 0$ a.e., so the function $F - f$ is both absolutely continuous and singular. Applying Corollary 6.3.4, we conclude that $F - f$ is constant. Consequently, for all $x \in [a,b]$ we have

$$F(x) - f(x) \;=\; F(a) - f(a) \;=\; 0 - f(a) \;=\; -f(a).$$

(c) $\Rightarrow$ (b). This follows by taking $g = f'$.

(b) $\Rightarrow$ (a). This follows from Lemma 6.4.1. $\square$

Combining Theorem 6.4.2 with the Banach–Zaretsky Theorem gives us a remarkable list of equivalent characterizations of absolute continuity of functions on $[a,b]$.

6.4.1 Applications of the FTC

We will give several implications of the Fundamental Theorem of Calculus. First, we use the FTC to prove that every function that has bounded variation can be written as the sum of an absolutely continuous function and a singular function.

Corollary 6.4.3. *If $f \in \mathrm{BV}[a,b]$, then $f = g + h$ where $g \in \mathrm{AC}[a,b]$ and h is singular on $[a,b]$. Moreover, g and h are unique up to an additive constant, and we can take*

$$g(x) \;=\; \int_a^x f'(t)\,dt, \qquad \textit{for } x \in [a,b]. \tag{6.10}$$

Proof. Since f has bounded variation on $[a,b]$, we know that f' exists a.e. and is integrable. Therefore the function g given by equation (6.10) is well-defined. Further, $g \in \mathrm{AC}[a,b]$ and $g' = f'$ a.e. by Lemma 6.4.1. Consequently f and g are each differentiable a.e., and $h = f - g$ satisfies $h' = 0$ a.e. Hence g is absolutely continuous and h is singular.

Suppose that we also had $f = g_1 + h_1$ where g_1 is absolutely continuous and h_1 is singular. Then $g - g_1 = h_1 - h$, which implies that $g - g_1$ is both absolutely continuous and singular. Hence $g - g_1$ is a constant, and therefore $h_1 - h$ is the same constant. □

Our second application of the Fundamental Theorem of Calculus relates the total variation of an absolutely continuous function f to the integral of $|f'|$. The special case where f belongs to $C^1[a,b]$ appeared earlier in Problem 5.2.27.

Theorem 6.4.4. *If $f \in \mathrm{AC}[a,b]$, then*

$$V[f;a,b] \;=\; \int_a^b |f'|. \tag{6.11}$$

Proof. Since f has bounded variation, the inequality $\int_a^b |f'| \le V[f;a,b]$ follows immediately from Corollary 5.4.3.

To prove the opposite inequality, we make use of the fact that f is absolutely continuous. The Fundamental Theorem of Calculus tells us that f' is integrable and

$$f(x) - f(a) \;=\; \int_a^x f', \qquad \text{for all } x \in [a,b].$$

Define

$$F(x) \;=\; \int_a^x f' \;=\; f(x) - f(a), \qquad \text{for } x \in [a,b].$$

Applying Lemma 5.2.9, we see that

$$V[F;a,b] \;\le\; \int_a^b |f'|.$$

But f and F only differ by a constant, so they have the same total variation. Therefore $V[f;a,b] = V[F;a,b] \le \int_a^b |f'|$. □

As a corollary, we will show that if f is absolutely continuous, then its total variation function is also absolutely continuous (for the converse implication, see Problem 6.4.18).

Corollary 6.4.5. *Choose $f \in \mathrm{AC}[a,b]$, and let $V(x) = V[f;a,x]$ be the total variation of f on the interval $[a,x]$. Then the following statements hold.*

(a) $V \in \mathrm{AC}[a,b]$.

(b) $V(x) = \int_a^x |f'|$ *for each* $x \in [a,b]$.

(c) $V' = |f'|$ *a.e.*

Proof. Applying Theorem 6.4.4 to f on the interval $[a,x]$, we see that $V(x) = \int_a^x |f'|$. Since $|f'| \in L^1[a,b]$, the Fundamental Theorem of Calculus therefore implies that V is absolutely continuous and $V' = |f'|$ almost everywhere. □

Even though $V' = |f'|$ a.e., the set of points where $V'(x)$ exists can be different than the set of points where $|f'(x)|$ exists (consider $f(x) = |x|$ on the interval $[-1,1]$).

6.4.2 Integration by Parts

As another application of the Fundamental Theorem of Calculus, we prove that integration by parts is valid for absolutely continuous functions.

Theorem 6.4.6 (Integration by Parts). *If f and g are absolutely continuous on $[a,b]$, then*

$$\int_a^b f(x)\, g'(x)\, dx \;=\; f(b)\, g(b) \;-\; f(a)\, g(a) \;-\; \int_a^b f'(x)\, g(x)\, dx. \tag{6.12}$$

Proof. The product $F = fg$ is absolutely continuous by Problem 6.1.8, so F is differentiable at almost every point. At any point t where f and g are both differentiable (which is a.e.), the product rule applies and we have

$$F'(t) \;=\; f(t)\, g'(t) \;+\; f'(t)\, g(t).$$

Since f' and g' are integrable and f and g are bounded, we know that fg' and $f'g$ are each integrable. Applying the Fundamental Theorem of Calculus to the absolutely continuous function F, it follows that for each point $x \in [a,b]$ we have

$$\int_a^x f(t)\, g'(t)\, dt + \int_a^x f'(t)\, g(t)\, dt \;=\; \int_a^x F'(t)\, dt \;=\; F(x) \;-\; F(a).$$

Rearranging, substituting $F = fg$, and taking $x = b$, we obtain equation (6.12). □

We will use integration by parts to prove the following theorem (also compare Problems 7.4.5 and 9.1.32).

Theorem 6.4.7. *If $f \in L^1[a,b]$ satisfies*

$$\int_a^b f(x)\, g(x)\, dx \;=\; 0, \qquad \text{for all } g \in C[a,b], \tag{6.13}$$

then $f = 0$ a.e.

Proof. Before beginning the proof, we observe that if we were allowed to take $g \in L^\infty[a,b]$ in equation (6.13) instead of $g \in C[a,b]$, then the proof would be easy, because we could choose g so that $|g(x)| = 1$ and $f(x)\, g(x) = |f(x)|$. Unfortunately, such a function g need not be continuous, so we must be more careful.

Let $F(x) = \int_a^x f$ for $x \in [a,b]$. Then $F(a) = 0$, and also $F(b) = 0$ since the constant function 1 belongs to $C[a,b]$. Since F is continuous, the Weierstrass Approximation Theorem (Theorem 1.3.4) implies that there exists a polynomial p such that $\|F - p\|_{\mathrm{u}} < \varepsilon$. Set $P(x) = \int_a^x \overline{p(t)}\, dt$. Then P is itself a polynomial, and by using integration by parts we see that

$$\int_a^b F(x)\, \overline{p(x)}\, dx \;=\; F(b)\, P(b) \;-\; F(a)\, P(a) \;-\; \int_a^b f(x)\, P(x)\, dx \;=\; 0.$$

Therefore

$$\int_a^b |F - p|^2\, dx \;=\; \int_a^b |F|^2 \;-\; 2\,\mathrm{Re} \int_a^b F\, \overline{p} \;+\; \int_a^b |p|^2.$$

Since $\int_a^b F\, \overline{p} = 0$ and $\int_a^b |p|^2 \ge 0$, it follows that

$$\int_a^b |F|^2 \;\le\; \int_a^b |F - p|^2\, dx \;\le\; \int_a^b \|F - p\|_{\mathrm{u}}^2\, dx \;<\; \varepsilon^2\, (b - a).$$

But ε is arbitrary and F is continuous, so this implies that $F = 0$ and therefore $F' = 0$. However, $f = F'$ a.e. by the Fundamental Theorem of Calculus, so the result follows. $\square$

Problems

6.4.8. Show that $x^\alpha \in \mathrm{AC}[a,b]$ for each $\alpha > 0$ and $0 \le a < b < \infty$.

6.4.9. Exhibit functions $f \in \mathrm{BV}[a,b]$ and $g \in C^\infty[a,b]$ for which the integration by parts formula given in equation (6.12) fails.

6.4.10. Show that $f\colon [a,b] \to \mathbb{C}$ is Lipschitz if and only if $f \in \mathrm{AC}[a,b]$ and $f' \in L^\infty[a,b]$.

6.4.11. Let $P \subseteq [0,1]$ be a "fat Cantor set" with positive measure, of the type constructed in Problem 2.2.42. Set $U = [0,1]\backslash P$, and define

$$f(x) \;=\; \int_0^x \chi_U(t)\,dt, \qquad \text{for } x \in [0,1].$$

Show that f is absolutely continuous and strictly increasing on $[0,1]$, yet $f' = 0$ on a set that has positive measure.

6.4.12. Suppose that $f\colon [a,b] \to \mathbb{R}$ is differentiable a.e. on $[a,b]$ and $f' \ge 0$ a.e. Must f be monotone increasing on $[a,b]$? What if we also assume that f is absolutely continuous?

6.4.13. Suppose that $f \in L^1(\mathbb{R})$ is such that $f' \in L^1(\mathbb{R})$ and $f \in \mathrm{AC}[a,b]$ for every finite interval $[a,b]$. Show that $\lim_{|x|\to\infty} f(x) = 0 = \int_{-\infty}^{\infty} f'$.

6.4.14. Suppose that functions $f_n \in C^1[0,1]$ satisfy:

(a) $f_n(0) = 0$,

(b) $|f_n'(x)| \le x^{-1/2}$ a.e., and

(c) there is a measurable function h such that $f_n'(x) \to h(x)$ for $x \in [0,1]$.

Prove that there exists an absolutely continuous function f such that f_n converges uniformly to f as $n \to \infty$.

6.4.15. Given $f\colon [0,1] \to \mathbb{R}$, prove that the following two statements are equivalent.

(a) $f \in \mathrm{AC}[0,1]$, $f(0) = 0$, and $f'(x)$ is either 0 or 1 for almost every x.

(b) There is a measurable set $A \subseteq [0,1]$ such that $f(x) = |A \cap [0,x]|$ for all $x \in [0,1]$.

6.4.16. Suppose that $f \in \mathrm{AC}[a,b]$ satisfies $f(a) = 0$. Show that

$$\int_a^b |f(x)\,f'(x)|\,dx \;\le\; \frac{1}{2}\biggl(\int_a^b |f'(x)|\,dx\biggr)^2.$$

6.4.17. (a) Suppose that $f \in \mathrm{BV}[a,b]$ is continuous and real-valued, f' is integrable on $[a,b]$, and

$$\int_a^b f' \;=\; f(b) - f(a).$$

Must f be absolutely continuous? What if f is monotone increasing on $[a,b]$?

(b) Suppose that $g\colon [a,b] \to [c,d]$ is a monotone increasing function that maps $[a,b]$ onto $[c,d]$. Let A be the set of points where g is not differentiable. Prove that $g \in \mathrm{AC}[a,b]$ if and only if $|g(A)| = 0$.

6.4.18. Fix $f \in \mathrm{BV}[a,b]$, and let $V(x) = V[f;a,x]$ for $x \in [a,b]$. Prove that the following three statements are equivalent.

(a) $f \in \mathrm{AC}[a,b]$.

(b) $V \in \mathrm{AC}[a,b]$.

(c) $\int_a^b |f'| = V[f;a,b]$.

Also prove that if the above statements hold, then the positive and negative variations $V^+(x) = V^+[f;a,x]$ and $V^-(x) = V^-[f;a,x]$ are absolutely continuous, $V^+(x) = \int_a^x (f')^+$, and $V^-(x) = \int_a^x (f')^-$.

6.4.19. Define $f(x) = |x|^{3/2} \sin \frac{1}{x}$ for $x \neq 0$, and set $f(0) = 0$. Prove the following facts.

(a) f is differentiable at every point,

(b) $f' \in L^1[-1,1] \setminus L^\infty[-1,1]$,

(c) $f \in \mathrm{AC}[-1,1] \setminus \mathrm{Lip}[-1,1]$.

Remark: This is a special case of both Problems 6.3.13 and 6.4.20, but it may be instructive to work it first.

6.4.20. Fix a, $b > 0$ and define $f(x) = |x|^a \sin |x|^{-b}$ for $x \neq 0$ and $f(0) = 0$. According to Problem 6.3.13, f belongs to $\mathrm{AC}[-1,1]$ if and only if $a > b$. Prove the following statements.

(a) f is differentiable everywhere on $[-1,1]$ if and only if $a > 1$.

(b) $f \in \mathrm{Lip}[-1,1]$ if and only if $a \geq b+1$.

(c) $f \in C^1[-1,1]$ if and only if $a > b+1$.

6.4.21. (a) Given $f \in L^1[a,b]$ and $\varepsilon > 0$, prove that there exists a polynomial $p(x) = \sum_{k=0}^n a_k x^k$ such that $\|f - p\|_1 < \varepsilon$.

(b) Suppose that $f \in L^1[a,b]$ satisfies $\int_a^b f(x)\, x^k\, dx = 0$ for all $k \geq 0$. Prove that $f = 0$ a.e.

(c) Suppose that $f \in L^1[0,1]$ is such that $\int_0^1 f(x)\, x^{2k}\, dx = 0$ for all $k \geq 0$. Prove that $f = 0$ a.e.

6.4.22.* Suppose that f is monotone increasing on $[a,b]$. Prove the following statements.

(a) If we set $f(a+) = \lim_{x\to a^+} f(x)$ and $f(b-) = \lim_{x\to b^-} f(x)$, then

$$\int_a^b f' \;\leq\; f(b-) \;-\; f(a+).$$

(b) $f = g + h$ where $g \in \mathrm{AC}[a,b]$, $h' = 0$ a.e., and both g and h are monotone increasing.

(c) If I is an interval contained in $[f(a), f(b)]$, then $f^{-1}(I)$ is either an interval, a single point, or empty. Further, $|g(f^{-1}(I))| \leq |I|$.

(d) If A is a measurable subset of $[a,b]$, then $|g(A)| \le |f(A)|_e$.

(e) If $E = \{x \in [a,b] : f \text{ is differentiable at } x\}$, then $\int_a^b f' = |f(E)|_e$.

(f) $\int_A g' = |g(A)| = |g(A \cap E)|$ for all measurable $A \subseteq [a,b]$.

(g) $\int_A f' = |f(A \cap E)|_e \le |f(A)|_e$ for all measurable $A \subseteq [a,b]$.

6.5 The Chain Rule and Changes of Variable

For functions that are differentiable at a point, we have the following fundamental result (for a proof, see [Rud76, Thm. 5.5] or [BS11, Thm. 6.1.6]).

Theorem 6.5.1 (Chain Rule). *Let $g\colon [a,b] \to [c,d]$ and $F\colon [c,d] \to \mathbb{C}$ be given. If g is differentiable at $t_0 \in [a,b]$, and F is differentiable at $g(t_0)$, then $F \circ g$ is differentiable at t_0 and*

$$(F \circ g)'(t_0) \;=\; F'(g(t_0))\, g'(t_0). \qquad \diamondsuit$$

As a corollary, if g and F are both differentiable everywhere on their domains, then $F \circ g$ is differentiable everywhere on $[a,b]$. The situation is more complicated if there are points where g or F are not differentiable. Let Z_g be the set of points in $[a,b]$ where g is not differentiable and let Z_F be the set of points in $[c,d]$ where F is not differentiable. Then $F \circ g$ will be differentiable for all t that do not belong to

$$Z_g \cup g^{-1}(Z_F) \;=\; \bigl\{t \in [a,b] : g'(t) \text{ does not exist or } F'(g(t)) \text{ does not exist}\bigr\}.$$

Unfortunately, even if Z_g and Z_F both have measure zero, it need not be the case that $g^{-1}(Z_F)$ has measure zero, even if g is absolutely continuous. Therefore, in general we have the unpleasant fact that

$$F \text{ and } g \text{ both differentiable a.e.} \quad \not\Longrightarrow \quad F \circ g \text{ is differentiable a.e.}$$

This makes the Chain Rule for functions that are only differentiable almost everywhere a more subtle matter than it is for functions that are differentiable everywhere. The following theorem, from [SV69], whose proof makes clever use of Corollary 6.2.3, gives us a fairly general version of the Chain Rule as long as we *assume* in the hypotheses that $F \circ g$ is differentiable a.e. After the theorem, we will derive several corollaries that do not require us to assume differentiability of $F \circ g$.

Theorem 6.5.2 (Chain Rule). *Assume that:*

(a) $g\colon [a,b] \to [c,d]$ *is differentiable a.e. on* $[a,b]$,

(b) $F\colon [c,d] \to \mathbb{C}$ *is differentiable a.e. on* $[c,d]$,

(c) $F \circ g\colon [a,b] \to \mathbb{C}$ *is differentiable a.e. on* $[a,b]$, *and*

(d) *if* $Z \subseteq [c,d]$ *satisfies* $|Z| = 0$, *then* $|F(Z)| = 0$.

Let $h\colon [c,d] \to \mathbb{C}$ *be any function such that* $h = F'$ *a.e. Then*

$$(F \circ g)' \;=\; (h \circ g)\, g' \;\text{ a.e.} \tag{6.14}$$

Proof. Since we can deal with the complex case by splitting F into real and imaginary parts, it suffices to assume that F is real-valued.

Let Z_g be the set of points in $[a,b]$ where g is not differentiable. Let Z_F be the set of all points $x \in [c,d]$ where either $F'(x)$ does not exist or $h(x) \neq F'(x)$. By hypothesis, $|Z_g| = 0$ and $|Z_F| = 0$. Define

$$B \;=\; g^{-1}(Z_F) \qquad \text{and} \qquad A \;=\; Z_g \cup B.$$

If $t \notin A$, then g is differentiable at t, F is differentiable at $g(t)$, and $h(g(t)) = F'(g(t))$. Applying the pointwise Chain Rule (Theorem 6.5.1), it follows that $F \circ g$ is differentiable at t and

$$(F \circ g)'(t) \;=\; F'(g(t))\, g'(t) \;=\; h(g(t))\, g'(t). \tag{6.15}$$

Now, g is differentiable a.e., so in particular it is differentiable at almost every point of B. Further,

$$g(B) \;=\; g(g^{-1}(Z_F)) \;\subseteq\; Z_F,$$

so $|g(B)| = 0$. Corollary 6.2.3 therefore implies that $g' = 0$ a.e. on B. Since Z_g has measure zero, it follows that $g' = 0$ a.e. on $A = Z_g \cup B$.

Since $|g(B)| = 0$ and F maps sets with measure zero to sets with measure zero, we have $|F(g(B))| = 0$. By hypothesis, $F \circ g$ is differentiable a.e., so if we apply Corollary 6.2.3 to $F \circ g$ then we see that $(F \circ g)' = 0$ a.e. on B, and therefore $(F \circ g)' = 0$ a.e. on $A = Z_g \cup B$. Consequently, for a.e. $t \in A$ we have

$$(F \circ g)'(t) \;=\; 0 \;=\; h(g(t))\, g'(t). \tag{6.16}$$

Finally, since equation (6.15) holds for all $t \notin A$ and equation (6.16) holds for a.e. $t \in A$, we obtain equation (6.14). $\square$

Remark 6.5.3. If $F\colon [c,d] \to \mathbb{C}$ is absolutely continuous, then hypotheses (b) and (d) of Theorem 6.5.2 are automatically satisfied. $\diamondsuit$

Looking at the proof of Theorem 6.5.2, we can see that a considerable simplification is possible if it so happens that the set $A = Z_g \cup g^{-1}(Z_F)$ has measure zero. Our first corollary makes this precise.

Corollary 6.5.4. *If* $g\colon [a,b] \to [c,d]$ *is differentiable a.e.,* $F\colon [c,d] \to \mathbb{C}$ *is differentiable a.e., and* $g'(t) \neq 0$ *for a.e.* t, *then* $F \circ g$ *is differentiable a.e. and equation* (6.15) *holds for any function* h *that satisfies* $h = F'$ *a.e.*

Proof. Repeating the proof of Theorem 6.5.2, we see that equation (6.15) holds for all t that do not belong to the set A, and $g' = 0$ a.e. on A. Since we are now assuming that $g'(t) \neq 0$ for a.e. t, it follows that $|A| = 0$. Therefore equation (6.15) holds for almost every t. □

Our second corollary gives two sufficient conditions under which the hypotheses of Theorem 6.5.2 will be satisfied.

Corollary 6.5.5. *Let $g\colon [a,b] \to [c,d]$ and $F\colon [c,d] \to \mathbb{C}$ be given. If either:*

(a) *F is absolutely continuous and g is monotone increasing, or*

(b) *F is Lipschitz and g has bounded variation,*

then $F \circ g$ is differentiable a.e. and equation (6.15) *holds for any function h that satisfies $h = F'$ a.e.*

Proof. Using either of the hypotheses in statements (a) or (b), it follows from Problem 5.2.20 that $F \circ g$ has bounded variation and consequently is differentiable a.e. Since either statement (a) or (b) implies that F is absolutely continuous, all of the hypotheses of Theorem 6.5.2 are satisfied and the result follows. □

By integrating the Chain Rule, we obtain the following general change of variables formula.

Theorem 6.5.6 (Change of Variable). *Assume that:*

(a) *$g\colon [a,b] \to [c,d]$ is differentiable a.e. on $[a,b]$,*

(b) *$f \in L^1[c,d]$, and*

(c) *$F \circ g \in \mathrm{AC}[a,b]$, where $F(x) = \int_c^x f$ for $x \in [c,d]$.*

Then $(f \circ g)\, g' \in L^1[a,b]$, and

$$\int_{g(u)}^{g(v)} f(x)\, dx \;=\; \int_u^v f(g(t))\, g'(t)\, dt, \qquad \text{for all } \; a \le u \le v \le b. \tag{6.17}$$

Proof. The function F is absolutely continuous and $F' = f$ a.e., so Theorem 6.5.2 implies that $(F \circ g)' = (f \circ g)\, g'$ a.e. Since F and $F \circ g$ are both absolutely continuous, it follows that

$$\begin{aligned}
\int_{g(u)}^{g(v)} f(x)\, dx \;=\; \int_{g(u)}^{g(v)} F'(x)\, dx \;&=\; F(g(v)) - F(g(u)) \\
&=\; \int_u^v (F \circ g)'(t)\, dt \\
&=\; \int_u^v f(g(t))\, g'(t)\, dt. \qquad \square
\end{aligned}$$

The next example shows that it is possible for the hypotheses of Theorem 6.5.6 to be satisfied even when g is not absolutely continuous.

Example 6.5.7. Consider the functions $f(x) = x$ and $g(t) = t\sin\frac{1}{t}$, both on the domain $[-1,1]$. We have $F(x) = \int_{-1}^{x} f = \frac{1}{2}(x^2-1)$. Although g is not absolutely continuous, the composition $(F\circ g)(t) = \frac{1}{2}\big(t^2\sin^2\frac{1}{t} - 1\big)$ is absolutely continuous (see Problem 6.3.7). Since g is differentiable a.e. and f is integrable, the hypotheses of Theorem 6.5.6 are satisfied, and the change of variable formula holds. Consequently, if $[u,v]\subseteq[-1,1]$, then

$$\begin{aligned}
\frac{1}{2}\big(v^2\sin^2\tfrac{1}{v} - u^2\sin^2\tfrac{1}{u}\big) \;=\; \int_{u\sin\frac{1}{u}}^{v\sin\frac{1}{v}} x\,dx \;&=\; \int_{g(u)}^{g(v)} f(x)\,dx\\
&= \int_u^v (f\circ g)(t)\,g'(t)\,dt\\
&= \int_u^v t\sin\tfrac{1}{t}\,\big(\sin\tfrac{1}{t} - \tfrac{1}{t}\cos\tfrac{1}{t}\big)\,dt.\\
&= \int_u^v \big(t\sin^2\tfrac{1}{t} - \sin\tfrac{1}{t}\cos\tfrac{1}{t}\big)\,dt. \quad \diamondsuit
\end{aligned}$$

Unfortunately, in order to invoke Theorem 6.5.6 we must know that $F\circ g$ is absolutely continuous. The following corollary gives some sufficient conditions which ensure that the hypotheses of Theorem 6.5.6 are satisfied.

Corollary 6.5.8. *Let $g\colon[a,b]\to[c,d]$ and $f\colon[c,d]\to\mathbb{C}$ be given. If either:*

(a) *f is integrable and $g\in \mathrm{AC}[a,b]$ is monotone increasing, or*

(b) *$f\in L^\infty[c,d]$ and $g\in\mathrm{AC}[a,b]$,*

then $(f\circ g)\,g'\in L^1[a,b]$ and equation (6.17) *holds.*

Proof. Let $F(x) = \int_c^x f$ for $x\in[c,d]$.

(a) If f is integrable, then F is absolutely continuous. Since g is absolutely continuous and monotone increasing, Problem 6.3.6 implies that $F\circ g$ is absolutely continuous. The hypotheses of Theorem 6.5.6 are therefore satisfied, and the result follows.

(b) If f is essentially bounded, then F is Lipschitz (see Problem 6.4.10). Since g is absolutely continuous, Problem 6.3.6 implies that $F\circ g$ is absolutely continuous. The hypotheses of Theorem 6.5.6 are therefore satisfied, and again the result follows. $\square$

Problems

6.5.9. Suppose that f is a strictly increasing map of $[a,b]$ onto $[c,d]$, and let $g\colon[c,d]\to[a,b]$ be its inverse function. Prove the following statements.

(a) f and g are continuous, and g is strictly increasing.

(b) If $f \in \mathrm{AC}[a,b]$, then $f'(g(t))\, g'(t) = 1$ for a.e. $t \in [c,d]$, and

$$\int_c^d g(t)\, dt \;=\; \int_a^b x f'(x)\, dx.$$

(c) If $g = f^{-1} \in \mathrm{AC}[a,b]$, then $g'(f(x))\, f'(x) = 1$ for a.e. $x \in [a,b]$, and

$$\int_a^b f(x)\, dx \;=\; \int_c^d t g'(t)\, dt.$$

6.5.10. Prove that if $f \in L^1[1,\infty)$ satisfies $\int_1^\infty f(x)\, x^{-2k}\, dx = 0$ for all $k \in \mathbb{N}$, then $f = 0$ a.e.

6.5.11. Exhibit a continuous function $g\colon [a,b] \to [c,d]$ and measurable functions $f_n, f\colon [c,d] \to \mathbb{R}$ such that $f_n \to f$ pointwise a.e., but $f_n \circ g$ does not converge to $f \circ g$ pointwise a.e.

6.5.12. Assume that $g\colon [a,b] \to [c,d]$ is absolutely continuous, $f \in L^1[c,d]$, and $(f \circ g)\, g' \in L^1[a,b]$. Prove that the change of variable formula given in equation (6.17) holds.

6.5.13. This problem will sketch an alternative direct proof of part (a) of Corollary 6.5.8. Assume that $g\colon [a,b] \to [c,d]$ is absolutely continuous and monotone increasing, and let $\mathcal{F}$ be the set of all functions $f \in L^1[c,d]$ such that $f(g(t))\, g'(t)$ is measurable and

$$\int_{g(a)}^{g(b)} f(x)\, dx \;=\; \int_a^b f(g(t))\, g'(t)\, dt. \tag{6.18}$$

Prove the following statements.

(a) If $[u,v] \subseteq [c,d]$, then $\chi_{[u,v]} \in \mathcal{F}$.

(b) If $f = 0$ a.e. on $[c,d]$, then $f \in \mathcal{F}$.

(c) If $E \subseteq [c,d]$ is measurable, then $\chi_E \in \mathcal{F}$.

(d) $\mathcal{F} = L^1[c,d]$.

6.6 Convex Functions and Jensen's Inequality

In this section we will derive an important inequality for convex functions known as *Jensen's Inequality.* Although Jensen's Inequality can be quite useful, the material of this section will only rarely be referred to in the remainder of this volume.

The following definition introduces convex functions. The reason for the terminology "convex" is best understood by considering the graph of a convex function, one of which is shown in Figure 6.1.

Definition 6.6.1 (Convex Function). Let $-\infty \le a < b \le \infty$ be given. We say that a function $\phi\colon (a,b) \to \mathbb{R}$ is *convex* on the open interval (a,b) if for all $x,\, y \in (a,b)$ and all $0 < t < 1$ we have

$$\phi\bigl(tx + (1-t)y\bigr) \;\le\; t\,\phi(x) + (1-t)\,\phi(y).$$

In other words, on any subinterval $[x,y]$ of (a,b), the graph of ϕ lies on or below the line segment that joins the points $(x,\phi(x))$ and $(y,\phi(y))$. An analogous definition is made for *concave functions.* $\diamondsuit$

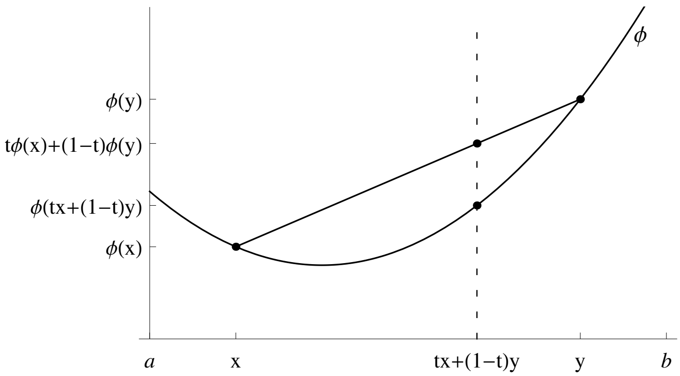


Fig. 6.1 Graph of a convex function.

We allow (a,b) to be an infinite open interval in Definition 6.6.1. Throughout this section we will implicitly assume that $-\infty \le a < b \le \infty$.

By repeatedly applying the definition of convexity, we obtain the discrete version of Jensen's Inequality.

Exercise 6.6.2 (Discrete Jensen Inequality). Assume that $\phi\colon (a,b) \to \mathbb{R}$ is a convex function. If $N \ge 2$, then for any points $x_1, \dots, x_N \in (a,b)$ and positive weights $t_1, \dots, t_N$ that satisfy $t_1 + \cdots + t_N = 1$, we have

$$\phi\biggl(\sum_{j=1}^N t_j x_j\biggr) \;\le\; \sum_{j=1}^N t_j \phi(x_j). \qquad \diamondsuit \tag{6.19}$$

We can also write the Discrete Jensen Inequality in an "unnormalized" form. Suppose ϕ is convex, $x_1, \dots, x_N$ are points in (a,b), and $t_1, \dots, t_N > 0$. Set $t = t_1 + \cdots + t_N$. Then equation (6.19) implies that

$$\phi\left(\frac{\sum t_j x_j}{\sum t_j}\right) \;=\; \phi\Bigl(\sum t^{-1} t_j x_j\Bigr) \;\le\; \sum t^{-1} t_j \phi(x_j) \;=\; \frac{\sum t_j \phi(x_j)}{\sum t_j}.$$

We will derive several properties of convex functions below. The following lemma will play an important role.

Lemma 6.6.3. *If ϕ is convex on (a,b) and $x \in (a,b)$ is fixed, then*

$$\beta(y) \;=\; \frac{\phi(y)-\phi(x)}{y-x}, \qquad y \in (a,b),\ y \neq x, \tag{6.20}$$

is monotone increasing on $(a,x)\cup(x,b)$.

Proof. Suppose that $x < y < z < b$, and write $y = tx + (1-t)z$ where $0 < t < 1$. Let g be the linear function whose graph passes through the points $(x,\phi(x))$ and $(z,\phi(z))$. This function satisfies $f(x) = \phi(x)$ and

$$\frac{g(u)-g(x)}{u-x} \;=\; \frac{\phi(z)-\phi(x)}{z-x} \qquad \text{for all } u \neq x.$$

Since $\phi(x) = g(x)$, by taking $u = y$ we see that

$$g(y) \;=\; (y-x)\,\frac{\phi(z)-\phi(x)}{z-x} \;+\; \phi(x).$$

Also, $\phi(y) \le g(y)$ by the definition of convexity, so

$$(y-x)\,\frac{\phi(y)-\phi(x)}{y-x} + \phi(x) \;=\; \phi(y) \;\le\; g(y) \;=\; (y-x)\,\frac{\phi(z)-\phi(x)}{z-x} + \phi(x).$$

Since $y - x > 0$, it follows that

$$\beta(y) \;=\; \frac{\phi(y)-\phi(x)}{y-x} \;\le\; \frac{\phi(z)-\phi(x)}{z-x} \;=\; \beta(z).$$

Thus β is increasing on (x,b). A similar argument applies on the interval (a,x), and another similar argument establishes that $\beta(z) \le \beta(y)$ when $z < x < y$. Hence β is monotone increasing on $(a,x)\cup(x,b)$. ☐

Next we derive an equivalent characterization of convexity.

Lemma 6.6.4. *A function $\phi\colon (a,b) \to \mathbb{R}$ is convex if and only if for all $a < x < y < z < b$ we have*

$$\frac{\phi(y)-\phi(x)}{y-x} \;\le\; \frac{\phi(z)-\phi(x)}{z-x}. \tag{6.21}$$

Proof. $\Rightarrow$. Assume that ϕ is convex, fix $x \in (a,b)$, and let $\beta(x)$ be defined by equation (6.20). Then equation (6.21) follows immediately from the fact that β is monotone increasing to the right of x.

$\Leftarrow$. Assume that equation (6.21) holds whenever $a < x < y < z < b$. Suppose that $a < x < z < b$ and $0 < t < 1$. Then $y = tx + (1-t)z$ satisfies $x < y < z$. Since

$$y - x \;=\; (t-1)x \,+\, (1-t)z \;=\; (1-t)\,(z-x),$$

equation (6.21) therefore implies that

$$\begin{aligned}\phi(tx+(1-t)z) \;=\; \phi(y) \;&\le\; (y-x)\,\frac{\phi(z)-\phi(x)}{z-x} \,+\, \phi(x)\\ &=\; (1-t)\,\bigl(\phi(z)-\phi(x)\bigr) \,+\, \phi(x)\\ &=\; (1-t)\,\phi(z) \,+\, t\,\phi(x). \qquad \square\end{aligned}$$

This provides us with a convenient sufficient condition for convexity.

Theorem 6.6.5. *If $\phi\colon (a,b) \to \mathbb{R}$ is differentiable at every point of (a,b) and ϕ' is monotone increasing on (a,b), then ϕ is convex.*

Proof. The reader should check that if $b_1,\, b_2 > 0$ and $a_1,\, a_2 \in \mathbb{R}$, then

$$\min\Bigl\{\frac{a_1}{b_1},\, \frac{a_2}{b_2}\Bigr\} \;\le\; \frac{a_1+a_2}{b_1+b_2} \;\le\; \max\Bigl\{\frac{a_1}{b_1},\, \frac{a_2}{b_2}\Bigr\}. \tag{6.22}$$

Fix $a < x < y < z < b$. Then ϕ is continuous on $[x,y]$ and differentiable on (x,y), so the Mean Value Theorem implies that there exists a point $\xi_1 \in (x,y)$ such that

$$\frac{\phi(y)-\phi(x)}{y-x} \;=\; \phi'(\xi_1).$$

Similarly, there exists a point $\xi_2 \in (y,z)$ such that

$$\frac{\phi(z)-\phi(y)}{z-y} \;=\; \phi'(\xi_2).$$

Since ϕ' is increasing, by applying equation (6.22) we see that

$$\begin{aligned}\frac{\phi(y)-\phi(x)}{y-x} \;=\; \phi'(\xi_1) \;&=\; \min\bigl\{\phi'(\xi_1), \phi'(\xi_2)\bigr\}\\ &=\; \min\biggl\{\frac{\phi(y)-\phi(x)}{y-x},\, \frac{\phi(z)-\phi(y)}{z-y}\biggr\}\\ &\le\; \frac{\bigl(\phi(y)-\phi(x)\bigr)+\bigl(\phi(z)-\phi(y)\bigr)}{(y-x)+(z-y)}\\ &=\; \frac{\phi(z)-\phi(x)}{z-x}.\end{aligned}$$

Lemma 6.6.4 therefore implies that ϕ is convex. $\square$

Corollary 6.6.6. (a) *If $1 \le p < \infty$, then x^p is convex on $(0,\infty)$.*
(b) *If $a \in \mathbb{R}$, then e^{ax} is convex on $(-\infty,\infty)$.*
(c) *$-\ln x$ is convex on $(0,\infty)$.* $\diamondsuit$

A convex function need not be differentiable at every point of (a,b), but we prove next that it will be differentiable at all but countably many points, right-differentiable at every point, and left-differentiable at every point. Here the right and left derivatives are defined, respectively, by

$$\phi'_+(x) \;=\; \lim_{y\to x^+} \frac{\phi(y)-\phi(x)}{y-x} \qquad \text{and} \qquad \phi'_-(x) \;=\; \lim_{y\to x^-} \frac{\phi(y)-\phi(x)}{y-x}.$$

Theorem 6.6.7. *If ϕ is a convex function on (a,b), then the following statements hold.*

(a) *$\phi'_+(x)$ and $\phi'_-(x)$ both exist (and are finite) at each point $x \in (a,b)$.*

(b) *ϕ is continuous on (a,b).*

(c) *If $a < x < y < b$, then*

$$\phi'_+(x) \;\le\; \frac{\phi(y)-\phi(x)}{y-x} \;\le\; \phi'_-(y). \tag{6.23}$$

(d) *If $a < x < b$, then $\phi'_-(x) \le \phi'_+(x)$.*

(e) *ϕ'_+ and ϕ'_- are monotone increasing on (a,b).*

(f) *ϕ is differentiable at all but at most countably many points in (a,b).*

Proof. (a) Fix $x \in (a,b)$. By Lemma 6.6.3, the function β defined by equation (6.20) is increasing on $(a,x) \cup (x,b)$. Consequently β is bounded above on (a,x), since if we fix any $z \in (x,b)$ then $\beta(y) \le \beta(z)$ for $y \in (a,x)$. Since β is monotone increasing and bounded on (a,x), it therefore has a finite limit as y approaches x from the left. That is,

$$\phi'_-(x) \;=\; \lim_{y\to x^-} \frac{\phi(y)-\phi(x)}{y-x} \;=\; \lim_{y\to x^-} \beta(y)$$

exists. A similar argument shows that $\phi'_+(x)$ exists.

(b) Since ϕ is both left and right differentiable at each point, it is both left and right continuous at each point.

(c) Since β is increasing on (x,b), if we fix $x < y < b$ then

$$\phi'_+(x) \;=\; \lim_{t\to x^+} \beta(t) \;\le\; \beta(y) \;=\; \frac{\phi(y)-\phi(x)}{y-x}.$$

A symmetric argument yields the other inequality.

(d) Since β is increasing on $(a,x) \cup (x,b)$, the values β takes to the left of x are less than or equal to the values that β takes to the right of x. Therefore

$$\phi'_-(x) \;=\; \lim_{t\to x^-} \beta(t) \;\le\; \lim_{t\to x^+} \beta(t) \;=\; \phi'_+(x).$$

(e) Combining parts (c) and (d), if $x < y < b$ then $\phi'_+(x) \le \phi'_-(y) \le \phi'_+(y)$. Therefore ϕ'_+ is monotone increasing, and a similar argument applies to ϕ'_-.

(f) Since ϕ'_+ is monotone increasing on (a,b), it can have at most countably many discontinuities. If y is not one of those points, then y is a point of continuity for ϕ'_+ and therefore, by part (c),

$$\phi'_+(y) \;\ge\; \phi'_-(y) \;=\; \lim_{x\to y^-} \frac{\phi(y)-\phi(x)}{y-x} \;\ge\; \lim_{x\to y^-} \phi'_+(x) \;=\; \phi'_+(y).$$

Hence $\phi'_+(y) = \phi'_-(y)$, so ϕ is differentiable at y. □

In order to prove Jensen's Inequality, we will need the following notion.

Definition 6.6.8 (Supporting Line). Let ϕ be a convex function on (a,b). A *supporting line* for ϕ at $x \in (a,b)$ is any line that passes through the point $(x,\phi(x))$ and lies on or below the graph of ϕ. ♢

Here is a way to recognize supporting lines.

Lemma 6.6.9. *Suppose that ϕ is convex on (a,b). Then any line that passes through $(x,\phi(x))$ and has a slope m that lies in the range $\phi'_-(x) \le m \le \phi'_+(x)$ is a supporting line for ϕ at x.*

Proof. Assume that L is such a line. If $x < y < b$, then

$$\begin{aligned} L(y) \;&=\; (y-x)\,m \;+\; \phi(x) \\ &\le\; (y-x)\,\phi'_+(x) \;+\; \phi(x) \\ &\le\; (y-x)\,\frac{\phi(y)-\phi(x)}{y-x} \;+\; \phi(x) \qquad \text{(by equation (6.23))} \\ &=\; \phi(y). \end{aligned}$$

Combining this with a similar argument for points y that lie to the left of x, we conclude that the graph of L lies on or below the graph of ϕ. □

Finally, we prove Jensen's Inequality.

Theorem 6.6.10 (Jensen's Inequality). *Let E be a measurable subset of $\mathbb{R}^d$ such that $0 < |E| < \infty$. If $g\colon E \to (a,b)$ is integrable and ϕ is convex on (a,b), then*

$$\phi\biggl(\frac{1}{|E|}\int_E g\biggr) \;\le\; \frac{1}{|E|}\int_E \phi\circ g. \tag{6.24}$$

Proof. Since g is integrable, $t = \frac{1}{|E|}\int_E g$ is a finite real number. Also, since $g(x) < b$ for every x,

$$t \;=\; \frac{1}{|E|}\int_E g \;\le\; \frac{1}{|E|}\int_E b \;=\; b. \tag{6.25}$$

Suppose for the moment that b is finite. If $t = b$, then equation (6.25) implies that $\int_E (b - g) = 0$. But $b - g \geq 0$, so this implies that $g = b$ a.e. This contradicts our assumption that $g(x) < b$ for every x. Consequently we must have $t < b$ if b is finite. On the other hand, if $b = \infty$ then we certainly have $t < b$ in that case as well. A similar argument shows that $a < t$, so we conclude that the number t belongs to the open interval (a, b).

Let L be any supporting line for ϕ at the point t, and let m be its slope. By definition $L(t) = \phi(t)$, so the equation for L is

$$L(y) \;=\; m\,(y - t) \,+\, \phi(t), \qquad \text{for } y \in \mathbb{R}.$$

Since L lies on or below the graph of ϕ,

$$L(y) \;=\; m\,(y - t) \,+\, \phi(t) \;\leq\; \phi(y), \qquad \text{for } y \in (a, b).$$

Choose any point $x \in E$. Then $g(x) \in (a, b)$, so by applying the preceding inequality to the point $y = g(x)$ we see that

$$L(g(x)) \;=\; m\big(g(x) - t\big) \,+\, \phi(t) \;\leq\; \phi(g(x)). \tag{6.26}$$

If we are allowed to integrate this equation over x then we obtain

$$\begin{aligned}
\int_E \phi(g(x))\,dx \;&\geq\; \int_E m\big(g(x) - t\big)\,dx \,+\, \int_E \phi(t)\,dx \\
&=\; m\int_E g \;-\; mt\,|E| \,+\, \phi(t)\,|E| \\
&=\; mt\,|E| \;-\; mt\,|E| \,+\, \phi(t)\,|E| \\
&=\; \phi\bigg(\frac{1}{|E|}\int_E g\bigg)\,|E|,
\end{aligned} \tag{6.27}$$

and by rearranging this we arrive at equation (6.24).

However, there is a technical issue. Although $\phi \circ g$ is measurable, we do not know that $\phi \circ g$ is nonnegative or that it is integrable. Therefore, it is possible that $\int_E (\phi \circ g)$ might not exist, in which case the calculations above do not make sense. To see that this integral does exist, we use the inequality in equation (6.26) and the integrability of g to compute that

$$\begin{aligned}
\int_E (\phi \circ g)^- \;&\leq\; \int_E \big|m\big(g(x) - t\big) + \phi(t)\big|\,dx \\
&\leq\; |m|\bigg(\int_E |g|\bigg) \,+\, |mt|\,|E| \,+\, |\phi(t)|\,|E| \;<\; \infty.
\end{aligned}$$

Hence $\int_E (\phi \circ g)^-$ and $\int_E (\phi \circ g)^+$ cannot both be infinite, so $\int_E (\phi \circ g)$ exists in the extended real sense. Our calculations in equation (6.27) are therefore valid even if it should be the case that $\int_E (\phi \circ g) = \infty$. $\square$

Problems

6.6.11. Prove the following statements.

(a) If ϕ and ψ are convex on (a,b), then $\phi+\psi$ is convex on (a,b).

(b) If ϕ is convex on (a,b) and $c>0$, then $c\phi$ is convex on (a,b).

(c) If $\{\phi_n\}_{n\in\mathbb{N}}$ is a sequence of convex functions on (a,b) and $\phi_n \to \phi$ pointwise, then ϕ is convex on (a,b).

6.6.12. Let $a,\, b \ge 0$ and $1<p<\infty$ be given, and let p' be the *dual index* to p, i.e., p' is the unique real number that satisfies $\frac{1}{p}+\frac{1}{p'}=1$. Write $a=e^{x/p}$ and $b=e^{y/p'}$, and use the Discrete Jensen Inequality and the fact that e^x is convex to prove that

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

6.6.13. Given numbers $0<a_n\le 1$, prove that $\displaystyle\sum_{n=1}^{\infty}\frac{\ln a_n}{2^n} \le \ln\left(\sum_{n=1}^{\infty}\frac{a_n}{2^n}\right)$.

6.6.14. Let E be a measurable subset of $\mathbb{R}^d$ such that $0<|E|<\infty$, and suppose that $f\colon E\to\mathbb{R}$ is measurable. Prove that

$$\exp\!\left(\frac{1}{|E|}\int_E f\right) \;\le\; \frac{1}{|E|}\int_E e^{f(x)}\,dx, \quad \text{where } \exp(t)=e^t,$$

and

$$\frac{1}{|E|}\int_E \ln|f| \;\le\; \ln\!\left(\frac{1}{|E|}\int_E |f|\right).$$

6.6.15. Prove that a function $\phi\colon (a,b)\to\mathbb{R}$ is convex if and only if ϕ is continuous and

$$\phi\!\left(\frac{x+y}{2}\right) \;\le\; \frac{\phi(x)+\phi(y)}{2}, \qquad \text{for all } x,y\in(a,b).$$

6.6.16. Assume that f is monotone increasing and integrable on (a,b). Prove that the indefinite integral $\phi(x)=\int_a^x f(t)\,dt$ is convex on (a,b).

6.6.17. Suppose that ϕ is convex on (a,b). Prove that ϕ is Lipschitz on each closed interval $[c,d]\subseteq(a,b)$.

Chapter 7
The L^p Spaces

The *Lebesgue spaces* provide us with a way to quantify integrability properties of functions. We have already seen two particular examples. The space $L^\infty(E)$, which consists of all essentially bounded functions on the domain E, was introduced in Section 3.3, and $L^1(E)$, which consists of the Lebesgue integrable functions on E, was defined in Section 4.4. Now we will consider an entire family of spaces $L^p(E)$ with $0 < p \le \infty$.

To illustrate the properties of $L^p(E)$, we first introduce a discrete version, the *ℓ^p-spaces*, in Section 7.1. We derive two fundamental results, *Hölder's Inequality* and *Minkowski's Inequality*, which establish that ℓ^p is a normed space when $p \ge 1$, and we prove that ℓ^p is *complete* with respect to that norm and therefore is a *Banach space* (at least for $p \ge 1$; for $p < 1$ it turns out that ℓ^p is a complete *metric space*, but is not a normed space).

We introduce the Lebesgue spaces $L^p(E)$ in Section 7.2. Some properties of the Lebesgue spaces parallel those of the ℓ^p spaces, but we find a technical difficulty in that a function that has zero L^p-norm need only be zero at *almost every* point. However, once we identify functions that are equal almost everywhere, we can prove that $L^p(E)$ is a Banach space for each index p in the range $1 \le p \le \infty$. We study convergence in L^p-norm in Section 7.3, and show in Section 7.4 that $L^p(E)$ is *separable* when p is finite, but not when $p = \infty$.

Norms and seminorms have appeared at various times in earlier chapters. In particular, we saw in Section 3.3 that

$$\|f\|_\infty \;=\; \operatorname*{ess\,sup}_{x \in E} |f(x)|$$

is a seminorm on $L^\infty(E)$, and we similarly observed in Section 4.4 that

$$\|f\|_1 \;=\; \int_E |f(x)|\,dx$$

C. Heil, *Introduction to Real Analysis*, Graduate Texts in Mathematics 280,
https://doi.org/10.1007/978-3-030-26903-6_7

is a seminorm on $L^1(E)$. We will make frequent use of norms and seminorms (and, to a lesser extent, metrics) in this chapter. Many of the important notions will be discussed as they are presented here, but the reader may wish to review Chapter 1 before proceeding further.

7.1 The ℓ^p Spaces

The ℓ^p *spaces* are vector spaces whose elements are infinite sequences of scalars that are either p-summable or bounded in the sense that we will make precise in the next definition. For simplicity of presentation, we will take the complex plane $\mathbb{C}$ to be our field of scalars throughout this section, but the reader can check that entirely analogous results hold if we restrict to just real scalars.

Definition 7.1.1 (p-Summable and Bounded Sequences).

(a) Let $0 < p < \infty$ be a finite real number. A sequence of scalars $x = (x_k)_{k\in\mathbb{N}}$ is *p-summable* if $\sum_{k=1}^{\infty} |x_k|^p < \infty$. In this case we set

$$\|x\|_p = \|(x_k)_{k\in\mathbb{N}}\|_p = \left(\sum_{k=1}^{\infty} |x_k|^p\right)^{1/p}.$$

If the sequence x is not p-summable, then we take $\|x\|_p = \infty$.

(b) A sequence of scalars $x = (x_k)_{k\in\mathbb{N}}$ is *bounded* if $\sup_{k\in\mathbb{N}} |x_k| < \infty$. In this case we set

$$\|x\|_\infty = \sup_{k\in\mathbb{N}} |x_k|.$$

If the sequence x is not bounded, then $\|x\|_\infty = \infty$. $\diamondsuit$

If $p = 1$ then we usually just write *summable* (or sometimes *absolutely summable*) instead of 1-*summable,* and for $p = 2$ we write *square summable* instead of 2-*summable.* Problem 7.1.22 shows that $\|\cdot\|_\infty$ is the limit of $\|\cdot\|_p$ in the sense that if x is p-summable for some finite p, then $\|x\|_p \to \|x\|_\infty$ as $p \to \infty$.

We collect the p-summable or bounded sequences to form the ℓ^p spaces, as follows.

Definition 7.1.2 (The ℓ^p Spaces).

(a) If $0 < p < \infty$, then the space ℓ^p consists of all p-summable sequences of scalars. That is, a sequence $x = (x_k)_{k\in\mathbb{N}}$ belongs to ℓ^p if and only if

$$\|x\|_p = \|(x_k)_{k\in\mathbb{N}}\|_p = \left(\sum_{k=1}^{\infty} |x_k|^p\right)^{1/p} < \infty.$$

(b) For $p = \infty$, the space ℓ^∞ consists of all *bounded* sequences of scalars. That is, a sequence $x = (x_k)_{k\in\mathbb{N}}$ belongs to ℓ^∞ if and only if

$$\|x\|_\infty \;=\; \|(x_k)_{k\in\mathbb{N}}\|_\infty \;=\; \sup_{k\in\mathbb{N}} |x_k| \;<\; \infty. \qquad \diamondsuit$$

For example, the sequence

$$x \;=\; \bigl(\tfrac{1}{k}\bigr)_{k\in\mathbb{N}} \;=\; \bigl(1, \tfrac{1}{2}, \tfrac{1}{3}, \dots\bigr)$$

belongs to ℓ^p for each index $1 < p \le \infty$, but x does not belong to ℓ^p for any $0 < p \le 1$. On the other hand, the constant sequence

$$y \;=\; (1)_{k\in\mathbb{N}} \;=\; \bigl(1, 1, 1, \dots\bigr)$$

belongs to ℓ^∞, but does not belong to ℓ^p for any finite p. Problem 7.1.21 asks for a proof that the ℓ^p spaces are *nested and distinct* in the following sense:

$$0 \;<\; p \;<\; q \;\le\; \infty \quad\Longrightarrow\quad \ell^p \;\subsetneq\; \ell^q. \tag{7.1}$$

Remark 7.1.3. By making appropriate changes in the preceding definitions, we can consider spaces of sequences that are indexed by sets other than the natural numbers $\mathbb{N}$. For example, if I is a countable index set, then we say that a sequence $x = (x_k)_{k\in I}$ is p-summable if and only if $\sum_{k\in I} |x_k|^p < \infty$. For finite p, we let $\ell^p(I)$ be the space of all p-summable sequences indexed by I, and we define $\ell^\infty(I)$ to be the space of all bounded sequences indexed by I. If $I = \mathbb{N}$, then this reduces to the definition of ℓ^p that we gave before, i.e., $\ell^p = \ell^p(\mathbb{N})$.

A common choice of index set is $I = \mathbb{Z}$. A sequence indexed by $\mathbb{Z}$ is a bi-infinite sequence of the form

$$x \;=\; (x_k)_{k\in\mathbb{Z}} \;=\; (\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots).$$

The space $\ell^p(\mathbb{Z})$ is the set of all bi-infinite sequences that are p-summable (if p is finite) or bounded (if $p = \infty$). For example, $x = \bigl(2^{-|k|}\bigr)_{k\in\mathbb{Z}}$ belongs to $\ell^p(\mathbb{Z})$ for every index $0 < p \le \infty$. Problem 7.1.27 shows how to define $\ell^p(I)$ when I is uncountable.

We can also let the index set be finite. If $I = \{1, \dots, d\}$ then a sequence indexed by I is simply a vector $x = (x_1, \dots, x_d) \in \mathbb{C}^d$. Every such sequence is p-summable and bounded, so for $I = \{1, \dots, d\}$ we have $\ell^p(I) = \mathbb{C}^d$ for every index $0 < p \le \infty$. $\diamondsuit$

We will prove in Theorem 7.1.15 that $\|\cdot\|_p$ is an norm on ℓ^p for all indices $1 \le p \le \infty$. Therefore we refer to $\|\cdot\|_p$ as the *ℓ^p-norm* when $p \ge 1$. For $p = 2$ we usually call $\|\cdot\|_2$ the *Euclidean norm*, and for $p = \infty$ we often refer to $\|\cdot\|_\infty$ as the *sup-norm.*

For $0 < p < 1$ we will see in Section 7.1.5 that $\|\cdot\|_p$ is not a norm. On the other hand, Theorem 7.1.18 will provide a substitute result, namely that $\mathrm{d}(x,y) = \|x-y\|_p^p$ defines a *metric* on ℓ^p when $0 < p < 1$.

Addition of sequences is performed componentwise, i.e., if $x = (x_k)_{k\in\mathbb{N}}$ and $y = (y_k)_{k\in\mathbb{N}}$, then $x + y = (x_k + y_k)_{k\in\mathbb{N}}$. The sum of two bounded sequences is bounded, so ℓ^∞ is closed under addition. The next lemma shows that ℓ^p is closed under addition when p is finite.

Lemma 7.1.4. *Let* $0 < p < \infty$ *be given.*

(a) *If* $a, b \geq 0$, *then* $(a+b)^p \leq 2^p\,(a^p + b^p)$.

(b) *If* $x = (x_k)_{k\in\mathbb{N}}$ *and* $y = (y_k)_{k\in\mathbb{N}}$ *are any two sequences of scalars, then*

$$\|x+y\|_p^p \;\leq\; 2^p\,\Big(\|x\|_p^p + \|y\|_p^p\Big).$$

(c) *If* $x, y \in \ell^p$, *then* $x + y \in \ell^p$.

Proof. If $a, b \geq 0$, then

$$(a+b)^p \;\leq\; \Big(\max\{a,b\} \;+\; \max\{a,b\}\Big)^p \;=\; 2^p \max\{a^p, b^p\} \;\leq\; 2^p\big(a^p + b^p\big).$$

Parts (b) and (c) follow immediately from this. □

Combining Lemma 7.1.4 with the fact that ℓ^p is closed under multiplication by scalars, we see that ℓ^p is a vector space. For this reason, we often refer to an element x of ℓ^p as a *vector in* ℓ^p. The zero vector in ℓ^p is the *zero sequence* $0 = (0,0,0,\dots)$. We use the same symbol 0 to denote both the zero sequence and the number zero, but the meaning should always be clear from context.

7.1.1 Hölder's Inequality

It is clear that $\|\cdot\|_p$ satisfies the nonnegativity, homogeneity, and uniqueness properties of a norm, but it is not obvious whether the Triangle Inequality is satisfied. We will prove that $\|\cdot\|_p$ is a norm on ℓ^p when $p \geq 1$, but first we need to establish a fundamental result known as *Hölder's Inequality.* This gives us a relationship between ℓ^p and $\ell^{p'}$, where p' is the *dual index* to p, the unique extended real number that satisfies

$$\frac{1}{p} + \frac{1}{p'} \;=\; 1. \tag{7.2}$$

In equation (7.2), we follow the standard real analysis convention that

$$\frac{1}{\infty} \;=\; 0.$$

Some examples of dual indices are

$$1' = \infty, \quad \left(\frac{4}{3}\right)' = 4, \quad \left(\frac{3}{2}\right)' = 3, \quad 2' = 2, \quad 3' = \frac{3}{2}, \quad 4' = \frac{4}{3}, \quad \infty' = 1.$$

The dual of p' is p, i.e., $(p')' = p$ for $1 \le p \le \infty$. For $1 < p < \infty$ we can write p' explicitly as

$$p' = \frac{p}{p-1}, \qquad 1 < p < \infty.$$

The key to Hölder's Inequality is the inequality for scalars established in the following exercise.

Exercise 7.1.5. (a) Show that if $0 < \theta < 1$, then $t^\theta \le \theta t + (1-\theta)$ for all $t \ge 0$, and equality holds if and only if $t = 1$.

(b) Suppose that $1 < p < \infty$ and $a, b \ge 0$. Apply part (a) with $t = a^p b^{-p'}$ and $\theta = 1/p$ to show that

$$ab \le \frac{a^p}{p} + \frac{b^{p'}}{p'}, \tag{7.3}$$

and prove that equality holds if and only if $b = a^{p-1}$. $\diamondsuit$

Remark 7.1.6. For $p = 2$, equation (7.3) reduces to $ab \le \frac{a^2}{2} + \frac{b^2}{2}$. Replacing a by $\sqrt{a}$ and b by $\sqrt{b}$ we obtain

$$\sqrt{ab} \le \frac{a+b}{2}, \qquad \text{for } a, b \ge 0,$$

which is the inequality that relates the arithmetic and geometric means of a and b. Hence equation (7.3) is a generalization of the arithmetic-geometric mean inequality to other values of p. $\diamondsuit$

Exercise 7.1.5 gives one proof of equation (7.3), but there are other approaches. For example, a proof based on *Jensen's inequality* appeared earlier in Problem 6.6.12. Alternatively, observe that x^{p-1} is continuous and strictly increasing on the interval $[0, a]$, and its inverse function is $y^{\frac{1}{p-1}}$. Figure 7.1 gives a "proof by picture" that

$$ab \le \int_0^a x^{p-1}\,dx + \int_0^b y^{\frac{1}{p-1}}\,dy. \tag{7.4}$$

Evaluating the right-hand side of equation (7.4), we obtain another proof of equation (7.3).

Now we prove Hölder's Inequality, which bounds the ℓ^1-norm of a "componentwise product sequence" $xy = (x_k y_k)_{k\in\mathbb{N}}$ in terms of the ℓ^p-norm of x and the $\ell^{p'}$-norm of y.

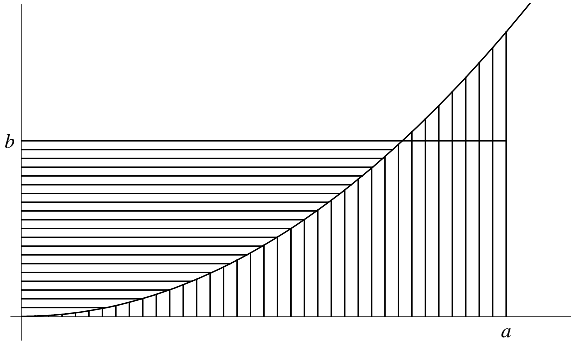


Fig. 7.1 The curved line is the graph of $y = x^{p-1}$. The area of the vertically hatched region is $\int_0^a x^{p-1}\,dx$, the area of the horizontally hatched region is $\int_0^b y^{\frac{1}{p-1}}\,dy$, and the area of the rectangle $[0,a] \times [0,b]$ is ab.

Theorem 7.1.7 (Hölder's Inequality). *Fix $1 \le p \le \infty$ and let p' be the dual index to p. If $x = (x_k)_{k\in\mathbb{N}} \in \ell^p$ and $y = (y_k)_{k\in\mathbb{N}} \in \ell^{p'}$, then the sequence $xy = (x_k y_k)_{k\in\mathbb{N}}$ belongs to ℓ^1, and*

$$\|xy\|_1 \le \|x\|_p \, \|y\|_{p'}. \tag{7.5}$$

If $1 < p < \infty$, then equation (7.5) is

$$\sum_{k=1}^{\infty} |x_k y_k| \;\le\; \biggl(\sum_{k=1}^{\infty} |x_k|^p\biggr)^{1/p} \biggl(\sum_{k=1}^{\infty} |y_k|^{p'}\biggr)^{1/p'}. \tag{7.6}$$

If $p = 1$, then equation (7.5) is

$$\sum_{k=1}^{\infty} |x_k y_k| \;\le\; \biggl(\sum_{k=1}^{\infty} |x_k|\biggr) \Bigl(\sup_{k\in\mathbb{N}} |y_k|\Bigr). \tag{7.7}$$

If $p = \infty$, then equation (7.5) is

$$\sum_{k=1}^{\infty} |x_k y_k| \;\le\; \Bigl(\sup_{k\in\mathbb{N}} |x_k|\Bigr) \biggl(\sum_{k=1}^{\infty} |y_k|\biggr). \tag{7.8}$$

Proof. Case $p = 1$. In this case $p' = \infty$, so y is bounded. Since $|y_k| \le \|y\|_\infty$ for every k, we see that

$$\sum_{k=1}^{\infty} |x_k y_k| \;\le\; \sum_{k=1}^{\infty} |x_k| \, \|y\|_\infty \;=\; \|y\|_\infty \sum_{k=1}^{\infty} |x_k|,$$

which is equation (7.7). The case $p = \infty$ is symmetrical, because $p' = 1$ when $p = \infty$.

Case $1 < p < \infty$. If either x or y is the zero sequence, then equation (7.6) holds trivially, so we may assume that $x \neq 0$ and $y \neq 0$.

Suppose first that $x \in \ell^p$ and $y \in \ell^{p'}$ are unit vectors in their respective spaces, i.e., $\|x\|_p = 1$ and $\|y\|_{p'} = 1$. Then by applying equation (7.3), we see that

$$\begin{aligned} \|xy\|_1 \;=\; \sum_{k=1}^{\infty} |x_k y_k| \;&\leq\; \sum_{k=1}^{\infty} \biggl(\frac{|x_k|^p}{p} + \frac{|y_k|^{p'}}{p'}\biggr) \\ &=\; \frac{\|x\|_p^p}{p} + \frac{\|y\|_{p'}^{p'}}{p'} \;=\; \frac{1}{p} + \frac{1}{p'} \;=\; 1. \end{aligned} \tag{7.9}$$

Now let x be any nonzero sequence in ℓ^p, and let y be any nonzero sequence in $\ell^{p'}$. Define

$$u \;=\; \frac{x}{\|x\|_p} \qquad \text{and} \qquad v \;=\; \frac{y}{\|y\|_{p'}}.$$

Then u is a unit vector in ℓ^p, and v is a unit vector in $\ell^{p'}$, so equation (7.9) implies that $\|uv\|_1 \leq 1$. However,

$$uv \;=\; \frac{xy}{\|x\|_p \, \|y\|_{p'}},$$

so by homogeneity we obtain

$$\frac{\|xy\|_1}{\|x\|_p \, \|y\|_{p'}} \;=\; \|uv\|_1 \;\leq\; 1.$$

Rearranging yields $\|xy\|_1 \leq \|x\|_p \, \|y\|_{p'}$. $\square$

7.1.2 Minkowski's Inequality

Our next goal is to show that $\|\cdot\|_p$ is a norm on ℓ^p when $1 \leq p \leq \infty$. The only difficulty is showing that the Triangle Inequality on ℓ^p (which is often called *Minkowski's Inequality*) is satisfied. For $p = 1$ and $p = \infty$ this is not difficult, so we assign those cases as an exercise.

Exercise 7.1.8 (Minkowski's Inequality). Prove that the following statements hold.

(a) If $x, y \in \ell^1$, then $\|x + y\|_1 \leq \|x\|_1 + \|y\|_1$.

(b) If $x, y \in \ell^\infty$, then $\|x + y\|_\infty \leq \|x\|_\infty + \|y\|_\infty$. $\diamondsuit$

The Triangle Inequality is more challenging to prove when $1 < p < \infty$. We will use Hölder's Inequality to derive Minkowski's Inequality for these cases.

Theorem 7.1.9 (Minkowski's Inequality). *Fix $1 < p < \infty$. If $x, y \in \ell^p$, then*

$$\|x+y\|_p \;\le\; \|x\|_p \;+\; \|y\|_p. \tag{7.10}$$

If $x = (x_k)_{k\in\mathbb{N}}$ and $y = (y_k)_{k\in\mathbb{N}}$, then equation (7.10) *is*

$$\biggl(\sum_{k=1}^{\infty} |x_k + y_k|^p\biggr)^{1/p} \;\le\; \biggl(\sum_{k=1}^{\infty} |x_k|^p\biggr)^{1/p} \;+\; \biggl(\sum_{k=1}^{\infty} |y_k|^p\biggr)^{1/p}.$$

Proof. Since $p > 1$, we can write

$$\begin{aligned}
\|x+y\|_p^p \;&=\; \sum_{k=1}^{\infty} |x_k+y_k|^p \\
&=\; \sum_{k=1}^{\infty} |x_k+y_k|\,|x_k+y_k|^{p-1} \\
&\le\; \sum_{k=1}^{\infty} |x_k|\,|x_k+y_k|^{p-1} \;+\; \sum_{k=1}^{\infty} |y_k|\,|x_k+y_k|^{p-1} \\
&=\; S_1 + S_2.
\end{aligned}$$

To simplify the series S_1, set $z_k = |x_k+y_k|^{p-1}$, so

$$S_1 \;=\; \sum_{k=1}^{\infty} |x_k|\,|z_k|.$$

We apply Hölder's Inequality, and then substitute $p' = p/(p-1)$, to compute as follows:

$$\begin{aligned}
S_1 \;=\; \sum_{k=1}^{\infty} |x_k|\,|z_k| \;&\le\; \biggl(\sum_{k=1}^{\infty} |x_k|^p\biggr)^{1/p} \biggl(\sum_{k=1}^{\infty} |z_k|^{p'}\biggr)^{1/p'} \\
&=\; \biggl(\sum_{k=1}^{\infty} |x_k|^p\biggr)^{1/p} \biggl(\sum_{k=1}^{\infty} |x_k+y_k|^{p}\biggr)^{(p-1)/p} \\
&=\; \|x\|_p\,\|x+y\|_p^{p-1}.
\end{aligned}$$

A similar calculation shows that

$$S_2 \;\le\; \|y\|_p\,\|x+y\|_p^{p-1}.$$

Combining these inequalities,

$$\|x+y\|_p^p \;\le\; S_1+S_2 \;\le\; \|x+y\|_p^{p-1}\bigl(\|x\|_p + \|y\|_p\bigr).$$

Dividing both sides by $\|x+y\|_p^{p-1}$ yields $\|x+y\|_p \le \|x\|_p + \|y\|_p$. $\square$

Now that we have established Minkowski's Inequality, we can show that $\|\cdot\|_p$ is a norm on ℓ^p.

Theorem 7.1.10. *If $1 \le p \le \infty$, then $\|\cdot\|_p$ is a norm on ℓ^p. That is, the following four statements are satisfied for all $x, y \in \ell^p$ and all scalars $c \in \mathbb{C}$.*

(a) Nonnegativity: $0 \le \|x\|_p < \infty$.

(b) Homogeneity: $\|cx\|_p = |c|\,\|x\|_p$.

(c) The Triangle Inequality: $\|x+y\|_p \le \|x\|_p + \|y\|_p$.

(d) Uniqueness: $\|x\|_p = 0$ *if and only if* $x = 0$. $\diamondsuit$

Proof. The nonnegativity requirement is satisfied by definition, and the homogeneity and uniqueness requirements follow easily. For $p = 1$ or $p = \infty$, the Triangle Inequality is established in Exercise 7.1.8, and for $1 < p < \infty$ it is proved in Theorem 7.1.9. $\square$

Our proofs of Hölder's and Minkowski's Inequalities can be easily adapted to sequences indexed by any other countable index set I. For example, if $I = \mathbb{Z}$ then

$$\|x\|_p \;=\; \biggl(\sum_{k=-\infty}^{\infty} |x_k|^p\biggr)^{1/p}, \qquad x = (x_k)_{k\in\mathbb{Z}} \in \ell^p(\mathbb{Z}),$$

defines a norm on $\ell^p(\mathbb{Z})$ for $1 \le p < \infty$, and $\|x\|_\infty = \sup_{k\in\mathbb{Z}} |x_k|$ is a norm on $\ell^\infty(\mathbb{Z})$. On the other hand, if we let $I = \{1, \dots, d\}$ then $\ell^p(I)$ is d-dimensional Euclidean space $\mathbb{C}^d$. This gives us the following collection of norms on $\mathbb{C}^d$. By restricting to real scalars, an entirely analogous result holds for $\mathbb{R}^d$.

Corollary 7.1.11. *For each $x = (x_1, \dots, x_d) \in \mathbb{C}^d$, define*

$$\|x\|_p \;=\; \begin{cases} \bigl(|x_1|^p + \cdots + |x_d|^p\bigr)^{1/p}, & \text{if } 1 \le p < \infty, \\[1ex] \max\bigl\{|x_1|, \dots, |x_d|\bigr\}, & \text{if } p = \infty. \end{cases}$$

Then $\|\cdot\|_p$ is a norm on $\mathbb{C}^d$ for each index $1 \le p \le \infty$. $\diamondsuit$

Open balls play an important role in any normed space. In ℓ^p, the *open ball centered at $x \in \ell^p$ with radius r* is

$$B_r(x) \;=\; \bigl\{y \in \ell^p : \|x-y\|_p < r\bigr\}.$$

Since ℓ^p is a normed space when $p \ge 1$, it shares all of the properties that any normed space enjoys. In particular, it follows from the Triangle Inequality that open balls in a normed space are convex (see Problem 1.2.11). The unit open balls in $\mathbb{R}^2$ corresponding to several choices of $p \ge 1$ are shown in Figure 7.2. All of these are indeed convex, although only the ball corresponding to $p = 2$ is "spherical" in the colloquial sense.

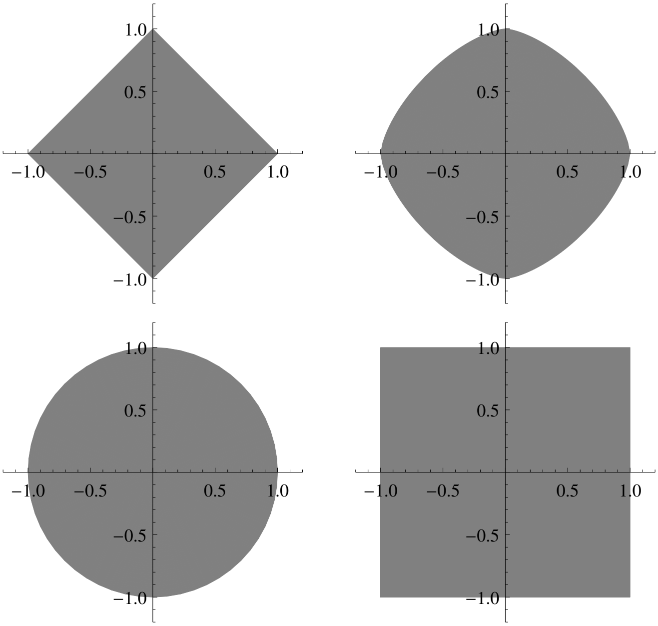


Fig. 7.2 Unit open balls $B_1(0)$ with respect to four norms $\|\cdot\|_p$ on $\mathbb{R}^2$. Top left: $p = 1$. Top right: $p = 3/2$. Bottom left: $p = 2$. Bottom right: $p = \infty$.

7.1.3 Convergence in the ℓ^p Spaces

When we speak of convergence in a normed space, unless we explicitly state otherwise we mean convergence with respect to the norm of that space. We spell this out precisely for ℓ^p in the following definition.

Definition 7.1.12 (Convergence in ℓ^p). A sequence of vectors $\{x_n\}_{n\in\mathbb{N}}$ in ℓ^p *converges* to a vector $x \in \ell^p$ if

$$\lim_{n\to\infty} \|x - x_n\|_p \;=\; 0.$$

In this case we write $x_n \to x$ *in* ℓ^p, and we say that x_n *converges to* x *in* ℓ^p*-norm.* $\diamondsuit$

Each vector x_n in Definition 7.1.12 is itself a sequence of scalars, as is the vector x. In order to describe the meaning of convergence in ℓ^p more explicitly, let us write x_n and x as

$$x_n \;=\; \bigl(x_n(k)\bigr)_{k\in\mathbb{N}} \;=\; \bigl(x_n(1), x_n(2), \dots\bigr)$$

and

$$x \;=\; \bigl(x(k)\bigr)_{k\in\mathbb{N}} \;=\; \bigl(x(1), x(2), \dots\bigr).$$

$$\begin{array}{rcllllll}
x_1 & = & (x_1(1), & x_1(2), & x_1(3), & x_1(4), & \dots) & \text{components of } x_1 \\
x_2 & = & (x_2(1), & x_2(2), & x_2(3), & x_2(4), & \dots) & \text{components of } x_2 \\
x_3 & = & (x_3(1), & x_3(2), & x_3(3), & x_3(4), & \dots) & \text{components of } x_3 \\
 & & \vdots & \vdots & \vdots & \vdots & & \vdots \\
 & & \downarrow & \downarrow & \downarrow & \downarrow & & \downarrow \\
x & = & (x(1), & x(2), & x(3), & x(4), & \dots) & \text{components of } x
\end{array}$$

Fig. 7.3 Illustration of componentwise convergence. For each k, the kth component of x_n converges to the kth component of x.

That is, $x_n(k)$ denotes the kth component of x_n, and $x(k)$ is the kth component of x. Using this notation, if p is finite then $x_n \to x$ in ℓ^p if and only if

$$\lim_{n\to\infty} \|x - x_n\|_p^p \;=\; \lim_{n\to\infty} \Biggl(\sum_{k=1}^{\infty} |x(k) - x_n(k)|^p\Biggr) \;=\; 0, \tag{7.11}$$

while if $p = \infty$ then $x_n \to x$ in ℓ^p if and only if

$$\lim_{n\to\infty} \|x - x_n\|_\infty \;=\; \lim_{n\to\infty} \Bigl(\sup_{k\in\mathbb{N}} |x(k) - x_n(k)|\Bigr) \;=\; 0. \tag{7.12}$$

Looking at equations (7.11) or (7.12), we see that if we choose a particular k and focus our attention on just the kth components of x_n and x, then we have

$$\lim_{n\to\infty} |x(k) - x_n(k)| \;\le\; \lim_{n\to\infty} \|x - x_n\|_p \;=\; 0. \tag{7.13}$$

That is, for each fixed k, the kth component of x_n converges to the kth component of x. As formalized in the next definition (and illustrated in Figure 7.3), this is called *componentwise convergence* of x_n to x.

Definition 7.1.13 (Componentwise Convergence). For each $n \in \mathbb{N}$ let $x_n = \bigl(x_n(k)\bigr)_{k\in\mathbb{N}}$ be a sequence of scalars, and let $x = \bigl(x(k)\bigr)_{k\in\mathbb{N}}$ be another sequence of scalars. We say that x_n *converges componentwise to* x if

$$\lim_{n\to\infty} x_n(k) \;=\; x(k) \quad \text{for every } k \in \mathbb{N}. \qquad \diamondsuit$$

Using this terminology, equation (7.13) establishes that convergence in ℓ^p implies componentwise convergence. We state this explicitly as follows.

Lemma 7.1.14. *Fix $0 < p \le \infty$. If $x_n, x \in \ell^p$ and $x_n \to x$ in ℓ^p, then x_n converges componentwise to x.* $\diamondsuit$

However, componentwise convergence *need not* imply convergence in ℓ^p-norm. For example, let

$$\delta_n \;=\; (0, \dots, 0, 1, 0, 0, \dots) \tag{7.14}$$

denote the sequence that has a 1 in the nth component and zeros elsewhere. We call δ_n the nth *standard basis vector*, and refer to $\mathcal{E} = \{\delta_n\}_{n\in\mathbb{N}}$ as the *sequence of standard basis vectors*, or simply the *standard basis*. Given k we have $\delta_n(k) = 0$ for all $n > k$, so δ_n converges componentwise to the zero sequence as $n \to \infty$. However, δ_n does not converge to 0 in ℓ^p-norm because $\|0 - \delta_n\|_p = 1$ for every n.

7.1.4 Completeness of the ℓ^p Spaces

The notion of a Cauchy sequence in a generic normed or metric space was introduced in Definition 1.1.2. Specializing to ℓ^p, a sequence $\{x_n\}_{n\in\mathbb{N}}$ in ℓ^p is *Cauchy in ℓ^p-norm*, or simply *Cauchy* for short, if for every $\varepsilon > 0$ there exists an integer $N > 0$ such that

$$m, n \geq N \quad \Longrightarrow \quad \|x_m - x_n\|_p < \varepsilon.$$

By applying the Triangle Inequality, we immediately see that every convergent sequence is Cauchy. A metric or normed space in which every Cauchy sequence converges to an element of the space is said to be *complete*, and a complete normed space is also called a *Banach space*. For example, $\mathbb{R}$ and $\mathbb{C}$ are Banach spaces with respect to absolute value.

We will prove that ℓ^p is complete for each index $1 \leq p \leq \infty$. The proof is a typical example of a completeness argument: We assume that $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence, then construct a "candidate vector" x that the sequence appears to converge to, and finally show that we do indeed have $\|x - x_n\|_p \to 0$ as $n \to \infty$.

Theorem 7.1.15 (Completeness of ℓ^p). *If $1 \leq p \leq \infty$, then ℓ^p is a Banach space with respect to the norm $\|\cdot\|_p$.*

Proof. We will present the proof for finite p, as the proof for $p = \infty$ is similar.

Assume that $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in ℓ^p, and write the components of x_n as

$$x_n = \bigl(x_n(1), x_n(2), \dots\bigr) = \bigl(x_n(k)\bigr)_{k\in\mathbb{N}}.$$

If $\varepsilon > 0$, then there is an integer $N > 0$ such that $\|x_m - x_n\|_p < \varepsilon$ for all $m, n \geq N$. Therefore, if we fix a particular index $k \in \mathbb{N}$ then for all $m, n \geq N$ we have

$$|x_m(k) - x_n(k)| \leq \|x_m - x_n\|_p < \varepsilon.$$

Thus, for this fixed k, we see that $\bigl(x_n(k)\bigr)_{n\in\mathbb{N}}$ is a Cauchy sequence of *scalars*. It must therefore converge since $\mathbb{C}$ is a Banach space. Define

$$x(k) = \lim_{n\to\infty} x_n(k), \tag{7.15}$$

and set $x = \bigl(x(1), x(2), \dots\bigr)$. Then, by construction, x_n *converges componentwise to* x. We must prove that $x \in \ell^p$, and that x_n *converges to* x *in* ℓ^p*-norm.*

Given $\varepsilon > 0$, there is an $N > 0$ such that $\|x_m - x_n\|_p < \varepsilon$ for all $m, n \ge N$. Applying the series version of Fatou's Lemma (see Problem 4.2.18), it follows that

$$\begin{aligned}
\|x - x_n\|_p^p &= \sum_{k=1}^{\infty} |x(k) - x_n(k)|^p \\
&= \sum_{k=1}^{\infty} \liminf_{m\to\infty} |x_m(k) - x_n(k)|^p \qquad \Bigl(\text{since } x(k) = \lim_{m\to\infty} x_m(k)\Bigr) \\
&\le \liminf_{m\to\infty} \sum_{k=1}^{\infty} |x_m(k) - x_n(k)|^p \qquad (\text{Fatou for Series}) \\
&= \liminf_{m\to\infty} \|x_m - x_n\|_p^p \\
&\le \varepsilon^p.
\end{aligned} \tag{7.16}$$

Even though we do not know yet that $x \in \ell^p$, this tells us that the vector $x - x_n$ has finite ℓ^p-norm and therefore belongs to ℓ^p. Since ℓ^p is closed under addition, it follows that $x = (x - x_n) + x_n \in \ell^p$. Thus, our candidate sequence x does belong to ℓ^p. Further, equation (7.16) establishes that $\|x - x_n\|_p \le \varepsilon$ for all $n \ge N$, so we have shown that $\|x - x_n\|_p \to 0$ as $n \to \infty$. Hence $x_n \to x$ in ℓ^p-norm, and therefore ℓ^p is complete. □

Similarly, if I is any countable index set then $\ell^p(I)$ is complete for each $1 \le p \le \infty$. In particular, taking $I = \{1, \dots, d\}$ gives the following corollary (and a similar result holds for $\mathbb{R}^d$).

Corollary 7.1.16. *If* $1 \le p \le \infty$, *then* $\mathbb{C}^d$ *is a Banach space with respect to the norm* $\|\cdot\|_p$ *defined in Corollary 7.1.11.* ◇

7.1.5 ℓ^p for $p < 1$

The ℓ^p spaces with indices $0 < p < 1$ do play an important role in certain applications, such as those requiring "sparse representations." Unfortunately, $\|\cdot\|_p$ is not a norm when $p < 1$. For example, using the first two standard basis vectors $\delta_1 = (1, 0, 0, 0, \dots)$ and $\delta_2 = (0, 1, 0, 0, \dots)$ we compute that

$$\|\delta_1 + \delta_2\|_p \;=\; 2^{1/p} \;>\; 2 \;=\; \|\delta_1\|_p + \|\delta_2\|_p.$$

Hence $\|\cdot\|_p$ fails the Triangle Inequality when $p < 1$. Even so, the following exercise shows that we can define a *metric* d_p on ℓ^p (see Definition 1.1.1 for the definition of a metric).

Exercise 7.1.17. Given $0 < p < 1$, prove the following statements.

(a) $(1+t)^p \le 1 + t^p$ for all $t > 0$.

(b) If $a, b > 0$, then $(a+b)^p \le a^p + b^p$.

(c) $\|x+y\|_p^p \le \|x\|_p^p + \|y\|_p^p$ for all $x, y \in \ell^p$.

(d) $\mathrm{d}_p(x,y) = \|x-y\|_p^p = \sum_{k=1}^\infty |x_k - y_k|^p$ defines a metric on ℓ^p. $\diamondsuit$

When $p < 1$, convergence and other notions in ℓ^p are defined with respect to the metric d_p. For example, $x_n \to x$ *in* ℓ^p if $\mathrm{d}_p(x, x_n) \to 0$ as $n \to \infty$. An argument virtually identical to the proof of Theorem 7.1.15 shows that every Cauchy sequence in ℓ^p converges to an element of ℓ^p. Thus ℓ^p is a complete metric space (but we do not call it a Banach space because it is not a *normed* space). We summarize this discussion as follows.

Theorem 7.1.18. *If* $0 < p < 1$, *then* d_p *is a metric on* ℓ^p, *and* ℓ^p *is a complete metric space with respect to* d_p. $\diamondsuit$

A direct computation shows that if $p < 1$ then the open ball

$$B_r(x) \;=\; \bigl\{y \in \ell^p : \mathrm{d}_p(x,y) < r\bigr\}$$

is *not convex* (compare Figure 7.4).

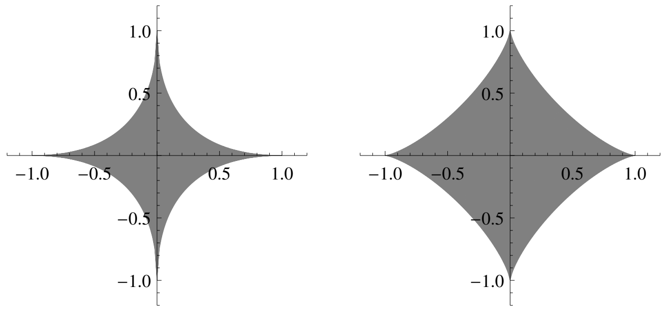


Fig. 7.4 Unit open balls $B_1(0)$ with respect to two metrics d_p on $\mathbb{R}^2$. Left: $p = 1/2$. Right: $p = 3/4$.

7.1.6 c_0 and c_{00}

We introduce two additional sequence spaces. These spaces, which are discrete analogues of the function spaces $C_0(\mathbb{R})$ and $C_c(\mathbb{R})$, are

$$c_0 \;=\; \Bigl\{x = (x_k)_{k\in\mathbb{N}} : \lim_{k\to\infty} x_k = 0\Bigr\}$$

and
$$c_{00} = \Big\{x = (x_k)_{k\in\mathbb{N}} : \text{only finitely many } x_k \neq 0\Big\}.$$

The elements of c_0 are sequences whose components "converge to zero at infinity," while the elements of c_{00} are sequences that "end with infinitely many zeros." If $0 < p < \infty$, then

$$c_{00} \subsetneq \ell^p \subsetneq c_0 \subsetneq \ell^\infty.$$

According to Problem 7.1.28, c_0 is a closed subspace of ℓ^∞ with respect to the norm $\|\cdot\|_\infty$, and hence is itself a Banach space with respect to the sup-norm. In contrast, Problem 7.1.29 shows that c_{00} is not complete with respect to any norm $\|\cdot\|_p$.

The elements of c_{00} are sometimes called *finite sequences* because they contain at most finitely many nonzero components. If we recall the standard basis vectors δ_n introduced in equation (7.14), we see that c_{00} is the set of all *finite linear combinations* of the set of standard basis vectors $\mathcal{E} = \{\delta_n\}_{n\in\mathbb{N}}$, because

$$\begin{aligned} c_{00} &= \Big\{x = (x_1, \dots, x_N, 0, 0, \dots) \;:\; N > 0,\, x_1, \dots, x_N \in \mathbb{C}\Big\} \\ &= \Big\{\sum_{k=1}^{N} x_k\delta_k \;:\; N > 0,\, x_1, \dots, x_N \in \mathbb{C}\Big\} \;=\; \operatorname{span}(\mathcal{E}). \end{aligned}$$

Since $\mathcal{E}$ spans c_{00} and $\mathcal{E}$ is linearly independent, we conclude that $\mathcal{E}$ is a basis for c_{00} in the usual vector space sense. Such a "vector space basis" is also called a *Hamel basis* (see Definition 1.2.2). However, $\mathcal{E}$ is not a Hamel basis for c_0 or ℓ^p because its span is only c_{00}, which is a proper subset of c_0 and ℓ^p.

Problems

7.1.19. Assume that $1 \le p < \infty$. Given $x = (x_k)_{k\in\mathbb{N}} \in \ell^p$, prove that $\sum_{k=1}^\infty \frac{|x_k|}{k} < \infty$. Show by example that this can fail if $x \in \ell^\infty$.

7.1.20. Observe that $\|x\|_\infty \le \|x\|_1$ for every $x \in \ell^1$. Prove that there does not exist a finite constant $B > 0$ such that the inequality $\|x\|_1 \le B\,\|x\|_\infty$ holds for every $x \in \ell^1$.

7.1.21. Show that if $0 < p < q \le \infty$ then $\ell^p \subsetneq \ell^q$, and $\|x\|_q \le \|x\|_p$ for all $x \in \ell^p$.

7.1.22. Prove that if $x \in \ell^q$ for some finite index q, then $\|x\|_p \to \|x\|_\infty$ as $p \to \infty$. Give an example of a sequence $x \in \ell^\infty$ for which this fails.

7.1.23. Given $1 < p < \infty$, show that equality holds in Hölder's Inequality (Theorem 7.1.7) if and only if there exist scalars α and β, not both zero, such that $\alpha\,|x_k|^p = \beta\,|y_k|^{p'}$ for each $k \in \mathbb{N}$. What about the cases $p = 1$ or $p = \infty$?

7.1.24. Prove the following generalization of Hölder's Inequality. Assume that $1 \le p,\ q,\ r \le \infty$ satisfy $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Given $x = (x_k)_{k\in\mathbb{N}} \in \ell^p$ and $y = (y_k)_{k\in\mathbb{N}} \in \ell^q$, prove that $xy = (x_k y_k)_{k\in\mathbb{N}}$ belongs to ℓ^r, and

$$\|xy\|_r \le \|x\|_p\,\|y\|_q.$$

7.1.25. Choose $1 \le p \le \infty$, and let $D = \{x \in \ell^p : \|x\|_p \le 1\}$ be the "closed unit disk" in ℓ^p. Observe that D is a bounded subset of ℓ^p, since it is contained in an open ball of finite radius. Prove the following statements.

(a) D is a closed subset of ℓ^p, i.e., if $\{x_n\}_{n\in\mathbb{N}}$ is a sequence in D such that $x_n \to x$ in ℓ^p-norm, then $x \in D$.

(b) The sequence of standard basis vectors $\{\delta_n\}_{n\in\mathbb{N}}$ contains no convergent subsequences.

(c) D is not a compact subset of ℓ^p (consider Theorem 1.1.10).

7.1.26. Fix $1 \le p < \infty$.

(a) Let $x = (x_k)_{k\in\mathbb{N}}$ be a sequence of scalars that decays on the order of $k^{-\alpha}$ where $\alpha > 1/p$. That is, assume that $\alpha > 1/p$ and there exists a constant $C > 0$ such that

$$|x_k| \;\le\; C k^{-\alpha} \quad \text{for all } k \in \mathbb{N}. \tag{7.17}$$

Show that $x \in \ell^p$.

(b) Set $\alpha = 1/p$. Exhibit a sequence $x \notin \ell^p$ that satisfies equation (7.17) for some $C > 0$, and another sequence $x \in \ell^p$ that satisfies equation (7.17) for some $C > 0$.

(c) Given $\alpha > 0$, show that there exists a sequence $x = (x_k)_{k\in\mathbb{N}} \in \ell^p$ such that there is *no* constant $C > 0$ that satisfies equation (7.17). Conclude that no matter how small we choose α, there exist sequences in ℓ^p whose decay rate is slower than $k^{-\alpha}$.

(d) Suppose that the components of $x = (x_k)_{k\in\mathbb{N}} \in \ell^p$ are nonnegative and *monotonically decreasing.* Show that there exists some $\alpha \ge 1/p$ and some $C > 0$ such that equation (7.17) holds.

7.1.27. Given an arbitrary (possibly uncountable) index set I, let $\ell^\infty(I)$ be the space of all bounded sequences $x = (x_i)_{i\in I}$, and set $\|x\|_\infty = \sup_{i\in I} |x_i|$. For $1 \le p < \infty$ let $\ell^p(I)$ consist of all sequences $x = (x_i)_{i\in I}$ *with at most countably many nonzero components* such that $\|x\|_p^p = \sum |x_i|^p < \infty$. Prove that $\ell^p(I)$ is a Banach space with respect to $\|\cdot\|_p$.

7.1.28. Prove that c_0 is a closed subspace of ℓ^∞, i.e., if $x_n \in c_0$ and $x \in \ell^\infty$ are such that $\|x - x_n\|_\infty \to 0$, then $x \in c_0$. (Consequently, Problem 1.2.12 implies that c_0 is a Banach space with respect to $\|\cdot\|_\infty$.)

7.1.29. Prove the following statements (we implicitly assume that the norm on ℓ^p is $\|\cdot\|_p$, and the norm on c_0 is $\|\cdot\|_\infty$).

(a) If $1 \le p < \infty$, then c_{00} is a dense subspace of ℓ^p. Further, c_{00} is a dense subspace of c_0, but c_{00} is not dense in ℓ^∞.

(b) If $1 \le p \le \infty$, then c_{00} is not complete with respect to $\|\cdot\|_p$. That is, there exist vectors $x_n \in c_{00}$ such that $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy with respect to $\|\cdot\|_p$, but there is no vector $x \in c_{00}$ such that $x_n \to x$ in ℓ^p-norm.

7.1.30. (a) Suppose that $\sum x_n$ is an *absolutely convergent* series in c_0, i.e., $x_n \in c_0$ for every $n \in \mathbb{N}$ and $\sum \|x_n\|_\infty < \infty$. Prove directly that the series $\sum x_n$ converges with respect to the sup-norm, i.e., there exists a sequence $x \in c_0$ such that $\lim_{N\to\infty} \bigl\|x - \sum_{n=1}^N x_n\bigr\|_\infty = 0$.

(b) Use part (a) and Theorem 1.2.8 to give another proof that c_0 is complete with respect to $\|\cdot\|_\infty$.

7.2 The Lebesgue Space $L^p(E)$

According to Definition 4.4.1, a measurable function f is *integrable* on a set E if the integral of $|f|$ on E is finite. Now we refine that notion. Given an index $0 < p < \infty$, we say that f is *p-integrable* if the integral of $|f|^p$ is finite. We collect all of the p-integrable functions to form a space that we call $L^p(E)$. For $p = \infty$, we define $L^\infty(E)$ to be the space of all essentially bounded functions on E. These L^p spaces are the function space analogues of the ℓ^p spaces.

There are actually two versions of each space, one consisting of complex-valued functions and one consisting of extended real-valued functions. Entirely similar results hold for both cases. As before, we treat both possibilities together by letting the symbol $\overline{\mathbf{F}}$ denote choice of either $[-\infty,\infty]$ or $\mathbb{C}$. In conjunction with this, we let the word *scalar* denote a real number if $\overline{\mathbf{F}} = [-\infty,\infty]$, and a complex number if $\overline{\mathbf{F}} = \mathbb{C}$ (compare Notation 3.1.1).

Definition 7.2.1 (The Lebesgue Space $L^p(E)$). Let E be a measurable subset of $\mathbb{R}^d$.

(a) If $0 < p < \infty$ and $f\colon E \to \overline{\mathbf{F}}$ is measurable, then we say that f is p-integrable if $\int_E |f|^p < \infty$. In this case we set

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

If f is not p-integrable then we take $\|f\|_p = \infty$. We define $L^p(E)$ to be the set of all p-integrable functions on E, and call $L^p(E)$ the *Lebesgue space of p-integrable functions on E*.

(b) If $p = \infty$, then $L^\infty(E)$ is the set of all measurable functions $f\colon E \to \overline{\mathbf{F}}$ that are essentially bounded. That is, f belongs to $L^\infty(E)$ if

$$\|f\|_\infty \;=\; \operatorname*{ess\,sup}_{x\in E} |f(x)| \;<\; \infty.$$

We call $L^\infty(E)$ the *Lebesgue space of essentially bounded functions on* E. $\diamondsuit$

If E is an interval, then to avoid multiplicities of brackets and parentheses we usually write $L^p[a,b]$ instead of $L^p([a,b])$, $L^p[a,b)$ instead of $L^p([a,b))$, and so forth.

Remark 7.2.2. A complex-valued function never takes the values $\pm\infty$, so a complex-valued function is (by definition) finite at every point. An extended real-valued function f can take the values $\pm\infty$, but if f belongs to $L^p(E)$ then this can happen only on a set of measure zero. Hence every function in $L^p(E)$ is finite a.e. On the other hand, a function that is finite a.e. need not belong to $L^p(E)$. For example, $f(x) = 1/x$ is finite a.e. on $[0,\infty)$, but it does not belong to $L^p[0,\infty)$ for any p. $\diamondsuit$

In certain respects, the L^p spaces behave similarly to the ℓ^p spaces, and consequently several proofs from Section 7.1 carry over to $L^p(E)$ with only minor changes. For example, a small modification of Lemma 7.1.4 shows that $L^p(E)$ is closed under addition of functions. We will state as exercises some results for L^p whose proofs can be directly adapted from those for ℓ^p.

The similarity between ℓ^p and $L^p(E)$ is a reflection of the deeper fact that both of these are particular cases of a more general class of spaces $L^p(\mu)$, where μ is a *positive measure* defined on a *measurable space* (X,Σ) that consists of a set X and a σ-algebra Σ of subsets of X (compare Problem 4.5.33). If we take $X = \mathbb{N}$ and $\Sigma = \mathcal{P}(\mathbb{N})$, then ℓ^p is precisely $L^p(\mu)$ where μ is counting measure on $\mathbb{N}$. Likewise, $L^p(E) = L^p(\mu)$ where μ is Lebesgue measure on $X = E$ and $\Sigma = \mathcal{L}(E)$ is the set of all Lebesgue measurable subsets of E. For more details on abstract measure theory, we refer to texts such as [Fol99] or [Rud90].

Although ℓ^p and $L^p(E)$ are similar in certain ways, in other respects their properties are quite different. For example, while $\ell^1 \subseteq \ell^\infty$ (Problem 7.1.21), we have $L^\infty(E) \subseteq L^1(E)$ when $|E| < \infty$, and there is no inclusion between $L^\infty(E)$ and $L^1(E)$ when $|E| = \infty$ (see Problem 7.2.16). Another difference concerns convergence, because convergence with respect to the norm of ℓ^p implies componentwise convergence (Lemma 7.1.14), while convergence in L^p-norm only implies the existence of a *subsequence* that converges pointwise a.e. (see Theorem 7.3.4). Yet another difference is that the zero sequence is the only sequence whose ℓ^p norm is zero, while any function that is zero *almost everywhere* will have zero L^p norm, even though such a function need not be identically zero.

7.2.1 Seminorm Properties of $\|\cdot\|_p$

We will show that $\|\cdot\|_p$ is a seminorm (but not a norm) on $L^p(E)$ when $1 \le p \le \infty$. The nonnegativity requirement is satisfied by definition, because $0 \le \|f\|_p < \infty$ for all $f \in L^p(E)$, and the homogeneity property $\|cf\|_p = |c|\,\|f\|_p$ follows directly. The proof that $\|\cdot\|_p$ satisfies the Triangle Inequality for $p = 1$ and $p = \infty$ is straightforward (and in fact was already done in Exercises 3.3.4 and 4.4.5). To prove the Triangle Inequality for $1 < p < \infty$ we need Hölder's Inequality for the L^p spaces. The proof is similar to the corresponding result for ℓ^p, so we assign it as an exercise.

Exercise 7.2.3 (Hölder's Inequality). Assume that $E \subseteq \mathbb{R}^d$ is measurable, and fix $1 \le p \le \infty$. Prove that if $f \in L^p(E)$ and $g \in L^{p'}(E)$, then $fg \in L^1(E)$ and

$$\|fg\|_1 \le \|f\|_p \, \|g\|_{p'}. \qquad \diamondsuit \tag{7.18}$$

For indices in the range $1 < p < \infty$, we can write Hölder's Inequality in the form

$$\int_E |fg| \;\le\; \biggl(\int_E |f|^p\biggr)^{1/p} \biggl(\int_E |g|^{p'}\biggr)^{1/p'}.$$

Note that if $1 < p < 2$ then $2 < p' < \infty$, and similarly if $2 < p < \infty$ then $1 < p' < 2$. For $p = 2$ we have "self-duality," because $2' = 2$. This fact will be especially important when we explore the Hilbert space properties of $L^2(E)$ in Chapter 8.

If $p = 1$ then $p' = \infty$, and in this case Hölder's Inequality takes the form

$$\int_E |fg| \;\le\; \biggl(\int_E |f|\biggr) \Bigl(\operatorname*{ess\,sup}_{x \in E} |g(x)|\Bigr).$$

The case $p = \infty$, $p' = 1$ is entirely symmetrical and follows by interchanging the roles of f and g in the preceding line.

The Triangle Inequality for $\|\cdot\|_p$ is also known as *Minkowski's Inequality.* We saw how to use Hölder's Inequality to prove Minkowski's Inequality for the ℓ^p spaces in Theorem 7.1.15, and the proof for $L^p(E)$ is similar.

Exercise 7.2.4 (Minkowski's Inequality). Let $E \subseteq \mathbb{R}^d$ be a measurable set, and fix $1 \le p \le \infty$. Prove *Minkowski's Inequality*:

$$\|f+g\|_p \;\le\; \|f\|_p + \|g\|_p, \qquad \text{for all } f, g \in L^p(E). \tag{7.19}$$

Conclude that $\|\cdot\|_p$ is a seminorm on $L^p(E)$. $\diamondsuit$

Although $\|\cdot\|_p$ is a seminorm on $L^p(E)$, it is not a norm because the uniqueness requirement is not strictly satisfied. To be a norm, it would have to be the case that $\|f\|_p = 0$ if and only if f is identically zero. However, any function f that is zero *almost everywhere* satisfies $\|f\|_p = 0$. The next theorem summarizes the properties of $\|\cdot\|_p$.

Theorem 7.2.5. *If $E \subseteq \mathbb{R}^d$ is measurable and $1 \le p \le \infty$, then the following statements hold for all functions $f, g \in L^p(E)$ and all scalars c.*

(a) Nonnegativity: $\|f\|_p \ge 0$.

(b) Homogeneity: $\|cf\|_p = |c|\, \|f\|_p$.

(c) The Triangle Inequality: $\|f+g\|_p \le \|f\|_p + \|g\|_p$.

(d) Almost Everywhere Uniqueness: $\|f\|_p = 0$ *if and only if* $f = 0$ *a.e.*

Proof. We have already observed that $\|\cdot\|_p$ is a seminorm, so the only issue is to show that statement (d) holds. For $p = \infty$, this follows from Corollary 2.2.29. On the other hand, if p is finite and $\|f\|_p = 0$ then $\int_E |f|^p = 0$, so Exercise 4.1.10 implies that $|f|^p = 0$ a.e. □

Thus $\|\cdot\|_p$ is "almost" a norm on $L^p(E)$. The seminorm properties are satisfied, but the zero function is not the only function whose L^p-norm is zero. Instead, $\|f\|_p = 0$ if and only if $f = 0$ *almost everywhere.*

7.2.2 Identifying Functions That Are Equal Almost Everywhere

In most circumstances, the fact that $\|\cdot\|_p$ is a seminorm but not quite a norm is only a minor nuisance. Changing the value of a function on a set of measure zero does not change its integral, so as far as most purposes related to integration are concerned, functions that are equal almost everywhere behave identically. Consequently, if f and g are two measurable functions that are equal a.e., then it is natural to identify them and regard them as being the same object. For example, if $\|f\|_p = 0$ then $f = 0$ a.e., so with respect to this identification f is the same object as the zero function and hence *is* the zero element of $L^p(E)$. Using this informal identification we have that

$$\|f\|_p = 0 \iff f = 0 \text{ a.e.} \iff f \text{ is the zero element of } L^p(E).$$

Once we adopt this convention of identifying functions that are equal a.e. the uniqueness requirement is automatically satisfied, so $\|\cdot\|_p$ is a norm on $L^p(E)$.

Notation 7.2.6 (Informal Convention for Elements of $L^p(E)$). We take $L^p(E)$ to be the set of all p-integrable functions on E, but if f and g are two p-integrable functions that are equal almost everywhere then we regard f and g as being the same element of $L^p(E)$. In this case, we say that f and g are *representatives* of this element of $L^p(E)$. ◇

Problem 7.2.24 shows how to make this convention completely rigorous by forming equivalence classes of functions. However, for most purposes the informal approach of Notation 7.2.6 is sufficient. We must exercise some care;

in particular, we should check that any definitions that we make or operations that we perform on elements of $L^p(E)$ are well-defined in the sense that they do not depend on the choice of representative. Usually this is not difficult. For example, the norm $\|f\|_p = \left(\int_E |f|^p\right)^{1/p}$ does not depend on the choice of representative, because if $f = g$ almost everywhere then $\int_E |f|^p = \int_E |g|^p$. In the same spirit, the following exercise asks for a justification that pointwise a.e. convergence is well-defined on $L^p(E)$.

Exercise 7.2.7. Let E be a measurable subset of $\mathbb{R}^d$. Given $f_n, f \in L^p(E)$, prove that pointwise a.e. convergence is independent of the choice of representatives of f_n and f. That is, show that if $f_n \to f$ pointwise a.e., $g_n = f_n$ a.e., and $g = f$ a.e., then $g_n \to g$ a.e. $\diamondsuit$

The set of measure zero in Exercise 7.2.7 on which $g_n(x)$ does not converge to $g(x)$ could be different than the set of measure zero on which $f_n(x)$ does not converge to $f(x)$, but we still have pointwise a.e. convergence. Consequently, it makes sense to say that elements of $L^p(E)$ converge pointwise almost everywhere; this just means pointwise a.e. convergence of any representatives of these functions.

In contrast, it does not make literal sense to say that an element of $L^p(E)$ is continuous, because continuity can depend on the choice of representative. For example, 0 and $\chi_{\mathbb{Q}}$ are both representatives of the zero function in $L^p(\mathbb{R})$, yet 0 is continuous while $\chi_{\mathbb{Q}}$ is not. Consequently, we adopt the following conventions.

Notation 7.2.8 (Continuity for Elements of $L^p(E)$).

(a) If f is a continuous function that is p-integrable on E, then we say that "*f belongs to $L^p(E)$*" with the understanding that this means that any function that equals f a.e. is the same element of $L^p(E)$.

(b) We write "*a function $f \in L^p(E)$ is continuous*" if there is a representative of f that is continuous. That is, f is continuous if there exists some continuous function g such that $f = g$ a.e. $\diamondsuit$

For example, $f(x) = e^{-|x|}$ is continuous and p-integrable on $\mathbb{R}$, so we write $e^{-|x|} \in L^p(\mathbb{R})$, with the understanding that any function g that satisfies $g(x) = e^{-|x|}$ a.e. is the same element of $L^p(\mathbb{R})$.

7.2.3 $L^p(E)$ for $0 < p < 1$

We considered ℓ^p with $0 < p < 1$ in Section 7.1.5, and saw that if $p < 1$ then $\|\cdot\|_p$ does not satisfy the Triangle Inequality, and therefore is not a norm on ℓ^p. A similar phenomenon holds for $L^p(E)$ when $p < 1$ (unless $|E| = 0$, in which case $L^p(E)$ only contains the zero function).

Exercise 7.2.9. Let E be a measurable subset of $\mathbb{R}^d$ such that $|E| > 0$. Prove that if $0 < p < 1$, then the following statements hold.

(a) $L^p(E)$ is a vector space, and

$$\mathrm{d}_p(f,g) \;=\; \|f-g\|_p^p \;=\; \int_E |f-g|^p, \qquad \text{for } f,g \in L^p(E),$$

defines a metric on $L^p(E)$.

(b) $L^p(E)$ is complete with respect to the metric d_p.

(c) The unit open ball $B_1(0) = \bigl\{f \in L^p(E) : \mathrm{d}_p(f,0) = \|f\|_p^p < 1\bigr\}$ is not a convex subset of $L^p(E)$.

(d) The metric d_p is not induced from any norm on $L^p(E)$. That is, there does not exist any norm $|||\cdot|||$ on $L^p(E)$ such that $\mathrm{d}_p(f,g) = |||f-g|||$ for all $f,g \in L^p(E)$. $\diamondsuit$

7.2.4 The Converse of Hölder's Inequality

If we fix a function $f \in L^p(E)$, then Hölder's Inequality implies that

$$\sup_{\|g\|_{p'}=1} \left|\int_E fg\right| \;\le\; \sup_{\|g\|_{p'}=1} \|f\|_p\,\|g\|_{p'} \;=\; \|f\|_p. \tag{7.20}$$

Our next theorem shows that equality holds in this equation.

Theorem 7.2.10 (Converse of Hölder's Inequality). *Let E be a measurable subset of $\mathbb{R}^d$, and fix $1 \le p \le \infty$. Then for each function $f \in L^p(E)$ we have*

$$\sup_{\|g\|_{p'}=1} \left|\int_E fg\right| \;=\; \|f\|_p. \tag{7.21}$$

Furthermore, this supremum is achieved. In fact, there exists a function g in $L^{p'}(E)$ such that $\|g\|_{p'} = 1$ and $\int_E fg = \|f\|_p$.

Proof. Assume first that $1 < p < \infty$. Hölder's Inequality gives us equation (7.20), so we need to prove that equality holds. Fix $f \in L^p(E)$. If $f = 0$ a.e., then the result is trivial, so we can assume that f is not the zero vector in $L^p(E)$. By choosing an appropriate representative of f (i.e., redefine $f(x)$ at any point in the set of measure zero where it takes the value $\pm\infty$), we can further assume that f is finite at every point.

For each x, let $\alpha(x)$ be a scalar such that $|\alpha(x)| = 1$ and $\alpha(x)\,f(x) = |f(x)|$. Explicitly, we can take

$$\alpha(x) \;=\; \begin{cases} |f(x)|/f(x), & \text{if } f(x) \ne 0,\\ 0, & \text{if } f(x) = 0. \end{cases}$$

This function α is measurable and bounded. Set

$$g(x) \;=\; \frac{\alpha(x)\,|f(x)|^{p-1}}{\|f\|_p^{p-1}}, \qquad \text{for } x \in E.$$

Since $(p-1)\,p' = p$,

$$\|g\|_{p'}^{p'} \;=\; \int_E \biggl(\frac{|f(x)|^{p-1}}{\|f\|_p^{p-1}}\biggr)^{p'} dx \;=\; \frac{\int_E |f(x)|^p\,dx}{\|f\|_p^p} \;=\; 1.$$

Thus g is a unit vector in $L^{p'}(E)$. Also,

$$\begin{aligned}\int_E fg\,dx \;&=\; \int_E f(x)\,\frac{\alpha(x)\,|f(x)|^{p-1}}{\|f\|_p^{p-1}}\,dx\\ &=\; \frac{\int_E |f(x)|^p\,dx}{\|f\|_p^{p-1}} \;=\; \frac{\|f\|_p^p}{\|f\|_p^{p-1}} \;=\; \|f\|_p.\end{aligned}$$

This shows that equality holds in equation (7.21), and that the supremum in that equation is achieved.

Exercise: Complete the proof for the cases $p = 1$ and $p = \infty$. □

Problems

7.2.11. Fix $1 \le p \le \infty$. Determine all values of $\alpha, \beta \in \mathbb{R}$ for which $f_\alpha(x) = x^\alpha\,\chi_{[0,1]}(x)$ or $g_\beta(x) = x^\beta\,\chi_{[1,\infty)}(x)$ belong to $L^p(\mathbb{R})$.

7.2.12. Fix $1 \le p \le \infty$, and let E be any measurable subset of $\mathbb{R}^d$. Suppose that $f_n \in L^p(E)$ for $n \in \mathbb{N}$ and $f_n \to f$ a.e. Prove that if $\sup \|f_n\|_p < \infty$ then $f \in L^p(E)$, but show by example that the assumption that $\{f_n\}_{n\in\mathbb{N}}$ is a bounded sequence in $L^p(E)$ is necessary.

7.2.13. Prove the following L^p version of *Tchebyshev's Inequality*: If $E \subseteq \mathbb{R}^d$ is a measurable set and $f\colon E \to \overline{\mathbf{F}}$ is a measurable function, then for each $\alpha > 0$ we have

$$\bigl|\{|f| > \alpha\}\bigr| \;\le\; \frac{1}{\alpha^p}\int_{\{|f|>\alpha\}} |f|^p \;\le\; \frac{1}{\alpha^p}\int_E |f|^p.$$

7.2.14. Let $E \subseteq \mathbb{R}^d$ be a measurable set such that $|E| < \infty$. Prove that if f is measurable on E, then $\|f\|_p \to \|f\|_\infty$ as $p \to \infty$. Show by example that the hypothesis that E has finite measure is necessary.

7.2.15. Given $1 < p < \infty$, show that equality holds in Hölder's Inequality if and only if there exist scalars α and β, not both zero, such that $\alpha\,|f|^p = \beta\,|g|^{p'}$ a.e.

7.2.16. Let E be a measurable subset of $\mathbb{R}^d$, and fix $0 < p < q \le \infty$. Prove the following statements.

(a) If $0 < |E| < \infty$, then $L^q(E) \subsetneq L^p(E)$ and

$$\|f\|_p \;\le\; |E|^{\frac{1}{p}-\frac{1}{q}}\, \|f\|_q, \qquad \text{for all } f \in L^p(E).$$

(b) If $|E| = \infty$, then $L^p(E)$ is not contained in $L^q(E)$, and $L^q(E)$ is not contained in $L^p(E)$.

7.2.17. Let $E \subseteq \mathbb{R}^m$ and $F \subseteq \mathbb{R}^n$ be measurable sets, let $f(x,y)$ be a measurable function on $E \times F$, and fix $1 \le p < \infty$. Prove *Minkowski's Integral Inequality*:

$$\left(\int_E \left(\int_F |f(x,y)|\,dy\right)^p dx\right)^{1/p} \;\le\; \int_F \left(\int_E |f(x,y)|^p\,dx\right)^{1/p} dy. \tag{7.22}$$

Remark: Equation (7.22) may be more revealing if we rewrite it as

$$\left\|\int_F |f(\cdot,y)|\,dy\right\|_p \;\le\; \int_F \|f(\cdot,y)\|_p\,dy.$$

Thus, Minkowski's Integral Inequality is an integral version of the Triangle Inequality (also known as Minkowski's Inequality) on $L^p(E)$.

7.2.18. (a) Suppose that f is absolutely continuous on $[a,b]$ and $f' \in L^p[a,b]$, where $1 < p \le \infty$. Prove that f is Hölder continuous with exponent $1/p'$.

(b) Show that the function g defined in Problem 1.4.4(d) is absolutely continuous on $[0,\frac{1}{2}]$, even though it is not Hölder continuous for any positive exponent.

7.2.19. Let $1 \le p \le \infty$ be given. Suppose that ϕ is a measurable function on $\mathbb{R}$ such that $f\phi \in L^p(\mathbb{R})$ for every $f \in L^p(\mathbb{R})$. Prove that $\phi \in L^\infty(\mathbb{R})$.

7.2.20. Formulate an analogue of Problem 7.1.24 for the L^p spaces, and then prove the following extension of Hölder's Inequality. Assume that $1 \le p_1, \dots, p_n, r \le \infty$ satisfy

$$\frac{1}{p_1} + \cdots + \frac{1}{p_n} \;=\; \frac{1}{r}.$$

Given functions $f_j \in L^{p_j}(E)$, prove that the product $f_1 \cdots f_n$ belongs to $L^1(E)$, and $\|f_1 \cdots f_n\|_r \le \|f_1\|_{p_1} \cdots \|f_n\|_{p_n}$.

7.2.21. Given a measurable function f on a measurable set $E \subseteq \mathbb{R}^d$, let $\omega(t) = \bigl|\{|f| > t\}\bigr|$ be the *distribution function* defined in Problem 4.6.21. Fix $1 \le p < \infty$, and prove the following statements.

(a) $f \in L^p(E)$ if and only if $\sum_{k\in\mathbb{Z}} 2^{kp}\,\omega(2^k) < \infty$.

(b) $f \in L^p(E)$ if and only if $\int_0^\infty t^{p-1}\,\omega(t)\,dt < \infty$. Further, in this case,

$$\int_E |f(x)|^p\,dx \;=\; p\int_0^\infty t^{p-1}\,\omega(t)\,dt.$$

7.2.22. Fix $1 < p < \infty$, and let E be a measurable subset of $\mathbb{R}^d$. Suppose that there exists some constant $C > 0$ and some index $p < q < \infty$ such that

$$\int_A |f| \;\le\; C\,|A|^{1/p'} \qquad \text{and} \qquad \int_A |f| \;\le\; C\,|A|^{1/q'}$$

for every measurable set $A \subseteq E$. Prove that $f \in L^r(E)$ for $p < r < q$.

7.2.23. Assume that $E \subseteq \mathbb{R}^d$ is measurable with $|E| = 1$, and fix $f \in L^1(E)$.

(a) Use Jensen's Inequality to prove that $\int_E \ln|f| \le \ln\|f\|_p$ for $0 < p < \infty$.

(b) Prove that $\lim_{p\to 0^+} \|f\|_p = \exp\Bigl(\int_E \ln|f|\Bigr)$.

7.2.24. Let E be a measurable subset of $\mathbb{R}^d$, and fix $1 \le p \le \infty$.

(a) Define a relation $\sim$ on $L^p(E)$ by declaring that $f \sim g$ if and only if $f = g$ a.e. Show that $\sim$ is an equivalence relation on $L^p(E)$.

(b) Let $[f]$ denote the equivalence class of f in $L^p(E)$ with respect to the relation $\sim$, i.e.,

$$[f] \;=\; \bigl\{g \in L^p(E) : g = f \text{ a.e.}\bigr\}.$$

Any particular function $g \in [f]$ is called a representative of the equivalence class $[f]$. Show that the quantity $|\!|\!|\,[f]\,|\!|\!|_p = \|g\|_p$ is independent of the choice of representative $g \in [f]$, i.e., $\|g\|_p = \|h\|_p$ for every choice of $g, h \in [f]$.

(c) Let $\widetilde{L^p}(E)$ be the *quotient space* of $L^p(E)$ with respect to $\sim$. That is, $\widetilde{L^p}(E) = \bigl\{[f] : f \in L^p(E)\bigr\}$ is the set of all distinct equivalence classes of functions in $L^p(E)$. Prove that $|\!|\!|\cdot|\!|\!|_p$ is a norm on $\widetilde{L^p}(E)$, and $\widetilde{L^p}(E)$ is a Banach space with respect to this norm.

7.3 Convergence in L^p-norm

We have seen that, once we identify functions that are equal a.e., $\|\cdot\|_p$ is a norm on $L^p(E)$. Convergence in $L^p(E)$ is, by definition, convergence with respect to that norm, which we spell out precisely in the next definition.

Definition 7.3.1 (Convergence in $L^p(E)$). Let E be a measurable subset of $\mathbb{R}^d$ and fix $1 \le p \le \infty$.

(a) A sequence $\{f_n\}_{n\in\mathbb{N}}$ in $L^p(E)$ *converges* to $f \in L^p(E)$ if

$$\lim_{n\to\infty} \|f - f_n\|_p \;=\; 0. \tag{7.23}$$

In this case we write $f_n \to f$ *in* $L^p(E)$, or $f_n \to f$ *in* L^p*-norm*, or for emphasis we may say that $f_n \to f$ *with respect to* $\|\cdot\|_p$.

(b) A sequence $\{f_n\}_{n\in\mathbb{N}}$ in $L^p(E)$ is *Cauchy in* L^p*-norm* if for every $\varepsilon > 0$ there exists some $N > 0$ such that

$$m, n \ge N \implies \|f_m - f_n\|_p < \varepsilon. \qquad \diamondsuit$$

The reader should verify that convergence in L^p-norm is well-defined, i.e., it is independent of the choice of representatives of f_n or f (and likewise for the definition of a Cauchy sequence).

Remark 7.3.2. If $p = \infty$ then equation (7.23) says that

$$\lim_{n\to\infty} \Bigl(\operatorname*{ess\,sup}_{x\in E} |f(x) - f_n(x)| \Bigr) = 0.$$

On the other hand, if p is finite then equation (7.23) is equivalent to

$$\lim_{n\to\infty} \int_E |f - f_n|^p = 0. \qquad \diamondsuit \tag{7.24}$$

For finite p, many of the facts that we proved about L^1-norm convergence have analogues for L^p-norm convergence. We list a few of these below.

Example 7.3.3. Fix $1 \le p < \infty$.

(a) Convergence in L^p-norm does not imply pointwise a.e. convergence in general. For example, the Boxes Marching in Circles from Example 3.5.5 converge to zero in L^p-norm, but they do not converge pointwise a.e.

(b) Pointwise a.e. convergence does not imply convergence in L^p-norm in general. For example, $f_n = n^{1/p}\,\chi_{[0,\frac{1}{n}]}$ converges pointwise a.e. to the zero function on $[0,\infty)$, but $\|f_n\|_p = 1$ for every n so f_n does not converge to zero in L^p-norm. $\diamondsuit$

Theorem 7.3.4. *Let $E \subseteq \mathbb{R}^d$ be a measurable set and fix $1 \le p \le \infty$. If $f_n, f \in L^p(E)$ and $\|f - f_n\|_p \to 0$, then $f_n \xrightarrow{\mathrm{m}} f$, and consequently there exists a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ such that $f_{n_k} \to f$ pointwise a.e.*

Proof. Tchebyshev's Inequality for L^p-norms is formulated in Problem 7.2.13, and convergence in measure follows from Tchebyshev's Inequality much like it does in the proof of Lemma 4.4.8. Then Lemma 3.5.6 implies the existence of a subsequence that converges pointwise a.e. $\square$

Figure 7.5 shows the main implications that hold between L^p-norm convergence and other types of convergence criteria.

The next exercise establishes that $L^p(E)$ is *complete*, i.e., all Cauchy sequences converge. The argument is similar to the one that we used to prove

$$\begin{array}{ccccccc}
 & & \text{pointwise a.e. convergence} & & L^p\text{-norm convergence} & & \\
 & & \Downarrow\ (\text{if } |E|<\infty) & & \Downarrow & & \\
L^\infty\text{-norm convergence} & \Longrightarrow & \text{almost uniform convergence} & \Longrightarrow & \text{convergence in measure} & \Longrightarrow & \text{pointwise a.e. convergence of a subsequence}\\
 & & \Downarrow & & & & \\
 & & \text{pointwise a.e. convergence} & & & &
\end{array}$$

Fig. 7.5 Relations among certain convergence criteria (valid for sequences of functions that are either complex-valued or extended real-valued but finite a.e.).

that ℓ^p is complete, but there are some complications due to the fact that convergence in measure only implies the existence of a *subsequence* that converges pointwise a.e. This exercise sketches one approach for the case of finite p; another approach is given in Problem 7.3.22.

Exercise 7.3.5. Let E be a measurable subset of $\mathbb{R}^d$ and fix $1 \le p < \infty$. Prove that if $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^p(E)$, then it is *Cauchy in measure* in the sense of Definition 3.5.9. Therefore, by applying Theorem 3.5.10 and Lemma 3.5.6, there exists a measurable function f such that $f_n \xrightarrow{\mathrm{m}} f$, and a subsequence such that $f_{n_k} \to f$ pointwise a.e. Show that $f \in L^p(E)$ and $\|f - f_{n_k}\|_p \to 0$, and finally that $f_n \to f$ in L^p-norm. $\diamondsuit$

For $p = \infty$, convergence in L^∞-norm implies almost uniform convergence, which implies pointwise a.e. convergence (however, pointwise a.e. convergence does not imply L^∞-norm convergence in general). The reader should use these facts to prove that $L^\infty(E)$ is also a complete space.

In summary, we have the following result (which some authors refer to as the *Riesz–Fischer Theorem*).

Theorem 7.3.6 ($L^p(E)$ is a Banach Space). *Let E be a measurable subset of $\mathbb{R}^d$ and fix $1 \le p \le \infty$. If we identify functions that are equal almost everywhere, then $\|\cdot\|_p$ is a norm on $L^p(E)$ and $L^p(E)$ is complete with respect to this norm.* $\diamondsuit$

7.3.1 Dense Subsets of $L^p(E)$

When trying to prove that some particular fact holds for all functions in $L^p(E)$, it is not unusual to find that it is easy to establish that the fact holds for some special subclass of functions, but it is not so clear how to prove it for arbitrary functions in $L^p(E)$. A standard technique in this situation is to try to extend the result from the "easy" class to the entire space by applying some type of approximation argument. Specifically, if every function in a class S has a certain property **P**, if S is dense in $L^p(E)$, and if we can show that property **P** is preserved under limits in L^p-norm, then we can conclude that every function in $L^p(E)$ has property **P**. We used this technique to prove several results about $L^1(E)$ in Section 4.5; now we consider $L^p(E)$.

The abstract definition of density was given in Definition 1.1.5. For convenience, we restate some equivalent formulations of density for the specific case of the L^p-norm as the following result.

Lemma 7.3.7 (Dense Subsets of $L^p(E)$). *Let $E \subseteq \mathbb{R}^d$ be measurable, and fix $1 \le p \le \infty$. If S is a subset of $L^p(E)$, then the following three statements are equivalent.*

(a) *S is dense in $L^p(E)$, i.e., the closure of S equals $L^p(E)$.*

(b) *If f is any element of $L^p(E)$, then there exist functions $f_n \in S$ such that $f_n \to f$ in L^p-norm.*

(c) *If f is any element of $L^p(E)$, then for each $\varepsilon > 0$ there exists a function $g \in S$ such that $\|f - g\|_p < \varepsilon$.* $\diamondsuit$

To illustrate, we will prove that the set of functions in $L^p(E)$ that are *compactly supported* is dense in $L^p(E)$ when p is finite. We do need to be careful about the meaning of "support" in this context. The support of a *continuous* function is the closure of the set where f is nonzero. This definition cannot literally be applied to elements of $L^p(E)$ because it depends on the choice of representative. For example, $\chi_{\mathbb{Q}}$ and the zero function are representatives of the same element of $L^p(\mathbb{R})$, but the closure of the set where $\chi_{\mathbb{Q}}$ is nonzero is $\mathbb{R}$, whereas the closure of the set where 0 is nonzero is the empty set. The precise definition of the support of an element of $L^p(E)$ is laid out in Problem 7.3.24, but for most purposes it is sufficient to declare, as we do next, that an element of $L^p(E)$ is compactly supported if it is zero a.e. outside of some compact set.

Definition 7.3.8 (Compact Support). Let $E \subseteq \mathbb{R}^d$ be a measurable set, and fix $1 \le p \le \infty$. We say that a function $f \in L^p(E)$ is *compactly supported* if there exists a compact set $K \subseteq \mathbb{R}^d$ such that $f(x) = 0$ for almost every $x \in E \backslash K$. $\diamondsuit$

The reader should check that Definition 7.3.8 does not depend on the choice of representative, i.e., if f is compactly supported and $g = f$ a.e., then g is also compactly supported. Using this notation, we will prove that the set of compactly supported functions in $L^p(E)$ is a dense subset of $L^p(E)$ when p is finite. This is simply another way of saying that every element of $L^p(E)$ can be approximated as closely as we like in L^p-norm by a compactly supported function (compare Lemma 4.5.4 for the case $p = 1$).

Theorem 7.3.9 (Compactly Supported Functions Are Dense). *Let $E \subseteq \mathbb{R}^d$ be a measurable set. If $1 \le p < \infty$, then*

$$L^p_c(E) \;=\; \bigl\{f \in L^p(E) : f \textit{ is compactly supported}\bigr\}$$

is dense in $L^p(E)$.

Proof. Choose $f \in L^p(E)$, and for each $n \in \mathbb{N}$ define $f_n = f \cdot \chi_{E \cap [-n,n]^d}$. Then $f - f_n \to 0$ pointwise a.e., and

$$|f - f_n|^p \;=\; |f \cdot \chi_{E \setminus [-n,n]^d}|^p \;\le\; |f|^p \;\in\; L^1(E).$$

The Dominated Convergence Theorem therefore implies that $|f - f_n|^p \to 0$ in L^1-norm, which is precisely the same as saying that $f_n \to f$ in L^p-norm. Since each f_n is compactly supported, we conclude that the set of compactly supported functions in $L^p(E)$ is dense in $L^p(E)$. $\square$

The conclusion of Theorem 7.3.9 can fail if $p = \infty$. For example, if $f = 1$ is the function that is identically 1, then $\|f - g\|_\infty \ge 1$ for every compactly supported function g. The constant function 1 cannot be well-approximated in L^∞-norm by compactly supported functions.

We give several exercises which establish that certain subsets are dense in $L^p(E)$. Additional density results appear in the problems for this section.

Exercise 7.3.10 (Simple Functions Are Dense). Assume that $E \subseteq \mathbb{R}^d$ is measurable and fix $1 \le p \le \infty$. Prove the following statements.

(a) The set S of all simple functions in $L^p(E)$ is is dense in $L^p(E)$.

(b) If p is finite, then the set S_c of all compactly supported simple functions on E is dense in $L^p(E)$. $\diamondsuit$

Exercise 7.3.11 (Continuous Functions Are Dense). The space $C_c(\mathbb{R}^d)$ consists of all continuous, compactly supported functions on $\mathbb{R}^d$. Prove the following statements.

(a) $C_c(\mathbb{R}^d)$ is dense in $L^p(\mathbb{R}^d)$ for $1 \le p < \infty$.

(b) With respect to the L^∞-norm, $C_c(\mathbb{R}^d)$ is dense in

$$C_0(\mathbb{R}^d) \;=\; \Bigl\{f \in C(\mathbb{R}^d) : \lim_{\|x\| \to \infty} f(x) = 0\Bigr\},$$

where the limit means that for each $\varepsilon > 0$ there exists some compact set K such that $|f(x)| < \varepsilon$ for all $x \notin K$. $\diamondsuit$

Exercise 7.3.12 (Really Simple Functions Are Dense). Fix $1 \le p < \infty$.

(a) Let $\mathcal{R}$ be the set of all really simple functions on $\mathbb{R}$,

$$\mathcal{R} = \left\{ \sum_{k=1}^{N} c_k \, \chi_{[a_k,b_k)} \; : \; N > 0, \, c_k \text{ scalar}, \, a_k < b_k \in \mathbb{R} \right\}.$$

Prove that $\mathcal{R}$ is dense in $L^p(\mathbb{R})$ when p is finite. $\diamondsuit$

Problems

7.3.13. Fix $1 < p \le \infty$. Show that there exist functions $f_n \in L^p[0,1]$ such that $\|f_n\|_1 = 1$ for every n but $\|f_n\|_p \to \infty$ as $n \to \infty$.

7.3.14. Suppose that $E \subseteq \mathbb{R}^d$ is measurable, and $1 \le p < q \le \infty$. Show that if $f_n \in L^p(E) \cap L^q(E)$, $f_n \to f$ in L^p-norm, and $f_n \to g$ in L^q-norm, then $f = g$ a.e.

7.3.15. Let E be a measurable subset of $\mathbb{R}^d$ and fix $1 \le p < \infty$.

(a) Given $f, g \in L^p(E)$, show that $2^p \left(|f|^p + |g|^p\right) - |f - g|^p \ge 0$ a.e.

(b) Suppose that $f_n, f \in L^p(E)$ and $f_n \to f$ a.e. Prove that $f_n \to f$ in L^p-norm if and only if $\|f_n\|_p \to \|f\|_p$.

7.3.16. Prove that if $1 \le p < \infty$ and $f \in L^p(\mathbb{R}^d)$, then $\lim_{a\to 0} \|T_a f - f\|_p = 0$, where $T_a f(x) = f(x - a)$.

7.3.17. Formulate a definition of "really simple functions" on $\mathbb{R}^d$, and prove that the really simple functions are dense in $L^p(\mathbb{R}^d)$, but not in $L^\infty(\mathbb{R}^d)$.

7.3.18. Let E be a measurable subset of $\mathbb{R}^d$. Prove that if $1 \le p < r < q \le \infty$, then $L^p(E) \cap L^q(E)$ is a dense subset of $L^r(E)$.

7.3.19. Fix $1 \le p < \infty$, and let $[a, b]$ be a closed bounded interval. Prove that the set $\mathcal{P}$ of all polynomials is dense in $L^p[a, b]$. What space is $\mathcal{P}$ dense in with respect to the L^∞-norm?

7.3.20. Fix $1 \le p < \infty$. For all $j, k \in \mathbb{Z}$, let $I_{jk} = \left[\frac{k}{2^j}, \frac{k+1}{2^j}\right)$ be a *dyadic interval* and let $\mathcal{D} = \left\{\chi_{I_{jk}} : j, k \in \mathbb{Z}\right\}$ be the set of all characteristic functions of dyadic intervals. Prove that $\operatorname{span}(\mathcal{D})$ is dense in $L^p(\mathbb{R})$.

7.3.21. Let $E \subseteq \mathbb{R}^d$ be measurable, and choose $1 < p < \infty$. Assume that functions $f_n \in L^p(E)$ satisfy $f_n \to f$ a.e. and $\sup \|f_n\|_p < \infty$. Prove that $f \in L^p(E)$, and for each $g \in L^{p'}(E)$ we have that $\lim_{n\to\infty} \int_E f_n g = \int_E fg$. Does the same result hold if $p = 1$?

7.3.22. Let E be a measurable subset of $\mathbb{R}^d$ and fix $1 \le p < \infty$.

(a) Suppose that $\sum f_n$ is an *absolutely convergent* series in $L^p(E)$, i.e., $f_n \in L^p(E)$ for $n \in \mathbb{N}$ and $\sum_{n=1}^\infty \|f_n\|_p < \infty$. Prove that:

- the series $f(x) = \sum_{n=1}^\infty f_n(x)$ converges for a.e. $x \in E$,
- $f \in L^p(E)$, and
- the series $f = \sum_{n=1}^\infty f_n$ converges in L^p-norm, i.e., $\displaystyle\lim_{N\to\infty} \Bigl\| f - \sum_{n=1}^N f_n \Bigr\|_p = 0.$

(b) Use part (a) and Theorem 1.2.8 to give another proof that $L^p(E)$ is complete with respect to $\|\cdot\|_p$.

(c) Show that if $\sum f_n$ is an absolutely convergent series in $L^1(E)$, then

$$\int_E \sum_{n=1}^\infty f_n \;=\; \sum_{n=1}^\infty \int_E f_n.$$

7.3.23. Fix $1 \le p < \infty$. Given $f_n \in L^p(\mathbb{R}^d)$, prove that $f_n \to f$ in $L^p(\mathbb{R}^d)$ if and only if the following three conditions hold.

(a) $f_n \xrightarrow{\mathrm{m}} f$.

(b) For each $\varepsilon > 0$ there exists a $\delta > 0$ such that for every measurable set $E \subseteq \mathbb{R}^d$ satisfying $|E| < \delta$ we have $\int_E |f_n|^p < \varepsilon$ for every n.

(c) For each $\varepsilon > 0$ there exists a measurable set $E \subseteq \mathbb{R}^d$ such that $|E| < \infty$ and $\int_{E^{\mathrm{C}}} |f_n|^p < \varepsilon$ for every n.

7.3.24. Define the *support* of a function $f \in L^p(\mathbb{R}^d)$ to be

$$\operatorname{supp}(f) \;=\; \bigcap \bigl\{F \subseteq \mathbb{R}^d : F \text{ is closed and } f(x) = 0 \text{ for a.e. } x \notin F\bigr\}.$$

Prove the following statements.

(a) $\operatorname{supp}(f)$ does not depend on the choice of representative of f, i.e., if $f = g$ a.e., then $\operatorname{supp}(f) = \operatorname{supp}(g)$.

(b) f is compactly supported in the sense of Definition 7.3.8 if and only if $\operatorname{supp}(f)$ is compact.

(c) If f is continuous, then $\operatorname{supp}(f)$ coincides with the usual definition of the support of f (the closure of $\{f \ne 0\}$).

7.3.25. Let E be a measurable subset of $\mathbb{R}^d$ and fix $1 \le p < q \le \infty$. Prove the following statements.

(a) $\|f\| = \|f\|_p + \|f\|_q$ is a norm on $L^p(E) \cap L^q(E)$, and $L^p(E) \cap L^q(E)$ is a Banach space with respect to this norm.

(b) If $1 \le p < r < q \le \infty$, then $L^p(E) \cap L^q(E) \subseteq L^r(E)$ and

$$\|f\|_r \;\le\; \|f\|_p^\theta \, \|f\|_q^{1-\theta}, \quad \text{where } \frac{\theta}{p} + \frac{1-\theta}{q} = \frac{1}{r}.$$

7.3.26. Let $E \subseteq \mathbb{R}^d$ be a measurable set such that $|E| < \infty$, and let $\mathcal{M}(E)$ be the vector space of all Lebesgue measurable functions $f\colon E \to \overline{\mathbf{F}}$ that are finite a.e. Show that if we identify functions in $\mathcal{M}(E)$ that are equal almost everywhere, then the following statements hold.

(a) $\mathrm{d}(f,g) = \displaystyle\int_E \frac{|f(x)-g(x)|}{1+|f(x)-g(x)|}\,dx$ defines a metric on $\mathcal{M}(E)$.

(b) The convergence criterion induced by the metric d is convergence in measure, i.e., $f_k \overset{\mathrm{m}}{\to} f$ if and only if $\lim_{k\to\infty} \mathrm{d}(f, f_k) = 0$.

(c) $\mathcal{M}(E)$ is complete with respect to the metric d, i.e., if $\{f_k\}_{k\in\mathbb{N}}$ is a sequence that is Cauchy with respect to the metric d, then there exists some $f \in \mathcal{M}(E)$ such that $f_k \overset{\mathrm{m}}{\to} f$.

7.4 Separability of $L^p(E)$

We will prove in this section that $L^p(E)$ is *separable* when p is finite. To motivate the definition of separability, recall that although the set of rationals is a countable set and hence is "small" in terms of cardinality, it is a "large" subset of $\mathbb{R}$ in the topological sense, since $\mathbb{Q}$ is dense in $\mathbb{R}$. In higher dimensions, the set $\mathbb{Q}^d$ consisting of vectors with rational components is a countable yet dense subset of $\mathbb{R}^d$. It may seem unlikely that an infinite-dimensional space could contain a countable, dense subset, yet we will see that this is true of $L^p(E)$ when p is finite. In contrast, we will show that $L^\infty(E)$ does not contain a countable dense subset (unless $|E| = 0$). Loosely speaking, a nonseparable space is "much larger" than a separable space. We recall the precise definition from Section 1.1.2.

Definition 7.4.1 (Separable Space). A metric space that contains a countable dense subset is said to be *separable*. $\diamondsuit$

To show that $L^p(\mathbb{R})$ is separable when p is finite, let S be the set of all characteristic functions of the form $\chi_{[a,b)}$,

$$S = \{\chi_{[a,b)} : -\infty < a < b < \infty\},$$

and let $\mathcal{R}$ be its finite linear span, which is the set of all really simple functions:

$$\mathcal{R} = \mathrm{span}(S) = \left\{\sum_{k=1}^{N} c_k \chi_{[a_k,b_k)} : N > 0,\ c_k \text{ scalar},\ a_k < b_k \in \mathbb{R}\right\}.$$

Exercise 7.3.12 showed that $\mathcal{R}$ is dense in $L^p(\mathbb{R})$. However, $\mathcal{R}$ is an uncountable set. Can we find a *countable* subset of $\mathcal{R}$ that is still dense? To do this, let $S_\mathbb{Q}$ be the subset of S that consists of characteristic functions of intervals whose endpoints are rational:

$$S_\mathbb{Q} \;=\; \bigl\{\chi_{[a,b)} : a < b \in \mathbb{Q}\bigr\}.$$

This set is a countable, but it is not dense. We could consider the span of $S_\mathbb{Q}$, but that is uncountable because it contains all possible finite linear combinations of elements of $S_\mathbb{Q}$. Therefore, we instead consider the "rational span," which is the set of all finite linear combinations that only employ rational scalars. Recalling that we say that a complex scalar is rational if both its real and imaginary parts are rational, this rational span is

$$\mathcal{R}_\mathbb{Q} \;=\; \biggl\{\sum_{k=1}^{N} c_k\, \chi_{[a_k,b_k)} \,:\, N > 0,\ c_k \text{ is rational},\ a_k < b_k \in \mathbb{Q}\biggr\}.$$

We will prove that $\mathcal{R}_\mathbb{Q}$ is dense in $L^p(\mathbb{R})$. This implies that $L^p(\mathbb{R})$ is separable (alternative approaches are given in Problems 7.3.19 and 7.3.20).

Theorem 7.4.2 (Separability of $L^p(\mathbb{R})$). *If $1 \le p < \infty$, then $L^p(\mathbb{R})$ contains a countable dense subset and therefore is separable.*

Proof. Choose any $f \in L^p(\mathbb{R})$ and fix $\varepsilon > 0$. By Exercise 7.3.12, there exists a really simple function $g = \sum_{k=1}^N t_k\, \chi_{[c_k,d_k)} \in \mathcal{R}$ such that $\|f - g\|_p < \varepsilon$. Without loss of generality, we may assume that $t_k \ne 0$ for each k. Choose rational real numbers a_k, b_k with $a_k < c_k$ and $b_k > d_k$ such that

$$c_k - a_k \;<\; \frac{1}{2}\biggl(\frac{\varepsilon}{N\,|t_k|}\biggr)^{p} \qquad\text{and}\qquad b_k - d_k \;<\; \frac{1}{2}\biggl(\frac{\varepsilon}{N\,|t_k|}\biggr)^{p}.$$

Now choose rational scalars r_k such that

$$|t_k - r_k| \;<\; \frac{\varepsilon}{N\,(b_k-a_k)^{1/p}}.$$

Then the function $h = \sum_{k=1}^N r_k\, \chi_{[a_k,b_k)}$ belongs to $\mathcal{R}_\mathbb{Q}$, and we compute that

$$\begin{aligned}
&\|t_k\, \chi_{[c_k,d_k)} - r_k\, \chi_{[a_k,b_k)}\|_p \\
&\quad\le\; \|t_k\, \chi_{[c_k,d_k)} - t_k\, \chi_{[a_k,b_k)}\|_p \;+\; \|t_k\, \chi_{[a_k,b_k)} - r_k\, \chi_{[a_k,b_k)}\|_p \\
&\quad=\; |t_k|\, \|\chi_{[a_k,b_k)\setminus[c_k,d_k)}\|_p \;+\; |t_k - r_k|\, \|\chi_{[a_k,b_k)}\|_p \\
&\quad=\; |t_k|\, \bigl((c_k - a_k) + (b_k - d_k)\bigr)^{1/p} \;+\; |t_k - r_k|\, (b_k - a_k)^{1/p} \\
&\quad\le\; |t_k| \biggl(\frac{1}{2}\biggl(\frac{\varepsilon}{N\,|t_k|}\biggr)^{p} + \frac{1}{2}\biggl(\frac{\varepsilon}{N\,|t_k|}\biggr)^{p}\biggr)^{1/p} \;+\; \frac{\varepsilon}{N\,(b_k-a_k)^{1/p}}\,(b_k - a_k)^{1/p} \\
&\quad=\; \frac{\varepsilon}{N} + \frac{\varepsilon}{N} \;=\; \frac{2\varepsilon}{N}.
\end{aligned}$$

Therefore

$$\begin{aligned} \|g-h\|_p &= \Bigl\|\sum_{k=1}^{N}\Bigl(t_k\chi_{[c_k,d_k)} - r_k\chi_{[a_k,b_k)}\Bigr)\Bigr\|_p \\ &\le \sum_{k=1}^{N}\|t_k\chi_{[c_k,d_k)} - r_k\chi_{[a_k,b_k)}\|_p \le 2\varepsilon, \end{aligned}$$

and consequently

$$\|f-h\|_p \le \|f-g\|_p + \|g-h\|_p \le 3\varepsilon.$$

Thus $\mathcal{R}_{\mathbb{Q}}$ is dense in $L^p(\mathbb{R})$. Since $\mathcal{R}_{\mathbb{Q}}$ is countable, this shows that $L^p(\mathbb{R})$ is separable. $\square$

As a corollary, we prove that $L^p(E)$ is separable for all measurable $E \subseteq \mathbb{R}$.

Corollary 7.4.3 (Separability of $L^p(E)$). *Let $E \subseteq \mathbb{R}$ be measurable, and fix $1 \le p < \infty$. If $\mathcal{S}$ is any countable, dense subset of $L^p(\mathbb{R})$, then*

$$\mathcal{S}(E) = \bigl\{f\cdot\chi_E : f \in \mathcal{S}\bigr\}$$

is a countable dense subset of $L^p(E)$. Consequently, $L^p(E)$ contains a countable dense subset, and therefore is separable.

Proof. Choose any $f \in L^p(E)$, and fix $\varepsilon > 0$. Extend f to all of $\mathbb{R}$ by setting $f(x) = 0$ for $x \notin E$. Then $f \in L^p(\mathbb{R})$, so there exists a function $g \in \mathcal{S}$ such that $\|f-g\|_{L^p(\mathbb{R})} < \varepsilon$. But then $h = g\cdot\chi_E$ belongs to $\mathcal{S}(E)$, and it satisfies

$$\|f-h\|_{L^p(E)}^p = \int_E |f-h|^p = \int_{\mathbb{R}} |f-g|^p = \|f-g\|_{L^p(\mathbb{R})}^p < \varepsilon^p.$$

Hence $\mathcal{S}(E)$ is a countable, dense subset of $L^p(E)$. $\square$

Extensions of Theorem 7.4.2 and Corollary 7.4.3 to higher dimensions are given in Problem 7.4.10.

The situation for $p = \infty$ is quite different. To motivate this, note that in $\mathbb{R}^d$ we can find up to $d+1$ vectors that are each unit distance from each other (for example, consider the three vertices of an equilateral triangle in $\mathbb{R}^2$, or the four vertices of a regular tetrahedron in $\mathbb{R}^3$). Not surprisingly, in an infinite-dimensional normed space we can find infinitely many vectors such that any pair are at least unit distance apart. However, in some spaces we can find only *countably* many such vectors, while in others we can find *uncountably* many. The following result shows that any metric space that contains uncountably many "separated" elements must be nonseparable.

Theorem 7.4.4. *If X is a metric space and there exists an uncountable set $\mathcal{A} \subseteq X$ such that $\mathrm{d}(x,y) \ge 1$ for every $x \ne y \in \mathcal{A}$, then X is not separable.*

Proof. Let S be a dense subset of X. If we choose any point $t \in \mathcal{A}$ then, since S is dense, there must exist some $x_t \in S$ such that $\|t - x_t\|_\infty < \frac{1}{2}$. Consequently, if $y \neq z \in \mathcal{A}$, then

$$1 \leq \mathrm{d}(y,z) \leq \mathrm{d}(y,x_y) + \mathrm{d}(x_y,x_z) + \mathrm{d}(x_z,z) < \frac{1}{2} + \mathrm{d}(x_y,x_z) + \frac{1}{2}.$$

Therefore $\mathrm{d}(x_y,x_z) > 0$, which tells us that x_y and x_z are distinct elements of S. Hence $t \mapsto x_t$ is an injective mapping from $\mathcal{A}$ into S, so S must be uncountable. $\square$

We will use Theorem 7.4.4 to show that $L^\infty(\mathbb{R})$ is nonseparable. If we set $f_a = \chi_{[a,a+1]}$ for $a \in \mathbb{R}$, then $\|f_a - f_b\|_\infty = 1$ whenever $a \neq b$ (see Figure 7.6). Therefore $\{f_a\}_{a\in\mathbb{R}}$ is an uncountable separated family in $L^\infty(\mathbb{R})$, so Theorem 7.4.4 implies that $L^\infty(\mathbb{R})$ is nonseparable. The same is true of $L^\infty(E)$ for any measurable set $E \subseteq \mathbb{R}^d$ that has positive measure, although it takes a bit more work to construct an uncountable "separated" family for a generic set E (this is Problem 7.4.10).

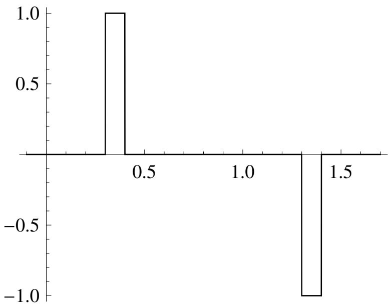


Fig. 7.6 Graph of $f_a - f_b$ for $f_a = \chi_{[0.3,1.3]}$ and $f_b = \chi_{[0.4,1.4]}$. Note that $f_a - f_b = \pm 1$ on a set with positive measure, and hence $\|f_a - f_b\|_\infty = 1$.

Problems

7.4.5. Fix $1 \leq p \leq \infty$. Suppose that $f \in L^p(\mathbb{R})$ and $\int_\mathbb{R} f\phi = 0$ for all $\phi \in C_c(\mathbb{R})$. Prove that $f = 0$ a.e.

7.4.6. Let X be a normed space, and suppose that there exists a countable sequence $\mathcal{F} = \{x_n\}_{n\in\mathbb{N}}$ in X such that

$$\mathrm{span}(\mathcal{F}) = \left\{ \sum_{n=1}^{N} c_n x_n \,:\, N \in \mathbb{N},\ c_n \text{ scalar} \right\}$$

is dense in X (such a sequence is said to be *complete* in X, see Definition 8.2.17). Prove that the rational span of $\mathcal{F}$,

$$S = \left\{\sum_{n=1}^{N} r_n x_n \,:\, N \in \mathbb{N},\ r_n \text{ rational}\right\},$$

is a countable, dense subset of X, and therefore X is separable.

7.4.7. Prove the following statements.

(a) c_0 is separable (with respect to the sup-norm), but ℓ^∞ is not separable.

(b) ℓ^p is separable for $1 \le p < \infty$.

(c) If I is an uncountable index set and $\ell^p(I)$ is the space defined in Problem 7.1.27, then $\ell^p(I)$ is nonseparable for every p.

7.4.8. Use Problems 7.3.19 and 7.3.20 to prove that $L^p[a,b]$ and $L^p(\mathbb{R})$ are separable.

7.4.9. Prove that $C[a,b]$ and $C_0(\mathbb{R})$ are separable (with respect to the uniform norm).

7.4.10. Given a measurable set $E \subseteq \mathbb{R}^d$ such that $|E| > 0$, prove that $L^p(E)$ is separable for $1 \le p < \infty$, but $L^\infty(E)$ is not separable.

7.4.11. A sequence $\{x_n\}_{n\in\mathbb{N}}$ is a *Schauder basis* for a Banach space X if for each vector $x \in X$ there exist *unique* scalars $c_n(x)$ such that

$$x = \sum_{n=1}^{\infty} c_n(x)\, x_n, \tag{7.25}$$

where this series converges in the norm of X. Prove the following statements.

(a) If $1 \le p < \infty$ then the standard basis $\mathcal{E} = \{\delta_n\}_{n\in\mathbb{N}}$ is a Schauder basis for ℓ^p.

(b) The standard basis $\mathcal{E} = \{\delta_n\}_{n\in\mathbb{N}}$ is a Schauder basis for c_0 (with respect to the sup-norm), but it is not a Schauder basis for ℓ^∞.

(c) $\{y_n\}_{n\in\mathbb{N}}$, where $y_n = (1,\dots,1,0,0,\dots)$, is a Schauder basis for c_0.

(d) If $\{x_n\}_{n\in\mathbb{N}}$ is a Schauder basis for a Banach space X, then $\{x_n\}_{n\in\mathbb{N}}$ is finitely linearly independent and $\operatorname{span}\{x_n\}_{n\in\mathbb{N}}$ is dense in X. Apply Problem 7.4.6 and conclude that X is separable.

(e) The set of monomials $\mathcal{M} = \{1, x, x^2, \dots\}$ is not a Schauder basis for the Banach space $C[0,1]$ (with respect to the uniform norm), but it is linearly independent and $\operatorname{span}(\mathcal{M})$ is dense in $C[0,1]$.

(f)* Can you construct a Schauder basis for $C[0,1]$ or $L^p[0,1]$?

Chapter 8
Hilbert Spaces and $L^2(E)$

We will see in this chapter that $L^2(E)$ holds a special place among the Lebesgue spaces $L^p(E)$, because the norm on $L^2(E)$ is *induced from an inner product*. An inner product allows us to determine the angle between vectors, not just the distance between them. Once we have angles, we have a notion of orthogonality, and from this we can define orthogonal projections and orthonormal bases. This provides us with an extensive set of tools for analyzing $L^2(E)$ (and ℓ^2) that are not available to us when $p \neq 2$.

We introduce inner products in an abstract setting in Section 8.1, and examine orthogonality in detail in Section 8.2. In Section 8.3 we prove that every separable Hilbert space has an *orthonormal basis*, which provides convenient, stable representations of vectors in the space. We construct some examples of orthonormal bases for $L^2[0,1]$ and $L^2(\mathbb{R})$ in that section, then examine in detail the *trigonometric system* (which is the basis for Fourier series) in Section 8.4.

8.1 Inner Products and Hilbert Spaces

In a normed vector space, each vector has an assigned *length*, and from this we obtain the *distance* from x to y as the length of the vector $x - y$. For vectors in $\mathbb{R}^d$ or $\mathbb{C}^d$ we also know how to measure the *angle* between vectors; in particular, two vectors x and y in Euclidean space are perpendicular, or *orthogonal*, if their dot product is zero. In this section we will study vector spaces that have an *inner product*, which is a generalization of the dot product. Using the inner product, we can develop the notion of orthogonality in abstract spaces.

C. Heil, *Introduction to Real Analysis*, Graduate Texts in Mathematics 280,
https://doi.org/10.1007/978-3-030-26903-6_8

8.1.1 The Definition of an Inner Product

Here are the defining properties of an inner product (recall that in this text we always take the scalar field associated with a vector space to be either the real line $\mathbb{R}$ or the complex plane $\mathbb{C}$).

Definition 8.1.1 (Semi-Inner Product, Inner Product). Let H be a vector space. A *semi-inner product* on H is a scalar-valued function $\langle\cdot,\cdot\rangle$ on $H \times H$ such that for all vectors x, y, $z \in H$ and all scalars a and b we have:

(a) Nonnegativity: $0 \le \langle x,x\rangle < \infty$,

(b) Conjugate Symmetry: $\langle x,y\rangle = \overline{\langle y,x\rangle}$, and

(c) Linearity in the First Variable: $\langle ax+by,z\rangle = a\,\langle x,z\rangle + b\,\langle y,z\rangle$.

If a semi-inner product $\langle\cdot,\cdot\rangle$ also satisfies:

(d) Uniqueness: $\langle x,x\rangle = 0$ if and only if $x = 0$,

then it is an *inner product* on H. In this case, H is called an *inner product space* or a *pre-Hilbert space.* ◇

The usual dot product

$$u \cdot v = u_1\overline{v_1} + \cdots + u_d\overline{v_d} \tag{8.1}$$

is an inner product on $\mathbb{R}^d$ or $\mathbb{C}^d$ (of course, on $\mathbb{R}^d$ the complex conjugate in equation (8.1) is superfluous; similarly, if H is a real vector space then the complex conjugate in the definition of conjugate symmetry is irrelevant).

If $\langle\cdot,\cdot\rangle$ is a semi-inner product on a vector space H, then for each $x \in H$ we set

$$\|x\| = \langle x,x\rangle^{1/2}.$$

By definition, $\|x\|$ is a nonnegative, finite real number. We will prove in Lemma 8.1.4 that $\|\cdot\|$ is a seminorm on H, and therefore we refer to $\|\cdot\|$ as the *seminorm induced by* $\langle\cdot,\cdot\rangle$. Likewise, we will see that if $\langle\cdot,\cdot\rangle$ is an inner product then $\|\cdot\|$ is a norm, so in this case we call $\|\cdot\|$ the *norm induced by* $\langle\cdot,\cdot\rangle$. It may be possible to place other norms on H, but unless we explicitly state otherwise, we assume that all norm-related statements on an inner product space are taken with respect to this induced norm.

8.1.2 Properties of an Inner Product

The following exercise gives some basic properties of an inner product.

Exercise 8.1.2. Prove that if $\langle\cdot,\cdot\rangle$ is a semi-inner product on a vector space H, then the following statements hold for all vectors x, y, $z \in H$ and all scalars a and b.

(a) Antilinearity in the Second Variable: $\langle x, ay+bz\rangle = \overline{a}\,\langle x,y\rangle + \overline{b}\,\langle x,z\rangle$.

(b) Polar Identity: $\|x+y\|^2 = \|x\|^2 + 2\,\mathrm{Re}\langle x,y\rangle + \|y\|^2$.

(c) Pythagorean Theorem: If $\langle x,y\rangle = 0$, then $\|x \pm y\|^2 = \|x\|^2 + \|y\|^2$.

(d) Parallelogram Law: $\|x+y\|^2 + \|x-y\|^2 = 2\big(\|x\|^2 + \|y\|^2\big)$. $\diamondsuit$

The next inequality is known by several names, including the *Schwarz Inequality*, the *Cauchy–Schwarz Inequality*, and the *Cauchy–Bunyakovski–Schwarz Inequality* (or simply the *CBS Inequality*).

Theorem 8.1.3 (Cauchy–Bunyakovski–Schwarz Inequality). *If $\langle\cdot,\cdot\rangle$ is a semi-inner product on a vector space H, then*

$$|\langle x,y\rangle| \le \|x\|\,\|y\|, \qquad \textit{for all } x,y \in H.$$

Proof. Assume that x and y are both nonzero, and let α be a scalar with modulus 1 such that $\langle x,y\rangle = \alpha\,|\langle x,y\rangle|$. Then for each $t \in \mathbb{R}$, by using the Polar Identity and antilinearity in the second variable, we compute that

$$\begin{aligned} 0 \le \|x-\alpha t y\|^2 &= \|x\|^2 - 2\,\mathrm{Re}\big(\langle x, \alpha t y\rangle\big) + t^2\,\|y\|^2 \\ &= \|x\|^2 - 2t\,\mathrm{Re}\big(\overline{\alpha}\,\langle x,y\rangle\big) + t^2\,\|y\|^2 \\ &= \|x\|^2 - 2t\,|\langle x,y\rangle| + t^2\,\|y\|^2 \\ &= at^2 + bt + c, \end{aligned}$$

where $a = \|y\|^2$, $b = -2\,|\langle x,y\rangle|$, and $c = \|x\|^2$. This is a real-valued quadratic polynomial in the variable t. Since this polynomial is nonnegative, it can have at most one real root. This implies that the discriminant $b^2 - 4ac$ cannot be strictly positive. Hence

$$b^2 - 4ac = \big(-2\,|\langle x,y\rangle|\big)^2 - 4\,\|x\|^2\,\|y\|^2 \le 0,$$

and the result follows by rearranging this inequality. $\square$

By combining the Polar Identity with the Cauchy–Bunyakovski–Schwarz Inequality, we can now prove that the induced seminorm $\|\cdot\|$ is indeed a seminorm on H.

Lemma 8.1.4. *Let H be a vector space. If $\langle\cdot,\cdot\rangle$ is a semi-inner product on H, then $\|\cdot\|$ is a seminorm on H, and if $\langle\cdot,\cdot\rangle$ is an inner product on H, then $\|\cdot\|$ is a norm on H.*

Proof. The only property that is not obvious is the Triangle Inequality. To prove this, we compute that

$$\begin{aligned}
\|x+y\|^2 &= \|x\|^2 + 2\,\mathrm{Re}\langle x,y\rangle + \|y\|^2 && \text{(Polar Identity)}\\
&\le \|x\|^2 + 2\,|\langle x,y\rangle| + \|y\|^2 && \text{($|\mathrm{Re}(z)| \le |z|$ for all scalars z)}\\
&\le \|x\|^2 + 2\,\|x\|\,\|y\| + \|y\|^2 && \text{(CBS Inequality)}\\
&= \big(\|x\| + \|y\|\big)^2. \qquad \square
\end{aligned}$$

Since the induced norm is a norm, all of the definitions and properties derived for norms in Chapter 1 apply to the induced norm. In particular, we have notions of convergence for sequences and infinite series. These can be used to derive the following further properties of inner products.

Exercise 8.1.5. Given an inner product space H, prove that the following statements hold.

(a) Continuity of the inner product: If $x_n \to x$ and $y_n \to y$ in H, then $\langle x_n, y_n\rangle \to \langle x, y\rangle$.

(b) If the series $\sum_{n=1}^{\infty} x_n$ converges in H, then

$$\Big\langle \sum_{n=1}^{\infty} x_n,\, y \Big\rangle = \sum_{n=1}^{\infty} \langle x_n, y\rangle, \qquad \text{for all } y \in H. \qquad \diamondsuit \tag{8.2}$$

Since an infinite series is a limit of the partial sums of the series, both the linearity of the inner product in the first variable *and* the continuity of the inner product are required to justify equation (8.2).

8.1.3 Hilbert Spaces

Just as in metric or normed spaces, in any inner product space it is important to know whether all Cauchy sequences converge. We give the following name to those inner product spaces that have this property.

Definition 8.1.6 (Hilbert Space). An inner product space H is called a *Hilbert space* if it is complete with respect to the induced norm. $\diamondsuit$

That is, an inner product space is a Hilbert space if and only if every Cauchy sequence in H converges to an element of H. Equivalently, a Hilbert space is an inner product space that is a Banach space with respect to its induced norm. For example, $\mathbb{R}^d$ and $\mathbb{C}^d$ are Hilbert spaces with respect to the usual dot product given in equation (8.1). We will show that ℓ^2 and $L^2(E)$ are also Hilbert spaces with respect to an appropriate inner product.

Example 8.1.7 (The ℓ^2-Inner Product). Recall that ℓ^2 is the space of all square-summable sequences of scalars. We proved in Section 7.1 that ℓ^2 is a Banach space with respect to the ℓ^2-norm. Now we will define an inner product on ℓ^2. By Hölder's Inequality, if $x = (x_k)_{k\in\mathbb{N}}$ and $y = (y_k)_{k\in\mathbb{N}}$ belong to ℓ^2, then

$$\sum_{k=1}^{\infty} |x_k y_k| \;\le\; \biggl(\sum_{k=1}^{\infty} |x_k|^2\biggr)^{1/2} \biggl(\sum_{k=1}^{\infty} |y_k|^2\biggr)^{1/2} \;=\; \|x\|_2\,\|y\|_2 \;<\; \infty. \tag{8.3}$$

Consequently, we can set

$$\langle x, y\rangle \;=\; \sum_{k=1}^{\infty} x_k\,\overline{y_k}, \tag{8.4}$$

because this is an absolutely convergent series of scalars. The reader should check that equation (8.4) defines an inner product on ℓ^2. We have

$$\langle x, x\rangle \;=\; \sum_{k=1}^{\infty} x_k\,\overline{x_k} \;=\; \sum_{k=1}^{\infty} |x_k|^2 \;=\; \|x\|_2^2,$$

so the norm induced from this inner product is precisely the ℓ^2-norm. Since we already know that ℓ^2 is complete with respect to this norm, we conclude that ℓ^2 is a Hilbert space with respect to this inner product. ♢

Example 8.1.8 (The L^2-Inner Product). Let E be a measurable subset of $\mathbb{R}^d$. The space $L^2(E)$ consists of all square-integrable functions on E (see Definition 7.2.1). If f and g belong to $L^2(E)$, then Hölder's Inequality implies that $f\overline{g}$ is integrable, so we can define

$$\langle f, g\rangle \;=\; \int_E f(x)\,\overline{g(x)}\,dx. \tag{8.5}$$

The reader can check that this defines an inner product on $L^2(E)$ (when we identify functions that are equal a.e.). The norm induced from this inner product is the L^2-norm $\|\cdot\|_2$. Since we know that $L^2(E)$ is complete with respect to this norm, it follows that $L^2(E)$ is a Hilbert space with respect to the inner product defined in equation (8.5). ♢

There are inner products on ℓ^2 or $L^2(E)$ other than the ones given above, but unless we explicitly state otherwise, we always assume that the inner products on ℓ^2 or $L^2(E)$ are the ones specified in equations (8.4) and (8.5).

Problems

8.1.9. Let $\langle\cdot,\cdot\rangle$ be a semi-inner product on a vector space H. Show that equality holds in the Cauchy–Bunyakovski–Schwarz Inequality if and only if there exist scalars α and β, not both zero, such that $\|\alpha x + \beta y\| = 0$. In particular, if $\langle\cdot,\cdot\rangle$ is an inner product, then either $x = cy$ or $y = cx$ where c is a scalar.

8.1.10. Let H be a Hilbert space. Given vectors x_n and x in H, we say that x_n *converges weakly to* x if $\langle x_n, y\rangle \to \langle x, y\rangle$ for every $y \in H$. Prove that $x_n \to x$ (convergence in norm) if and only if x_n converges weakly to x and $\|x_n\| \to \|x\|$.

8.1.11. Let E be a measurable subset of $\mathbb{R}^d$ such that $|E| > 0$. Show that if $1 \le p \le \infty$ and $p \ne 2$, then $\|\cdot\|_p$ does not satisfy the Parallelogram Law. Consequently, the norm on $L^p(E)$ is not induced from any inner product, i.e., there is no inner product $\langle\cdot,\cdot\rangle$ on $L^p(E)$ such that $\langle f, f\rangle = \|f\|_p^2$ for all $f \in L^p(E)$.

8.1.12. Suppose that f is positive and monotone increasing on $(0,\infty)$, and $f \in \mathrm{AC}[a,b]$ for every finite interval $[a,b]$. Suppose that there is a constant $C > 0$ such that $f(x) \le Cx^2$ for all $x > 0$. Prove that $\int_0^\infty 1/f' = \infty$.

8.1.13. Let H be the set of all absolutely continuous functions $f \in \mathrm{AC}[a,b]$ such that $f' \in L^2[a,b]$. Prove that H is a Hilbert space with respect to the inner product $\langle f, g\rangle = \int_a^b f(x)\,\overline{g(x)}\,dx + \int_a^b f'(x)\,\overline{g'(x)}\,dx$.

8.1.14. This problem will establish a special case of *Hardy's Inequalities.* Prove that if $f \in L^2[0,\infty)$, then

$$\left|\int_0^x f(t)\,dt\right|^2 \;\le\; 2x^{1/2}\int_0^x t^{1/2}\,|f(t)|^2\,dt, \qquad \text{for } x \ge 0.$$

Then define $F(x) = \frac{1}{x}\int_0^x f(t)\,dt$ for $x \ge 0$, and show that $F \in L^2[0,\infty)$ and $\|F\|_2 \le 2\,\|f\|_2$.

8.1.15. Assume that $f \in L^2(\mathbb{R}^d)$ and $g \in L^1(\mathbb{R}^d)$ are both nonnegative.

(a) Use Tonelli's Theorem to prove that the convolution $(f * g)(x) = \int f(y)\,g(x-y)\,dy$ exists a.e. and is measurable.

(b) Apply the CBS Inequality with factors $f(y)\,g(x-y)^{1/2}$ and $g(x-y)^{1/2}$ to prove that

$$|(f*g)(x)| \;\le\; \|g\|_1 \int_{\mathbb{R}^d} |f(y)|^2\,|g(x-y)|\,dy,$$

and from this show that $f * g \in L^2(\mathbb{R}^d)$ and $\|f*g\|_2 \le \|f\|_2\,\|g\|_1$.

(c) Prove that parts (a) and (b) hold for all functions $f \in L^2(\mathbb{R}^d)$ and $g \in L^1(\mathbb{R}^d)$, nonnegative or not.

8.1.16. Let $f_n, f \in L^2[a,b]$ be given, and for $x \in [a,b]$ define

$$F_n(x) \;=\; \int_a^x f_n(t)\,dt \qquad \text{and} \qquad F(x) \;=\; \int_a^x f(t)\,dt.$$

Prove the following statements.

(a) If $f_n \to f$ in L^2-norm, then $F_n \to F$ uniformly.

(b) F_n and F are Hölder continuous with exponent $1/2$.

(c) If f_n converges weakly to f in the sense of Problem 8.1.10 and if $\sup \|f_n\|_2 < \infty$, then $F_n \to F$ uniformly.

Remark: In fact, all weakly convergent sequences are bounded (for one proof, see [Heil11, Thm. 2.38]), so the assumption in part (c) that $\sup \|f_n\|_2$ is finite is redundant.

8.2 Orthogonality

The existence of a notion of orthogonality gives inner product spaces a much richer and more tractable structure than generic Banach spaces, and leads to many elegant results that have natural, constructive proofs. We will derive some of these in this chapter. First we define orthogonal vectors.

Definition 8.2.1. Let H be an inner product space, and let I be an arbitrary index set.

(a) Two vectors $x, y \in H$ are *orthogonal*, denoted $x \perp y$, if $\langle x, y\rangle = 0$.

(b) A sequence of vectors $\{x_i\}_{i\in I}$ is *orthogonal* if $\langle x_i, x_j\rangle = 0$ whenever $i \neq j$.

(c) A sequence of vectors $\{x_i\}_{i\in I}$ is *orthonormal* if it is orthogonal and each vector x_i is a unit vector. Using the Kronecker delta notation, $\{x_i\}_{i\in I}$ is an orthonormal set if for all $i, j \in I$ we have

$$\langle x_i, x_j\rangle \;=\; \delta_{ij} \;=\; \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \qquad \diamondsuit$$

For example, the sequence of standard basis vectors $\mathcal{E} = \{\delta_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence in ℓ^2.

The zero vector may be an element of a sequence of orthogonal vectors. Any orthogonal sequence $\{x_i\}_{i\in I}$ of nonzero vectors can be rescaled to form an orthonormal sequence, simply by dividing each vector by its length.

If $\{x_n\}_{n\in\mathbb{N}}$ is a countable sequence of linearly independent, but not necessarily orthogonal, vectors then the *Gram–Schmidt orthonormalization procedure* that we will describe in Section 8.3.5 can be used to construct an orthonormal sequence $\{e_n\}_{n\in\mathbb{N}}$ such that $\mathrm{span}\{x_1, \dots, x_k\} = \mathrm{span}\{e_1, \dots, e_k\}$ for every k.

The following lemma will be useful to us later.

Lemma 8.2.2. *Let x and y be vectors in an inner product space H. Then*

$$x \perp y \quad \Longleftrightarrow \quad \|x\| \le \|x + \lambda y\| \text{ for every scalar } \lambda.$$

Proof. $\Rightarrow$. If $\langle x, y\rangle = 0$ then, by the Pythagorean Theorem,

$$\|x+\lambda y\|^2 \;=\; \|x\|^2 \,+\, \|\lambda y\|^2 \;\geq\; \|x\|^2.$$

$\Leftarrow$. Assume that $\|x\| \le \|x+\lambda y\|$ for every λ. Replacing λ with $-\lambda$ and applying the Polar Identity, we see that

$$\|x\|^2 \;\le\; \|x-\lambda y\|^2 \;=\; \|x\|^2 \,-\, 2\operatorname{Re}\langle x, \lambda y\rangle \,+\, \|\lambda y\|^2.$$

Rearranging gives

$$2\operatorname{Re}\bigl(\overline{\lambda}\langle x, y\rangle\bigr) \;\le\; |\lambda|^2\,\|y\|^2. \tag{8.6}$$

In particular, let $\lambda = t > 0$ be a positive real number. Then equation (8.6) reduces to $2\operatorname{Re}\langle x, y\rangle \le t\,\|y\|^2$. Letting t approach zero through positive values, we therefore obtain

$$2\operatorname{Re}\langle x, y\rangle \;\le\; \lim_{t\to 0^+} t\,\|y\|^2 \;=\; 0.$$

Thus $\operatorname{Re}\langle x, y\rangle \le 0$. By considering $\lambda = t < 0$ we can similarly show that $\operatorname{Re}\langle x, y\rangle \ge 0$, and therefore $\operatorname{Re}\langle x, y\rangle = 0$. If H is a real inner product space, then this shows that $\langle x, y\rangle = 0$, and so we are done. On the other hand, if H is a complex inner product space, then by considering $\lambda = it$ with $t > 0$ and then $t < 0$ we can show that $\operatorname{Im}\langle x, y\rangle = 0$, and therefore $\langle x, y\rangle = 0$. $\square$

8.2.1 Orthogonal Complements

We have defined what it means for vectors to be orthogonal, but sometimes we need to consider subsets or subspaces that are orthogonal. For example, we often say that the z-axis in $\mathbb{R}^3$ is orthogonal to the x-y plane. What we mean by this statement is that every vector on the z-axis is orthogonal to every vector in the x-y plane. The following definition extends this idea to subsets of an inner product space.

Definition 8.2.3 (Orthogonal Subsets). Let H be an inner product space, and let A and B be subsets of H.

(a) We say that a vector $x \in H$ is *orthogonal to the set* A, and write $x \perp A$, if $x \perp y$ for every $y \in A$.

(b) We say that A and B are *orthogonal sets*, and write $A \perp B$, if $x \perp y$ for every $x \in A$ and $y \in B$. $\diamondsuit$

The *largest possible* set B that is orthogonal to a given set A is called the *orthogonal complement* of A, defined precisely as follows.

Definition 8.2.4 (Orthogonal Complement). Let A be a subset of an inner product space H. The *orthogonal complement* of A is

$$A^\perp = \{x \in H : x \perp A\} = \{x \in H : \langle x, y\rangle = 0 \text{ for all } y \in A\}. \qquad \diamondsuit$$

For example, although the x-axis in $\mathbb{R}^3$ is orthogonal to the z-axis, it is not the largest set that is orthogonal to the z-axis. The largest set that is orthogonal to the z-axis is the x-y plane, and this plane is the orthogonal complement of the z-axis in $\mathbb{R}^3$. To emphasize, the orthogonal complement $A^\perp$ contains *all* (not just *some*) of the vectors x in H that are orthogonal to all elements of A.

To give another example, we declare that a function $f \in L^2(\mathbb{R})$ is *even* if $f(x) = f(-x)$ for a.e. x, and similarly f is *odd* if $f(x) = -f(-x)$ for a.e. x.

Exercise 8.2.5. Let E be the set of all even functions in $L^2(\mathbb{R})$, and let O be the set of all odd functions in $L^2(\mathbb{R})$. Prove that E and O are closed subspaces of $L^2(\mathbb{R})$, and we have both $E^\perp = O$ and $O^\perp = E$. $\diamondsuit$

Often the set A will be a subspace of H (as in the preceding example), but it does not have to be.

Here are some properties of orthogonal complements.

Lemma 8.2.6. *If A is a subset of an inner product space H, then the following statements hold.*

(a) *$A^\perp$ is a closed subspace of H.*

(b) *$H^\perp = \{0\}$ and $\{0\}^\perp = H$.*

(c) *If $A \subseteq B$, then $B^\perp \subseteq A^\perp$.*

(d) *$A \subseteq (A^\perp)^\perp$.*

Proof. (a) Choose any vectors $y, z \in A^\perp$ and scalars a and b. Then for every $x \in A$ we have

$$\langle ay + bz, x\rangle = a\langle y, x\rangle + b\langle z, x\rangle = 0,$$

so $ay + bz \in A^\perp$. Therefore $A^\perp$ is a subspace of H.

Now suppose that vectors $y_n \in A^\perp$ are such that $y_n \to y$ in H. Then for every $x \in A$ we have by the continuity of the inner product that

$$\langle x, y\rangle = \lim_{n\to\infty} \langle x, y_n\rangle = 0.$$

Therefore $y \in A^\perp$, so $A^\perp$ is closed.

(b) Every vector in H is orthogonal to every vector in $\{0\}$, so $\{0\}^\perp = H$. Suppose $x \in H^\perp$. Then x is orthogonal to every vector in H, including itself. Therefore $\|x\|^2 = \langle x, x\rangle = 0$, which implies that $x = 0$. Hence $H^\perp = \{0\}$.

(c) Assume that $A \subseteq B \subseteq H$, and suppose that $x \in B^\perp$. Then x is orthogonal to every vector in B, and therefore it is orthogonal to every vector in A. Hence $x \in A^\perp$, which shows that $B^\perp \subseteq A^\perp$.

(d) Fix $x \in A$. Then x is orthogonal to every vector in $A^\perp$ (by the definition of $A^\perp$), so x belongs to $(A^\perp)^\perp$. Thus $A \subseteq (A^\perp)^\perp$. $\square$

In Lemma 8.2.14 we will prove that *if M is a closed subspace of a Hilbert space, then $(M^\perp)^\perp = M$.*

8.2.2 Orthogonal Projections

Finding a point that is closest to a given set is a type of optimization problem that arises in a wide variety of circumstances. Unfortunately, in a generic Banach space it can be difficult to compute the exact distance from a point x to a set S, or to determine if there is a vector in S that is closest to x. Even if a closest point exists, it need not be unique. The following theorem states that if S is a *closed and convex subset of a Hilbert space* H, then for each vector $x \in H$ there exists a *unique* vector $y \in S$ that is closest to x.

Theorem 8.2.7 (Closest Point Theorem). *Let H be a Hilbert space, and let S be a nonempty closed, convex subset of H. If $x \in H$, then there exists a unique vector $y \in S$ that is closest to x. That is, there is a unique vector $y \in S$ that satisfies*

$$\|x-y\| \;=\; \operatorname{dist}(x,S) \;=\; \inf\bigl\{\|x-z\| : z \in S\bigr\}.$$

Proof. Set $d = \operatorname{dist}(x,S)$. Then, by the definition of an infimum, there exist vectors $y_n \in S$ such that $\|x-y_n\| \to d$ as $n \to \infty$. For each of these vectors we have $\|x-y_n\| \ge d$. Therefore, if we fix an $\varepsilon > 0$ then we can find an integer $N > 0$ such that

$$d^2 \;\le\; \|x-y_n\|^2 \;\le\; d^2+\varepsilon^2, \qquad \text{for all } n \ge N.$$

We will show that the sequence $\{y_n\}_{n\in\mathbb{N}}$ is Cauchy, and hence converges to some point y, which we will then prove is the unique closest point to S.

To do this, choose any integers $m, n \ge N$, and let $w = (y_m+y_n)/2$ be the midpoint of the line segment joining y_m to y_n. Since S is convex we have $w \in S$, and therefore $\|x-w\| \ge d$. Using the Parallelogram Law, it follows that

$$\begin{aligned}
\|y_n-y_m\|^2 + 4d^2 \;&\le\; \|y_n-y_m\|^2 + 4\,\|x-w\|^2\\
&=\; \|(x-y_n)-(x-y_m)\|^2 + \|(x-y_n)+(x-y_m)\|^2\\
&=\; 2\bigl(\|x-y_n\|^2 + \|x-y_m\|^2\bigr) \qquad \text{(Parallelogram Law)}\\
&\le\; 4\,(d^2+\varepsilon^2).
\end{aligned}$$

Rearranging, we see that $\|y_m-y_n\| \le 2\varepsilon$. This holds for all $m, n \ge N$, so $\{y_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in H. Since H is complete, this sequence must converge, say to y. Since S is closed and $y_n \in S$ for every n, the vector y belongs to S. Also, since $x-y_n \to x-y$, the continuity of the norm implies that $\|x-y_n\| \to \|x-y\|$. Hence

$$\|x-y\| \;=\; \lim_{n\to\infty} \|x-y_n\| \;=\; d.$$

Therefore y is a point in S that is closest to x.

It only remains to show that y is the *unique* point in S that is closest to x. Suppose that $z \in S$ is also a closest point, i.e., $\|x-y\| = d = \|x-z\|$. Since the midpoint $w = (y+z)/2$ belongs to S, we have $\|x-w\| \ge d$. Applying the Parallelogram Law again, we see that

$$\begin{aligned} 4d^2 &= 2\left(\|x-y\|^2 + \|x-z\|^2\right) \\ &= \|(x-y)-(x-z)\|^2 + \|(x-y)+(x-z)\|^2 \\ &= \|y-z\|^2 + 4\,\|x-w\|^2 \\ &\ge \|y-z\|^2 + 4d^2. \end{aligned}$$

Rearranging yields $\|y-z\| \le 0$, which implies that $y = z$. □

In particular, every closed subspace M of H is nonempty, closed, and convex, so we can apply the Closest Point Theorem to M. For this setting we introduce a name for the point p in M that is closest to a given vector x. We also use the same name to denote the function that maps x to the point p in M that is closest to x.

Definition 8.2.8 (Orthogonal Projection). Let M be a closed subspace of a Hilbert space H.

(a) If $x \in H$, then the unique vector $p \in M$ that is closest to x is called the *orthogonal projection of x onto M*.

(b) The function $P\colon H \to H$ defined by $Px = p$, where p is the orthogonal projection of x onto M, is called the *orthogonal projection of H onto M*. ◇

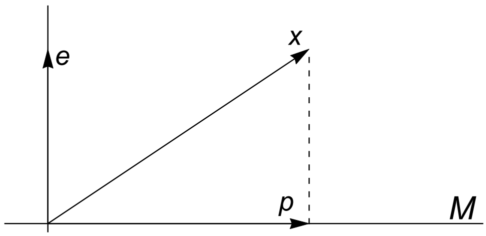


Fig. 8.1 The orthogonal projection of a vector x onto a subspace M. The vector p is the point in M that is closest to x, and $e = x - p$.

Since the orthogonal projection p is the vector in M that is closest to x, we can think of p as being the best approximation to x by vectors from M. The difference $e = x - p$ is the error in this approximation (see Figure 8.1).

Example 8.2.9. For simplicity, we take scalars to be real in this example. Let $M = \{(x_1, x_2, 0) : x_1, x_2 \in \mathbb{R}\}$ be the x_1-x_2 plane in $\mathbb{R}^3$, and choose any point $x = (x_1, x_2, x_3)$ in $\mathbb{R}^3$. We claim that $p = (x_1, x_2, 0) \in M$ is the orthogonal projection of x onto M. To prove this, choose an arbitrary point $w = (w_1, w_2, 0)$ in M. Then $x - w = (x_1 - w_1, x_2 - w_2, x_3)$ while $x - p = (0, 0, x_3)$, so

$$\|x - w\|_2^2 \;=\; |x_1 - w_1|^2 + |x_2 - w_2|^2 + |x_3|^2 \;\geq\; |x_3|^2 \;=\; \|x - p\|_2^2.$$

Thus p is closer to x than w, so p is the orthogonal projection. $\diamondsuit$

If we let $\{e_1, e_2, e_3\}$ be the standard basis for $\mathbb{R}^3$, then Example 8.2.9 tells us that the orthogonal projection of

$$x \;=\; (x_1, x_2, x_3) \;=\; x_1e_1 + x_2e_2 + x_3e_3$$

onto $M = \mathrm{span}\{e_1, e_2\}$ is

$$p \;=\; (x_1, x_2, 0) \;=\; x_1e_1 + x_2e_2.$$

Next we derive an analogous formula for the orthogonal projection of a vector onto any nontrivial finite-dimensional subspace (the trivial case is easy: the orthogonal projection of any vector onto $M = \{0\}$ is the zero vector).

Lemma 8.2.10. *Let $\{e_1, \dots, e_d\}$ be a finite set of orthonormal vectors in a Hilbert space H, and let $M = \mathrm{span}\{e_1, \dots, e_d\}$. Then the following statements hold.*

(a) *If $x \in M$, then*

$$x \;=\; \sum_{n=1}^{d} \langle x, e_n\rangle\, e_n$$

is the unique representation of x as a linear combination of $e_1, \dots, e_d$, and we have

$$\|x\|^2 \;=\; \sum_{n=1}^{d} |\langle x, e_n\rangle|^2. \tag{8.7}$$

(b) *M is a closed subspace of H.*

(c) *The orthogonal projection of an arbitrary vector $x \in H$ onto M is*

$$p \;=\; \sum_{n=1}^{d} \langle x, e_n\rangle\, e_n, \tag{8.8}$$

and we have

$$\|p\|^2 \;=\; \sum_{n=1}^{d} |\langle x, e_n\rangle|^2.$$

Proof. (a) By hypothesis, the vectors $e_1, \dots, e_d$ span M. Therefore, if $x \in M$ then $x = \sum_{n=1}^{d} c_n e_n$ for some scalars c_n. If $1 \leq k \leq n$, then the fact that the vectors $e_1, \dots, e_d$ are orthonormal implies that

$$\langle x, e_k \rangle = \left\langle \sum_{n=1}^{d} c_n e_n,\, e_k \right\rangle = \sum_{n=1}^{d} c_n \langle e_n, e_k \rangle = c_k.$$

Finally, equation (8.7) follows by applying the Pythagorean Theorem.

(b) It is a fact that every finite-dimensional subspace of a normed space is closed (see Section 1.2.4). Alternatively, part (a) can be used to give a direct proof that M is closed; we assign the details as an exercise for the reader.

(c) Let $e = x - p$. If we fix any integer k between 1 and d, then

$$\begin{aligned} \langle e, e_k \rangle = \langle x, e_k \rangle - \langle p, e_k \rangle &= \langle x, e_k \rangle - \left\langle \sum_{n=1}^{d} \langle x, e_n \rangle e_n,\, e_k \right\rangle \\ &= \langle x, e_k \rangle - \sum_{n=1}^{d} \langle x, e_n \rangle \langle e_n, e_k \rangle \\ &= \langle x, e_k \rangle - \langle x, e_k \rangle = 0. \end{aligned}$$

Hence e is orthogonal to each of $e_1, \dots, e_d$. Since every vector in M is a linear combination of these vectors, it follows that e is orthogonal to every element of M. In particular, if $w \in M$ then $w - p$ is also in M, so $e \perp w - p$ and therefore

$$\begin{aligned} \|x - w\|^2 &= \|e - (w-p)\|^2 \\ &= \|e\|^2 + \|w - p\|^2 \qquad \text{(Pythagorean Theorem)} \\ &\geq \|e\|^2 = \|x - p\|^2. \end{aligned}$$

Thus p is closer to x than w. Therefore p is the orthogonal projection of x onto M (see the illustration in Figure 8.2). $\square$

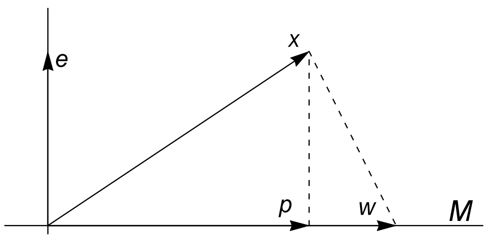


Fig. 8.2 The vector p in M is closer to x than the point $w \in M$. Each of p, w, and $w - p$ are orthogonal to $e = x - p$.

In Section 8.3 we will generalize Lemma 8.2.10 from finite-dimensional subspaces to arbitrary closed subspaces M of H.

8.2.3 Characterizations of the Orthogonal Projection

Now we give several equivalent reformulations of the orthogonal projection. In particular, we see that the orthogonal projection of x onto M is the unique vector $p \in M$ such that the error vector $e = x - p$ is orthogonal to M.

Theorem 8.2.11. *Let M be a closed subspace of a Hilbert space H. If x and p are vectors in H, then the following four statements are equivalent.*

(a) *p is the orthogonal projection of x onto M, i.e., p is the unique point in M that is closest to x.*

(b) *$p \in M$ and $x - p \perp M$.*

(c) *$x = p + e$, where $p \in M$ and $e \in M^\perp$.*

(d) *$e = x - p$ is the orthogonal projection of x onto $M^\perp$.*

Proof. (a) $\Rightarrow$ (b). Let p be the point in M that is closest to x, and let $e = x - p$. Choose any vector $y \in M$. We must show that $\langle y, e\rangle = 0$. Since M is a subspace, $p - \lambda y \in M$ for every scalar λ. But p is closer to x than $p - \lambda y$, so

$$\|e\| \;=\; \|x-p\| \;\le\; \|x-(p-\lambda y)\| \;=\; \|e+\lambda y\|.$$

Lemma 8.2.2 therefore implies that $y \perp e$.

Exercise: Prove the remaining implications. $\square$

8.2.4 The Closed Span

The span of a set A, denoted $\mathrm{span}(A)$, is the set of all *finite linear combinations* of elements of A. In order to characterize the orthogonal complement of the orthogonal complement of a set A, we will need to consider the closure of the span of A. We call this the *closed span* of A, and we introduce the following notation (which makes sense in any normed space).

Notation 8.2.12 (Closed Span). If A is a subset of a normed space X, then we denote the closure of the span of A by

$$\overline{\mathrm{span}}(A) \;=\; \overline{\mathrm{span}(A)}.$$

We call $\overline{\mathrm{span}}(A)$ the *closed span* of A. If $A = \{x_n\}_{n\in\mathbb{N}}$ is a sequence, then we often write $\overline{\mathrm{span}}\{x_n\}_{n\in\mathbb{N}}$ or just $\overline{\mathrm{span}}\{x_n\}$ for the closed span of the sequence $\{x_n\}_{n\in\mathbb{N}}$. $\diamondsuit$

By Exercise 1.1.7, the closed span of A consists of all *limits of elements of* $\mathrm{span}(A)$:

$$\overline{\mathrm{span}}(A) \;=\; \bigl\{y \in X \,:\, \exists\, y_n \in \mathrm{span}(A) \text{ such that } y_n \to y\bigr\}. \tag{8.9}$$

Note that in equation (8.9) we take limits of *finite linear combinations* of elements of A, not just limits of elements of A. The following exercise shows that we can equivalently characterize the closed span as *the smallest closed subspace of X that contains A.*

Exercise 8.2.13 (Smallest Closed Subspace). Suppose that A is a subset of a normed space X. Prove that

(a) $\overline{\mathrm{span}}(A)$ is a closed subspace of X, and

(b) if M is any closed subspace such that $A \subseteq M$, then $\overline{\mathrm{span}}(A) \subseteq M$.

Consequently, the closed span is the intersection of all of the closed subspaces that contain A:

$$\overline{\mathrm{span}}(A) \;=\; \bigcap\{M : M \text{ is a closed subspace and } M \supseteq A\}. \qquad \diamondsuit$$

8.2.5 The Complement of the Complement

Now we prove that the orthogonal complement of the orthogonal complement of a set A is the closed span of A. We begin with the case where our set is a closed subspace.

Lemma 8.2.14. *If H is a Hilbert space and M is a closed subspace of H, then $(M^\perp)^\perp = M$.*

Proof. We saw in Lemma 8.2.6 that $M \subseteq (M^\perp)^\perp$. Conversely, suppose that $x \in (M^\perp)^\perp$, and let p be the orthogonal projection of x onto M. Since M is a closed subspace, we have $x = p + e$ where $p \in M$ and $e \in M^\perp$. Since $x \in (M^\perp)^\perp$ and $p \in M \subseteq (M^\perp)^\perp$, it follows that $e = x - p \in (M^\perp)^\perp$. However, we also know that $e \in M^\perp$, so e is orthogonal to itself and therefore is zero. Hence $x = p + 0 \in M$. This shows that $(M^\perp)^\perp \subseteq M$. $\square$

The next exercise will allow us to generalize from closed subspaces M to arbitrary subsets A in H.

Exercise 8.2.15. Let A be a subset of a Hilbert space H, and suppose that $x \perp A$. Prove that $x \perp \mathrm{span}(A)$ and $x \perp \overline{\mathrm{span}}(A)$, and use this to show that

$$A^\perp \;=\; \mathrm{span}(A)^\perp \;=\; \overline{\mathrm{span}}(A)^\perp. \qquad \diamondsuit$$

Corollary 8.2.16. *If H is a Hilbert space and $A \subseteq H$, then*

$$(A^\perp)^\perp \;=\; \overline{\mathrm{span}}(A).$$

Proof. If we let $M = \overline{\mathrm{span}}(A)$, then Exercise 8.2.15 implies that $A^\perp = M^\perp$. But M is closed subspace, so $(M^\perp)^\perp = M$ by Lemma 8.2.14. $\square$

8.2.6 Complete Sequences

We often seek sequences whose closed span is as large as possible. We introduce the following terminology for such sequences. Note that the meaning of a "complete sequence" as given in this definition is entirely distinct from the meaning of a "complete space" as given in Definition 1.1.4.

Definition 8.2.17 (Complete Sequence). Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of vectors in a normed space X. We say that the sequence $\{x_n\}_{n\in\mathbb{N}}$ is *complete* in X if $\operatorname{span}\{x_n\}_{n\in\mathbb{N}}$ is dense in X, i.e., if

$$\overline{\operatorname{span}}\{x_n\}_{n\in\mathbb{N}} \;=\; X.$$

Complete sequences are also known as *total* or *fundamental* sequences. $\diamondsuit$

Applying this terminology to the results of Section 8.2.5 gives us the following characterization.

Corollary 8.2.18. *If $\{x_n\}_{n\in\mathbb{N}}$ is a sequence of vectors in a Hilbert space H, then the following two statements are equivalent.*

(a) *$\{x_n\}_{n\in\mathbb{N}}$ is a complete sequence in H.*

(b) *The only vector in H that is orthogonal to every x_n is the zero vector, i.e., if $x \in H$ and $\langle x, x_n\rangle = 0$ for every n, then $x = 0$.* $\diamondsuit$

Problems

8.2.19. Prove that any set of nonzero orthogonal vectors in an inner product space is finitely linearly independent.

8.2.20. Let M be a closed subspace of a Hilbert space H, and let P be the orthogonal projection of H onto M. Show that $I - P$ is the orthogonal projection of H onto $M^\perp$.

8.2.21. Assume that $E \subseteq \mathbb{R}^d$ is measurable with $|E| > 0$, and set

$$M \;=\; \bigl\{g \in L^2(\mathbb{R}^d) \,:\, g(x) = 0 \text{ for a.e. } x \notin E\bigr\}.$$

Prove that M is a closed subspace of $L^2(\mathbb{R}^d)$, and the orthogonal projection of $f \in L^2(\mathbb{R}^d)$ onto M is $p = f \cdot \chi_E$. What is the orthogonal complement of M?

8.2.22. (a) Let H be a finite-dimensional Hilbert space. Prove that a finite set of vectors $\{x_1, \dots, x_m\}$ is complete in H if and only if $x_1, \dots, x_m$ span H.

(b) Prove that the sequence of standard basis vectors $\{\delta_n\}_{n\in\mathbb{N}}$ is complete in ℓ^2, but it does not span ℓ^2.

8.2.23. Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in a Hilbert space H. Show that if $y \perp x_n$ for every n, then $y \in \bigl(\overline{\operatorname{span}}\{x_n\}_{n\in\mathbb{N}}\bigr)^{\perp}$.

8.2.24. Given a sequence $\{x_n\}_{n\in\mathbb{N}}$ in a Hilbert space H, prove that the following two statements are equivalent.

(a) For each $m \in \mathbb{N}$ we have $x_m \notin \overline{\operatorname{span}}\{x_n\}_{n\neq m}$, i.e., x_m does not lie in the closed span of the other vectors (such a sequence is said to be *minimal*).

(b) There exists a sequence $\{y_n\}_{n\in\mathbb{N}}$ in H such that $\langle x_m, y_n\rangle = \delta_{mn}$ for all $m, n \in \mathbb{N}$ (we say that sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ satisfying this condition are *biorthogonal*).

Show further that if statements (a) and (b) hold, then the sequence $\{y_n\}_{n\in\mathbb{N}}$ is unique if and only if $\{x_n\}_{n\in\mathbb{N}}$ is complete.

8.2.25. Prove that $\sin 2\pi x$ and $\cos 2\pi x$ are orthogonal functions in $L^2[0,1]$, and there is no function $f \in L^2[0,1]$ that satisfies

$$\int_0^1 \bigl|f(x) - \sin 2\pi x\bigr|^2\, dx \;<\; \frac{4}{9} \quad\text{and}\quad \int_0^1 \bigl|f(x) - \cos 2\pi x\bigr|^2\, dx \;<\; \frac{1}{9}.$$

8.2.26. Let M be a closed subspace of a Hilbert space H. Given $x \in H$, prove that

$$\operatorname{dist}(x, M) \;=\; \sup\bigl\{|\langle x, y\rangle| : y \in M^{\perp},\ \|y\| = 1\bigr\},$$

and the supremum is achieved.

8.3 Orthonormal Sequences and Orthonormal Bases

In this section we will take a closer look at orthonormal sequences, focusing especially on countably infinite orthonormal sequences $\{e_n\}_{n\in\mathbb{N}}$. The reader should check (this is Problem 8.3.18) that similar results hold for finite orthonormal sequences $\{e_1, \dots, e_d\}$; in fact the statements and proofs are easier in that case because there are no issues with convergence of infinite series.

8.3.1 Orthonormal Sequences

Suppose that $\{e_n\}_{n\in\mathbb{N}}$ is an arbitrary sequence in a Banach space X. In general, if we are given some scalars c_n then it can be extremely difficult to determine whether the infinite series $\sum c_n e_n$ converges in X. However, the next theorem shows that *if $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence in a Hilbert space H*, then we can completely characterize the scalars for which this happens. Recall that an infinite series converges if there is a vector x

such that the *partial sums* $s_N = \sum_{n=1}^N c_n e_n$ converge to x in the norm of H as $N \to \infty$.

Theorem 8.3.1. *If $\mathcal{E} = \{e_n\}_{n \in \mathbb{N}}$ is an orthonormal sequence in a Hilbert space H, then the following statements hold.*

(a) Bessel's Inequality: $\displaystyle\sum_{n=1}^{\infty} |\langle x, e_n \rangle|^2 \le \|x\|^2$ *for each $x \in H$.*

(b) *If the series $x = \displaystyle\sum_{n=1}^{\infty} c_n e_n$ converges, then $c_n = \langle x, e_n \rangle$ for each $n \in \mathbb{N}$.*

(c) $\displaystyle\sum_{n=1}^{\infty} c_n e_n$ *converges* $\iff$ $\displaystyle\sum_{n=1}^{\infty} |c_n|^2 < \infty$.

Proof. (a) Choose $x \in H$. If we fix $N \in \mathbb{N}$, then Lemma 8.2.10 tells us that

$$p_N = \sum_{n=1}^{N} \langle x, e_n \rangle \, e_n$$

is the orthogonal projection of x onto $\operatorname{span}\{e_1, \dots, e_N\}$. Consequently the "error vector" $q_N = x - p_N$ is orthogonal to p_N. Hence

$$\begin{aligned}
\|x\|^2 = \|p_N + q_N\|^2 &= \|p_N\|^2 + \|q_N\|^2 && \text{(Pythagorean Theorem)} \\
&\ge \|p_N\|^2 && (\text{since } \|q_N\|^2 \ge 0) \\
&= \sum_{n=1}^{N} |\langle x, e_n \rangle|^2 && \text{(Pythagorean Theorem).}
\end{aligned}$$

Letting $N \to \infty$, we obtain Bessel's Inequality.

(b) Suppose that the series $x = \sum c_n e_n$ converges, and fix $m \in \mathbb{N}$. Then, by applying equation (8.2), we compute that

$$\langle x, e_m \rangle = \left\langle \sum_{n=1}^{\infty} c_n e_n, \, e_m \right\rangle = \sum_{n=1}^{\infty} c_n \langle e_n, e_m \rangle = \sum_{n=1}^{\infty} c_n \delta_{mn} = c_m.$$

(c) If $x = \sum c_n e_n$ converges, then $c_n = \langle x, e_n \rangle$ by part (b), and therefore $\sum |c_n|^2 < \infty$ by Bessel's Inequality. Conversely, suppose that $\sum |c_n|^2 < \infty$ and set

$$s_N = \sum_{n=1}^{N} c_n e_n \qquad \text{and} \qquad t_N = \sum_{n=1}^{N} |c_n|^2.$$

If $N > M$ then, by the Pythagorean Theorem,

$$\|s_N - s_M\|^2 = \left\| \sum_{n=M+1}^{N} c_n e_n \right\|^2 = \sum_{n=M+1}^{N} \|c_n e_n\|^2 = |t_N - t_M|.$$

Since $\{t_N\}_{N\in\mathbb{N}}$ is a Cauchy sequence of scalars, it follows that $\{s_N\}_{N\in\mathbb{N}}$ is a Cauchy sequence in H. But H is complete (since it is a Hilbert space), so the sequence $\{s_N\}_{N\in\mathbb{N}}$ must converge. Therefore, by the definition of an infinite series, $\sum c_n e_n$ converges. $\square$

8.3.2 Unconditional Convergence

We have seen that if $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence, then the infinite series $\sum c_n x_n$ converges if and only if $(c_n)_{n\in\mathbb{N}} \in \ell^2$. We will show that the convergence is actually *unconditional* in the following sense.

Definition 8.3.2 (Unconditional Convergence). Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of vectors in a normed space X. If $\sum_{n=1}^\infty x_{\sigma(n)}$ converges for every bijection $\sigma\colon \mathbb{N} \to \mathbb{N}$, then we say that the infinite series $\sum_{n=1}^\infty x_n$ *converges unconditionally.* A series that converges but does not converge unconditionally is said to be *conditionally convergent.* $\diamondsuit$

That is, a series $\sum x_n$ converges unconditionally if it converges no matter what ordering we impose on the index set. The following theorem states that unconditional and absolute convergence are equivalent *for series of scalars* (for one proof, see [Heil11, Lemma 3.3]).

Theorem 8.3.3. *If $(c_n)_{n\in\mathbb{N}}$ is a sequence of scalars, then $\sum c_n$ converges absolutely if and only if it converges unconditionally. That is,*

$$\sum_{n=1}^\infty |c_n| < \infty \iff \sum_{n=1}^\infty c_{\sigma(n)} \text{ converges for each bijection } \sigma\colon \mathbb{N} \to \mathbb{N}. \quad \diamondsuit$$

For example, the alternating harmonic series $\sum_{n=1}^\infty (-1)^n/n$ does not converge absolutely, and therefore there must be some reordering $\sigma\colon \mathbb{N} \to \mathbb{N}$ such that $\sum_{n=1}^\infty (-1)^{\sigma(n)}/\sigma(n)$ diverges (exhibit such a permutation σ).

The equivalence given in Theorem 8.3.3 extends to infinite series in finite-dimensional normed spaces (see [Heil11, Sec. 3.6] for details). In any Banach space it is always true that *absolute convergence implies unconditional convergence* (this is Problem 8.3.23). However, as we will explain below, in an infinite-dimensional Hilbert space, unconditional convergence does not imply absolute convergence. On the other hand, for an *orthonormal sequence* we have the following connection between convergence and unconditional convergence.

Corollary 8.3.4. *If $\mathcal{E} = \{e_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence in a Hilbert space H, then*

$$\sum_{n=1}^\infty c_n e_n \text{ converges} \quad \iff \quad \sum_{n=1}^\infty c_n e_n \text{ converges unconditionally.}$$

Proof. If $\sigma\colon \mathbb{N} \to \mathbb{N}$ is a bijection, then

$$\begin{aligned}
\sum_{n=1}^{\infty} c_n e_n \text{ converges} \quad &\iff \quad \sum_{n=1}^{\infty} |c_n|^2 < \infty && \text{(Theorem 8.3.1)}\\
&\iff \quad \sum_{n=1}^{\infty} |c_{\sigma(n)}|^2 < \infty && \text{(Theorem 8.3.3)}\\
&\iff \quad \sum_{n=1}^{\infty} c_{\sigma(n)}\, e_{\sigma(n)} \text{ converges} && \text{(Theorem 8.3.1).}
\end{aligned}$$

Thus, if $\sum c_n e_n$ converges then so does any reordering of the series. □

We use this corollary to exhibit an infinite series that converges unconditionally but not absolutely.

Example 8.3.5. Let H be any infinite-dimensional Hilbert space, and let $\{e_n\}_{n\in\mathbb{N}}$ be an infinite orthonormal sequence in H. Since $(\frac{1}{n})_{n\in\mathbb{N}} \in \ell^2$, Theorem 8.3.1 and Corollary 8.3.4 imply that the series $\sum \frac{1}{n} e_n$ converges unconditionally. However,

$$\sum_{n=1}^{\infty} \left\| \frac{1}{n} e_n \right\| = \sum_{n=1}^{\infty} \frac{1}{n} = \infty,$$

so $\sum \frac{1}{n} e_n$ does not converge absolutely. ♢

The *Dvoretzky–Rogers Theorem* is a nontrivial result that implies that *every* infinite-dimensional normed space contains an infinite series that converges unconditionally but not absolutely (see [Heil11, Sec. 3.6] for discussion and details).

8.3.3 Orthogonal Projections Revisited

If $\{e_n\}_{n\in\mathbb{N}}$ is a sequence of orthonormal vectors in a Hilbert space, then its closed span is a closed subspace of H. The next theorem gives an explicit formula for the orthogonal projection of a vector onto a closed span.

Theorem 8.3.6. *Let H be a Hilbert space, let $\mathcal{E} = \{e_n\}_{n\in\mathbb{N}}$ be an orthonormal sequence in H, and let $M = \overline{\operatorname{span}}(\mathcal{E})$ be the closed span of $\mathcal{E}$. If $x \in H$, then the following statements hold.*

(a) *The orthogonal projection of x onto M is*

$$p = \sum_{n=1}^{\infty} \langle x, e_n\rangle\, e_n.$$

(b) *The norm of* p *satisfies*

$$\|p\|^2 \;=\; \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2.$$

(c) *We have*

$$x \in M \quad\Longleftrightarrow\quad x \;=\; \sum_{n=1}^{\infty} \langle x, e_n\rangle\, e_n \quad\Longleftrightarrow\quad \|x\|^2 \;=\; \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2.$$

Proof. (a) By Bessel's Inequality, $\sum |\langle x, e_n\rangle|^2 \le \|x\|^2 < \infty$. Part (b) of Theorem 8.3.1 therefore implies that the infinite series that defines p does converge. We must show that this vector p is the orthogonal projection of x onto M.

If we fix $k \in \mathbb{N}$ then, since the e_n are orthonormal,

$$\langle x - p, e_k\rangle \;=\; \langle x, e_k\rangle \;-\; \sum_{n=1}^{\infty} \langle x, e_n\rangle\, \langle e_n, e_k\rangle \;=\; \langle x, e_k\rangle \;-\; \langle x, e_k\rangle \;=\; 0.$$

Thus $x-p$ is orthogonal to every e_k. By linearity and by the continuity of the inner product, it follows that $x-p$ is orthogonal to every vector in M (see Exercise 8.2.15). Therefore we have both $p \in M$ and $x - p \perp M$, so Theorem 8.2.11 implies that p is the orthogonal projection of x onto M.

(b) Using the continuity of the inner product in the form of Exercise 8.1.5(b), we compute that

$$\|p\|^2 \;=\; \langle p, p\rangle \;=\; \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \langle x, e_m\rangle\, \overline{\langle x, e_n\rangle}\, \langle e_m, e_n\rangle \;=\; \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2.$$

(c) Let i, ii, iii denote the three statements that appear in statement (c). We must prove that i, ii, and iii are equivalent.

i $\Rightarrow$ ii. If $x \in M$, then the orthogonal projection of x onto M is x itself, so $x = p = \sum \langle x, e_n\rangle\, e_n$.

ii $\Rightarrow$ iii. If $x = p$ then $\|x\|^2 = \|p\|^2 = \sum |\langle x, e_n\rangle|^2$.

iii $\Rightarrow$ i. Suppose $\|x\|^2 = \sum |\langle x, e_n\rangle|^2$. Then, since $x - p \perp p$,

$$\begin{aligned}
\|x\|^2 \;&=\; \|(x-p) + p\|^2 \\
&=\; \|x-p\|^2 \;+\; \|p\|^2 \qquad\qquad \text{(Pythagorean Theorem)} \\
&=\; \|x-p\|^2 \;+\; \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2 \\
&=\; \|x-p\|^2 \;+\; \|x\|^2.
\end{aligned}$$

Hence $\|x - p\| = 0$, so $x = p \in M$. □

8.3.4 Orthonormal Bases

According to Definition 8.2.17, if $\{x_n\}_{n\in\mathbb{N}}$ is a countable sequence in a normed space X then we say that $\{x_n\}_{n\in\mathbb{N}}$ is *complete*, *total*, or *fundamental* if its closed span is all of X. Completeness by itself is typically a rather weak property, but Theorem 8.3.6 tells us that if H is a Hilbert space and $\{e_n\}_{n\in\mathbb{N}}$ is a sequence in H that is both **orthonormal and complete**, then every vector $x \in H$ can be written as $x = \sum \langle x, e_n \rangle\, e_n$. The following theorem gives us a converse to this fact, *assuming that* $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence, and additionally gives several other useful characterizations of complete orthonormal sequences.

Theorem 8.3.7. *If H is a Hilbert space and $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal sequence in H, then the following five statements are equivalent.*

(a) $\{e_n\}_{n\in\mathbb{N}}$ *is* complete, *i.e.,* $\overline{\operatorname{span}}\{e_n\}_{n\in\mathbb{N}} = H$.

(b) *For each $x \in H$ there exists a* unique *sequence of scalars $(c_n)_{n\in\mathbb{N}}$ such that $x = \sum c_n e_n$.*

(c) *Every $x \in H$ satisfies*

$$x \;=\; \sum_{n=1}^{\infty} \langle x, e_n \rangle\, e_n, \tag{8.10}$$

where this series converges in the norm of H.

(d) Plancherel's Equality *holds*:

$$\|x\|^2 \;=\; \sum_{n=1}^{\infty} |\langle x, e_n \rangle|^2 \qquad \text{for all } x \in H.$$

(e) Parseval's Equality *holds*:

$$\langle x, y \rangle \;=\; \sum_{n=1}^{\infty} \langle x, e_n \rangle\, \langle e_n, y \rangle \qquad \text{for all } x, y \in H.$$

Proof. For simplicity of notation, let $\mathcal{E} = \{e_n\}_{n\in\mathbb{N}}$ and set $M = \overline{\operatorname{span}}(\mathcal{E})$.

(a) ⇒ (c), (d). If $\mathcal{E}$ is complete, then $M = H$ by definition. Hence if $x \in H$ then $x \in M$, and therefore $x = \sum \langle x, e_n \rangle\, e_n$ and $\|x\|^2 = \sum |\langle x, e_n \rangle|^2$ by Theorem 8.3.6.

(b) ⇒ (c). If statement (b) holds, then $c_n = \langle x, x_n \rangle$ by Theorem 8.3.1(b).

(c) ⇒ (b). The uniqueness follows from the orthonormality of the e_n.

(c) $\Rightarrow$ (e). If statement (c) holds and $x, y \in H$, then

$$\langle x, y\rangle = \left\langle \sum_{n=1}^{\infty} \langle x, e_n\rangle\, e_n,\, y \right\rangle = \sum_{n=1}^{\infty} \Bigl\langle \langle x, e_n\rangle\, e_n,\, y \Bigr\rangle = \sum_{n=1}^{\infty} \langle x, e_n\rangle\, \langle e_n, y\rangle,$$

where we used Exercise 8.1.5 to move the infinite series out of the inner product.

(e) $\Rightarrow$ (d). This follows by taking $x = y$.

(d) $\Rightarrow$ (a). If statement (d) holds, then Theorem 8.3.6 implies that every $x \in H$ belongs to M. Hence $M = H$, so $\mathcal{E}$ is complete. $\square$

Since the Plancherel and Parseval Equalities are equivalent, those two names are often used interchangeably.

We refer to a sequence that satisfies the equivalent conditions in Theorem 8.3.7 as an *orthonormal basis.*

Definition 8.3.8 (Orthonormal Basis). Let H be a Hilbert space. A countably infinite orthonormal sequence $\{e_n\}_{n\in\mathbb{N}}$ that is complete in H is called an *orthonormal basis* for H. $\diamondsuit$

In particular, if $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for H then every $x \in H$ can be written uniquely as $x = \sum \langle x, e_n\rangle\, e_n$ (so $\{e_n\}_{n\in\mathbb{N}}$ is a *Schauder basis* for H in the sense of Problem 7.4.11). Further, by Corollary 8.3.4, this series converges *unconditionally* in H.

Example 8.3.9. The sequence of standard basis vectors $\{\delta_k\}_{k\in\mathbb{N}}$ is both complete and orthonormal in ℓ^2, so it is an orthonormal basis for ℓ^2. If $x = (x_k)_{k\in\mathbb{N}}$ is a vector in ℓ^2 then $\langle x, \delta_k\rangle = x_k$ for every k, so the representation of x with respect to the standard basis is simply

$$x = \sum_{k=1}^{\infty} \langle x, \delta_k\rangle\, \delta_k = \sum_{k=1}^{\infty} x_k\, \delta_k. \qquad \diamondsuit$$

If $\{e_1, \dots, e_d\}$ is a complete orthonormal sequence in a finite-dimensional Hilbert space H, then a modification of Theorem 8.3.7 (see Problem 8.3.18) shows that $\{e_1, \dots, e_d\}$ is a basis for H in the usual vector space sense (i.e., it is a Hamel basis), and for each $x \in H$ we have

$$x = \sum_{k=1}^{d} \langle x, e_n\rangle\, e_n.$$

Since $\{e_1, \dots, e_d\}$ is both orthonormal and a basis, we extend Definition 8.3.8 to cover this case as well, and refer to a complete orthonormal sequence $\{e_1, \dots, e_d\}$ as an *orthonormal basis* for H.

8.3.5 Existence of an Orthonormal Basis

A normed space is *separable* if it contains a countable dense subset (see Definition 7.4.1). All finite-dimensional normed spaces are separable, and $L^p(E)$ and ℓ^p are separable when p is finite. Hence $L^2(E)$ and ℓ^2 are infinite-dimensional separable Hilbert spaces. Not every Hilbert space is separable; one example is given in Problem 8.3.31.

We will show that every separable Hilbert space contains an orthonormal basis. We begin with finite-dimensional spaces, where we can use the same *Gram–Schmidt orthonormalization procedure* that is employed to construct orthonormal sequences in $\mathbb{R}^d$ or $\mathbb{C}^d$.

Theorem 8.3.10. *If H is a finite-dimensional Hilbert space then H contains an orthonormal basis $\{e_1, \dots, e_d\}$, where $d = \dim(H)$ is the dimension of the vector space H.*

Proof. Since H is a d-dimensional vector space, it has a Hamel basis, i.e., there is a set $\mathcal{B} = \{x_1, \dots, x_d\}$ that is both linearly independent and spans H. We will define a recursive procedure that constructs orthogonal vectors $y_1, \dots, y_d$ that span H.

First, set $y_1 = x_1$, and note that $x_1 \neq 0$ since $x_1, \dots, x_d$ are linearly independent. Define

$$M_1 \;=\; \operatorname{span}\{x_1\} \;=\; \operatorname{span}\{y_1\}.$$

If $d = 1$ then $M_1 = H$ and we stop here. Otherwise M_1 is a proper subspace of H, and $x_2 \notin M_1$ (because $\{x_1, \dots, x_d\}$ is linearly independent). Let p_2 be the orthogonal projection of x_2 onto M_1. Then $y_2 = x_2 - p_2$ is orthogonal to x_1, and $y_2 \neq 0$ since $x_2 \notin M_1$. Therefore, we can define

$$M_2 \;=\; \operatorname{span}\{x_1, x_2\} \;=\; \operatorname{span}\{y_1, y_2\},$$

where the second equality follows from the fact that y_1, y_2 are linear combinations of x_1, x_2, and vice versa. Continuing in this way, we obtain orthogonal vectors $y_1, \dots, y_d$ that span H. Hence $\{y_1, \dots, y_d\}$ is an orthogonal, but not necessarily orthonormal, basis for H. Setting $e_k = y_k/\|y_k\|$ therefore gives us an orthonormal basis $\{e_1, \dots, e_d\}$ for H. $\square$

Next we consider infinite-dimensional, but still separable, Hilbert spaces.

Theorem 8.3.11. *If H is a infinite-dimensional separable Hilbert space, then H contains an orthonormal basis of the form $\{e_n\}_{n\in\mathbb{N}}$.*

Proof. Since H is separable, it contains a countable dense subset $\{z_n\}_{n\in\mathbb{N}}$. The span of $\{z_n\}_{n\in\mathbb{N}}$ is dense in H, but $\{z_n\}_{n\in\mathbb{N}}$ need not be linearly independent. However, we can extract a subsequence that is independent and has the same span. Simply let x_1 be the first z_n that is nonzero. Then let x_2 be the first z_n after x_1 that is not a multiple of x_1. Then let x_3 be the first z_n after x_2 that does not belong to $\mathrm{span}\{x_1, x_2\}$, and so forth. In this way we obtain an independent sequence $\{x_n\}_{n\in\mathbb{N}}$ such that $\mathrm{span}\{x_n\}_{n\in\mathbb{N}} = \mathrm{span}\{z_n\}_{n\in\mathbb{N}}$. This span is dense in H by hypothesis.

Now we apply the Gram–Schmidt procedure utilized in the proof of Theorem 8.3.10 to the vectors $x_1, x_2, \dots$, but without stopping. This gives us orthonormal vectors $e_1, e_2, \dots$ such that for every n we have

$$\mathrm{span}\{e_1, \dots, e_n\} \;=\; \mathrm{span}\{x_1, \dots, x_n\}.$$

Consequently, $\mathrm{span}\{e_n\}_{n\in\mathbb{N}}$ equals $\mathrm{span}\{x_n\}_{n\in\mathbb{N}}$, which equals $\mathrm{span}\{z_n\}_{n\in\mathbb{N}}$, which is dense in H. Therefore $\{e_n\}_{n\in\mathbb{N}}$ is a complete orthonormal sequence, so it is, by definition, an orthonormal basis for H. $\square$

Theorems 8.3.10 and 8.3.11 show that every separable Hilbert space contains an orthonormal basis. This basis is finite if H is finite-dimensional, and countably infinite if H is infinite-dimensional. Conversely, Problem 7.4.6 implies that any Hilbert space that contains a countable orthonormal basis must be separable.

We will see several specific examples of orthonormal bases below.

8.3.6 The Legendre Polynomials

Let $[a, b]$ be a finite closed interval with $a < b$. The Weierstrass Approximation Theorem (Theorem 1.3.4) tells us that the set of monomials $\mathcal{M} = \{x^k\}_{k\geq 0}$ is a complete sequence in $C[a, b]$ with respect to the uniform norm (implicitly, k denotes a nonnegative integer here). Because $[a, b]$ has finite measure, it follows directly from this that the monomials are complete in $L^2[a, b]$ with respect to the L^2-norm (see Problem 8.3.27). However, they are not an orthogonal sequence, because

$$\langle x^j, x^k \rangle \;=\; \int_a^b x^j\, x^k\, dx \;=\; \frac{b^{j+k+1} - a^{j+k+1}}{j + k + 1},$$

which cannot be simultaneously zero for all $j \neq k$.

Although the monomials $\{x^k\}_{k\geq 0}$ are not orthogonal, they are linearly independent, so we can apply the Gram–Schmidt procedure to obtain an orthogonal or orthonormal basis for $L^2[a, b]$. In particular, the *Legendre polynomials* are the orthogonal basis $\{P_k\}_{k\geq 0}$ obtained by applying Gram–Schmidt to the monomials x^k on the interval $[-1, 1]$. Since P_k is defined to be a linear combination of $1, x, \dots, x^k$, it is a polynomial, and in fact it is a polynomial of degree k. Traditionally, these polynomials are not normalized so that their L^2-norm is 1, but rather are scaled so that $\|P_k\|_2^2 = \frac{2}{2k+1}$. Hence $\{P_k\}_{k\geq 0}$ is an orthogonal, but not orthonormal, basis for $L^2[-1, 1]$. Using this normalization, the first few Legendre polynomials are

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \tfrac{1}{2}(3x^2 - 1), \quad P_3(x) = \tfrac{1}{2}(5x^3 - 3x).$$

By making a change of variables we can easily obtain a similar orthogonal basis of polynomials for $L^2[a,b]$.

The Legendre polynomials arise naturally in a variety of applications. For example, they are solutions to *Legendre's differential equation*

$$\frac{d}{dx}\left((1-x^2)\,\frac{d}{dx}P_n(x)\right) \;+\; n(n+1)\,P_n(x) \;=\; 0.$$

There are many other types of orthogonal polynomials, and they have numerous applications in approximation theory and other areas. We refer to texts such as [Ask75] or [Sze75] for more details on orthogonal polynomials and related systems.

8.3.7 The Haar System

While the Gram–Schmidt procedure is appropriate for constructing some orthonormal bases, it may not suffice when we seek an orthonormal basis whose elements possess some special structure or have some particular properties. For example, in this section we will construct an orthonormal basis for $L^2(\mathbb{R})$ whose elements are obtained by translating and dilating two simple starting functions.

Let $\chi = \chi_{[0,1)}$ be the box function. The function

$$\psi \;=\; \chi_{[0,1/2)} - \chi_{[1/2,1)}$$

is called the *Haar wavelet* or the *square wave*. Given integers $n, k \in \mathbb{Z}$, we create a function $\psi_{n,k}$ by dilating and translating ψ as follows:

$$\psi_{n,k}(x) \;=\; 2^{n/2}\psi(2^n x - k) \;=\; 2^{n/2}\psi\bigl(2^n(x - 2^{-n}k)\bigr), \qquad x \in \mathbb{R}.$$

By direct calculation, $\psi_{n,k} \perp \psi_{n',k'}$ whenever $(n,k) \neq (n',k')$; see the "proof by picture" in Figure 8.3. Furthermore, $\psi_{n,k} \perp \chi(x-j)$ for all integers $n \geq 0$ and $k, j \in \mathbb{Z}$. The *Haar system* for $L^2(\mathbb{R})$ is the orthonormal collection

$$\bigl\{\chi(x-k)\bigr\}_{k\in\mathbb{Z}} \;\cup\; \bigl\{\psi_{n,k}\bigr\}_{n\geq 0,\, k\in\mathbb{Z}}.$$

We will use the Lebesgue Differentiation Theorem to prove that the Haar system is an orthonormal basis for $L^2(\mathbb{R})$.

Theorem 8.3.12. *The Haar system is an orthonormal basis for* $L^2(\mathbb{R})$.

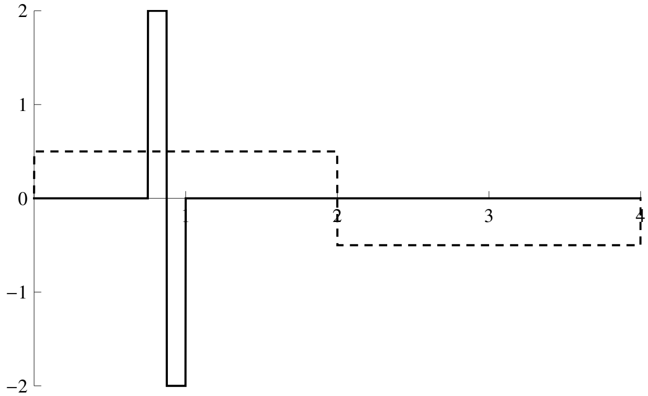


Fig. 8.3 Graphs of $\psi_{-2,0}$ (dashed) and $\psi_{2,3}$ (solid). The product of these two functions is $\psi_{-2,0} \cdot \psi_{2,3} = \frac{1}{2}\psi_{2,3}$, and therefore $\langle \psi_{-2,0}, \psi_{2,3} \rangle = \frac{1}{2} \int \psi_{2,3} = 0$.

Proof. We have already observed that the Haar system is an orthonormal sequence. Therefore, we need only prove that it is complete. Suppose that $f \in L^2(\mathbb{R})$ is orthogonal to every vector in the Haar system. Since the box function $\chi = \chi_{[0,1]}$ and all of its integer translates are elements of the Haar system, this implies that

$$\int_k^{k+1} f(t)\,dt \;=\; 0, \qquad \text{for all } k \in \mathbb{Z}.$$

In particular, since $f \perp \chi$ we have

$$\int_0^{1/2} f(t)\,dt \;+\; \int_{1/2}^1 f(t)\,dt \;=\; \int_0^1 f(t)\,dt \;=\; \langle f, \chi \rangle \;=\; 0.$$

Since f is also orthogonal to the Haar wavelet $\psi = \chi_{[0,1/2)} - \chi_{[1/2,1)}$, we have

$$\int_0^{1/2} f(t)\,dt \;-\; \int_{1/2}^1 f(t)\,dt \;=\; \langle f, \psi \rangle \;=\; 0.$$

Adding and subtracting, we see that

$$\int_0^{1/2} f(t)\,dt \;=\; 0 \;=\; \int_{1/2}^1 f(t)\,dt.$$

Continuing in this way using the other elements of the Haar system, we can show that

$$\int_{I_{n,k}} f(t)\,dt \;=\; 0 \quad \text{for every dyadic interval } I_{n,k} = \Big[\frac{k}{2^n}, \frac{k+1}{2^n}\Big].$$

Let $x \in \mathbb{R}$ be any Lebesgue point of f. For each $n \in \mathbb{N}$, let $J_n(x) = I_{n,k_n(x)}$ be a dyadic interval that contains x. Because of our work above, we have $\int_{J_n(x)} f = 0$. The collection of intervals $\{J_n(x)\}_{n\in\mathbb{N}}$ *shrinks regularly* to x in the sense of Definition 5.5.9, so Theorem 5.5.10 implies that

$$f(x) \;=\; \lim_{n\to\infty} \frac{1}{|J_n(x)|} \int_{J_n(x)} f(t)\,dt \;=\; 0.$$

Since almost every x is a Lebesgue point, it follows that $f = 0$ a.e. Applying Corollary 8.2.18, we conclude that the Haar system is complete in $L^2(\mathbb{R})$. $\square$

If we restrict the Haar system to elements that are supported within the interval $[0,1]$, then we obtain the collection

$$\{\chi\} \;\cup\; \{\psi_{n,k}\}_{n\ge 0,\, k=0,\dots,2^n-1}.$$

This family is an orthonormal basis for $L^2[0,1]$; in fact it is the system that was originally introduced by Haar in 1910 [Haar10]. An English translation of Haar's paper can be found in [HW06].

The Haar system is the simplest example of a *wavelet orthonormal basis* for $L^2(\mathbb{R})$. Wavelets play important roles in harmonic analysis, signal processing, image processing, and other applications. For more details on the construction and application of wavelet bases, we refer to texts such as [Dau92], [KV95], [HW96], [SN96], [Wal02], [Heil11].

8.3.8 Unitary Operators

Now we introduce some terminology and prove some results regarding operators. This material will be applied in Section 9.4, but is not otherwise used in the remainder of the text.

We begin with *isometries*, which are functions that *preserve the norms of vectors*. We will mostly be interested in operators on Hilbert spaces that additionally are linear, but we state the definition for general functions on normed spaces.

Definition 8.3.13 (Isometry). Let X and Y be normed spaces. A function $U\colon X \to Y$ is an *isometry* if

$$\|U(x)\| \;=\; \|x\|, \qquad \text{for all } x \in X. \qquad \diamondsuit$$

Every *linear* isometry is injective, because if $U(x) = U(y)$ then $U(x-y) = U(x) - U(y) = 0$ and therefore $\|x-y\| = \|U(x-y)\| = 0$.

The following example shows that a linear isometry need not be surjective.

Example 8.3.14. The *right-shift operator* is the function $R\colon \ell^2 \to \ell^2$ defined by

$$R(x) \;=\; (0, x_1, x_2, x_3, \dots), \qquad \text{for } x = (x_1, x_2, \dots) \in \ell^2.$$

Since $\|R(x)\|_2 = \|x\|_2$, this function is isometric. It is also linear, but it is not surjective. For example, the first standard basis vector $\delta_1 = (1, 0, 0, \dots)$ does not belong to the range of R.

There is also a *left-shift operator* $L\colon \ell^2 \to \ell^2$, defined by

$$L(x) \;=\; (x_2, x_3, \dots), \qquad \text{for } x = (x_1, x_2, \dots) \in \ell^2.$$

This function is linear and surjective, but it is not injective and it is not isometric since $L(\delta_1) = 0$. $\diamondsuit$

By making use of the Polar Identity, we will prove that a linear isometry on Hilbert spaces *preserves inner products* as well as norms.

Lemma 8.3.15. *Let H and K be Hilbert spaces. If $U\colon H \to K$ is a linear isometry, then $\big\langle U(x), U(y)\big\rangle = \langle x, y\rangle$ for all $x, y \in H$.*

Proof. If x and y are any two vectors in H, then

$$\begin{aligned}
&\|x\|^2 + 2\operatorname{Re}\langle x, y\rangle + \|y\|^2 \\
&\qquad = \|x + y\|^2 && \text{(Polar Identity)} \\
&\qquad = \|U(x) + U(y)\|^2 && \text{(isometry + linear)} \\
&\qquad = \|U(x)\|^2 + 2\operatorname{Re}\big\langle U(x), U(y)\big\rangle + \|U(y)\|^2 && \text{(Polar Identity)} \\
&\qquad = \|x\|^2 + 2\operatorname{Re}\big\langle U(x), U(y)\big\rangle + \|y\|^2 && \text{(isometry).}
\end{aligned}$$

Thus $\operatorname{Re}\big\langle U(x), U(y)\big\rangle = \operatorname{Re}\langle x, y\rangle$. If we are using real scalars, then we are done. If we are using complex scalars, then a similar calculation based on expanding $\|x + iy\|^2$ shows that $\operatorname{Im}\big\langle U(x), U(y)\big\rangle = \operatorname{Im}\langle x, y\rangle$. $\square$

Since a linear isometry is automatically injective, it is a bijection if and only if it is surjective. We have a special name for such operators on Hilbert spaces.

Definition 8.3.16 (Unitary Operators). Let H and K be Hilbert spaces.

(a) A function $U\colon H \to K$ that is *linear, isometric, and surjective* is called a *unitary operator*.

(b) We say that H and K are *unitarily equivalent* if there exists a unitary operator $U\colon H \to K$. $\diamondsuit$

Thus, a unitary operator is a linear bijection that preserves both *lengths of vectors* and *angles between vectors* (because it preserves both norms and inner products). For example, rotations and flips on the Euclidean space $\mathbb{R}^d$ are unitary operators. Here is an example of a unitary operator on an infinite-dimensional Hilbert space.

Theorem 8.3.17. *Every separable infinite-dimensional Hilbert space H is unitarily equivalent to ℓ^2. In particular, if $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for H then the function $U: H \to \ell^2$ defined by $U(x) = \big(\langle x, e_n\rangle\big)_{n\in\mathbb{N}}$ for $x \in H$ is a unitary operator.*

Proof. Theorem 8.3.11 tells us that a separable infinite-dimensional Hilbert space H has an orthonormal basis of the form $\{e_n\}_{n\in\mathbb{N}}$. If $x \in H$, then we have $\sum |\langle x, e_n\rangle|^2 < \infty$ by Bessel's Inequality, so the sequence $U(x) = (\langle x, e_n\rangle)_{n\in\mathbb{N}}$ belongs to ℓ^2. Indeed, the Plancherel Equality implies that

$$\|U(x)\|_2^2 \;=\; \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^2 \;=\; \|x\|^2.$$

Hence U is isometric, and it is clearly linear. Finally, if $c = (c_n)_{n\in\mathbb{N}}$ is any sequence in ℓ^2 then the series $x = \sum c_n e_n$ converges by Theorem 8.3.1, and $U(x) = c$. Therefore U is surjective, so it is unitary. $\square$

Problems

8.3.18. Let $\{e_1, \dots, e_d\}$ be a finite set of orthonormal vectors in a Hilbert space H. Formulate and prove analogues of Theorem 8.3.1, 8.3.6, and 8.3.7 for $\{e_1, \dots, e_d\}$.

8.3.19. Let H be an infinite-dimensional Hilbert space. Prove that H contains an infinite orthonormal sequence $\{e_n\}_{n\in\mathbb{N}}$.

8.3.20. Suppose that $\{e_n\}_{n\in\mathbb{N}}$ is an infinite orthonormal sequence in a Hilbert space H. Prove that $\{e_n\}_{n\in\mathbb{N}}$ contains no convergent subsequences, yet e_n *converges weakly* to 0, i.e., $\langle e_n, x\rangle \to 0$ for every $x \in H$.

8.3.21. Suppose that $\{e_1, \dots, e_d\}$ is an orthonormal basis for a finite-dimensional subspace M of a separable, infinite-dimensional Hilbert space H. Prove that there exist orthonormal vectors $e_{d+1}, e_{d+2}, \dots$ such that $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for H.

8.3.22. Suppose that H is an infinite-dimensional Hilbert space. Prove that the closed unit ball $D = \{x \in H : \|x\| \le 1\}$ is a closed and bounded subset of H that is not compact.

8.3.23. (a) Let X be a Banach space. Show that if an infinite series $\sum x_n$ converges absolutely in X, then it converges unconditionally.

(b) Prove that if H is an infinite-dimensional Hilbert space, then there exists an infinite series $\sum x_n$ in H that converges unconditionally but not absolutely.

8.3.24. Assume that $E \subseteq \mathbb{R}^d$ is measurable and $0 < |E| < \infty$. Prove the following statements.

(a) There exists an infinite orthogonal sequence in $L^2(E)$ of the form $\{\chi_{E_n}\}_{n\in\mathbb{N}}$, where each $E_n \subseteq E$ is measurable and $\sum |E_n| = |E|$.

(b) The rescaled sequence $\mathcal{E} = \bigl\{|E_n|^{-1/2}\,\chi_{E_n}\bigr\}_{n\in\mathbb{N}}$ is orthonormal, but it is not an orthonormal basis for $L^2(E)$.

8.3.25. Assume that $\{e_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for a Hilbert space H.

(a) Suppose that vectors $y_n \in H$ satisfy $\sum \|e_n - y_n\|^2 < 1$. Prove that $\{y_n\}_{n\in\mathbb{N}}$ is a complete sequence in H.

(b) Show that part (a) can fail if we only have $\sum \|e_n - y_n\|^2 = 1$.

8.3.26. The *Rademacher system* is the sequence $\{R_n\}_{n=0}^\infty$ in $L^2[0,1]$ defined by

$$R_n(x) \;=\; \operatorname{sign}(\sin 2^n \pi x),$$

where $\operatorname{sign}(t) = 1$ if $t > 0$, $\operatorname{sign}(0) = 0$, and $\operatorname{sign}(t) = -1$ if $t < 0$. Prove that $\{R_n\}_{n=0}^\infty$ is an orthonormal sequence in $L^2[0,1]$, but $R_1R_2 \perp R_n$ for every $n \geq 0$ and therefore $\{R_n\}_{n=0}^\infty$ is not complete.

Remark: The *Walsh system* is an extension of the Rademacher system that forms an orthonormal basis for $L^2[0,1]$.

8.3.27. Given a finite closed interval $[a,b]$, prove the following statements (in this problem, k implicitly denotes an integer).

(a) $\{x^k\}_{k\geq 0}$ is a complete and linearly independent sequence in $L^2[a,b]$.

(b) $\{x^k\}_{k\geq N}$ is a complete and linearly independent sequence in $L^2[a,b]$ for each integer $N \in \mathbb{N}$.

(c) The set of Legendre polynomials $\{P_k\}_{k\geq 0}$ is complete in $L^2[-1,1]$, but no proper subset is complete.

(d) $\{x^{2k}\}_{k\geq 0}$ is a complete and linearly independent sequence in $L^2[0,1]$.

(e) $\{x^{2k}\}_{k\geq N}$ is a complete and linearly independent sequence in $L^2[0,1]$ for each integer $N \in \mathbb{N}$.

8.3.28. (Vitali [Vit21]) Let $\{e_n\}_{n\in\mathbb{N}}$ be an orthonormal sequence in $L^2[a,b]$. Prove that $\{e_n\}_{n\in\mathbb{N}}$ is complete in $L^2[a,b]$ if and only if

$$\sum_{n=1}^\infty \biggl|\int_a^x e_n(t)\,dt\biggr|^2 \;=\; x - a, \qquad \text{for all } x \in [a,b].$$

8.3.29. (Dalzell [Dal45]) Let $\{f_n\}_{n\in\mathbb{N}}$ be an orthonormal sequence in $L^2[a,b]$. Show that $\{f_n\}_{n\in\mathbb{N}}$ is complete in $L^2[a,b]$ if and only if

$$\sum_{n=1}^\infty \int_a^b \biggl|\int_a^x f_n(t)\,dt\biggr|^2 \;=\; \frac{(b-a)^2}{2}.$$

8.3.30. (Boas and Pollard [BP48]) Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is an orthonormal basis for $L^2[a,b]$. Show that there is a function $m \in L^\infty[a,b]$ such that $\{mf_n\}_{n\ge 2}$ is complete in $L^2[a,b]$.

8.3.31. Let $\ell^2(\mathbb{R})$ consist of all sequences $x = (x_t)_{t\in\mathbb{R}}$ indexed by the real line such that at most countably many components x_t are nonzero and $\sum_{t\in\mathbb{R}} |x_t|^2 < \infty$. Prove that $\ell^2(\mathbb{R})$ is a nonseparable Hilbert space with respect to the inner product $\langle x, y\rangle = \sum_{t\in\mathbb{R}} x_t\,\overline{y_t}$.

8.3.32. (a) Prove that if E and F are measurable subsets of $\mathbb{R}^d$ with $|E|$, $|F| > 0$, then $L^2(E)$ and $L^2(F)$ are unitarily equivalent.

(b) Prove that two finite-dimensional Hilbert spaces H and K are unitarily equivalent if and only if they have the same dimension.

8.4 The Trigonometric System

In this section we will take $\mathbf{F} = \mathbb{C}$ and consider the complex Hilbert space $L^2[0,1]$. For each integer $n \in \mathbb{Z}$, let e_n denote the complex exponential function with frequency n:

$$e_n(x) \;=\; e^{2\pi i n x}, \qquad \text{for } x \in \mathbb{R}.$$

Each function e_n is square-integrable on $[0,1]$. The sequence

$$\{e_n\}_{n\in\mathbb{Z}} \;=\; \{e^{2\pi i n x}\}_{n\in\mathbb{Z}}$$

is called the (complex) *trigonometric system* in $L^2[0,1]$.

If $m \ne n$, then the inner product of e_m with e_n is

$$\langle e_m, e_n\rangle \;=\; \int_0^1 e_m(x)\,\overline{e_n(x)}\,dx \;=\; \int_0^1 e^{2\pi i(m-n)x}\,dx \;=\; \frac{e^{2\pi i(m-n)} - 1}{2\pi i(m-n)} \;=\; 0.$$

Therefore $\{e_n\}_{n\in\mathbb{Z}}$ is an infinite orthonormal sequence in $L^2[0,1]$. It is a much more subtle fact that $\{e_n\}_{n\in\mathbb{Z}}$ is complete in $L^2[0,1]$ and therefore is an orthonormal basis for $L^2[0,1]$. We state this as the following theorem.

Theorem 8.4.1. *The trigonometric system $\{e_n\}_{n\in\mathbb{Z}}$ is complete in $L^2[0,1]$, and therefore it is an orthonormal basis for $L^2[0,1]$.* $\diamondsuit$

After we have further developed the machinery of convolution in Chapter 9, we will prove that the trigonometric system is complete in $L^p[0,1]$ for every finite p, not just for $p = 2$ (this is Theorem 9.3.13). Alternatively, an exposition of a different proof based on the Stone–Weierstrass Theorem can be found in [Heil18, Sec. 5.11]. So, for now we will simply take Theorem 8.4.1

as given, and focus our attention on some implications of the fact that the trigonometric system is an orthonormal basis for $L^2[0,1]$.

If $f \in L^2[0,1]$, then the inner product of f with $e_n(x) = e^{2\pi i n x}$ is called the *nth Fourier coefficient* of f. These scalars are traditionally denoted by $\widehat{f}(n)$. Explicitly writing out the inner products, the Fourier coefficients are

$$\widehat{f}(n) \;=\; \langle f, e_n \rangle \;=\; \int_0^1 f(x)\, e^{-2\pi i n x}\, dx, \qquad \text{for } n \in \mathbb{Z}. \tag{8.11}$$

Applying Theorem 8.3.7, Corollary 8.3.4, and Theorem 8.3.17 to the trigonometric system, and using the notation of equation (8.11), therefore gives us the following result.

Theorem 8.4.2 (Fourier Series for $L^2[0,1]$).

(a) *If* $f \in L^2[0,1]$, *then*

$$f \;=\; \sum_{n\in\mathbb{Z}} \widehat{f}(n)\, e_n, \tag{8.12}$$

where this series converges unconditionally in the norm of $L^2[0,1]$.

(b) Plancherel's Equality: *If* $f \in L^2[0,1]$, *then*

$$\|f\|_2^2 \;=\; \sum_{n\in\mathbb{Z}} |\widehat{f}(n)|^2. \tag{8.13}$$

(c) Parseval's Equality: *If* $f, g \in L^2[0,1]$, *then*

$$\langle f, g \rangle \;=\; \sum_{n\in\mathbb{Z}} \widehat{f}(n)\, \overline{\widehat{g}(n)}. \tag{8.14}$$

(d) *The mapping*

$$U(f) \;=\; \bigl(\widehat{f}(n)\bigr)_{n\in\mathbb{Z}}$$

that sends a function $f \in L^2[0,1]$ *to its sequence of Fourier coefficients defines a unitary operator* $U\colon L^2[0,1] \to \ell^2(\mathbb{Z})$. $\diamondsuit$

Equation (8.12) is called the *Fourier series representation* of f. We often write the Fourier series representation in the form

$$f(x) \;=\; \sum_{n\in\mathbb{Z}} \widehat{f}(n)\, e^{2\pi i n x}, \tag{8.15}$$

but it is important to note that we know only that this series converges in L^2-norm. In general, it need not converge pointwise, even if f is continuous! Indeed, establishing the convergence of Fourier series in senses other than L^2-norm can be very difficult. Given any index $1 < p < \infty$ and any function $f \in L^p[0,1]$, it can be shown that the *symmetric partial sums*

$$S_N f(x) = \sum_{n=-N}^{N} \widehat{f}(n)\, e^{2\pi i n x}$$

converge to f in L^p-norm, but convergence can fail if $p = 1$ or $p = \infty$, even if f is continuous (e.g., see [Kat04, Chap. II] or [Heil11, Chap. 14] for proofs). The *Carleson–Hunt Theorem* states that the symmetric partial sums of the Fourier series of $f \in L^p[0,1]$ converge pointwise almost everywhere to f when $1 < p < \infty$ (see Theorem 9.3.18).

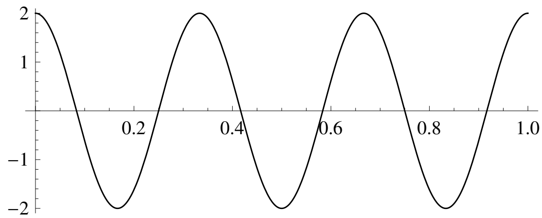


Fig. 8.4 Graph of $\varphi(x) = 2\cos(2\pi 3x)$.

We expand on the meaning of equation (8.15). The graph of the complex exponential function $e^{2\pi i n x}$ is pictured in Figure 9.5. This function is a *pure tone*, and the function $\widehat{f}(n)\, e^{2\pi i n x}$ is a pure tone that has been scaled so that its amplitude is $\widehat{f}(n)$. In general, $\widehat{f}(n)$ is a complex number, but if $\widehat{f}(n)$ is real then the real part of this function is $\widehat{f}(n)\, \cos(2\pi n x)$; see Figure 8.4. This could represent the displacement of the center of an ideal string vibrating at the frequency n with amplitude $\widehat{f}(n)$. It could also represent the displacement of the center of an ideal stereo speaker from its rest position at time x. If you were listening to this ideal speaker, you would hear a "pure tone." Of course, real strings and speakers are quite complicated and do not vibrate as pure tones—there are overtones and other issues. Still, the function $e^{2\pi i n x}$ represents a "pure tone," and the idea of Fourier series is that we can use these pure tones as elementary building blocks for the construction of other, more complicated, signals.

Given two frequencies m and n and amplitudes $\widehat{f}(m)$ and $\widehat{f}(n)$, a function φ of the form

$$\varphi(x) = \widehat{f}(m)\, e^{2\pi i m x} + \widehat{f}(n)\, e^{2\pi i n x}$$

is a superposition of two pure tones. An illustration of the real part of such a superposition appears in Figure 8.5. The real part of a superposition of 75 pure tones with randomly chosen amplitudes is shown in Figure 8.6.

Equation (8.15) says that any function $f \in L^2[0,1]$ can be represented as a sum of pure tones $\widehat{f}(n)\, e^{2\pi i n x}$ over all possible frequencies $n \in \mathbb{Z}$. By superimposing all the pure tones with the correct amplitudes, we create any square-integrable function that we like. The pure tones are our simple "building blocks," and by combining them we can create any sound, or signal, or

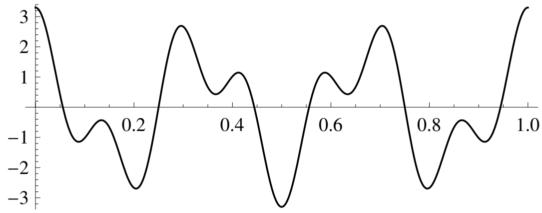


Fig. 8.5 Graph of $\varphi(x) = 2\cos(2\pi 3x) + 1.3\cos(2\pi 7x)$.

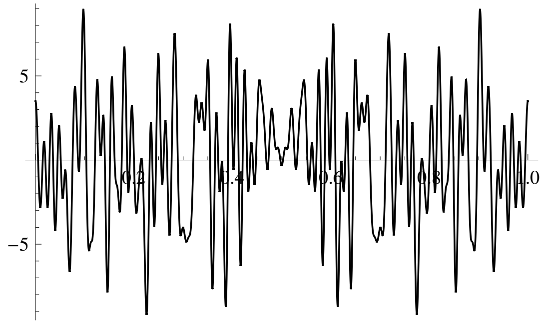


Fig. 8.6 Graph of 75 superimposed pure tones: $\varphi(x) = \sum_{n=1}^{75} \widehat{f}(n)\, \cos(2\pi nx)$.

function. Of course, the "superposition" is an infinite sum and the convergence is in the L^2-norm sense, but still the point is that by combining our very simple special functions $e^{2\pi inx}$ we create very complicated functions f.

We have focused on the domain $[0,1]$. If we like, we can also view $e_n(x) = e^{2\pi inx}$ as a 1-periodic function that is defined on the entire real line. If we take this point of view, then the trigonometric system $\{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis for the space $L^2(\mathbb{T})$ that consists of all 1-*periodic functions* f that satisfy

$$\|f\|_2^2 \;=\; \int_0^1 |f(x)|^2\, dx \;<\; \infty.$$

That is, if we take the domain of e_n to be the entire real line, then we can only represent 1-periodic functions using the trigonometric system. We have an orthonormal basis for $L^2(\mathbb{T})$, but not for $L^2(\mathbb{R})$.

On the other hand, if we separately restrict each of the functions e_n to each of the finite intervals $[k, k+1]$ with $k \in \mathbb{Z}$, then we can piece together trigonometric systems in the following way to create an orthonormal basis for $L^2(\mathbb{R})$.

Exercise 8.4.3. Show that $\mathcal{G} = \bigl\{e^{2\pi inx}\,\chi_{[k,k+1]}\bigr\}_{k,n\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{R})$. $\diamondsuit$

The basis $\mathcal{G}$ is the simplest example of a *Gabor frame* for $L^2(\mathbb{R})$. Gabor frames play an important role in time-frequency analysis, signal processing, and other applications. We refer to [Grö01], [Chr16], or [Heil11, Chap. 11] for more details on Gabor frames and other types of frames and bases. The Gabor frame given in Exercise 8.4.3 is not very pleasant because its elements are discontinuous functions. Examples of Gabor frames whose elements are continuous are given in Problem 8.4.11.

Problems

8.4.4. This problem provides a real-valued analogue of the trigonometric system $\{e^{2\pi inx}\}_{n\in\mathbb{Z}}$. For this problem we assume that scalars are real, so $L^2[0,1]$ is the set of all square-integrable extended real-valued functions on $[0,1]$. Prove that

$$\{1\} \cup \{\sqrt{2}\,\cos 2\pi nx\}_{n\in\mathbb{N}} \cup \{\sqrt{2}\,\sin 2\pi nx\}_{n\in\mathbb{N}}$$

forms an orthonormal basis for $L^2[0,1]$.

8.4.5. Prove that if $f \in L^2[0,1]$, then

$$\lim_{n\to\infty} \int_0^1 f(x)\,\cos 2\pi nx\,dx \;=\; 0 \;=\; \lim_{n\to\infty} \int_0^1 f(x)\,\sin 2\pi nx\,dx.$$

8.4.6. (a) Compute the Fourier coefficients of the Haar wavelet, and use this to show that

$$\frac{\pi^2}{8} \;=\; \sum_{n=1}^{\infty} \frac{1}{(2n-1)^2}.$$

(b) Prove *Euler's Formula*: $\displaystyle \frac{\pi^2}{6} \;=\; \sum_{n=1}^{\infty} \frac{1}{n^2}.$

8.4.7. Let $f(x) = x$ for $x \in [0,1]$. Compute the Fourier coefficients of f, and use this to give another proof of Euler's Formula.

8.4.8. Use the Vitali Criterion (Problem 8.3.28) to prove that the following three statements are equivalent.

(a) The trigonometric system $\{e^{2\pi inx}\}_{n\in\mathbb{Z}}$ is complete in $L^2[0,1]$.

(b) $\displaystyle \sum_{n=1}^{\infty} \frac{1-\cos 2\pi nx}{\pi^2 n^2} \;=\; x - x^2$ for every $x \in [0,1]$.

(c) $\displaystyle \sum_{n=1}^{\infty} \frac{\cos 2\pi nx}{\pi^2 n^2} \;=\; x^2 - x + \frac{1}{6}$ for every $x \in [0,1]$.

8.4.9. Prove that if $f \in L^2[0,1]$ and $\widehat{f} \in \ell^1(\mathbb{Z})$, then f is continuous.

8.4.10. Let $b > 0$ be a fixed positive scalar. This problem will consider the properties of the sequence $\mathcal{E}_b = \{e^{2\pi ibnx}\}_{n\in\mathbb{Z}}$ in the two spaces $L^2[0, b^{-1}]$ and $L^2[0, 1]$. Prove the following statements.

(a) $\mathcal{E}_b$ is an orthogonal (but not orthonormal) basis for $L^2[0, b^{-1}]$.

(b) If $b > 1$, then $\mathcal{E}_b$ is not complete in $L^2[0, 1]$. Explicitly exhibit a nonzero function in $L^2[0, 1]$ that is orthogonal to $e^{2\pi ibnx}$ for every $n \in \mathbb{Z}$.

(c) If $0 < b < 1$, then the following statements hold.

- If $f \in L^2[0, 1]$, then

$$\sum_{n\in\mathbb{Z}} |\langle f, e^{2\pi ibnx}\rangle|^2 \;=\; \frac{1}{b}\,\|f\|_2^2. \tag{8.16}$$

- If $f \in L^2[0, 1]$, then

$$f(x) \;=\; b \sum_{n\in\mathbb{Z}} \langle f, e^{2\pi ibnx}\rangle\, e^{2\pi ibnx},$$

  where this series converges unconditionally in the norm of $L^2[0, 1]$.

- $\{e^{2\pi ibnx}\}_{n\in\mathbb{Z}}$ is not an orthogonal sequence in $L^2[0, 1]$.
- There are at least two distinct choices of coefficients $(c_n)_{n\in\mathbb{Z}}$ such that $1 = \sum_{n\in\mathbb{Z}} c_n e^{2\pi ibnx}$, where these series converge in L^2-norm. (Consequently, $\mathcal{E}_b$ is not a Schauder basis for $L^2[0, 1]$ in the sense of Problem 7.4.11.)

Remark: Using terminology from frame theory, equation (8.16) says that $\mathcal{E}_b$ is a *tight frame* for $L^2[0, 1]$. The *Classical* (or *Shannon*) *Sampling Theorem* is a consequence of this fact; see [Heil11, Thm. 10.7].

8.4.11. (a) Let $a, b > 0$ be fixed. Suppose that $g \in L^2(\mathbb{R})$ is such that

- $g = 0$ a.e. outside of the interval $[0, \frac{1}{b}]$, and
- $\sum_{k\in\mathbb{Z}} |g(x - ak)|^2 = 1$ a.e.

Set $g_{kn}(x) = e^{2\pi ibnx} g(x - ak)$ for $k, n \in \mathbb{Z}$, and prove that the *Gabor system* $\mathcal{G} = \{g_{kn}\}_{k,n\in\mathbb{Z}}$ satisfies

$$\sum_{k,n\in\mathbb{Z}} |\langle f, g_{kn}\rangle|^2 \;=\; \frac{1}{b}\,\|f\|_2^2, \qquad \text{for all } f \in L^2(\mathbb{R}). \tag{8.17}$$

Remark: Using the language of frame theory, equation (8.17) says that $\mathcal{G}$ is a *tight frame* for $L^2(\mathbb{R})$; see [Grö01] or [Heil11].

(b) Exhibit a continuous function g and corresponding constants $a, b > 0$ such that the hypotheses of part (a) are satisfied. Prove that for this choice of g, a, and b, the Gabor system $\mathcal{G}$ is not an orthogonal sequence.

8.4.12. For each $\xi \in \mathbb{R}$, define $e_\xi(t) = e^{2\pi i \xi t}$ for $t \in \mathbb{R}$. Let $H = \mathrm{span}\{e_\xi\}_{\xi\in\mathbb{R}}$ be the finite linear span of the family $\{e_\xi\}_{\xi\in\mathbb{R}}$. Show that

$$\langle f, g\rangle \;=\; \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(t)\,\overline{g(t)}\,dt, \qquad f, g \in H,$$

defines an inner product on H, and $\{e_\xi\}_{\xi\in\mathbb{R}}$ is an uncountable orthonormal system in H.

Remark: H is not complete, but its *completion* $\widetilde{H}$ is an important nonseparable Hilbert space that contains the class of *almost periodic functions*, see [Kat04].

8.4.13. For each $n \in \mathbb{Z}$ let $e_n(x) = e^{2\pi i n x}$. For $n \neq 0$, define

$$f_n(x) \;=\; x\, e_n(x) \qquad \text{and} \qquad g_n(x) \;=\; \frac{e_n(x) - 1}{x}.$$

Let $\mathcal{F} = \{f_n\}_{n\neq 0}$ and $\mathcal{G} = \{g_n\}_{n\neq 0}$. For this problem, we order $\mathbb{Z}\backslash\{0\}$ as

$$\mathbb{Z}\backslash\{0\} \;=\; \{1, -1, 2, -2, 3, -3, \dots\}.$$

This means that a series of the form $h = \sum_{n\neq 0} h_n$ converges if and only if the partial sums of

$$h_1 + h_{-1} + h_2 + h_{-2} + h_3 + h_{-3} + \cdots$$

converge to h in L^2-norm. Prove the following statements.

(a) f_n and g_n belong to $L^2[0,1]$, and their norms satisfy $\|f_n\|_2 \;=\; 3^{-1/2}$ and $\lim_{n\to\infty} \|g_n\|_2 \;=\; \infty$.

(b) $\mathcal{F}$ and $\mathcal{G}$ are *biorthogonal*, i.e.,

$$\langle f_m, g_n\rangle \;=\; \delta_{mn}, \qquad \text{all } m \neq 0 \text{ and } n \neq 0.$$

(c) $\mathcal{F}$ is *minimal*, i.e., for each $m \neq 0$ the function f_m does not belong to the closed span of the remaining functions f_n:

$$f_m \;\notin\; \overline{\mathrm{span}}\big(\{f_n\}_{n\neq m, n\neq 0}\big), \qquad \text{for all } m \neq 0$$

(see Problem 8.2.24). As a consequence, $\mathcal{F}$ is finitely linearly independent.

(d) $\mathcal{F}$ is complete, i.e., $\overline{\mathrm{span}}(\mathcal{F}) = L^2[0,1]$.

(e) If c_n are scalars and the series $f = \sum_{n\neq 0} c_n f_n$ converges, then $c_n = \langle f, g_n\rangle$ for every $n \neq 0$, and $c_n \to 0$ as $n \to \pm\infty$.

(f)* The constant function 1 belongs to $\overline{\mathrm{span}}(\mathcal{F})$, but there do not exist any scalars c_n such that

$$1 \;=\; \sum_{n\neq 0} c_n f_n.$$

Chapter 9
Convolution and the Fourier Transform

In this chapter we will present several mathematical applications of the Lebesgue integral and the L^p spaces. In Section 9.1 we study the *convolution* of functions. Using this operation we will prove, for example, that the space $C_c^\infty(\mathbb{R})$ of infinitely differentiable, compactly supported functions is dense in $L^p(\mathbb{R})$ for all finite p. Then in Section 9.2 we introduce the *Fourier transform*, which is the central operation of harmonic analysis for functions on the real line. In Section 9.3 we study *Fourier series*, which is the analogue of the Fourier transform for periodic functions. In particular, we prove that the trigonometric system $\{e^{2\pi inx}\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $L^2[0,1]$. Finally, in Section 9.4 we prove that the Fourier transform can be extended from $L^1(\mathbb{R})$ to $L^2(\mathbb{R})$. In particular, the Fourier transform is a *unitary mapping* of $L^2(\mathbb{R})$ onto itself, and we explain why this is the correct analogy for the Fourier transform of the fact that the trigonometric system is an orthonormal basis for $L^2(\mathbb{T})$.

9.1 Convolution

We introduced the convolution of integrable functions on $\mathbb{R}^d$ in Section 4.6.3, and now we will consider this operation in detail. Convolution is an extremely useful operation that plays important roles in harmonic analysis, physics, signal processing, and many other areas. For more details on convolution and its applications beyond what is presented here we refer to texts such [DM72], [Ben97], [Kat04], or [Heil11].

For simplicity, in this section we will take the domain of our functions to be the real line $\mathbb{R}$, but entirely similar results hold for functions on $\mathbb{R}^d$. Later we will also consider convolution of sequences indexed by $\mathbb{Z}$ (see Problem 9.1.18) and convolution of 1-periodic functions (in Section 9.3.3). In fact, convolution can be defined much more generally; all we require is that the domain of our functions be a *locally compact group* (although if the group is not commuta-

C. Heil, *Introduction to Real Analysis*, Graduate Texts in Mathematics 280,
https://doi.org/10.1007/978-3-030-26903-6_9

tive then there is a difference between *left* and *right* convolution). We refer to [HR79] or [Rud90] for more details on convolution on abstract groups.

9.1.1 The Definition of Convolution

We defined convolution in Section 4.6.3, but for convenience we recall the formal definition here.

Definition 9.1.1 (Convolution). Let f and g be measurable functions on the real line $\mathbb{R}$. The *convolution* of f and g is the function $f * g$ defined by

$$(f * g)(x) = \int_{-\infty}^{\infty} f(y)\, g(x-y)\, dy, \tag{9.1}$$

as long as this integral exists. $\diamondsuit$

The convolution of two arbitrary measurable functions f and g will not always exist. For example, if $f(x) = x$ and $g(x) = 1$ then $(f * g)(x)$ is not defined for any x. Consequently, when we speak of a convolution, we must be careful to prove that $f * g$ exists in some sense—perhaps for all x, or perhaps only for almost every x. We will give several different conditions on f and g that imply that their convolution exists.

It is instructive to compute at least one convolution by hand. The following exercise shows that the convolution of the box function $\chi_{[-\frac{1}{2},\frac{1}{2}]}$ with itself is the hat function on the interval $[-1, 1]$.

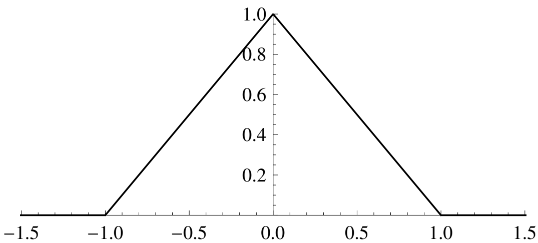


Fig. 9.1 Graph of the hat function W.

Exercise 9.1.2. Let $\chi = \chi_{[-\frac{1}{2},\frac{1}{2}]}$, and let

$$W(x) = \max\{1 - |x|, 0\}$$

be the hat function on $[-1, 1]$ that is pictured in Figure 9.1. Show that

$$\chi * \chi = W. \qquad \diamondsuit$$

Note that $\chi * \chi$ is continuous, while χ is discontinuous. This is typical—convolution tends to be a type of smoothing procedure.

9.1.2 Existence

In Section 4.6.3, we used Fubini's and Tonelli's Theorems to establish one sufficient condition for the existence of a convolution. Specifically, we saw in Theorem 4.6.11 that if f and g are both integrable, then $f * g$ is defined a.e. and is integrable. Some other properties of the convolution of integrable functions were obtained in Problem 4.6.26. For convenience, we summarize those facts as the following theorem.

Theorem 9.1.3. *If f, g, $h \in L^1(\mathbb{R})$, then the following statements hold.*

(a) $F(x,y) = f(y)\,g(x-y)$ *is an integrable function on* $\mathbb{R}^2$.

(b) $(f*g)(x)$ *exists for almost every* $x \in \mathbb{R}$.

(c) $f*g$ *is measurable, and* $f*g \in L^1(\mathbb{R})$.

(d) $\|f*g\|_1 \le \|f\|_1\,\|g\|_1$.

(e) $f*g = g*f$ *a.e.*

(f) $(f*g)*h = f*(g*h)$ *a.e.*

(g) $f*(ag+bh) = a(f*g) + b(f*h)$ *a.e. for all scalars a and b.*

(h) *Convolution commutes with translation, i.e.,*

$$f*(T_a g) \;=\; (T_a f)*g \;=\; T_a(f*g) \quad \textit{for all } a \in \mathbb{R}. \qquad \diamondsuit$$

In summary, Theorem 9.1.3 tells us that $L^1(\mathbb{R})$ is closed with respect to convolution, convolution is commutative and associative and satisfies the distributive laws, and it also satisfies the submultiplicative norm inequality $\|f*g\|_1 \le \|f\|_1\,\|g\|_1$. Using the language of functional analysis, this says that $L^1(\mathbb{R})$ is a *commutative Banach algebra* with respect to convolution. One interesting feature of this algebra that we will prove in Section 9.2 is that *there is no identity element for convolution in* $L^1(\mathbb{R})$.

Next we will give a different type of sufficient condition for the existence of a convolution. Since $(f*g)(x)$ is the integral of $f(y)\,g(x-y)$ with respect to the variable y, in order for $(f*g)(x)$ to exist at a particular point x, the product $f(y)\,g(x-y)$ must be an integrable function of y. The simplest sufficient condition that ensures that a product is integrable is provided by Hölder's Inequality, which says that the product of a function in $L^p(\mathbb{R})$ with a function in $L^{p'}(\mathbb{R})$ is integrable. The next exercise develops this idea, and derives some of the properties of $f*g$ when f and g lie in dual Lebesgue spaces. The special case $p=1$ was considered earlier in Problem 4.6.27.

Exercise 9.1.4. Fix $1 \le p \le \infty$. Prove that if $f \in L^p(\mathbb{R})$ and $g \in L^{p'}(\mathbb{R})$, then the following statements hold.

(a) $(f*g)(x)$ is defined at *every* point $x \in \mathbb{R}$, and $(f*g)(x) = (g*f)(x)$.

(b) $f*g$ is bounded, and $\|f*g\|_\infty \le \|f\|_p \, \|g\|_{p'}$.

(c) For all $a, x \in \mathbb{R}$,

$$\big|(f*g)(x) - (f*g)(x-a)\big| \le \|f\|_p \, \|g - T_a g\|_{p'}. \tag{9.2}$$

(d) $f*g$ is continuous and bounded. Hence $f*g \in C_b(\mathbb{R})$, and we have $\|f*g\|_{\mathrm{u}} \le \|f\|_p \, \|g\|_{p'}$.

(e) $f*g$ is *uniformly* continuous on $\mathbb{R}$. $\diamondsuit$

Thus, if $f \in L^p(\mathbb{R})$ and $g \in L^{p'}(\mathbb{R})$, then the convolution $f*g$ is defined at every point and $f*g$ is bounded and uniformly continuous. As we will discuss below, this is a reflection of the fact that convolution tends to be a smoothing process.

For indices in the range $1 < p < \infty$, we can prove a bit more.

Theorem 9.1.5. *Assume $1 < p < \infty$. If $f \in L^p(\mathbb{R})$ and $g \in L^{p'}(\mathbb{R})$, then $f*g \in C_0(\mathbb{R})$.*

Proof. We know from Exercise 9.1.4 that $f*g$ belongs to $C_b(\mathbb{R})$. In order to prove that $f*g$ belongs to the smaller space $C_0(\mathbb{R})$, we will show that there exist functions $h_n \in C_0(\mathbb{R})$ that converge uniformly to $f*g$. Since $C_0(\mathbb{R})$ is closed under uniform limits, this will imply that $f*g$ belongs to $C_0(\mathbb{R})$.

Since p is finite, Exercise 7.3.11 tells us that $C_c(\mathbb{R})$ is dense in $L^p(\mathbb{R})$. Therefore, there exist functions $f_n \in C_c(\mathbb{R})$ such that $f_n \to f$ in L^p-norm. Since convergent sequences in a normed space are bounded, we have

$$M = \sup_{n\in\mathbb{N}} \|f_n\|_p < \infty.$$

On the other hand, we have $1 < p' < \infty$, so $C_c(\mathbb{R})$ is dense in $L^{p'}(\mathbb{R})$ as well. Therefore there exist functions $g_n \in C_c(\mathbb{R})$ such that $g_n \to g$ in $L^{p'}$-norm.

By Problem 4.6.28, $C_c(\mathbb{R})$ is closed under convolution. Hence the function $h_n = f_n * g_n$ belongs to $C_c(\mathbb{R})$, which is a subspace of $C_0(\mathbb{R})$. Since

$$\begin{aligned}
\|f*g - h_n\|_{\mathrm{u}} &\le \|f*g - f_n*g\|_{\mathrm{u}} + \|f_n*g - f_n*g_n\|_{\mathrm{u}} \\
&\le \|f-f_n\|_p \, \|g\|_{p'} + \|f_n\|_p \, \|g-g_n\|_{p'} \\
&\le \|f-f_n\|_p \, \|g\|_{p'} + M \, \|g-g_n\|_{p'} \\
&\to 0 \quad \text{as } n \to \infty,
\end{aligned}$$

we see that $h_n \to f*g$ uniformly. But each function h_n belongs to $C_0(\mathbb{R})$, so it follows that $f*g \in C_0(\mathbb{R})$. $\square$

Theorem 9.1.5 does not extend to $p = 1$. For example, if $f = \chi_{[0,1]}$ and $g = 1$ then $f \in L^1(\mathbb{R})$ and $g \in L^\infty(\mathbb{R})$, but their convolution is

$$(f * g)(x) \;=\; \int_{-\infty}^{\infty} f(y)\, g(x-y)\, dy \;=\; \int_{-\infty}^{\infty} \chi_{[0,1]}(y)\, dy \;=\; 1.$$

We do have $f*g \in C_b(\mathbb{R})$, but $f*g \notin C_0(\mathbb{R})$. On the other hand, the following exercise shows that Theorem 9.1.5 does extend to $p = 1$ if we replace the hypothesis $g \in L^\infty(\mathbb{R})$ with $g \in C_0(\mathbb{R})$.

Exercise 9.1.6. Show that if $f \in L^1(\mathbb{R})$ and $g \in C_0(\mathbb{R})$, then $f * g$ belongs to $C_0(\mathbb{R})$. $\diamondsuit$

9.1.3 Convolution as Averaging

Now we take a closer look at the meaning of convolution. For each number $T > 0$, let

$$\chi_T \;=\; \frac{1}{2T}\chi_{[-T,T]}.$$

This is a characteristic function that has been rescaled so that $\int \chi_T = 1$ for every T. The convolution of a function $f \in L^1(\mathbb{R})$ with χ_T at a point $x \in \mathbb{R}$ is

$$(f * \chi_T)(x) \;=\; \int_{-\infty}^{\infty} f(y)\, \chi_T(x-y)\, dy \;=\; \frac{1}{2T}\int_{x-T}^{x+T} f(y)\, dy. \tag{9.3}$$

This is precisely the average of f on the interval $[x-T, x+T]$ (see Figure 9.2). Since χ_T is bounded, Exercise 9.1.4 implies that $f * \chi_T$ is continuous. Thus $f * \chi_T$ is a smoothed, averaged version of f.

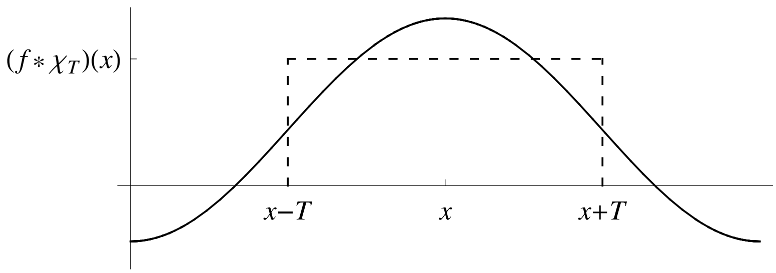


Fig. 9.2 The height of the dashed box is $(f * \chi_T)(x)$. The area of the dashed box is $\int_{x-T}^{x+T} f(y)\, dy$, which equals the area under the graph of f between $x - T$ and $x + T$.

For a generic function g, the convolution of f and g can be interpreted as a weighted average of f, with g weighting some parts of the domain more than others. Technically, it may be better to think of the function $g^*(x) = g(-x)$ as the weighting function, since g^* is the function being translated when we compute

$$(f * g)(x) \;=\; \int_{-\infty}^{\infty} f(y)\, g^*(y-x)\, dy \;=\; \int_{-\infty}^{\infty} f(y)\, T_x g^*(y)\, dy.$$

In any case, $(f * g)(x)$ is a weighted average of f around the point x. Alternatively, since convolution is commutative, we can equally view it as an averaging of g using the weighting corresponding to $f^*(x) = f(-x)$.

We usually think of averaging as a smoothing process, and the next exercise presents a quantitative version of this statement. To motivate this exercise, note that if we formally interchange an integral and a derivative, then we obtain

$$\begin{aligned}
\frac{d}{dx}(f * g)(x) \;&=\; \frac{d}{dx}\int_{-\infty}^{\infty} f(y)\, g(x-y)\, dy && \text{(definition)}\\
&=\; \int_{-\infty}^{\infty} f(y)\, \frac{d}{dx} g(x-y)\, dy && \text{(unjustified step)}\\
&=\; \int_{-\infty}^{\infty} f(y)\, g'(x-y)\, dy && \text{(chain rule)}\\
&=\; (f * g')(x) && \text{(definition).}
\end{aligned}$$

This is only a formal calculation, but it suggests that if g is differentiable, then $f*g$ should be differentiable as well and we should have $(f*g)' = f*g'$. The next exercise asks for a justification of this argument (one approach is to treat the derivative as a limit and apply the Dominated Convergence Theorem). Once this is done, it is straightforward to extend to higher derivatives by induction. Recall that $C_b^m(\mathbb{R})$ denotes the space of all m-times differentiable functions g such that each of $g, g', \dots, g^{(m)}$ is continuous and bounded. Similarly, $C_b^\infty(\mathbb{R})$ is the space of all infinitely differentiable functions g such that $g^{(k)}$ is bounded for every k.

Exercise 9.1.7. (a) Prove that differentiation commutes with convolution in the following sense: If $f \in L^1(\mathbb{R})$ and $g \in C_b^1(\mathbb{R})$, then $f * g \in C_b^1(\mathbb{R})$ and

$$(f * g)' \;=\; f * g'.$$

(b) Extend part (a) to higher derivatives. Specifically, prove that if $f \in L^1(\mathbb{R})$ and $g \in C_b^m(\mathbb{R})$ for some $m \in \mathbb{N}$, then $f * g \in C_b^m(\mathbb{R})$ and

$$(f * g)^{(k)} \;=\; f * g^{(k)}, \qquad \text{for } k = 0, \dots, m.$$

(c) Prove that if $f \in L^1(\mathbb{R})$ and $g \in C_b^\infty(\mathbb{R})$, then $f * g \in C_b^\infty(\mathbb{R})$ and

$$(f * g)^{(k)} = f * g^{(k)}, \qquad \text{for all } k \geq 0. \qquad \diamondsuit$$

In summary, the convolution $f * g$ "inherits" the smoothness of g. Since convolution is commutative, $f * g$ similarly inherits smoothness from f.

9.1.4 Approximate Identities

Consider again equation (9.3), which states that $(f * \chi_T)(x)$ is the average of f over the interval $[x-T, x+T]$. What happens to this average as $T \to 0$? As T decreases, the function $\chi_T = \frac{1}{2T}\chi_{[-T,T]}$ becomes a taller and taller "spike" centered at the origin, with the height of the spike chosen so that the integral of χ_T is always 1. Intuitively, averaging over smaller and smaller intervals should give values $(f * \chi_T)(x)$ that are closer and closer to $f(x)$. This intuition is made precise in the *Lebesgue Differentiation Theorem* (Theorem 5.5.7), which states that if $f \in L^1(\mathbb{R})$, or even if f is merely locally integrable, then

$$f(x) = \lim_{T \to 0} (f * \chi_T)(x) \quad \text{for almost every } x \in \mathbb{R}.$$

Thus $f \approx f * \chi_T$ when T is small. Although there is no identity element for convolution in $L^1(\mathbb{R})$, the function χ_T is *approximately* an identity for convolution, and this approximation becomes better and better the smaller that T becomes (although the rate of convergence will be different for each function f).

The Lebesgue Differentiation Theorem deals with pointwise a.e. convergence. Here we will concentrate on convergence in L^1-norm. We will prove that we can create many different sequences of functions $\{k_N\}_{N\in\mathbb{N}}$ such that $f * k_N \to f$ in L^1-norm for every $f \in L^1(\mathbb{R})$. The following definition specifies the exact properties that we need the functions k_N to possess.

Definition 9.1.8 (Approximate Identity). An *approximate identity* or *summability kernel* on $\mathbb{R}$ is a family $\{k_N\}_{N\in\mathbb{N}}$ of functions in $L^1(\mathbb{R})$ such that the following three conditions are satisfied.

(a) L^1*-normalization*: $\int_{-\infty}^{\infty} k_N(x)\,dx = 1$ for every N.

(b) L^1*-boundedness*: $\sup \|k_N\|_1 = \sup \int_{-\infty}^{\infty} |k_N(x)|\,dx < \infty$.

(c) L^1*-concentration*: For every $\delta > 0$,

$$\lim_{N\to\infty} \int_{|x|\geq\delta} |k_N(x)|\,dx = 0. \qquad \diamondsuit$$

Property (a) of this definition says that each function k_N has the same total "signed mass" in the sense that its integral is 1, and property (c) says that most of this mass is being squeezed into smaller and smaller intervals around the origin as N increases. Property (b) requires the "absolute mass" of k_N to be bounded independently of N. If $k_N \ge 0$ for every N, then property (a) implies that $\|k_N\|_1 = 1$ for every N, so property (b) is automatically satisfied in this case.

The next exercise describes the "easy" way to construct an approximate identity: Simply choose any integrable function k whose integral is 1, and then dilate k appropriately to create k_N.

Exercise 9.1.9. Let $k \in L^1(\mathbb{R})$ be any function that satisfies

$$\int_{-\infty}^{\infty} k(x)\,dx \;=\; 1.$$

Define k_N by an L^1-normalized dilation:

$$k_N(x) \;=\; N\,k(Nx), \qquad \text{for } N \in \mathbb{N}.$$

Prove that the family $\{k_N\}_{N\in\mathbb{N}}$ forms an approximate identity. $\diamondsuit$

Thus, to create an approximate identity, all we need to do is to choose an integrable function k whose integral is 1, and set $k_N(x) = N\,k(Nx)$. We can impose whatever extra properties on k that are convenient for our application. For example, if we let k be smooth, then every k_N will be smooth, and this smoothness will be inherited by $f * k_N$.

Here is one particular approximate identity that appears often in applications of convolution in harmonic analysis.

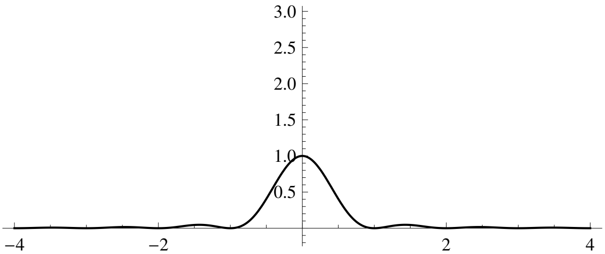
Exercise 9.1.10 (The Fejér Kernel). The *Fejér function* is

$$w(x) \;=\; \left(\frac{\sin \pi x}{\pi x}\right)^2,$$

and the *Fejér kernel* is $\{w_N\}_{N\in\mathbb{N}}$ where $w_N(x) = Nw(Nx)$. Prove that w is integrable and $\int w = 1$. Conclude that the Fejér kernel is an approximate identity. $\diamondsuit$

The letter "w" is for "Weiss," which was Fejér's surname at birth. Plots of w and w_3 appears in Figure 9.3. We can see in that figure that w_N becomes more spike-like as N increases, just as χ_T becomes more spike-like as $T \to 0$.

Now we prove our claim that if $\{k_N\}_{N\in\mathbb{N}}$ is an approximate identity, then $f * k_N \to f$ in L^1-norm for every function $f \in L^1(\mathbb{R})$. The proof of this theorem illustrates two "standard tricks." First, we introduce k_N into one term of the computation by using the fact that $\int k_N = 1$. Second, we divide the domain of integration into small and large parts in order to make use of the L^1-concentration property of an approximate identity.



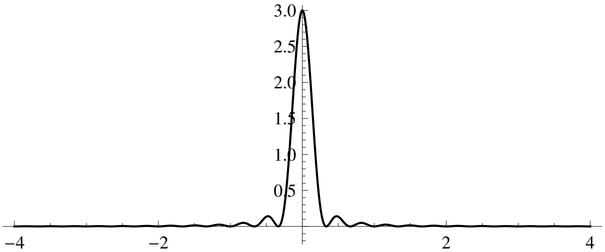


Fig. 9.3 Top: The *Fejér function* $w(x) = \left(\frac{\sin \pi x}{\pi x}\right)^2$. Bottom: The dilation $w_3(x) = 3w(3x)$ of the Fejér function.

Theorem 9.1.11. *If $\{k_N\}_{N\in\mathbb{N}}$ is an approximate identity, then*

$$\lim_{N\to\infty} \|f - f * k_N\|_1 \;=\; 0 \quad \textit{for every } f \in L^1(\mathbb{R}).$$

Proof. Fix any $f \in L^1(\mathbb{R})$. Since $k_N \in L^1(\mathbb{R})$, we know that $f * k_N \in L^1(\mathbb{R})$, and we wish to show that $f * k_N$ approximates f well in L^1-norm. Using the fact that $\int k_N = 1$, we compute that

$$\begin{aligned}
\|f - f * k_N\|_1 &= \int_{-\infty}^{\infty} |f(x) - (f * k_N)(x)|\, dx \\
&= \int_{-\infty}^{\infty} \left| f(x) \int_{-\infty}^{\infty} k_N(t)\, dt \;-\; \int_{-\infty}^{\infty} f(x-t)\, k_N(t)\, dt \right| dx \\
&\le \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x) - f(x-t)|\, |k_N(t)|\, dt\, dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x) - f(x-t)|\, |k_N(t)|\, dx\, dt \qquad \text{(by Tonelli)} \\
&= \int_{-\infty}^{\infty} |k_N(t)| \int_{-\infty}^{\infty} |f(x) - T_t f(x)|\, dx\, dt \\
&= \int_{-\infty}^{\infty} |k_N(t)|\, \|f - T_t f\|_1\, dt. \qquad (9.4)
\end{aligned}$$

We were allowed to interchange the order of integration in this calculation because the integrands are nonnegative. We want to show that the quantity in equation (9.4) is small when N is large.

Choose any $\varepsilon > 0$. Problem 7.3.16 tells us that translation is *strongly continuous* on $L^1(\mathbb{R})$, i.e., there exists a $\delta > 0$ such that

$$|t| < \delta \quad \Longrightarrow \quad \|f - T_t f\|_1 < \varepsilon.$$

The L^1-boundedness property of an approximate identity implies that

$$K = \sup_{N \in \mathbb{N}} \|k_N\|_1 < \infty,$$

and by the L^1-concentration property we know that there is some $N_0 > 0$ such that $\int_{|t| \geq \delta} |k_N(t)|\, dt < \varepsilon$ for all $N \geq N_0$. Therefore, for $N \geq N_0$ we can continue equation (9.4) as follows:

$$\begin{aligned}
(9.4) &= \int_{|t|<\delta} |k_N(t)|\, \|f - T_t f\|_1\, dt + \int_{|t| \geq \delta} |k_N(t)|\, \|f - T_t f\|_1\, dt \\
&\leq \int_{|t|<\delta} |k_N(t)|\, \varepsilon\, dt + \int_{|t| \geq \delta} |k_N(t)|\, \big(\|f\|_1 + \|T_t f\|_1\big)\, dt \\
&\leq \varepsilon \int_{-\infty}^{\infty} |k_N(t)| + 2\, \|f\|_1 \int_{|t| \geq \delta} |k_N(t)|\, dt \\
&\leq \varepsilon K + 2\, \|f\|_1\, \varepsilon.
\end{aligned}$$

Thus $\|f - f * k_N\|_1 \to 0$ as $N \to \infty$. □

Figure 9.4 illustrates the convergence derived in the preceding theorem. We use the Fejér kernel $\{w_N\}_{N \in \mathbb{N}}$ constructed in Exercise 9.1.10, and depict the convolution of the box function $\chi_{[0,1]}$ with some elements of the Fejér kernel. Specifically, in Figure 9.4 we see the convolutions $\chi_{[0,1]} * w_N$ for $N = 1$, 5, and 25. From Exercise 9.1.7 we know that $\chi_{[0,1]} * w_N$ is a continuous function, and Theorem 9.1.11 tells us that $\chi * w_N$ converges to χ in L^1-norm as N increases. This is in agreement with what we see in Figure 9.4.

We proved in Theorem 4.5.8 that $C_c(\mathbb{R})$ is dense in $L^1(\mathbb{R})$. We will use Theorem 9.1.11 to show that the seemingly "much tinier" space $C_c^\infty(\mathbb{R})$ is also dense in $L^1(\mathbb{R})$.

Theorem 9.1.12. *$C_c^\infty(\mathbb{R})$ is dense in $L^1(\mathbb{R})$.*

Proof. Let $k \in C_c^\infty(\mathbb{R})$ be any function that satisfies $\int k = 1$ (see Problem 9.1.26 for one construction of such a function). If we set $k_N(x) = N\, k(Nx)$, then $\{k_N\}_{N \in \mathbb{N}}$ is an approximate identity, and $\|k_N\|_1 = \|k\|_1$ for every N.

Choose any function $f \in L^1(\mathbb{R})$. Since k_N is infinitely differentiable, Exercise 9.1.7 implies that $f * k_N$ is also infinitely differentiable. However, $f * k_N$

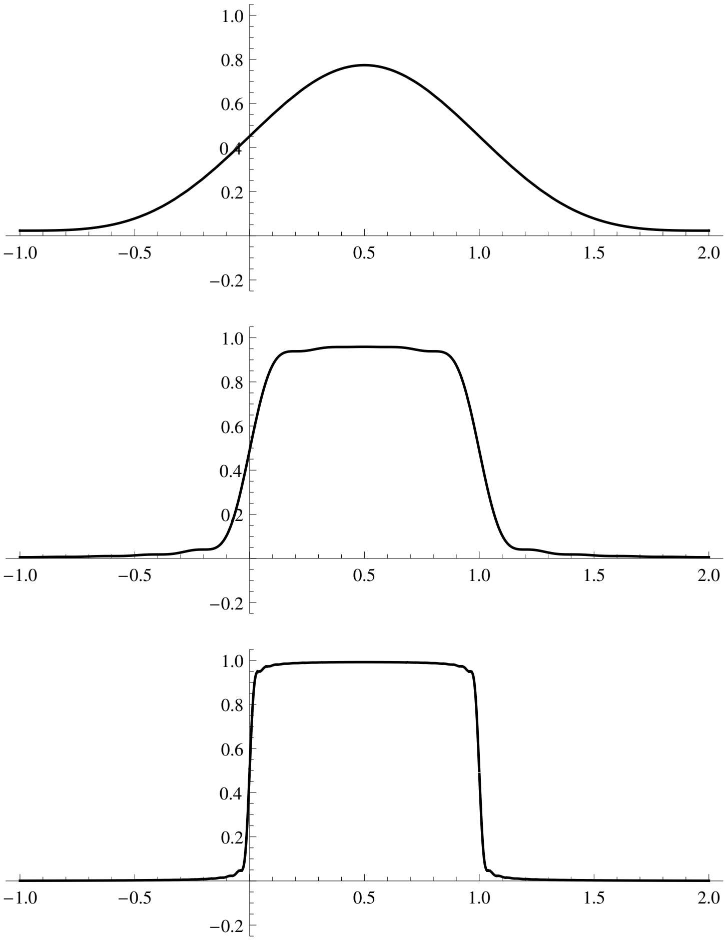


Fig. 9.4 Convolution of the characteristic function $\chi_{[0,1]}$ with some elements of the Fejér kernel $\{w_N\}_{N\in\mathbb{N}}$. Top: $\chi_{[0,1]} * w$. Middle: $\chi_{[0,1]} * w_5$. Bottom: $\chi_{[0,1]} * w_{25}$.

need not be compactly supported. Therefore, we instead consider the functions

$$f_N = (f \cdot \chi_{[-N,N]}) * k_N, \qquad \text{for } N \in \mathbb{N}. \tag{9.5}$$

Because $f \cdot \chi_{[-N,N]}$ is integrable and k_N is infinitely differentiable, f_N is also infinitely differentiable. Since $f \cdot \chi_{[-N,N]}$ is zero a.e. outside of $[-N, N]$ and k_N is identically zero outside of some interval $[a, b]$, a direct calculation shows that their convolution, which is f_N, is identically zero outside of the interval $[-N + a, N + b]$. Therefore f_N belongs to $C_c^\infty(\mathbb{R})$.

Now, Theorem 9.1.11 tells us that $f * k_N \to f$ in L^1-norm. Further, the Dominated Convergence Theorem implies that $f \cdot \chi_{[-N,N]} \to f$ in L^1-norm. Consequently,

$$\begin{aligned}
\|f - f_N\|_1 &\le \|f - f * k_N\|_1 + \|f * k_N - (f \cdot \chi_{[-N,N]}) * k_N\|_1 \\
&= \|f - f * k_N\|_1 + \|(f - f \cdot \chi_{[-N,N]}) * k_N\|_1 \\
&\le \|f - f * k_N\|_1 + \|f - f \cdot \chi_{[-N,N]}\|_1 \, \|k_N\|_1 \\
&= \|f - f * k_N\|_1 + \|f - f \cdot \chi_{[-N,N]}\|_1 \, \|k\|_1 \\
&\to 0 \quad \text{as } N \to \infty.
\end{aligned}$$

Therefore $C_c^\infty(\mathbb{R})$ is dense in $L^1(\mathbb{R})$. $\square$

Since $C_c^\infty(\mathbb{R}) \subseteq C_c^m(\mathbb{R})$, a corollary is that $C_c^m(\mathbb{R})$ is dense in $L^1(\mathbb{R})$ for every integer $m \in \mathbb{N}$.

9.1.5 Young's Inequality

Now we will show that most of the results of Section 9.1.4 can be extended from $L^1(\mathbb{R})$ to $L^p(\mathbb{R})$ for indices in the range $1 \le p < \infty$. There is also an extension for $p = \infty$, but for that case the appropriate extension space is $C_0(\mathbb{R})$ rather than $L^\infty(\mathbb{R})$. The key to the extension is given in the following exercise.

Exercise 9.1.13. Fix $1 < p < \infty$, and let $f \in L^p(\mathbb{R})$ and $g \in L^1(\mathbb{R})$ be given. Assume first that f and g are nonnegative, and apply Tonelli's Theorem to show that the integral defining $(f * g)(x)$ exists for a.e. x and $f * g$ is measurable. Observe that

$$|(f * g)(x)| \le \int_{-\infty}^{\infty} \Bigl(|f(y)| \, |g(x-y)|^{1/p}\Bigr) \, |g(x-y)|^{1/p'} \, dy. \tag{9.6}$$

Apply Hölder's Inequality with exponents p and p' to the two factors that appear on the right-hand side of equation (9.6) to show that

$$|(f * g)(x)| \le \|g\|_1^{1/p'} \left(\int_{-\infty}^{\infty} |f(y)|^p \, |g(x-y)| \, dy \right)^{1/p}.$$

Then use Tonelli again to show that

$$\|f * g\|_p \le \|f\|_p \, \|g\|_1. \tag{9.7}$$

Finally, extend from nonnegative functions to arbitrary functions $f \in L^p(\mathbb{R})$ and $g \in L^1(\mathbb{R})$. $\diamondsuit$

The inequality in equation (9.7) is known as *Young's Inequality*. Exercise 9.1.13 establishes Young's Inequality for $1 < p < \infty$, but Exercise 9.1.4 and Theorem 9.1.3 show that it also holds for $p = 1$ and $p = \infty$. We formalize this as the following theorem.

Theorem 9.1.14 (Young's Inequality). *Fix $1 \le p \le \infty$. If $f \in L^p(\mathbb{R})$ and $g \in L^1(\mathbb{R})$, then $f * g \in L^p(\mathbb{R})$ and*

$$\|f * g\|_p \;\le\; \|f\|_p \, \|g\|_1. \qquad \diamondsuit$$

An alternative proof of Theorem 9.1.14 based on Minkowski's Integral Inequality is sketched in Problem 9.1.20. Additionally, Problem 9.1.21 presents a more general version of Young's Inequality: $f * g \in L^r(\mathbb{R})$ whenever $f \in L^p(\mathbb{R})$, $g \in L^q(\mathbb{R})$, and $1 \le p, q, r \le \infty$ satisfy the relationship

$$\frac{1}{r} \;=\; \frac{1}{p} + \frac{1}{q} - 1.$$

According to Theorem 9.1.11, if $\{k_N\}_{N\in\mathbb{N}}$ is an approximate identity, then $f * k_N \to f$ in L^1-norm for every $f \in L^1(\mathbb{R})$. Suppose that we instead take $f \in L^p(\mathbb{R})$. The functions k_N belong to $L^1(\mathbb{R})$ (this is part of the definition of an approximate identity), so Young's Inequality ensures that $f * k_N$ belongs to $L^p(\mathbb{R})$. Will we have $f * k_N \to f$ in L^p-norm when $p > 1$? The following result states that this is the case, as long as p is finite.

Theorem 9.1.15. *Fix $1 \le p < \infty$, and let $\{k_N\}_{N\in\mathbb{N}}$ be an approximate identity. Then*

$$\lim_{N\to\infty} \|f - f * k_N\|_p \;=\; 0, \qquad \text{for all } f \in L^p(\mathbb{R}).$$

Proof. The case $p = 1$ is Theorem 9.1.11, so we focus on $1 < p < \infty$.

An approximate identity is uniformly bounded above in L^1-norm, so let $K = \sup \|k_N\|_1 < \infty$. Using Hölder's Inequality and Tonelli's Theorem, we compute that

$$\begin{aligned}
&\|f - f * k_N\|_p^p \\
&\quad= \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} \bigl(f(x) - f(x-t)\bigr)\, k_N(t)\, dt \right|^p dx \\
&\quad\le \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |f(x) - f(x-t)|\, |k_N(t)|^{1/p}\, |k_N(t)|^{1/p'}\, dt \right)^p dx \\
&\quad\le \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |f(x) - f(x-t)|^p\, |k_N(t)|\, dt \right)^{p/p} \left(\int_{-\infty}^{\infty} |k_N(t)|\, dt \right)^{p/p'} dx \\
&\quad= \|k_N\|_1^{p/p'} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x) - f(x-t)|^p\, |k_N(t)|\, dt\, dx \\
&\quad\le K^{p/p'} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} |f(x) - f(x-t)|^p\, dx \right) |k_N(t)|\, dt \\
&\quad= K^{p/p'} \int_{-\infty}^{\infty} \|f - T_t f\|_p^p\, |k_N(t)|\, dt.
\end{aligned}$$

From this point onwards, the proof is nearly identical to the proof of Theorem 9.1.11, using the fact that translation is strongly continuous in $L^p(\mathbb{R})$ when p is finite. $\square$

Theorem 9.1.15 suggests that if we choose our approximate identity so that k_N is smooth, then we should be able to show that $C_c^\infty(\mathbb{R})$ is dense in $L^p(\mathbb{R})$, just as we showed earlier that $C_c^\infty(\mathbb{R})$ is dense in $L^1(\mathbb{R})$. In order to do this, we need to know that $f * k_N$ inherits the smoothness of k_N. Exercise 9.1.7 showed that this is the case if f is integrable and k_N and its derivatives are bounded. However, if we assume instead that $f \in L^p(\mathbb{R})$, then boundedness of k_N is not enough to ensure that $f * k_N$ will exist. On the other hand, if we impose the stronger assumption that k_N is compactly supported, then $f * k_N$ will exist and it will inherit the smoothness of k_N. This kind of flexibility in imposing properties on an approximate identity can be useful in many situations.

Exercise 9.1.16. Fix $1 \le p < \infty$, and prove the following statements.

(a) If $m \in \mathbb{N}$, $f \in L^p(\mathbb{R})$, and $g \in C_c^m(\mathbb{R})$, then $f * g \in C_0^m(\mathbb{R})$ and

$$(f * g)^{(k)} \;=\; f * g^{(k)}, \qquad \text{for } k = 0, \dots, m. \tag{9.8}$$

(b) $C_c^\infty(\mathbb{R})$ is a dense subspace of $L^p(\mathbb{R})$. $\diamondsuit$

Similar results hold for $p = \infty$ if we replace $L^\infty(\mathbb{R})$ with $C_0(\mathbb{R})$.

Exercise 9.1.17. Prove the following statements.

(a) If $\{k_N\}_{N \in \mathbb{N}}$ is an approximate identity and f is bounded and uniformly continuous on $\mathbb{R}$ (for example, if $f \in C_0(\mathbb{R})$), then $f * k_N \to f$ uniformly.

(b) If $f \in C_0(\mathbb{R})$ and $g \in C_c^m(\mathbb{R})$, then $f * g \in C_0^m(\mathbb{R})$ and equation (9.8) holds.

(c) $C_c^\infty(\mathbb{R})$ is a dense subspace of $C_0(\mathbb{R})$. $\diamondsuit$

Problems

9.1.18. The convolution of two sequences $a = (a_k)_{k \in \mathbb{Z}}$ and $b = (b_k)_{k \in \mathbb{Z}}$ is the sequence $a * b = \big((a * b)_k\big)_{k \in \mathbb{Z}}$ whose components are

$$(a * b)_k \;=\; \sum_{j=-\infty}^{\infty} a_j\, b_{k-j}, \qquad \text{for } k \in \mathbb{Z}, \tag{9.9}$$

as long as this series converges for each $k \in \mathbb{Z}$.

(a) Fix $1 \le p \le \infty$. Prove the following version of Young's Inequality for convolution of sequences: If $a \in \ell^p(\mathbb{Z})$ and $b \in \ell^1(\mathbb{Z})$, then $a * b \in \ell^p(\mathbb{Z})$ and $\|a * b\|_p \le \|a\|_p \, \|b\|_1$.

(b) Set $\delta = \delta_0 = \bigl(\delta_{0n}\bigr)_{n\in\mathbb{Z}}$. Show that δ is an identity for convolution on $\ell^p(\mathbb{Z})$, i.e., $x * \delta = x$ for every sequence $x \in \ell^p(\mathbb{Z})$.

Remark: In contrast, we will see in Corollary 9.2.7 that there is no function in $L^1(\mathbb{R})$ that is an identity element for convolution of functions.

9.1.19. Show that if $f, g \in L^1(\mathbb{R})$ and $f, g \ge 0$ a.e., then $\|f*g\|_1 = \|f\|_1\,\|g\|_1$. Find a function $h \in L^1(\mathbb{R})$ such that $\|h*h\|_1 < \|h\|_1^2$.

9.1.20. This problem gives an alternative proof of Young's Inequality. Given $f \in L^p(\mathbb{R})$ and $g \in L^1(\mathbb{R})$, write out $\|f*g\|_p$ as an iterated integral, and apply Minkowski's Integral Inequality (Problem 7.2.17) to obtain another proof of equation (9.7).

9.1.21. This problem gives a general version of Young's Inequality. Assume that $1 < p,\, q,\, r < \infty$ satisfy

$$\frac{1}{r} \;=\; \frac{1}{p} + \frac{1}{q} - 1. \tag{9.10}$$

Let $f \in L^p(\mathbb{R})$ and $g \in L^q(\mathbb{R})$ be given.

(a) Show that

$$|(f*g)(x)| \;\le\; \int_{-\infty}^{\infty} \Bigl(\bigl|f(y)\bigr|^{p/r}\,\bigl|g(x-y)\bigr|^{q/r}\Bigr)\,\bigl|f(y)\bigr|^{p(\frac{1}{p}-\frac{1}{r})}\,\bigl|g(x-y)\bigr|^{q(\frac{1}{q}-\frac{1}{r})}\,dy.$$

(b) Define

$$\frac{1}{p_1} \;=\; \frac{1}{p} - \frac{1}{r} \qquad\text{and}\qquad \frac{1}{p_2} \;=\; \frac{1}{q} - \frac{1}{r}.$$

Use Hölder's Inequality for a product of three functions (Problem 7.2.20), with exponents r, p_1, p_2, to prove *Young's Inequality*:

$$\|f*g\|_r \;\le\; \|f\|_p\,\|g\|_q.$$

(c) Show that Young's Inequality also holds for any numbers r, p, q in the range $1 \le p, q, r \le \infty$ that satisfy equation (9.10).

9.1.22. Fix $1 \le p < q < \infty$, and suppose that $f \in L^p(\mathbb{R}) \cap L^q(\mathbb{R})$. Prove that there exist functions $g_n \in C_c^\infty(\mathbb{R})$ such that $g_n \to f$ in L^p-norm *and* $g_n \to f$ in L^q-norm.

9.1.23. Let $\{k_N\}_{N\in\mathbb{N}}$ be an approximate identity. Show that if a function $f \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ is continuous at a point $x \in \mathbb{R}$, then

$$\lim_{N\to\infty} (f*k_N)(x) \;=\; f(x).$$

9.1.24. Let $k\colon \mathbb{R} \to \mathbb{R}$ be a bounded measurable function such that $\int k = 1$ and $k(x) = 0$ for $|x| \ge 1$, and define $k_N(x) = N\,k(Nx)$. Given $f \in L^1(\mathbb{R})$, prove that $(f*k_N)(x) \to f(x)$ at every Lebesgue point x of f.

9.1.25. (a) Exhibit functions $f \in L^1(\mathbb{R})$ and $g \in L^\infty(\mathbb{R})$ such that

$$\lim_{x\to\infty} (f * g)(x) \quad \text{does not exist.}$$

(b) Prove that if $f \in L^1(\mathbb{R})$ and $g \in L^\infty(\mathbb{R})$, then

$$\lim_{x\to\infty} \int_{-\infty}^{b} f(x-y)\, g(y)\, dy \;=\; 0, \qquad \text{for all } b \in \mathbb{R}.$$

(c) Suppose that $g \in L^\infty(\mathbb{R})$ is such that $L = \lim_{x\to\infty} g(x)$ exists. Given $f \in L^1(\mathbb{R})$, show that $\lim_{x\to\infty} (f * g)(x) = L \int_{-\infty}^{\infty} f$.

9.1.26. Let $\gamma(x) = e^{-1/x}\, \chi_{[0,\infty)}(x)$ and $\beta(x) = \gamma\big(1-x^2\big)$. Prove the following statements.

(a) $\gamma(x) = 0$ for all $x \le 0$, and $\gamma(x) > 0$ for all $x > 0$.

(b) For each $n \in \mathbb{N}$, there exists a polynomial p_n of degree $n-1$ such that

$$\gamma^{(n)}(x) \;=\; \frac{p_n(x)}{x^{2n}}\, \gamma(x).$$

(c) $\gamma \in C^\infty(\mathbb{R})$ and $\gamma^{(n)}(0) = 0$ for every $n \ge 0$.

(d) $\beta \in C_c^\infty(\mathbb{R})$, $\beta(x) > 0$ for $|x| < 1$, and $\beta(x) = 0$ for $|x| \ge 1$.

9.1.27. Choose any function $k \in C_c^\infty(\mathbb{R})$ that satisfies $\int k = 1$, $k \ge 0$, and $k(x) = 0$ for $|x| > 1$. Show that the convolution $\theta = \chi_{[-2,2]} * k$ has the following properties:

(a) $\theta \in C_c^\infty(\mathbb{R})$,

(b) $0 \le \theta \le 1$,

(c) $\theta(x) = 1$ for $|x| \le 1$,

(d) $\theta(x) = 0$ for $|x| > 3$.

9.1.28. Suppose that f is differentiable everywhere on $\mathbb{R}$, and $f, f' \in L^1(\mathbb{R})$. Let $\theta \in C_c^\infty(\mathbb{R})$ be the function constructed in Problem 9.1.27, and for each $n \in \mathbb{N}$ define $\theta_n(x) = \theta(\frac{x}{n})$. Prove the following statements.

(a) $\sup \|\theta_n'\|_\infty < \infty$.

(b) $f'\theta_n \to f'$ and $f\theta_n' \to 0$ in L^1-norm.

(c) $\int_{-\infty}^{\infty} f' = 0$.

9.1.29. This problem will derive a C^∞-analogue of Urysohn's Lemma for functions on $\mathbb{R}$. Let K be a compact subset of $\mathbb{R}$, and assume that $U \supseteq K$ is open. Define $d = \operatorname{dist}(K, \mathbb{R}\backslash U) = \inf\big\{|x-y| : x \in K, y \notin U\big\}$, and set

$$V \;=\; \Big\{y \in \mathbb{R} : \operatorname{dist}(y, K) < \frac{d}{3}\Big\}.$$

Choose any function $k \in C_c^\infty(\mathbb{R})$ that satisfies $\int k = 1,\ k \geq 0,$ and $k(x) = 0$ for $|x| > d/3$. Show that the convolution $\theta = \chi_V * k$ has the following properties:

(a) $\theta \in C_c^\infty(\mathbb{R})$,

(b) $0 \leq \theta \leq 1$,

(c) $\theta(x) = 1$ for $x \in K$,

(d) $\theta(x) = 0$ for $x \notin U$.

9.1.30. Fix $1 \leq p \leq \infty$. If $f \in L^p(\mathbb{R})$ and there exists a function $h \in L^p(\mathbb{R})$ such that

$$\lim_{a \to 0} \left\| h - \frac{f - T_a f}{a} \right\|_p = 0,$$

then we call h a *strong L^p-derivative of f* and denote it by $h = \partial_p f$. Assume that $f \in L^p(\mathbb{R})$ has a strong L^p-derivative. Given $g \in L^{p'}(\mathbb{R})$, prove that $f * g$ is differentiable at every point, and $(f * g)' = \partial_p f * g$.

9.1.31. Show that if $f \in C_c^m(\mathbb{R})$ and $g \in L^\infty(\mathbb{R})$, then $f * g \in C_b^m(\mathbb{R})$ and $(f * g)^{(k)} = f^{(k)} * g$ for $k = 1, \dots, m$.

9.1.32. Redo Problem 7.4.5, but with $C_c(\mathbb{R})$ replaced by $C_c^\infty(\mathbb{R})$.

9.1.33. Suppose that $f \in L^\infty(\mathbb{R})$ satisfies $\lim_{a \to 0} \|T_a f - f\|_\infty = 0$. Prove that there exists a uniformly continuous function g such that $f = g$ a.e.

9.1.34. We say that a function $f\colon \mathbb{R} \to \mathbb{R}$ is *additive* if $f(x+y) = f(x) + f(y)$ for all $x, y \in \mathbb{R}$.

(a) Show that if f is additive, then $f(rx) = rf(x)$ for all $x \in \mathbb{R}$ and $r \in \mathbb{Q}$.

(b) Prove that a continuous function f is additive if and only if f has the form $f(x) = cx$ for some $c \in \mathbb{R}$.

(c) Since the set $\mathbb{Q}$ of rational numbers is a field, we can consider the vector space $\mathbb{R}$ over the field $\mathbb{Q}$. A consequence of the Axiom of Choice is that every vector space has a Hamel basis (in fact, this statement is equivalent to the Axiom of Choice). Consequently, there exists a Hamel basis $\{x_i\}_{i \in I}$ for $\mathbb{R}$ over $\mathbb{Q}$ (note that this index set I will necessarily be uncountable). That is, every nonzero number $x \in \mathbb{R}$ can be written uniquely as $x = \sum_{k=1}^N c_k x_{i_k}$ for some distinct indices $i_1, \dots, i_N \in I$ and nonzero *rational* scalars $c_1, \dots, c_N$. Use this to show that there exists a function $f\colon \mathbb{R} \to \mathbb{R}$ that is additive yet f *does not* satisfy $f(cx) = cf(x)$ for all $c, x \in \mathbb{R}$. Thus f is not linear, even though f respects addition.

(d) Suppose that f is additive and $f(x) = 0$ for all x in the Cantor set C. Prove that $f = 0$.

9.1.35. Assume that $f\colon \mathbb{R} \to \mathbb{R}$ is additive, i.e., $f(x+y) = f(x) + f(y)$ for all $x, y \in \mathbb{R}$, and suppose further that f is *measurable*.

(a) Prove that the function $g(x) = e^{2\pi i f(x)}$ has the following properties.

- $g(x+y) = g(x)\,g(y)$ for all $x, y \in \mathbb{R}$.
- If $\phi \in C_c(\mathbb{R})$, then there exists a scalar C_ϕ such that $\phi * g = C_\phi\, g$.
- There exists some $\phi \in C_c^1(\mathbb{R})$ such that $C_\phi \neq 0$.
- g is differentiable and $g'(x) = \beta\, g(x)$ for some constant $\beta \in \mathbb{C}$.
- There exists an $\alpha \in \mathbb{R}$ such that $g(x) = e^{2\pi i \alpha x}$ for all $x \in \mathbb{R}$.

(b) To emphasize that care will be needed in the next step, exhibit a discontinuous function $h\colon \mathbb{R} \to \mathbb{R}$ such that $e^{2\pi i h(x)}$ is continuous on $\mathbb{R}$.

(c) Prove that $f(x) = \alpha x$ for all $x \in \mathbb{R}$.

9.2 The Fourier Transform

The Fourier transform is the cornerstone of *harmonic analysis.* We will give a brief introduction to the Fourier transform on the space $L^1(\mathbb{R})$. For more detailed introductions to harmonic analysis, we refer to texts such as [DM72], [Ben97], [SS03] or [Kat04].

The complex exponential functions $e_\xi(x) = e^{2\pi i \xi x}$ play a fundamental role in the definition of the Fourier transform. The graph of e_ξ is

$$\bigl\{(x, e^{2\pi i \xi x}) : x \in \mathbb{R}\bigr\} \;\subseteq\; \mathbb{R} \times \mathbb{C}.$$

Identifying $\mathbb{R} \times \mathbb{C}$ with $\mathbb{R} \times \mathbb{R}^2 = \mathbb{R}^3$, this graph is a helix in $\mathbb{R}^3$ coiling around the x-axis, which runs down the center of the helix (see Figure 9.5). In higher dimensions, the frequency is a vector $\xi \in \mathbb{R}^d$, and the complex exponential function $e_\xi\colon \mathbb{R}^d \to \mathbb{C}$ is given by

$$e_\xi(x) \;=\; e^{2\pi i \xi \cdot x}, \qquad \text{for } x \in \mathbb{R}^d,$$

where $\xi \cdot x$ is the usual dot product of vectors in $\mathbb{R}^d$.

We define the Fourier transform of an integrable function on $\mathbb{R}$ as follows.

Definition 9.2.1 (Fourier Transform on $L^1(\mathbb{R})$). The *Fourier transform* of $f \in L^1(\mathbb{R})$ is the function $\widehat{f}\colon \mathbb{R} \to \mathbb{C}$ defined by

$$\widehat{f}(\xi) \;=\; \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i \xi x}\, dx, \qquad \text{for } \xi \in \mathbb{R}. \tag{9.11}$$

For notational clarity, we sometimes write $f^\wedge$ or $(f)^\wedge$ instead of $\widehat{f}$. $\diamondsuit$

If f is integrable, then $\widehat{f}(\xi)$ exists for every $\xi \in \mathbb{R}$ because

$$\int_{-\infty}^{\infty} |f(x)\, e^{-2\pi i \xi x}|\, dx \;=\; \int_{-\infty}^{\infty} |f(x)|\, dx \;=\; \|f\|_1 \;<\; \infty. \tag{9.12}$$

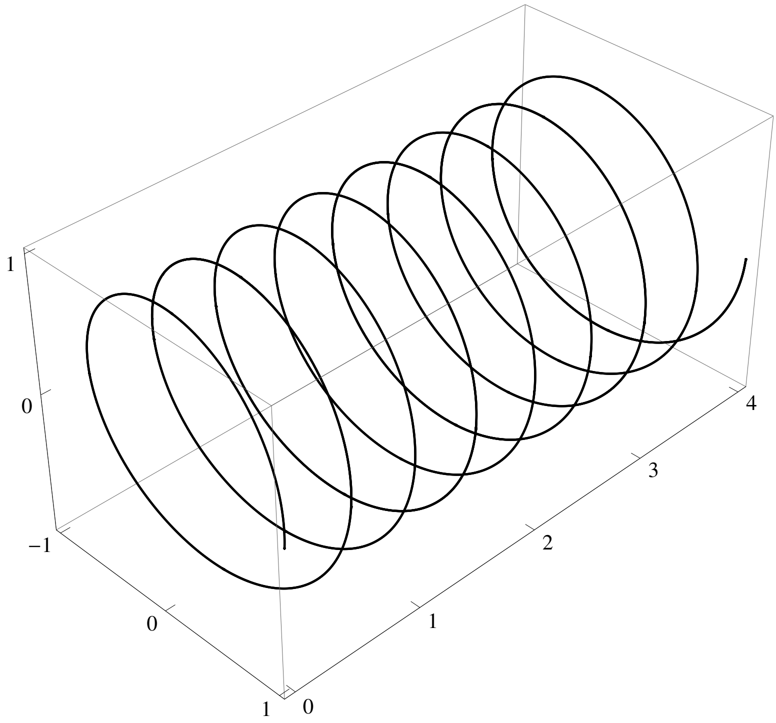


Fig. 9.5 Graph of $e_\xi(x) = e^{2\pi i \xi x}$ for $\xi = 2$ and $0 \le x \le 4$.

Thus $\widehat{f}(\xi)$ is defined at every point, even though $f(x)$ need only be defined almost everywhere. Additionally, $\widehat{f}(\xi)$ is complex in general, even if f is purely real-valued. Therefore, for the remainder of this chapter we will assume that all functions are complex-valued. That is,

$$\textit{from now on we take } \overline{\mathbf{F}} = \mathbb{C}.$$

Remark 9.2.2. The definition of the Fourier transform of $f \in L^1(\mathbb{R})$ closely resembles the definition of the Fourier coefficients of a function $f \in L^2[0,1]$ that are given in equation (8.11). The nth Fourier coefficient $\widehat{f}(n)$ of a function $f \in L^2[0,1]$ is

$$\widehat{f}(n) \;=\; \int_0^1 f(x)\, e^{-2\pi i n x}\, dx,$$

which is the inner product in the Hilbert space $L^2[0,1]$ of f with the function $e_n(x) = e^{2\pi i n x}$. In contrast, even if we took f in $L^2(\mathbb{R})$ instead of $L^1(\mathbb{R})$, the formula for $\widehat{f}(\xi)$ given in equation (9.11) is not an inner product of two functions in the Hilbert space $L^2(\mathbb{R})$ because $e_\xi(x) = e^{2\pi i \xi x}$ does not belong to $L^2(\mathbb{R})$. Even so, the apparent similarities between Fourier coefficients and the Fourier transform are real indications that there is a fundamental underlying connection between these two objects. Indeed, Fourier series and the Fourier transform are two special cases of abstract Fourier transforms on locally compact abelian groups. Another special case is the *discrete Fourier transform*, or DFT, which plays a key role in digital signal processing. The DFT is the Fourier transform on the finite cyclic group $\mathbb{Z}_N = \{0, 1, \dots, N-1\}$ (under addition mod N). More details on abstract Fourier transforms can be found in the texts referenced at the beginning of this section. $\diamondsuit$

We prove next that $\widehat{f}$ is continuous on $\mathbb{R}$.

Lemma 9.2.3. *If $f \in L^1(\mathbb{R})$, then $\widehat{f}$ is bounded and uniformly continuous on $\mathbb{R}$, and*

$$\|\widehat{f}\|_\infty \;\le\; \|f\|_1. \tag{9.13}$$

Proof. Since

$$|\widehat{f}(\xi)| \;=\; \left|\int_{-\infty}^{\infty} f(x)\, e^{-2\pi i\xi x}\, dx\right| \;\le\; \int_{-\infty}^{\infty} |f(x)\, e^{-2\pi i\xi x}|\, dx \;=\; \|f\|_1,$$

we see that $\widehat{f}$ is bounded and $\|\widehat{f}\|_\infty \le \|f\|_1$.

To prove that $\widehat{f}$ is continuous, fix $\xi \in \mathbb{R}$ and choose any $\eta \in \mathbb{R}$. Then

$$\begin{aligned}
\bigl|\widehat{f}(\xi+\eta) - \widehat{f}(\xi)\bigr| \;&=\; \left|\int_{-\infty}^{\infty} f(x)\, e^{-2\pi i(\xi+\eta)x}\, dx \;-\; \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i\xi x}\, dx\right| \\
&\le\; \int_{-\infty}^{\infty} |f(x)|\, |e^{-2\pi i\xi x}|\, |e^{-2\pi i\eta x} - 1|\, dx \\
&=\; \int_{-\infty}^{\infty} |f(x)|\, |e^{-2\pi i\eta x} - 1|\, dx.
\end{aligned} \tag{9.14}$$

Note that the quantity after the equality in equation (9.14) is independent of ξ. Now, for almost every x (specifically, any x where $f(x)$ is defined), we have that

$$\lim_{\eta\to 0} |f(x)|\, |e^{-2\pi i\eta x} - 1| \;=\; 0.$$

Additionally,

$$|f(x)|\, |e^{-2\pi i\eta x} - 1| \;\le\; 2\,|f(x)| \;\in\; L^1(\mathbb{R}).$$

Therefore we can apply the Dominated Convergence Theorem and compute that

$$\sup_{\xi\in\mathbb{R}} \bigl|\widehat{f}(\xi+\eta) - \widehat{f}(\xi)\bigr| \;\le\; \int_{-\infty}^{\infty} |f(x)|\, |e^{-2\pi i\eta x} - 1|\, dx \;\to\; 0 \qquad \text{as } \eta \to 0.$$

This implies that $\widehat{f}$ is uniformly continuous. $\square$

We will compute the Fourier transform of the characteristic function of the symmetric interval $[-T, T]$.

Example 9.2.4. By direct computation,

$$(\chi_{[-T,T]})^\wedge(\xi) \;=\; \int_{-T}^{T} e^{-2\pi i\xi x}\,dx \;=\; \begin{cases} \frac{\sin 2\pi T\xi}{\pi\xi}, & \text{if } \xi \neq 0, \\ 2T, & \text{if } \xi = 0. \end{cases} \tag{9.15}$$

This is a continuous function, so we usually implicitly assume that it is defined appropriately at the origin and just write $(\chi_{[-T,T]})^\wedge(\xi) = \frac{\sin 2\pi T\xi}{\pi\xi}$.

An important special case is the (normalized) *sinc function*

$$\operatorname{sinc}(\xi) \;=\; \frac{\sin \pi\xi}{\pi\xi} \;=\; (\chi_{[-\frac{1}{2},\frac{1}{2}]})^\wedge(\xi). \tag{9.16}$$

If we compare the sinc function to the Fejér function w defined in Exercise 9.1.10, we see that

$$w(\xi) \;=\; \operatorname{sinc}(\xi)^2.$$

The Fejér function is both integrable and nonnegative, while the sinc function is neither. $\diamondsuit$

Since $\chi_{[-T,T]}$ is integrable while the sinc function is not, the preceding example shows that the Fourier transform of an integrable function need not be integrable. On the other hand, the sinc function belongs to $C_0(\mathbb{R})$, and we prove next that $\widehat{f}$ always belongs to $C_0(\mathbb{R})$ whenever f is integrable. An alternative proof of Theorem 9.2.5 is outlined in Problem 9.2.24.

Theorem 9.2.5 (Riemann–Lebesgue Lemma). *If $f \in L^1(\mathbb{R})$, then $\widehat{f} \in C_0(\mathbb{R})$.*

Proof. We saw in Lemma 9.2.3 that $\widehat{f}$ is continuous, so it only remains to show that $\widehat{f}$ decays to zero at $\pm\infty$. Since $e^{-\pi i} = -1$, for $\xi \neq 0$ we have

$$\begin{aligned}
\widehat{f}(\xi) \;&=\; \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i\xi x}\, dx && (9.17)\\
&=\; -\int_{-\infty}^{\infty} f(x)\, e^{-2\pi i\xi x}\, e^{-2\pi i\xi(\frac{1}{2\xi})}\, dx \\
&=\; -\int_{-\infty}^{\infty} f(x)\, e^{-2\pi i\xi(x+\frac{1}{2\xi})}\, dx \\
&=\; -\int_{-\infty}^{\infty} f\bigl(x - \tfrac{1}{2\xi}\bigr)\, e^{-2\pi i\xi x}\, dx. && (9.18)
\end{aligned}$$

Averaging equalities (9.17) and (9.18) yields

$$\widehat{f}(\xi) \;=\; \frac{1}{2}\int_{-\infty}^{\infty} \Bigl(f(x) - f\bigl(x - \tfrac{1}{2\xi}\bigr)\Bigr)\, e^{-2\pi i\xi x}\, dx.$$

Hence, using the strong continuity of translation derived in Exercise 4.5.9, we compute that

$$|\widehat{f}(\xi)| \;\le\; \frac{1}{2}\int_{-\infty}^{\infty} \Bigl|f(x) - f\bigl(x - \tfrac{1}{2\xi}\bigr)\Bigr|\,dx \;=\; \frac{1}{2}\,\|f - T_{\frac{1}{2\xi}} f\|_1 \;\to\; 0 \;\text{ as } |\xi| \to \infty.$$

Therefore $\widehat{f} \in C_0(\mathbb{R})$. $\square$

The Riemann–Lebesgue Lemma tells us that the Fourier transform maps $L^1(\mathbb{R})$ into $C_0(\mathbb{R})$. In Corollary 9.2.12, we will prove that the Fourier transform is injective on $L^1(\mathbb{R})$. The range of the Fourier transform is the set

$$A(\mathbb{R}) \;=\; \bigl\{\widehat{f} : f \in L^1(\mathbb{R})\bigr\}. \tag{9.19}$$

We will see in Corollary 9.2.16 that $A(\mathbb{R})$ is a dense, but proper, subspace of $C_0(\mathbb{R})$. Thus, although the Fourier transform is injective and has dense range, it is not a surjective mapping of $L^1(\mathbb{R})$ onto $C_0(\mathbb{R})$.

The next exercise, which is an application of Fubini's Theorem, shows that the Fourier transform converts convolution into pointwise multiplication.

Exercise 9.2.6. If $f, g \in L^1(\mathbb{R})$, then it follows from Theorem 9.1.3 that their convolution $f * g$ belongs to $L^1(\mathbb{R})$. Prove that the Fourier transform of $f * g$ is the product of the Fourier transforms of f and g:

$$(f * g)^\wedge(\xi) \;=\; \widehat{f}(\xi)\,\widehat{g}(\xi), \qquad \text{for all } \xi \in \mathbb{R}. \qquad \diamondsuit \tag{9.20}$$

Now we use Exercise 9.2.6 to prove that there is no identity element for convolution in $L^1(\mathbb{R})$.

Corollary 9.2.7. *There is no function $\delta \in L^1(\mathbb{R})$ such that $f * \delta = f$ for every $f \in L^1(\mathbb{R})$.*

Proof. Suppose that there were such a function δ in $L^1(\mathbb{R})$. Then for every $f \in L^1(\mathbb{R})$ we would have

$$\widehat{f}(\xi) \;=\; (f * \delta)^\wedge(\xi) \;=\; \widehat{f}(\xi)\,\widehat{\delta}(\xi).$$

In particular, the function $f = \chi_{[-1,1]}$ is integrable and $\widehat{f}(\xi) \ne 0$ for a.e. ξ. Therefore $\widehat{\delta}(\xi) = 1$ for a.e. ξ. But this contradicts the Riemann–Lebesgue Lemma, so no such function δ can exist. $\square$

9.2.1 The Inversion Formula

Our next goal is to prove the *Inversion Formula* for the Fourier transform. This theorem will show that if $f \in L^1(\mathbb{R})$ is such that its Fourier transform $\widehat{f}$ is also integrable, then we can recover f from $\widehat{f}$. This is similar in spirit to Theorem 8.4.2, which states that the Fourier coefficients of a function $f \in L^2[0,1]$ can be used to recover f. That result follows from the fact that the trigonometric system $\{e^{2\pi inx}\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $L^2[0,1]$. Here the situation is different, because the uncountable system $\{e^{2\pi i\xi x}\}_{\xi\in\mathbb{R}}$ is not an orthonormal basis for $L^2(\mathbb{R})$. Indeed, $e^{2\pi i\xi x}$ does not belong to $L^2(\mathbb{R})$ for any ξ. Even so, we will be able to use convolution and approximate identities to prove the Inversion Formula.

In order to state our results more succinctly, we introduce the following notation.

Definition 9.2.8 (Inverse Fourier Transform on $L^1(\mathbb{R})$). The *inverse Fourier transform* of $f \in L^1(\mathbb{R})$ is

$$\check{f}(\xi) \;=\; \int_{-\infty}^{\infty} f(x)\, e^{2\pi i\xi x}\, dx, \qquad \text{for } \xi \in \mathbb{R}. \qquad \diamondsuit \tag{9.21}$$

The inverse Fourier transform behaves much like the Fourier transform. Indeed, if $f \in L^1(\mathbb{R})$ then both $\widehat{f}$ and $\check{f}$ are well-defined continuous functions, and

$$\check{f}(\xi) \;=\; \widehat{f}(-\xi), \qquad \text{for all } \xi \in \mathbb{R}.$$

Therefore, by making an appropriate change of variables, every result that we have stated so far for the Fourier transform has an analogue for the inverse Fourier transform.

The word "inverse" in Definition 9.2.8 needs to be interpreted with some care. Even if f is integrable, its Fourier transform $g = \widehat{f}$ need not be integrable, and so its "inverse Fourier transform" $\check{g}$ might not even exist, much less equal f. However, we will prove in this section that if it is the case that f and $\widehat{f}$ are both integrable then $\big(\widehat{f}\big)^{\vee} = f$. It is only in this restricted sense that the inverse Fourier transform is the inverse of the Fourier transform for integrable functions. We state that theorem next, but then must develop some machinery before we can prove it.

Theorem 9.2.9 (Inversion Formula). *If $f, \widehat{f} \in L^1(\mathbb{R})$, then both f and $\widehat{f}$ are continuous, and*

$$f(x) \;=\; \big(\widehat{f}\big)^{\vee}(x) \;=\; \int_{-\infty}^{\infty} \widehat{f}(\xi)\, e^{2\pi i\xi x}\, d\xi, \qquad \textit{for every } x \in \mathbb{R}. \tag{9.22}$$

Similarly,

$$f(x) \;=\; \big(\check{f}\big)^{\wedge}(x) \;=\; \int_{-\infty}^{\infty} \check{f}(\xi)\, e^{-2\pi i\xi x}\, d\xi, \qquad \textit{for every } x \in \mathbb{R}. \qquad \diamondsuit$$

These equations give us some insight into why the Fourier transform is such an important operator. As long as f and $\widehat{f}$ are both integrable, equation (9.22) says that f can be represented as an integral (in effect, a continuous sum or superposition) of $\widehat{f}(\xi)\, e^{2\pi i\xi x}$ over all frequencies $\xi \in \mathbb{R}$. The Fourier transform $\widehat{f}(\xi)$ tells us what amplitude to assign to the pure tone $e^{2\pi i\xi x}$

of frequency ξ, and by summing all of these pure tones with the correct amplitudes we obtain f. In essence, the pure tones are a set of very simple building blocks that we can use to build very complicated functions f.

In order to prove the Inversion Formula, we will make use of the *Fejér kernel* $\{w_N\}_{N\in\mathbb{N}}$ that was introduced in Exercise 9.1.10. Explicitly, $w_N(x) = Nw(Nx)$ where w is the *Fejér function*

$$w(x) \;=\; \left(\frac{\sin \pi x}{\pi x}\right)^2 \;=\; \operatorname{sinc}(x)^2.$$

Exercise 9.1.10 showed that the Fejér kernel is an approximate identity. The Fejér kernel is not the only approximate identity that we could use to prove the Inversion Formula, but it does have some convenient properties. Specifically, w is continuous, integrable, even, and nonnegative, and the following lemma shows that it is the Fourier transform of a continuous, compactly supported, even, nonnegative function.

Lemma 9.2.10. *Let* $W(x) = \max\bigl\{1 - |x|,\, 0\bigr\}$ *denote the hat function supported on the interval* $[-1,1]$*. Then* $\widehat{W} = w = \check{W}$.

Proof. Exercise 9.1.2 showed that $W = \chi * \chi$ where $\chi = \chi_{[-\frac{1}{2},\frac{1}{2}]}$. Further, we saw in Example 9.2.4 that $\widehat{\chi}$ is the sinc function. Since the Fourier transform converts convolution into multiplication (Exercise 9.2.6), it follows that

$$\widehat{W} \;=\; (\chi * \chi)^\wedge \;=\; (\widehat{\chi})^2 \;=\; \operatorname{sinc}^2 \;=\; w. \tag{9.23}$$

Finally, since W is even, a change of variables shows that

$$\begin{aligned} w(x) \;=\; \widehat{W}(x) \;&=\; \int_{-\infty}^{\infty} W(x)\, e^{-2\pi i \xi x}\, d\xi \\ &=\; \int_{-\infty}^{\infty} W(\xi)\, e^{2\pi i \xi x}\, d\xi \;=\; \check{W}(x). \qquad \square \end{aligned} \tag{9.24}$$

Since $w = \check{W}$, we have $\widehat{w} = \bigl(\check{W}\bigr)^\wedge$. Once we prove the Inversion Formula, we will see that $\bigl(\check{W}\bigr)^\wedge = W$ and therefore $\widehat{w} = W$, but we have not proved this yet.

As a first step toward proving the Inversion Formula, we will consider a modified version of equation (9.22) obtained by inserting the "convergence factor"

$$W(\xi/N) \;=\; \max\biggl\{1 - \frac{|\xi|}{N},\, 0\biggr\},$$

which is the hat function with height 1 supported on the interval $[-N,N]$ (see Figure 9.6). Inserting this factor will allow us to prove that the convolution $f * w_N$ can be reconstructed from $\widehat{f}$. This is quite similar to how Cesàro

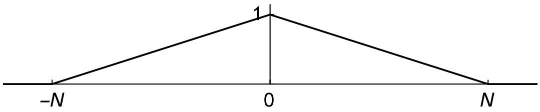


Fig. 9.6 Graph of $W(\xi/N) = \max\{1 - |\xi|/N, 0\}$.

summation can sometimes be used to deal with infinite series that do not converge. Indeed, when we consider the analogous theorem for Fourier series in Section 9.3, we will see that using the Fejér kernel in that setting is precisely the same as using Cesàro summation on a Fourier series.

Lemma 9.2.11. *If $f \in L^1(\mathbb{R})$, then for each $N > 0$ we have*

$$\begin{aligned}(f * w_N)(x) &= \int_{-\infty}^{\infty} \widehat{f}(\xi)\, W(\xi/N)\, e^{2\pi i \xi x}\, d\xi \\ &= \int_{-N}^{N} \widehat{f}(\xi) \left(1 - \frac{|\xi|}{N}\right) e^{2\pi i \xi x}\, d\xi. \qquad (9.25)\end{aligned}$$

Proof. Since f is integrable and $w_N \in C_c(\mathbb{R})$, we know from Exercise 9.1.4 that $f * w_N$ is continuous. Making a change of variables in equation (9.24), we have

$$w_N(x) = N w(Nx) = \int_{-\infty}^{\infty} W(\xi/N)\, e^{2\pi i \xi x}\, d\xi.$$

Therefore, assuming that we can interchange integrals in the following calculation, we compute that

$$\begin{aligned}(f * w_N)(x) &= \int_{-\infty}^{\infty} f(y)\, w_N(x-y)\, dy \\ &= \int_{-\infty}^{\infty} f(y) \int_{-\infty}^{\infty} W(\xi/N)\, e^{2\pi i \xi (x-y)}\, d\xi\, dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(y)\, e^{-2\pi i \xi y}\, dy\right) W(\xi/N)\, e^{2\pi i \xi x}\, d\xi \\ &= \int_{-\infty}^{\infty} \widehat{f}(\xi)\, W(\xi/N)\, e^{2\pi i \xi x}\, d\xi.\end{aligned}$$

Exercise: Use Fubini's Theorem to justify the interchange of integrals. □

Now we obtain the Inversion Formula by taking the limit on both sides of equation (9.25). Before reading the proof, it may be helpful to review Notation 7.2.8, which gives our conventions for the meaning of continuity of elements of $L^1(\mathbb{R})$.

Proof (of Theorem 9.2.9). Suppose that $f \in L^1(\mathbb{R})$ is such that $\widehat{f} \in L^1(\mathbb{R})$. Since f is integrable, $\widehat{f}$ is continuous. On the other hand, since $\widehat{f}$ is integrable,

$\big(\widehat{f}\,\big)^\vee$ is continuous. The function $f * w_N$ is also continuous, because it is the convolution of the integrable function f with the continuous, compactly supported function w_N.

Fix $x \in \mathbb{R}$. Then for every $\xi \in \mathbb{R}$ we have

$$\lim_{N\to\infty} \widehat{f}(\xi)\, W(\xi/N)\, e^{2\pi i\xi x} \;=\; \widehat{f}(\xi)\, e^{2\pi i\xi x} \Big(\lim_{N\to\infty} W(\xi/N)\Big) \;=\; \widehat{f}(\xi)\, e^{2\pi i\xi x}.$$

Also, since $0 \le W \le 1$,

$$|\widehat{f}(\xi)\, W(\xi/N)\, e^{2\pi i\xi x}| \;\le\; |\widehat{f}(\xi)| \;\in\; L^1(\mathbb{R}).$$

Holding x fixed, we can therefore apply the Dominated Convergence Theorem to obtain

$$\begin{aligned} \lim_{N\to\infty} (f * w_N)(x) \;&=\; \lim_{N\to\infty} \int_{-\infty}^{\infty} \widehat{f}(\xi)\, W(\xi/N)\, e^{2\pi i\xi x}\, d\xi \\ &=\; \int_{-\infty}^{\infty} \widehat{f}(\xi)\, e^{2\pi i\xi x}\, d\xi \;=\; \big(\widehat{f}\,\big)^\vee(x). \end{aligned} \tag{9.26}$$

On the other hand, since the Fejér kernel is an approximate identity we know that $f * w_N \to f$ in L^1-norm. Consequently there is a subsequence such that

$$\lim_{k\to\infty} (f * w_{N_k})(x) \;=\; f(x), \qquad \text{for a.e. } x. \tag{9.27}$$

Combining equations (9.26) and (9.27), we see that

$$\big(\widehat{f}\,\big)^\vee(x) \;=\; \lim_{k\to\infty} (f * w_{N_k})(x) \;=\; f(x) \quad \text{a.e.}$$

Thus f is equal a.e. to the continuous function $\big(\widehat{f}\,\big)^\vee$. Hence f and $\big(\widehat{f}\,\big)^\vee$ are the same element of $L^1(\mathbb{R})$, and so we can redefine f on a set of measure zero in such a way that $f(x) = \big(\widehat{f}\,\big)^\vee(x)$ for every x. $\square$

As a corollary, since both w and W are integrable, by combining the Inversion Formula with Lemma 9.2.10 we see that

$$\widehat{w} \;=\; (\check{W})^\wedge \;=\; W \;=\; (\widehat{W})^\vee \;=\; \check{w}.$$

Next, we use the Inversion Formula to prove that integrable functions are uniquely determined by their Fourier transforms.

Corollary 9.2.12 (Uniqueness Theorem). *If $f, g \in L^1(\mathbb{R})$, then*

$$f = g \text{ a.e.} \quad\iff\quad \widehat{f} = \widehat{g} \text{ a.e.}$$

In particular,

$$f = 0 \text{ a.e.} \quad\iff\quad \widehat{f} = 0 \text{ a.e.}$$

Proof. The first equivalence is a consequence of the second (consider $f - g$). If $f = 0$ a.e., then we immediately obtain $\widehat{f} = 0$ everywhere. On the other hand, if $\widehat{f} = 0$ a.e., then we have both f, $\widehat{f} \in L^1(\mathbb{R})$, so the Inversion Formula applies and we obtain $f = \bigl(\widehat{f}\bigr)^\vee = \check{0} = 0$. □

9.2.2 Smoothness and Decay

One of the important properties of the Fourier transform is that it interchanges smoothness and decay. For our first theorem in this direction, we will assume that $f \in L^1(\mathbb{R})$ has a certain amount of decay in the sense that $x^m f(x) \in L^1(\mathbb{R})$. This is not a pointwise decay requirement, but rather a kind of "average decay" assumption. As x increases, the value of $|x^m f(x)|$ becomes increasingly large compared to the value of $|f(x)|$, yet if $x^m f(x)$ is integrable then the area under the graph of $|x^m f(x)|$, and not merely the area under $|f(x)|$, must be finite. We will prove that if f satisfies this decay hypothesis, then $\widehat{f}$ is smooth in the sense that it is m-times differentiable. Although it is a slight abuse of notation, we will write $\bigl((-2\pi i x)^k f(x)\bigr)^\wedge$ to denote the Fourier transform of the function $g(x) = (-2\pi i x)^k f(x)$.

Theorem 9.2.13. *Let $f \in L^1(\mathbb{R})$ and $m \in \mathbb{N}$ be given. Then*

$$x^m f(x) \in L^1(\mathbb{R}) \implies \widehat{f} \in C_0^m(\mathbb{R}),$$

i.e., $\widehat{f}$ is m-times differentiable and $\widehat{f}, \widehat{f}\,', \dots, \widehat{f}^{\,(m)} \in C_0(\mathbb{R})$. Furthermore, in this case we have $x^k f(x) \in L^1(\mathbb{R})$ for $k = 0, \dots, m$, and the kth derivative of $\widehat{f}$ is the Fourier transform of $(-2\pi i x)^k f(x)$:

$$\widehat{f}^{\,(k)} = \frac{d^k}{d\xi^k}\widehat{f} = \bigl((-2\pi i x)^k f(x)\bigr)^\wedge, \qquad \text{for } k = 0, \dots, m. \tag{9.28}$$

Proof. We will proceed by induction. The base step is $m = 1$. To motivate equation (9.28) and its proof, note that if we were allowed to interchange a derivative with an integral, then we could write

$$\begin{aligned}
\frac{d}{d\xi}\widehat{f}(\xi) &= \frac{d}{d\xi}\int_{-\infty}^{\infty} f(x)\, e^{-2\pi i \xi x}\, dx \\
&= \int_{-\infty}^{\infty} f(x)\, \frac{d}{d\xi} e^{-2\pi i \xi x}\, dx \\
&= \int_{-\infty}^{\infty} f(x)\, (-2\pi i x)\, e^{-2\pi i \xi x}\, dx \\
&= \bigl(-2\pi i x f(x)\bigr)^\wedge(\xi).
\end{aligned}$$

Essentially, our task is to justify this interchange.

Since $m = 1$, our hypothesis is that f and $xf(x)$ both belong to $L^1(\mathbb{R})$. For simplicity of notation, set $e_x(\xi) = e^{-2\pi i\xi x}$. Then

$$\begin{aligned}\frac{\widehat{f}(\xi+\eta) - \widehat{f}(\xi)}{\eta} &= \int_{-\infty}^{\infty} f(x)\, \frac{e^{-2\pi i(\xi+\eta)x} - e^{-2\pi i\xi x}}{\eta}\, dx \\ &= \int_{-\infty}^{\infty} f(x)\, \frac{e_x(\xi+\eta) - e_x(\xi)}{\eta}\, dx.\end{aligned}$$

The integrand converges pointwise (as $\eta \to 0$) for almost every x, because for every x where $f(x)$ is defined we have

$$\lim_{\eta\to 0} f(x)\, \frac{e_x(\xi+\eta) - e_x(\xi)}{\eta} \;=\; f(x)\, e_x'(\xi) \;=\; f(x)\,(-2\pi i x)\, e^{-2\pi i\xi x}.$$

Since $|e^{i\theta} - 1| \le |\theta|$ for every real number θ, we also have that the integrand is bounded by an integrable function:

$$\begin{aligned}\left| f(x)\, \frac{e_x(\xi+\eta) - e_x(\xi)}{\eta} \right| &= \left| f(x)\, \frac{e^{-2\pi i(\xi+\eta)x} - e^{-2\pi i\xi x}}{\eta} \right| \\ &= |f(x)|\, |e^{-2\pi i\xi x}| \left| \frac{e^{-2\pi i\eta x} - 1}{\eta} \right| \\ &\le |f(x)| \left| \frac{-2\pi i\eta x}{\eta} \right| \\ &= 2\pi\, |xf(x)| \in L^1(\mathbb{R}).\end{aligned}$$

Therefore we can apply the Dominated Convergence Theorem to obtain

$$\begin{aligned}\widehat{f}\,'(\xi) &= \lim_{\eta\to 0} \frac{\widehat{f}(\xi+\eta) - \widehat{f}(\xi)}{\eta} \\ &= \lim_{\eta\to 0} \int_{-\infty}^{\infty} f(x)\, \frac{e_x(\xi+\eta) - e_x(\xi)}{\eta}\, dx \\ &= \int_{-\infty}^{\infty} \lim_{\eta\to 0} f(x)\, \frac{e_x(\xi+\eta) - e_x(\xi)}{\eta}\, dx \qquad \text{(DCT)} \\ &= \int_{-\infty}^{\infty} f(x)\,(-2\pi i x)\, e^{-2\pi i\xi x}\, dx \\ &= \big((-2\pi i x) f(x)\big)^{\wedge}(\xi).\end{aligned}$$

Thus, $\widehat{f}$ is differentiable, and since $\widehat{f}\,'$ is the Fourier transform of the integrable function $(-2\pi i x)f(x)$, the Riemann–Lebesgue Lemma implies that $\widehat{f}\,' \in C_0(\mathbb{R})$. This establishes the base step.

Now we proceed to the inductive step. Suppose that the result holds for some $m \ge 1$, and suppose that $f \in L^1(\mathbb{R})$ is such that $x^{m+1}f(x) \in L^1(\mathbb{R})$.

Fix any integer $1 \le k \le m+1$. Note that

$$|x| \le 1 \quad \Longrightarrow \quad |x^k f(x)| \le |f(x)|,$$

and

$$|x| > 1 \quad \Longrightarrow \quad |x^k f(x)| \le |x^{m+1} f(x)|.$$

Since both f and $x^{m+1}f(x)$ are integrable, it follows that $x^k f(x) \in L^1(\mathbb{R})$.

In particular, $x^k f(x)$ is integrable for $k = 1, \dots, m$, so the inductive hypothesis implies that $\widehat{f} \in C_0^m(\mathbb{R})$. Further, if we set $g(x) = (-2\pi i x)^m f(x)$, then

$$\widehat{g} = \bigl((-2\pi i x)^m f(x)\bigr)^\wedge = \widehat{f}^{\,(m)}.$$

Now, $g(x), xg(x) \in L^1(\mathbb{R})$, so by the base step we have $\widehat{g} \in C_0^1(\mathbb{R})$ and

$$\widehat{f}^{\,(m+1)} = \widehat{g}\,' = \bigl(-2\pi i x\, g(x)\bigr)^\wedge = \bigl((-2\pi i x)^{m+1} f(x)\bigr)^\wedge.$$

The completes the induction. $\square$

Next we will prove a complementary result showing that smoothness of f implies decay of $\widehat{f}$. The proof will apply the Banach–Zaretsky Theorem and the Fundamental Theorem of Calculus.

Theorem 9.2.14. *Let $f \in L^1(\mathbb{R})$ and $m \in \mathbb{N}$ be given. If f is everywhere m-times differentiable and if $f, f', \dots, f^{(m)} \in L^1(\mathbb{R})$, then*

$$(f^{(k)})^\wedge(\xi) = (2\pi i \xi)^k \widehat{f}(\xi), \qquad \text{for } k = 0, \dots, m.$$

Consequently,

$$|\widehat{f}(\xi)| \le \frac{\|f^{(m)}\|_1}{|2\pi\xi|^m}, \qquad \text{for all } \xi \ne 0. \tag{9.29}$$

Proof. We proceed by induction. The base step is $m = 1$, i.e., we assume that f is everywhere differentiable and $f, f' \in L^1(\mathbb{R})$. Then Corollary 6.3.3, which is a consequence of the Banach–Zaretsky Theorem, implies that f is absolutely continuous on every finite interval. Therefore the Fundamental Theorem of Calculus (Theorem 6.4.2) applies to f, so we have

$$f(x) - f(0) = \int_0^x f'(t)\,dt, \qquad \text{for all } x \in \mathbb{R}.$$

Since f' is integrable, the following limit exists:

$$\lim_{x\to\infty} f(x) = f(0) + \lim_{x\to\infty} \int_0^x f'(t)\,dt = f(0) + \int_0^\infty f'(t)\,dt.$$

Since f is both continuous and integrable, the only way that this limit can exist is if it is zero. Therefore $f(x) \to 0$ as $x \to \infty$. A symmetric argument applies as $x \to -\infty$, so we conclude that $f \in C_0(\mathbb{R})$.

Integration by parts is valid for absolutely continuous functions by Theorem 6.4.6, so for every $a < b$ we have

$$\int_a^b f'(x)\, e^{-2\pi i\xi x}\, dx \;=\; e^{-2\pi i\xi b} f(b) \,-\, e^{-2\pi i\xi a} f(a) \,+\, 2\pi i\xi \int_a^b f(x)\, e^{-2\pi i\xi x}\, dx.$$

Consequently, since f and f' are integrable, we see that

$$\begin{aligned}
\widehat{f'}(\xi) &= \int_{-\infty}^{\infty} f'(x)\, e^{-2\pi i\xi x}\, dx \\
&= \lim_{\substack{a\to-\infty \\ b\to\infty}} \int_a^b f'(x)\, e^{-2\pi i\xi x}\, dx \\
&= \lim_{\substack{a\to-\infty \\ b\to\infty}} \biggl(e^{-2\pi i\xi b} f(b) \,-\, e^{-2\pi i\xi a} f(a) \,+\, 2\pi i\xi \int_a^b f(x)\, e^{-2\pi i\xi x}\, dx \biggr) \\
&= 2\pi i\xi \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i\xi x}\, dx \\
&= 2\pi i\xi\, \widehat{f}(\xi).
\end{aligned}$$

Finally, for $\xi \neq 0$ we have

$$|\widehat{f}(\xi)| \;=\; \frac{|\widehat{f'}(\xi)|}{|2\pi i\xi|} \;\le\; \frac{\|\widehat{f'}\|_\infty}{|2\pi\xi|} \;\le\; \frac{\|f'\|_1}{|2\pi\xi|}.$$

For the inductive step, suppose that the result is valid for some $m \ge 1$, and suppose that f is $(m+1)$-times everywhere differentiable and all of $f, f', \dots, f^{(m)}, f^{(m+1)}$ are integrable. Set $g = f^{(m)}$. Then both g and g' are integrable so, by the base step,

$$(f^{(m+1)})^\wedge(\xi) \;=\; \widehat{g'}(\xi) \;=\; 2\pi i\xi\, \widehat{g}(\xi) \;=\; (2\pi i\xi)^{m+1}\, \widehat{f}(\xi).$$

Therefore the result holds for $m+1$. $\square$

In general, the Fourier transform $\widehat{f}$ of an integrable function f need not itself be integrable. The following corollary gives us a simple sufficient condition that implies that $\widehat{f}$ is integrable.

Corollary 9.2.15. *If $f \in L^1(\mathbb{R})$ is twice differentiable and $f'' \in L^1(\mathbb{R})$, then $\widehat{f} \in L^1(\mathbb{R})$. In particular,*

$$f \in C_c^2(\mathbb{R}) \quad\Longrightarrow\quad \widehat{f} \in L^1(\mathbb{R}).$$

Proof. Since f is integrable, the Riemann–Lebesgue Lemma tells us that $\widehat{f} \in C_0(\mathbb{R})$. Therefore $\widehat{f}$ is continuous, so it is bounded near the origin. Also, since f'' is integrable, Theorem 9.2.14 implies that $|\widehat{f}(\xi)| \le C/|\xi|^2$ away from the origin. Combining these facts, we conclude that $\widehat{f}$ is integrable. $\square$

We introduced a space $A(\mathbb{R})$ in equation (9.19). Restating that equation,

$$A(\mathbb{R}) \;=\; \bigl\{\widehat{f} : f \in L^1(\mathbb{R})\bigr\},$$

i.e., $A(\mathbb{R})$ is the range of the Fourier transform as an operator on the domain $L^1(\mathbb{R})$. We know that $A(\mathbb{R}) \subseteq C_0(\mathbb{R})$, and we will use Corollary 9.2.15 to prove that $A(\mathbb{R})$ is dense in $C_0(\mathbb{R})$ (with respect to the uniform norm, which is the standard norm on $C_0(\mathbb{R})$).

Corollary 9.2.16. *We have*

$$C_c^2(\mathbb{R}) \;\subseteq\; A(\mathbb{R}) \;\subseteq\; C_0(\mathbb{R}).$$

Consequently, $A(\mathbb{R})$ is dense in $C_0(\mathbb{R})$.

Proof. The Riemann–Lebesgue Lemma implies that $A(\mathbb{R})$ is contained in $C_0(\mathbb{R})$. For the other inclusion, let g be any function in $C_c^2(\mathbb{R})$. Then g is continuous and compactly supported, so $g \in L^1(\mathbb{R})$. Further, Corollary 9.2.15 ensures that $\widehat{g} \in L^1(\mathbb{R})$, and by a change of variables we also have $\check{g} \in L^1(\mathbb{R})$. Setting $f = \check{g}$ and applying the Inversion Formula, it follows that

$$g \;=\; (\check{g})^\wedge \;=\; \widehat{f} \in A(\mathbb{R}).$$

Thus $C_c^2(\mathbb{R}) \subseteq A(\mathbb{R})$. Exercise 9.1.16 tells us that the even smaller space $C_c^\infty(\mathbb{R})$ is dense in $C_0(\mathbb{R})$ with respect to the uniform norm, so we conclude that $A(\mathbb{R})$ is dense in $C_0(\mathbb{R})$ as well. $\square$

However, $A(\mathbb{R})$ is a proper subset of $C_0(\mathbb{R})$. According to Problem 9.2.31, one specific function $F \in C_0(\mathbb{R}) \backslash A(\mathbb{R})$ is

$$F(x) \;=\; \begin{cases} 1/\ln x, & \text{if } x > e, \\ x/e, & \text{if } -e \le x \le e, \\ -1/\ln(-x), & \text{if } x < -e. \end{cases} \tag{9.30}$$

There even exist functions in $C_c(\mathbb{R})$ that do not belong to $A(\mathbb{R})$. One example, constructed in [Her85], is

$$B(x) \;=\; \begin{cases} \frac{1}{n}\sin(2\pi 4^n x), & \text{if } \frac{1}{2^{n+1}} \le |x| \le \frac{1}{2^n}, \\ 0, & \text{if } x = 0 \text{ or } |x| > \frac{1}{2}. \end{cases}$$

The letter B is for “butterfly”; see the illustration in Figure 9.7.

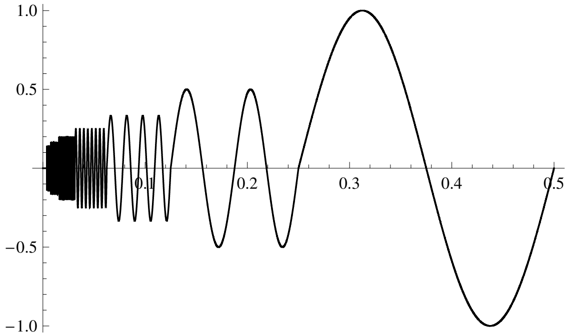


Fig. 9.7 Graph of the butterfly function.

Problems

9.2.17. Show that the Fourier transform is linear on $L^1(\mathbb{R})$, i.e., if $f, g \in L^1(\mathbb{R})$ and $a, b \in \mathbb{C}$, then $(af+bg)^\wedge = a\widehat{f} + b\widehat{g}$.

9.2.18. (a) Prove that if $f \in L^1(\mathbb{R})$ is even, then $\widehat{f}$ is even, and if $f \in L^1(\mathbb{R})$ is odd, then $\widehat{f}$ is odd.

(b) Fix $f \in L^1(\mathbb{R})$. Prove that if $\widehat{f}$ is even then f is even, and if $\widehat{f}$ is odd then f is odd.

9.2.19. Show that the Fourier transforms of the *one-sided exponential* $f(x) = e^{-x}\chi_{[0,\infty)}(x)$ and *two-sided exponential* $g(x) = e^{-|x|}$ are

$$\widehat{f}(\xi) \;=\; \frac{1}{2\pi i\xi + 1} \qquad \text{and} \qquad \widehat{g}(\xi) \;=\; \frac{2}{4\pi^2\xi^2 + 1}.$$

Show further that $\|\widehat{f}\|_\infty = \|f\|_1$ and $\|\widehat{g}\|_\infty = \|g\|_1$.

9.2.20. Let ψ be the *square wave* function $\psi = \chi_{[0,\frac{1}{2})} - \chi_{[-\frac{1}{2},0]}$. Show that

$$\widehat{\psi}(\xi) \;=\; -2i\,\frac{\sin^2(\pi\xi/2)}{\pi\xi},$$

and use this to prove that $\|\widehat{\psi}\|_\infty < \|\psi\|_1 = 1$.

9.2.21. Define the following operations on functions $f\colon \mathbb{R} \to \mathbb{C}$.

$$\begin{array}{llll}
\text{Translation:} & (T_a f)(x) \;=\; f(x-a), & & a \in \mathbb{R}. \\
\text{Modulation:} & (M_b f)(x) \;=\; e^{2\pi i b x} f(x), & & b \in \mathbb{R}. \\
\text{Dilation:} & (D_\lambda) f(x) \;=\; \lambda f(\lambda x), & & \lambda > 0. \\
\text{Involution:} & \widetilde{f}(x) \;=\; \overline{f(-x)}. & &
\end{array}$$

Given a function $f \in L^1(\mathbb{R})$, prove the following statements, and also derive analogous statements for the inverse Fourier transform.

(a) $(T_a f)^\wedge(\xi) = (M_{-a}\widehat{f})(\xi) = e^{-2\pi i a\xi}\widehat{f}(\xi)$.

(b) $(M_b f)^\wedge(\xi) = (T_b\widehat{f})(\xi) = \widehat{f}(\xi - b)$.

(c) $(D_\lambda f)^\wedge(\xi) = \widehat{f}(\xi/\lambda)$.

(d) $\bigl(\widetilde{f}\bigr)^\wedge(\xi) = \overline{\widehat{f}(\xi)}$.

(e) $\bigl(f * \widetilde{f}\bigr)^\wedge(\xi) = |\widehat{f}(\xi)|^2$.

9.2.22. Show that the only function in $L^1(\mathbb{R})$ that satisfies $f = f * f$ is $f = 0$ a.e.

9.2.23. Suppose that $f \in L^1(\mathbb{R})$ is such that $\widehat{f} \in L^1(\mathbb{R})$. Prove the following statements.

(a) $f, \widehat{f} \in C_0(\mathbb{R})$.

(b) $f^{\wedge\wedge}(x) = f(-x)$ for every $x \in \mathbb{R}$.

(c) $f^{\wedge\wedge\wedge\wedge}(x) = f(x)$ for every $x \in \mathbb{R}$.

9.2.24. (a) Prove directly that $(\chi_{[a,b]})^\wedge \in C_0(\mathbb{R})$.

(b) Use part (a) and the density of the really simple functions in $L^1(\mathbb{R})$ to give another proof of the Riemann–Lebesgue Lemma.

9.2.25. Prove that the Fourier transform is a *continuous* mapping of $L^1(\mathbb{R})$ into $C_0(\mathbb{R})$. That is, show that if $f_n, f \in L^1(\mathbb{R})$ and $f_n \to f$ in L^1-norm, then $\widehat{f_n} \to \widehat{f}$ uniformly.

9.2.26. Prove that if $\{k_N\}_{N\in\mathbb{N}}$ is an approximate identity, then $\widehat{k_N}(\xi) \to 1$ pointwise as $N \to \infty$.

9.2.27. Given $f \in L^1(\mathbb{R})$, show that

$$\{T_a f\}_{a\in\mathbb{R}} \text{ is complete in } L^1(\mathbb{R}) \implies \widehat{f}(\xi) \neq 0 \text{ for all } \xi \in \mathbb{R}. \tag{9.31}$$

Remark: The converse of equation (9.31) is also true, but this is a deeper fact that is a consequence of *Wiener's Tauberian Theorem.*

9.2.28. Show that if $f, g \in L^1(\mathbb{R})$ and $\widehat{f} \in L^1(\mathbb{R})$, then $fg \in L^1(\mathbb{R})$ and $(fg)^\wedge = \widehat{f} * \widehat{g}$.

9.2.29. Suppose $f \in L^1(\mathbb{R})$ and there exist constants $C > 0$ and $0 < \alpha < 1$ such that $|\widehat{f}(\xi)| \le C/|\xi|^{1+\alpha}$ for all $\xi \neq 0$. Prove that f is Hölder continuous with exponent α.

9.2.30. Show that $\displaystyle\int_{-\infty}^{\infty} \frac{\sin \pi x}{x}\, e^{-2\pi|x|+\pi i x}\, dx \;=\; \frac{\pi}{4}$.

9.2.31. Prove the following statements.

(a) If $f \in L^1(\mathbb{R})$ is odd, then $\sup_{b\geq 1} \left|\int_1^b \frac{\widehat{f}(\xi)}{\xi}\,d\xi\right| < \infty$.

(b) If $f \in L^1(\mathbb{R})$ is odd, $\widehat{f}$ is differentiable at $\xi = 0$, and $\widehat{f} \geq 0$ on $(0,\infty)$, then $\widehat{f}(\xi)/\xi \in L^1(\mathbb{R})$.

(c) The function F defined in equation (9.30) belongs to $C_0(\mathbb{R})$ but does not belong to $A(\mathbb{R})$.

9.2.32. Let $Df = f'$, and for $k \in \mathbb{N}$ let $D^k f = f^{(k)}$.

(a) Show that if f is n-times differentiable and $x^j f^{(k)}(x) \in L^1(\mathbb{R})$ for $j = 0, \dots, m$ and $k = 0, \dots, n$, then

$$\Big(D^n\big((-2\pi i x)^m f(x)\big)\Big)^{\wedge}(\xi) \;=\; (2\pi i\xi)^n\, D^m \widehat{f}(\xi), \qquad \text{for all } \xi \in \mathbb{R}.$$

(b) The *Schwartz space* is

$$\mathcal{S}(\mathbb{R}) \;=\; \big\{f \in C^\infty(\mathbb{R}) : x^m f^{(n)}(x) \in L^\infty(\mathbb{R}) \text{ for all } m,n \geq 0\big\}.$$

Exhibit a nonzero function in $\mathcal{S}(\mathbb{R})$, and show that if $f \in \mathcal{S}(\mathbb{R})$, then $f^{(n)}$ is integrable for every $n \geq 0$. Prove that $\mathcal{S}(\mathbb{R})$ is dense in $L^1(\mathbb{R})$.

(c) Show that if $f \in \mathcal{S}(\mathbb{R})$, then $\widehat{f} \in \mathcal{S}(\mathbb{R})$.

(d) Prove that the Fourier transform maps $\mathcal{S}(\mathbb{R})$ bijectively onto itself.

9.3 Fourier Series

We proved in Section 8.4 that the trigonometric system $\{e^{2\pi i n x}\}_{n\in\mathbb{Z}}$ is an orthonormal sequence in $L^2[0,1]$, and we stated that we would later prove that the trigonometric system is complete in $L^2[0,1]$ and hence is an orthonormal basis for that Hilbert space. We will complete that proof in this section (and also establish other interesting results).

Throughout this section we continue to take $\overline{\mathbf{F}} = \mathbb{C}$, and for notational convenience we set

$$e_n(x) \;=\; e^{2\pi i n x}, \qquad \text{for } n \in \mathbb{Z}.$$

Also, we let

$$\mathcal{E} \;=\; \{e^{2\pi i n x}\}_{n\in\mathbb{Z}} \;=\; \{e_n\}_{n\in\mathbb{Z}}$$

denote the trigonometric system.

As noted, one of our main goals is to prove that $\mathcal{E}$ is an orthonormal basis for $L^2[0,1]$. Once we have established that, it will follow from Theorem 8.3.7 that every function $f \in L^2[0,1]$ can be uniquely written as

$$f = \sum_{n\in\mathbb{Z}} \langle f, e_n\rangle\, e_n, \tag{9.32}$$

where this series converges unconditionally in L^2-norm. Equation (9.32) is referred to as the *Fourier series* for f. The inner products $\langle f, e_n\rangle$ are called the *Fourier coefficients* of f, and are traditionally denoted by

$$\widehat{f}(n) = \langle f, e_n\rangle = \int_0^1 f(x)\, e^{-2\pi i n x}\, dx, \qquad \text{for } n \in \mathbb{Z}. \tag{9.33}$$

When we want to refer to the entire sequence of Fourier coefficients, we denote it by

$$\widehat{f} = \bigl(\widehat{f}(n)\bigr)_{n\in\mathbb{Z}}.$$

Although much of our interest is in $L^2[0,1]$, every integrable function f in $L^1[0,1]$ (which contains $L^2[0,1]$) has Fourier coefficients $\widehat{f}(n)$ that are defined by equation (9.33) for $n \in \mathbb{Z}$. However, while we will prove that the Fourier series representation in equation (9.32) holds for $f \in L^2[0,1]$, there are integrable functions f for which equation (9.32) does not hold. The convergence of Fourier series in senses other than L^2-norm can be a very subtle issue, which we will explore in Section 9.3.6.

Fourier series and the Fourier transform have many similarities, and we will see that many of the facts that we proved in Section 9.2 for the Fourier transform have analogues for Fourier coefficients (in fact, historically speaking, Fourier series came first). In particular, the techniques that we will use to prove that the trigonometric system is complete in $L^2[0,1]$ are similar to the ones that we employed when we proved the Inversion Formula for the Fourier transform. On the other hand, while there are many similarities, there are interesting differences as well.

9.3.1 Periodic Functions

When we discussed Fourier series and the trigonometric system in Section 8.4 we considered $L^2[0,1]$, the space of square-integrable functions on the domain $[0,1]$. However, it is entirely equivalent and often more convenient to instead consider the space of functions that are 1-periodic on $\mathbb{R}$ and are square-integrable on $[0,1]$, where 1-*periodic* means that

$$f(x+1) = f(x) \quad \text{for } x \in \mathbb{R}.$$

We will denote this space by

$$L^2(\mathbb{T}) = \biggl\{ f\colon\mathbb{R}\to\mathbb{C} \;:\; f \text{ is 1-periodic and } \int_0^1 |f(x)|^2\, dx < \infty \biggr\}.$$

As usual, we identify any two functions in $L^2(\mathbb{T})$ that are equal a.e. The norm on $L^2(\mathbb{T})$ is

$$\|f\|_2 \;=\; \biggl(\int_0^1 |f(x)|^2\,dx\biggr)^{1/2}.$$

We define $L^p(\mathbb{T})$ similarly for finite p, and we let $L^\infty(\mathbb{T})$ be the set of all essentially bounded 1-periodic functions. Since the interval $[0,1]$ has finite measure, we have

$$L^p(\mathbb{T}) \subseteq L^1(\mathbb{T}), \qquad \text{for } 1 \le p \le \infty.$$

In contrast, $L^p(\mathbb{R})$ is not contained in $L^1(\mathbb{R})$ for any $p > 1$, nor is $L^1(\mathbb{R})$ contained in $L^p(\mathbb{R})$.

Other spaces of functions on $\mathbb{T}$ are defined in the same way. For example, $C(\mathbb{T})$ is the space of all continuous, 1-periodic functions, and $C^m(\mathbb{T})$ is the space of all m-times differentiable, 1-periodic functions f such that $f, f', \dots, f^{(m)}$ are all continuous.

A trivial, but important, fact about 1-periodic functions is that if f is an element of $L^1(\mathbb{T})$, then

$$\int_0^1 f(x-y)\,dx \;=\; \int_0^1 f(x)\,dx, \qquad \text{for every } y \in \mathbb{R}. \tag{9.34}$$

Thus, integrals on $\mathbb{T}$ are invariant under the change of variable $x \mapsto x - y$.

Remark 9.3.1. A 1-periodic function is entirely determined by its values on the interval $[0,1)$ (note that if we are considering almost-everywhere properties then we can use whichever of $[0,1)$ or $[0,1]$ is more convenient). In essence, when dealing with 1-periodic functions we are really considering functions on the group $[0,1)$ endowed with the operation of addition modulo 1. Letting a mod 1 denote the fractional part of a, we can write the group operation on $[0,1)$ as

$$x \oplus y \;=\; x + y \bmod 1 \;=\; \begin{cases} x+y, & \text{if } 0 \le x+y < 1,\\ x+y-1, & \text{if } 1 \le x+y < 2. \end{cases}$$

This group is isomorphic to the circle group $S^1 = \{e^{i\theta} : \theta \in \mathbb{R}\}$ under multiplication of complex scalars. The circle is the one-dimensional torus; hence our use of the letter $\mathbb{T}$ in this context. $\diamondsuit$

9.3.2 Decay of Fourier Coefficients

We begin by proving some facts about Fourier coefficients that are reminiscent of results that we established for the Fourier transform. For example, Lemma 9.2.3 showed that if $f \in L^1(\mathbb{R})$, then its Fourier transform $\widehat{f}$ is both bounded and continuous. Now suppose that f is a 1-periodic integrable function, i.e., $f \in L^1(\mathbb{T})$. Then its Fourier coefficients $\widehat{f}(n)$ are defined only for integer values of n, so it no longer makes sense to ask whether $\widehat{f}$ is continuous, but we see from the computation

$$|\widehat{f}(n)| \;=\; \left|\int_0^1 f(x)\, e^{-2\pi i n x}\, dx\right| \;\le\; \int_0^1 |f(x)\, e^{-2\pi i n x}|\, dx \;=\; \|f\|_1 \tag{9.35}$$

that $\widehat{f}(n)$ is bounded in n. In fact, equation (9.35) shows that if $f \in L^1(\mathbb{T})$ then the sequence of Fourier coefficients $\widehat{f}$ belongs to $\ell^\infty(\mathbb{Z})$, and

$$\|\widehat{f}\|_\infty \;\le\; \|f\|_1.$$

The next exercise gives a refinement of this fact.

Exercise 9.3.2 (Riemann–Lebesgue Lemma). Show that if $f \in L^1(\mathbb{T})$, then $\widehat{f} \in c_0$, i.e.,

$$\lim_{|n|\to\infty} \widehat{f}(n) \;=\; 0. \qquad \diamondsuit$$

However, the fact that $\widehat{f}$ belongs to c_0 does not give us any quantitative information on *how quickly* (or slowly) $\widehat{f}(n)$ decays to zero. Our next result gives a connection between the total variation of f and the decay of its Fourier coefficients. Here, $\mathrm{BV}(\mathbb{T})$ denotes the set of 1-periodic functions that have bounded variation on the interval $[0,1]$. The total variation of a 1-periodic function f is $V[f;\mathbb{T}] = V[f;0,1]$.

Theorem 9.3.3. *If $f \in \mathrm{BV}(\mathbb{T})$, then*

$$|\widehat{f}(n)| \;\le\; \frac{V[f;\mathbb{T}]}{|n|}, \qquad \text{for all } n \ne 0.$$

Proof. Fix any integer $n > 0$, and for each integer k let I_k be the interval

$$I_k \;=\; \left[\tfrac{k-1}{n}, \tfrac{k}{n}\right).$$

Let g be the step function on $[0,1)$ defined by

$$g \;=\; \sum_{k=1}^{n} f\left(\tfrac{k}{n}\right) \chi_{I_k}.$$

If we assume that g is extended 1-periodically to $\mathbb{R}$, then $g \in L^1(\mathbb{T})$. Therefore the Fourier coefficients $\widehat{g}(j)$ exist for all $j \in \mathbb{Z}$. In particular, the nth Fourier coefficient of g is

$$\widehat{g}(n) \;=\; \sum_{k=1}^{n} f\bigl(\tfrac{k}{n}\bigr) \int_{\frac{k-1}{n}}^{\frac{k}{n}} e^{-2\pi i n x}\, dx \;=\; \sum_{k=1}^{n} f\bigl(\tfrac{k}{n}\bigr) \int_{k-1}^{k} e^{-2\pi i x}\, \frac{dx}{n} \;=\; 0.$$

Recall that if $a \le x < y \le b$, then

$$|f(x) - f(y)| \;\le\; V[f; x, y] \;\le\; V[f; a, b].$$

Therefore,

$$\begin{aligned}
|\widehat{f}(n)| \;=\; |\widehat{f}(n) - \widehat{g}(n)| \;&=\; \left| \int_0^1 \bigl(f(x) - g(x)\bigr)\, e^{-2\pi i n x}\, dx \right| \\
&\le\; \int_0^1 |f(x) - g(x)|\, dx \\
&=\; \sum_{k=1}^{n} \int_{\frac{k-1}{n}}^{\frac{k}{n}} \bigl|f(x) - f\bigl(\tfrac{k}{n}\bigr)\bigr|\, dx \\
&\le\; \sum_{k=1}^{n} \frac{1}{n}\, V\bigl[f; \tfrac{k-1}{n}, \tfrac{k}{n}\bigr] \\
&\le\; \frac{1}{n}\, V[f; 0, 1],
\end{aligned}$$

where at the last step we have used the additivity property of the variation given in Lemma 5.2.12. □

Thus, the Fourier coefficients of a function with bounded variation decay on the order of $1/n$. The next exercise gives a decay estimate for differentiable functions, similar to the relationship between smoothness and decay for the Fourier transform that was obtained in Theorem 9.2.14.

Exercise 9.3.4. Let $m \in \mathbb{N}$ be given. Prove that if $f \in C^m(\mathbb{T})$ then

$$(f^{(k)})^\wedge(n) \;=\; (2\pi i n)^k\, \widehat{f}(n), \qquad \text{for } n \in \mathbb{Z} \text{ and } k = 0, \dots, m.$$

Use this to show that

$$|\widehat{f}(n)| \;\le\; \frac{\|f^{(m)}\|_1}{|2\pi n|^m}, \qquad \text{for all } n \neq 0. \qquad \diamondsuit$$

In particular, it follows that if $f \in C^2(\mathbb{T})$, then its Fourier coefficients $\widehat{f}(n)$ are summable. Consequently, if we set

$$A(\mathbb{T}) \;=\; \bigl\{ f \in L^1(\mathbb{T}) : \widehat{f} \in \ell^1 \bigr\},$$

then

$$C^2(\mathbb{T}) \;\subseteq\; A(\mathbb{T}).$$

However, this is not the best result. Let $C^\alpha(\mathbb{T})$ be the space of 1-periodic functions that are Hölder continuous with exponent α. Then *Bernstein's Theorem* says that $C^\alpha(\mathbb{T}) \subseteq A(\mathbb{T})$ for all $\alpha > 1/2$. This result is sharp, i.e., $C^{1/2}(\mathbb{T})$ is not contained in $A(\mathbb{T})$. For proofs of these facts, see [Kat04, Thm. 6.3].

9.3.3 Convolution of Periodic Functions

One reason that we prefer $L^p(\mathbb{T})$ over $L^p[0,1]$ is that it is notationally simpler to define the convolution of 1-periodic functions than functions on $[0,1]$, because we can avoid the use of the mod 1 operator. We give the formal definition next; note how the assumption that g is 1-periodic comes into play when we translate g to obtain $g(x-y)$. If we wanted to define the convolution of functions on the domain $[0,1]$, we would replace $g(x-y)$ in equation (9.36) with $g(x-y \bmod 1)$.

Definition 9.3.5 (Convolution). Assume that f and g are measurable, 1-periodic functions. Their *convolution* is the function $f * g$ formally defined by

$$(f * g)(x) \;=\; \int_0^1 f(y)\, g(x-y)\, dy, \tag{9.36}$$

if this integral exists. $\diamondsuit$

Here is Young's Inequality for convolution of 1-periodic functions.

Exercise 9.3.6 (Young's Inequality). Fix $1 \le p \le \infty$, and assume that $f \in L^p(\mathbb{T})$ and $g \in L^1(\mathbb{T})$. Prove that

(a) $f * g$ is defined a.e.,

(b) $f * g$ is 1-periodic,

(c) $f * g$ is measurable and $f * g \in L^p(\mathbb{T})$,

(d) $\|f * g\|_p \;\le\; \|f\|_p\, \|g\|_1$, and

(e) $(f*g)^\wedge(n) \;=\; \widehat{f}(n)\, \widehat{g}(n)$ for all $n \in \mathbb{Z}$. $\diamondsuit$

9.3.4 Approximate Identities and the Inversion Formula

We define approximate identities for periodic functions similarly to how we defined them for functions on the real line (compare Definition 9.1.8).

Definition 9.3.7 (Approximate Identity). An *approximate identity* or a *summability kernel* on $\mathbb{T}$ is a family $\{k_N\}_{N\in\mathbb{N}}$ of functions in $L^1(\mathbb{T})$ such that the following three conditions are satisfied.

(a) L^1-*normalization*: $\int_0^1 k_N(x)\,dx = 1$ for every $N \in \mathbb{N}$.

(b) L^1-*boundedness*: $\sup \|k_N\|_1 < \infty$.

(c) L^1-*concentration*: For every $0 < \delta < \frac{1}{2}$,

$$\lim_{N\to\infty} \int_{\delta \le |x| < \frac{1}{2}} |k_N(x)|\,dx \;=\; 0. \qquad \diamondsuit$$

Here is the analogue of Theorem 9.1.15 for 1-periodic functions.

Exercise 9.3.8. Let $\{k_N\}_{N\in\mathbb{N}}$ be an approximate identity for $\mathbb{T}$. Prove the following statements.

(a) If $1 \le p < \infty$ and $f \in L^p(\mathbb{T})$, then $f * k_N \to f$ in L^p-norm as $N \to \infty$.

(b) If $f \in C(\mathbb{T})$, then $f * k_N \to f$ uniformly as $N \to \infty$. $\diamondsuit$

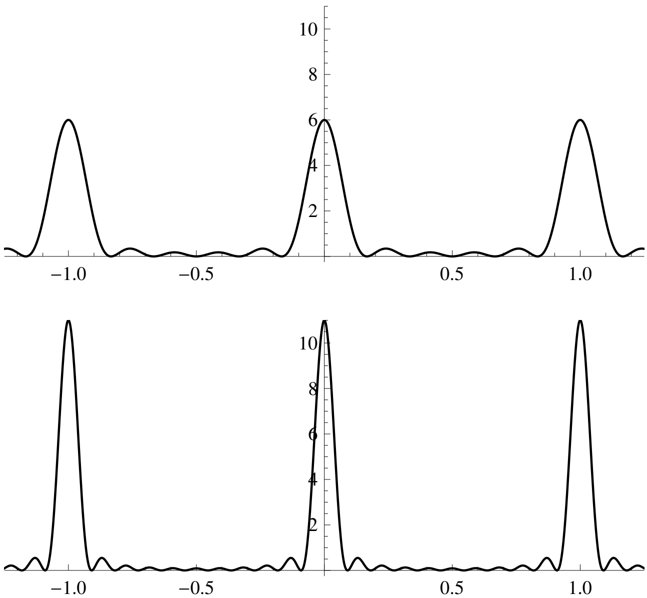


Fig. 9.8 Two elements of the Fejér kernel. Top: w_5. Bottom: w_{10}.

We will need a periodic analogue of the Fejér kernel. We obtained the Fejér kernel on $\mathbb{R}$ by starting with the Fejér function w, which is the Fourier transform of the hat function $W(x) = \max\{1-|x|, 0\}$. We dilated w to obtain the elements w_N of the Fejér kernel. Unfortunately, there is no convenient dilation that we can apply to 1-periodic functions, but still we can create w_N as the transform of a hat function. Specifically, the "discrete hat function" supported on the set of integers $\{-N-1, \dots, N+1\}$ is

$$W_N(n) \;=\; \max\biggl\{1 - \frac{|n|}{N+1},\, 0\biggr\}, \qquad n \in \mathbb{Z}. \tag{9.37}$$

Just as we obtained the Fejér function by taking the Fourier transform of the hat function, we now define w_N by using $W_N(n)$ as coefficients in a Fourier series. That is, we define

$$w_N(x) \;=\; \sum_{n\in\mathbb{Z}} W_N(n)\, e^{2\pi inx} \;=\; \sum_{n=-N}^{N} \biggl(1 - \frac{|n|}{N+1}\biggr)\, e^{2\pi inx}. \tag{9.38}$$

The Fejér kernel for $\mathbb{T}$ is $\{w_N\}_{N\in\mathbb{N}}$. Some elements of the Fejér kernel are shown in Figure 9.8. We can see in the figure that w_N appears to become more like a "1-periodic spike train" as N increases, which is qualitatively what we expect of an approximate identity. However, in contrast to the Fejér kernel for the real line defined in Exercise 9.1.10, these functions w_N are not obtained by a dilation of some single function, and as a result it takes more work to prove that $\{w_N\}_{N\in\mathbb{N}}$ is an approximate identity for $\mathbb{T}$.

Exercise 9.3.9. Given scalars a_k for $k \in \mathbb{Z}$, let $s_N = \sum_{k=-N}^{N} a_k$ denote the (symmetric) partial sums of these scalars. Their *Cesàro means* are the averages

$$\sigma_N \;=\; \frac{s_0 + \cdots + s_N}{N+1}$$

of the partial sums. Prove the following statements.

(a) $\displaystyle \sigma_N \;=\; \sum_{n=-N}^{N} \biggl(1 - \frac{|n|}{N+1}\biggr)\, a_n \;=\; \sum_{n=-N}^{N} W_N(n)\, a_n.$

(b) $\displaystyle \sum_{n=-N}^{N} e^{2\pi inx} \;=\; \frac{\sin{(2N+1)\pi x}}{\sin \pi x}.$

(c) The function w_N defined by equation (9.38) satisfies

$$w_N(x) \;=\; \frac{1}{N+1}\, \biggl(\frac{\sin{(N+1)\pi x}}{\sin \pi x}\biggr)^2.$$

(d) $\{w_N\}_{N\in\mathbb{N}}$ is an approximate identity for $\mathbb{T}$. $\diamondsuit$

The Fejér kernel is certainly not the only approximate identity for $\mathbb{T}$, but it will be useful for our purposes. One kernel that we *cannot* use is the *Dirichlet kernel* $\{d_N\}_{N\in\mathbb{N}}$, whose elements are the Fourier transforms of the "discrete box function" on $\{-N, \dots, N\}$. Specifically, d_N is defined by

$$d_N(x) \;=\; \sum_{n=-N}^{N} e^{2\pi inx} \;=\; \frac{\sin{(2N+1)\pi x}}{\sin \pi x}. \tag{9.39}$$

Each function d_N is integrable on $\mathbb{T}$, and its graph does appear to become more like a "1-periodic spike train" as $N \to \infty$ (see Figure 9.9). However, the oscillations of d_N decay so slowly with N (see Problem 9.3.35) that we end up with

$$\sup_{N\in\mathbb{N}} \|d_N\|_1 \;=\; \sup_{N\in\mathbb{N}} \int_0^1 |d_N(x)|\, dx \;=\; \infty.$$

That is, the "absolute mass" of d_N grows with N. The "signed mass" of d_N is

$$\int_0^1 d_N \;=\; 1, \qquad \text{for every } N \in \mathbb{N},$$

but we achieve this only because the large oscillations of d_N produce "miraculous cancellations" in the integral. Consequently, the Dirichlet kernel is not an approximate identity for $\mathbb{T}$.

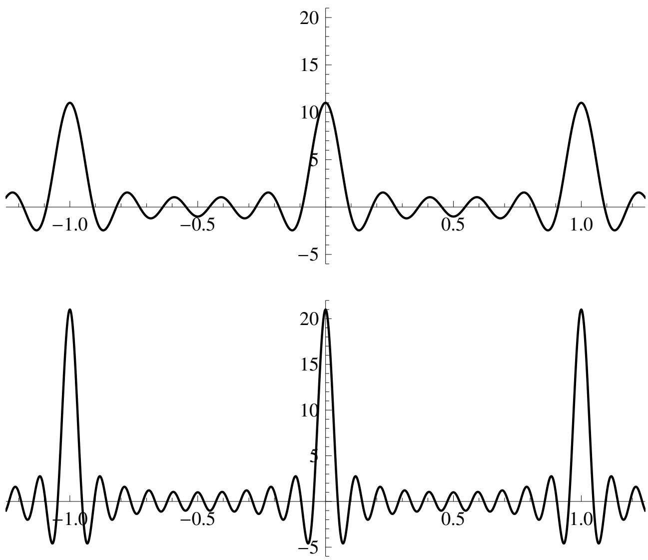


Fig. 9.9 Two elements of the Dirichlet kernel. Top: d_5. Bottom: d_{10}.

The fact that the Dirichlet kernel is not an approximate identity is unpleasant but very important. To see why, recall that we are hoping to prove that the trigonometric system $\mathcal{E} = \{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{T})$, which implies in particular that for all $f \in L^2(\mathbb{T})$ we will have

$$f \;=\; \sum_{n\in\mathbb{Z}} \widehat{f}(n)\, e_n.$$

The partial sums of this series are therefore crucial, since we must show that they converge to f. The *symmetric partial sums* $S_N f = \sum_{n=-N}^{N} \widehat{f}(n)\, e_n$ of this series are precisely the convolutions of f with d_N! This is because

$$\begin{aligned}
(f * d_N)(x) &= \int_0^1 f(t)\, d_N(x-t)\, dt \\
&= \int_0^1 f(t) \sum_{n=-N}^{N} e^{2\pi i n(x-t)}\, dt \\
&= \sum_{n=-N}^{N} \biggl(\int_0^1 f(t)\, e^{-2\pi i n t}\, dt\biggr) e^{2\pi i n x} \\
&= \sum_{n=-N}^{N} \widehat{f}(n)\, e_n(x) \;=\; S_N f(x). \qquad (9.40)
\end{aligned}$$

If it were the case that the Dirichlet kernel $\{d_N\}_{N\in\mathbb{N}}$ was an approximate identity, then Exercise 9.3.8 would immediately imply that the partial sums $S_N f = f * d_N$ converge to f in L^p-norm for every $f \in L^p(\mathbb{T})$ and every index $1 \le p \le \infty$. This is precisely what we are hoping to prove when $p = 2$. And we will prove this for $p = 2$, but the point is that we cannot use the Dirichlet kernel to do it because $\{d_N\}_{N\in\mathbb{N}}$ is not an approximate identity. (Moreover, this is *not* true for $p = 1$ or $p = \infty$, yet it would have to be true for all $1 \le p \le \infty$ if $\{d_N\}_{N\in\mathbb{N}}$ were an approximate identity.)

Instead of trying to deal with $f * d_N$, which is the actual Nth symmetric partial sum of the Fourier series, we will instead consider the convolution of f with elements of the Fejér kernel. A computation similar to the one that led to equation (9.40) shows that if $f \in L^1(\mathbb{T})$, then

$$\begin{aligned}
(f * w_N)(x) &= \int_0^1 f(t)\, w_N(x-t)\, dt \\
&= \int_0^1 f(t) \sum_{n=-N}^{N} \Bigl(1 - \frac{|n|}{N+1}\Bigr) e^{2\pi i n(x-t)}\, dt \\
&= \sum_{n=-N}^{N} \Bigl(1 - \frac{|n|}{N+1}\Bigr) \biggl(\int_0^1 f(t)\, e^{-2\pi i n t}\, dt\biggr) e^{2\pi i n x} \\
&= \sum_{n=-N}^{N} \Bigl(1 - \frac{|n|}{N+1}\Bigr) \widehat{f}(n)\, e^{2\pi i n x} \\
&= \sum_{n=-N}^{N} W_N(n)\, \widehat{f}(n)\, e_n(x). \qquad (9.41)
\end{aligned}$$

Thus $f * w_N$ is precisely the Nth Cesàro mean of the symmetric partial sums of the Fourier series of f. Since $\{w_N\}_{N\in\mathbb{N}}$ is an approximate identity, these Cesàro means $f * w_N$ are much better behaved than the actual partial sums $f * d_N$. Indeed, by applying Exercise 9.3.8 we immediately deduce the following convergence results.

Lemma 9.3.10. (a) *If* $1 \le p < \infty$ *and* $f \in L^p(\mathbb{T})$, *then* $f * w_N \to f$ *in* L^p*-norm as* $N \to \infty$.

(b) *If* $f \in C(\mathbb{T})$, *then* $f * w_N \to f$ *uniformly as* $N \to \infty$. $\diamondsuit$

Lemma 9.3.10 tells us only that the *Cesàro means* $f * w_N$ of the symmetric partial sums converge to f. Still, we will use this to prove the following Inversion Formula for 1-periodic functions, which says that if f is integrable *and* $\widehat{f}$ is summable, then the partial sums of the Fourier series converge uniformly to f (and therefore, since $[0,1]$ has finite measure, they also converge in L^p-norm for every p). This result is analogous to the Inversion Formula for the Fourier transform that we obtained in in Theorem 9.2.9.

Theorem 9.3.11 (Inversion Formula). *If* $f \in L^1(\mathbb{T})$ *and* $\widehat{f} \in \ell^1(\mathbb{Z})$, *then* f *is continuous and*

$$f(x) \;=\; \sum_{n\in\mathbb{Z}} \widehat{f}(n)\, e^{2\pi i n x}, \qquad \textit{for all } x \in \mathbb{R},$$

where this series converges with respect to the uniform norm (*in fact, it converges absolutely, and therefore unconditionally, with respect to* $\|\cdot\|_{\mathrm{u}}$).

Proof. Since $\widehat{f} \in \ell^1(\mathbb{Z})$ and the uniform norm of $e_n(x) = e^{2\pi i n x}$ is $\|e_n\|_{\mathrm{u}} = 1$, the sum of the norms of the terms in the Fourier series for f is

$$\sum_{n\in\mathbb{Z}} \|\widehat{f}(n)\, e_n\|_{\mathrm{u}} \;=\; \sum_{n\in\mathbb{Z}} |\widehat{f}(n)| \;=\; \|\widehat{f}\|_1 \;<\; \infty.$$

Hence the series

$$\bigl(\widehat{f}\,\bigr)^{\vee} \;=\; \sum_{n\in\mathbb{Z}} \widehat{f}(n)\, e_n \tag{9.42}$$

converges absolutely with respect to the uniform norm $\|\cdot\|_{\mathrm{u}}$. Since $C(\mathbb{T})$ is a Banach space, an absolutely convergent series in $C(\mathbb{T})$ must converge (in fact, it converges unconditionally). Therefore the series in equation (9.42) converges uniformly, and $\bigl(\widehat{f}\,\bigr)^{\vee} \in C(\mathbb{T})$. Our task is to show that $\bigl(\widehat{f}\,\bigr)^{\vee}$ equals f (as an element of $L^1(\mathbb{T})$).

Equation (9.41) tells us that

$$(f * w_N)(x) \;=\; \sum_{n\in\mathbb{Z}} W_N(n)\, \widehat{f}(n)\, e_n(x).$$

Fix any particular x. If $n \in \mathbb{Z}$, then $W_N(n) \to 1$ as $N \to \infty$, so

$$\lim_{N\to\infty} W_N(n)\, \widehat{f}(n)\, e_n(x) \;=\; \widehat{f}(n)\, e_n(x).$$

Further, $|W_N(n)\,\widehat{f}(n)\,e_n(x)| \le |\widehat{f}(n)|$ and $\widehat{f} \in \ell^1(\mathbb{Z})$. Therefore, we can apply the series version of the Dominated Convergence Theorem to obtain

$$\begin{aligned}
\lim_{N\to\infty} (f * w_N)(x) &= \lim_{N\to\infty} \sum_{n\in\mathbb{Z}} W_N(n)\,\widehat{f}(n)\,e^{2\pi inx} \\
&= \sum_{n\in\mathbb{Z}} \lim_{N\to\infty} W_N(n)\,\widehat{f}(n)\,e^{2\pi inx} \\
&= \sum_{n\in\mathbb{Z}} \widehat{f}(n)\,e^{2\pi inx} \;=\; \big(\widehat{f}\,\big)^{\vee}(x).
\end{aligned}$$

On the other hand, Lemma 9.3.10 implies that $f * w_N \to f$ in L^1-norm, so there is a subsequence such that $(f * w_{N_k})(x) \to f(x)$ pointwise a.e. Therefore $\big(\widehat{f}\,\big)^{\vee}(x) = f(x)$ for a.e. x. Thus f is equal almost everywhere to the continuous function $\big(\widehat{f}\,\big)^{\vee}$, which is what we mean when we say that an element of $L^1(\mathbb{R})$ is continuous. $\square$

As a corollary, we see that integrable functions are uniquely determined by their Fourier coefficients.

Corollary 9.3.12 (Uniqueness Theorem). *If $f, g \in L^1(\mathbb{T})$, then*

$$f = g \ \text{a.e.} \qquad \Longleftrightarrow \qquad \widehat{f}(n) = \widehat{g}(n) \ \text{for every } n \in \mathbb{Z}.$$

In particular,

$$f = 0 \ \text{a.e.} \qquad \Longleftrightarrow \qquad \widehat{f}(n) = 0 \ \text{for every } n \in \mathbb{Z}. \qquad \diamondsuit$$

9.3.5 Completeness of the Trigonometric System

We know that the trigonometric system $\mathcal{E} = \{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal *sequence* in $L^2(\mathbb{T})$, and now we want to prove that it is an orthonormal *basis* for $L^2(\mathbb{T})$. Because $L^2(\mathbb{T})$ is a Hilbert space and because $\mathcal{E}$ is *orthonormal*, Theorem 8.3.7 tells us that in order to prove that $\mathcal{E}$ is a basis we need only prove that $\mathcal{E}$ is complete. That is, if we can simply show that the finite linear span of $\mathcal{E}$ is dense in $L^2(\mathbb{T})$, then we can immediately conclude that every $f \in L^2(\mathbb{T})$ can actually be written as $f = \sum_{n\in\mathbb{Z}} \widehat{f}(n)\,e_n$, where the series converges unconditionally in L^2-norm.

We will use the Fejér kernel to prove that $\mathcal{E}$ is complete in $L^2(\mathbb{T})$. In fact, the same proof shows that $\mathcal{E}$ is complete in $L^p(\mathbb{T})$ for every finite p, and also that it is complete in $C(\mathbb{T})$. Unfortunately, only for $p = 2$ does this allow us to draw any extra conclusion about the basis properties of $\mathcal{E}$. At the end of this section we will comment more on the differences between the cases $p = 2$ and $p \neq 2$.

Theorem 9.3.13. (a) $\mathcal{E} = \{e_n\}_{n\in\mathbb{Z}}$ *is complete in* $L^p(\mathbb{T})$ *for each* $1 \le p < \infty$.
(b) $\mathcal{E} = \{e_n\}_{n\in\mathbb{Z}}$ *is complete in* $C(\mathbb{T})$ *with respect to the uniform norm.*

Proof. (a) Since p is finite, if $f \in L^p(\mathbb{T})$ then $f * w_N \to f$ in L^p-norm (see Lemma 9.3.10). From equation (9.41),

$$f * w_N \;=\; \sum_{n=-N}^{N} W_N(n)\,\widehat{f}(n)\,e_n \;\in\; \mathrm{span}(\mathcal{E}),$$

so we conclude that f is the limit in L^p-norm of a sequence of elements of $\mathrm{span}(\mathcal{E})$. This implies that $\mathrm{span}(\mathcal{E})$ is dense in $L^p(\mathbb{T})$, and therefore $\mathcal{E}$ is a complete sequence in $L^p(\mathbb{T})$.

(b) The proof is similar, using the fact that Lemma 9.3.10 implies that $f * w_N \to f$ uniformly for every $f \in C(\mathbb{T})$. $\square$

For $p = 2$, we obtain the following corollary.

Corollary 9.3.14 (The Trigonometric System is an ONB). *The trigonometric system* $\mathcal{E} = \{e_n\}_{n\in\mathbb{Z}}$ *is an orthonormal basis for* $L^2(\mathbb{T})$. *Consequently, if* $f \in L^2(\mathbb{T})$ *then*

$$f \;=\; \sum_{n\in\mathbb{Z}} \widehat{f}(n)\,e_n, \tag{9.43}$$

where this series converges unconditionally in L^2*-norm. Further, we have the* Plancherel Equality,

$$\|f\|_2^2 \;=\; \sum_{n\in\mathbb{Z}} |\widehat{f}(n)|^2, \qquad \text{for all } f \in L^2(\mathbb{T}), \tag{9.44}$$

and the Parseval Equality,

$$\langle f, g\rangle \;=\; \sum_{n\in\mathbb{Z}} \widehat{f}(n)\,\overline{\widehat{g}(n)}, \qquad \text{for all } f, g \in L^2(\mathbb{T}).$$

Proof. Since $\mathcal{E}$ is both orthonormal and complete in $L^2(\mathbb{T})$, Theorem 8.3.7 implies that $\mathcal{E}$ is an orthonormal basis for $L^2(\mathbb{T})$. $\square$

Thus, the L^2-norm of a function $f \in L^2(\mathbb{T})$ is exactly equal to the ℓ^2-norm of its sequence of Fourier coefficients $\widehat{f} = \big(\widehat{f}(n)\big)_{n\in\mathbb{Z}}$. Moreover, equation (9.43) shows that every square-integrable 1-periodic function f can be represented as a countable superposition of the "pure tones" $e_n(x) = e^{2\pi i n x}$ over integers $n \in \mathbb{Z}$.

Example 9.3.15. Let $f = \chi_{[0,1/2)} - \chi_{[1/2,1)}$ be the square wave function (also known as the *Haar wavelet*). This function is square-integrable on $[0,1]$, so

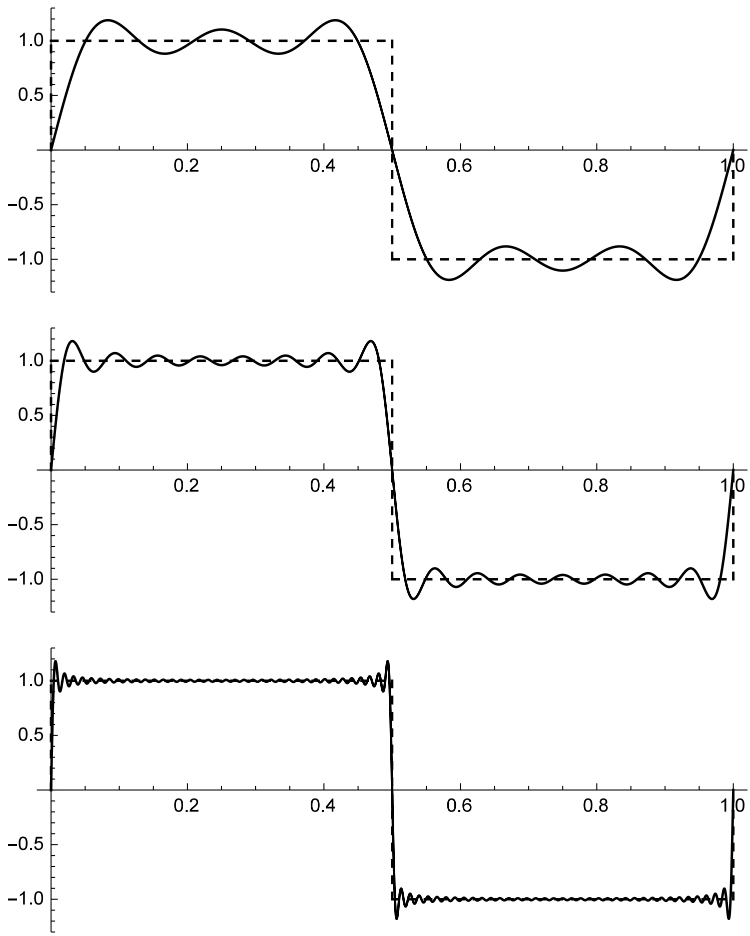


Fig. 9.10 Symmetric partial sums of the Fourier series of the square wave. Top: S_5. Middle: S_{15}. Bottom: S_{75}. The square wave itself is shown with dashed lines.

Corollary 9.3.14 implies that its Fourier series converges unconditionally in L^2-norm. In particular, the symmetric partial sums $f * d_N$ converge to f in L^2-norm. Figure 9.10 shows $f * d_N$ for $N = 5$, 15, and 75. It does appear from the diagram that $\|f - f * d_N\|_2 \to 0$, but we can also see *Gibbs' phenomenon* in this figure, which is that the partial sums do not converge *uniformly* to f. Instead, $f * d_N$ always overshoots f at its points of discontinuity by an amount (about 9%) that does not decrease with N. For a proof of Gibbs' phenomenon, see [DM72] or other texts on harmonic analysis. $\diamondsuit$

Although the series in equation (9.43) converges unconditionally for every $f \in L^2(\mathbb{T})$, it need not converge absolutely in L^2-norm. For example, if f is the 1-periodic function defined on $[0, 1)$ by $f(x) = x$, then a direct calculation shows that

$$\widehat{f}(0) = \frac{1}{2} \qquad \text{and} \qquad \widehat{f}(n) = -\frac{1}{2\pi i n} \quad \text{for } n \neq 0.$$

Since $f \in L^2(\mathbb{T})$, its Fourier series $f = \sum_{n\in\mathbb{Z}} \widehat{f}(n)\, e_n$ converges unconditionally in L^2-norm. However, this series does not converge absolutely, because

$$\sum_{n\in\mathbb{Z}} \|\widehat{f}(n)\, e_n\|_2 = \sum_{n\in\mathbb{Z}} |\widehat{f}(n)| = \infty.$$

9.3.6 Convergence of Fourier Series for $p \neq 2$

We have seen two cases where the partial sums (and not just the Cesàro means) of a Fourier series converge to f. First, by the Inversion Formula, if $f \in L^1(\mathbb{T})$ is such that $\widehat{f} \in \ell^1(\mathbb{Z})$ then the Fourier series of f converges uniformly to f. Second, Corollary 9.3.14 tells us that if $f \in L^2(\mathbb{T})$ then the Fourier series converges to f in L^2-norm. In both of these cases, the convergence is unconditional.

The general situation is far more delicate. For a generic function $f \in L^1(\mathbb{T})$, even if we restrict our attention to just the symmetric partial sums $f * d_N$, then there exist functions in $L^1(\mathbb{T})$ such that $f * d_N$ does not converge to f in L^1-norm. Likewise, there exist functions $f \in C(\mathbb{T})$ such that $f * d_N$ does not converge uniformly. We state this as the following theorem. We have not developed the tools needed to prove this result, but one proof can be found in [Heil11, Thm. 14.3].

Theorem 9.3.16. (a) *There exists an integrable function $f \in L^1(\mathbb{T})$ whose Fourier series does not converge in L^1-norm (i.e., $f*d_N$ does not converge in L^1-norm as $N \to \infty$).*

(b) *There exists a continuous function $f \in C(\mathbb{T})$ whose Fourier series does not converge uniformly (i.e., $f * d_N$ does not converge uniformly as $N \to \infty$).* $\diamondsuit$

As a consequence, *the trigonometric system is not a Schauder basis for either $L^1(\mathbb{T})$ or $C(\mathbb{T})$*. The fact that there are continuous functions whose Fourier series do not converge uniformly is surprising, but even more surprising is that there exist continuous functions $f \in C(\mathbb{T})$ such that $(f * d_N)(x)$ diverges for *almost every x* (for one proof, see [Kat04, Thm. 3.5]). On the other hand, if $f \in C(\mathbb{T})$ is a continuous function *that has bounded variation*, then the symmetric partial sums $f * d_N$ will converge uniformly to f (see [Kat04, Cor. 2.2]).

Turning to indices in the range $1 < p < \infty$, it can be shown—albeit with considerably more work than was needed to prove Corollary 9.3.14—that the symmetric partial sums $f * d_N$ do converge in L^p-norm when $1 < p < \infty$.

We state this as the following result; one proof can be found in [Heil11, Thm. 14.8].

Theorem 9.3.17. *If $1 < p < \infty$, then for every $f \in L^p(\mathbb{T})$ the symmetric partial sums*

$$f * d_N \;=\; \sum_{n=-N}^{N} \widehat{f}(n)\, e_n$$

converge to f in L^p-norm as $N \to \infty$. $\diamondsuit$

Consequently, the trigonometric system $\mathcal{E}$ is a Schauder basis for $L^p(\mathbb{T})$, but even in this statement there is a subtlety. When $p = 2$, the Fourier series

$$f \;=\; \sum_{n\in\mathbb{Z}} \widehat{f}(n)\, e_n \tag{9.45}$$

converges unconditionally. Hence, no matter how we choose to order $\mathbb{Z}$, the partial sums with respect to that ordering will converge. In contrast, when $1 < p < \infty$ and $p \neq 2$, we know only that *the symmetric partial sums converge in L^p-norm.* If $p \neq 2$, then there exist functions in $L^p(\mathbb{T})$ whose Fourier series converge *conditionally* in L^p-norm—only partial sums of certain orderings of $\mathbb{Z}$ will converge (such as the symmetric partial sums, which are partial sums corresponding to the ordering $\mathbb{Z} = \{0, -1, 1, -2, 2, -3, 3, \dots\}$). We refer to [Heil11, Chap. 14] for details.

There are even more layers of subtlety when we consider other types of convergence. One of the deepest results in Fourier analysis is the following theorem on pointwise almost everywhere convergence of Fourier series, proved by Lennart Carleson for $p = 2$ [Car66] and extended to $1 < p < \infty$ by Richard Hunt [Hunt68].

Theorem 9.3.18 (Carleson–Hunt Theorem). *If $1 < p < \infty$, then for each $f \in L^p(\mathbb{T})$, the symmetric partial sums $f * d_N$ converge to f pointwise a.e. That is,*

$$f(x) \;=\; \lim_{N\to\infty} \sum_{n=-N}^{N} \widehat{f}(n)\, e^{2\pi i n x} \quad a.e. \qquad \diamondsuit$$

Problems

9.3.19. Given a sequence of scalars $a = (a_k)_{k\in\mathbb{Z}}$, let $s_N = \sum_{k=-N}^{N} a_k$ denote the partial sums and $\sigma_N = (s_0 + \cdots + s_N)/(N+1)$ the Cesàro means of this sequence (compare Exercise 9.3.9).

(a) Show that if the partial sums s_N converge, then the Cesàro means σ_N converge to the same limit, i.e.,

$$\lim_{N\to\infty} \sum_{n=-N}^{N} \Bigl(1 - \frac{|n|}{N+1}\Bigr)\, a_n \;=\; \lim_{N\to\infty} s_N \;=\; \sum_{n=-\infty}^{\infty} a_n.$$

(b) Set $a_n = (-1)^n$ for $n \ge 0$ and $a_n = 0$ for $n < 0$. Show that the series $\sum_{n\in\mathbb{Z}} a_n$ is Cesàro summable even though the partial sums do not converge, and find the limit of the Cesàro means.

9.3.20. (a) Prove that every function in $C(\mathbb{T})$ is uniformly continuous, and use this to prove that translation is *strongly continuous* on $C(\mathbb{T})$, i.e.,

$$\lim_{a\to 0} \|T_a f - f\|_\infty \;=\; 0, \qquad \text{for all } f \in C(\mathbb{T}).$$

(b) Fix $1 \le p < \infty$. Prove that $C(\mathbb{T})$ is dense in $L^p(\mathbb{T})$. Use this to show that translation is *strongly continuous* on $L^p(\mathbb{T})$, i.e., $\lim_{a\to 0} \|T_a f - f\|_p \;=\; 0$ for all $1 \le p < \infty$ and all $f \in L^p(\mathbb{T})$.

9.3.21. Prove that $C^\infty(\mathbb{T})$ is dense in $L^p(\mathbb{T})$ for each index $1 \le p < \infty$, and $C^\infty(\mathbb{T})$ is dense in $C(\mathbb{T})$ with respect to the uniform norm.

9.3.22. Prove that there is no function in $L^1(\mathbb{T})$ that is an identity for convolution on $L^1(\mathbb{T})$.

9.3.23. Given $f \in L^1(\mathbb{T})$, prove that $f * e_n = \widehat{f}(n)\, e_n$ for every $n \in \mathbb{Z}$, where $e_n(x) = e^{2\pi i n x}$ (thus the complex exponentials with integer frequencies are eigenvectors for convolution).

9.3.24. (a) Show that if $f \in L^1(\mathbb{T})$ and $\widehat{f} \in \ell^2(\mathbb{Z})$, then $f \in L^2(\mathbb{T})$.

(b) Use part (a) to show that the Plancherel Equality given in equation (8.13) remains true if we assume only that f belongs to $L^1(\mathbb{T})$ rather than requiring it to belong to the smaller space $L^2(\mathbb{T})$. In other words, show that if $f \in L^1(\mathbb{T})$, then

$$\sum_{n\in\mathbb{Z}} |\widehat{f}(n)|^2 \;=\; \|f\|_2^2,$$

in the sense that one side is finite if and only if the other side is finite and in this case they are equal; otherwise, both sides are infinite.

9.3.25. Let $f(x) \;=\; x^2 - x + \frac{1}{6}$ for $x \in [0,1)$. Note that if we extend f 1-periodically to $\mathbb{R}$, then $f \in C(\mathbb{T})$.

(a) Compute $\widehat{f}$ and show that $\widehat{f} \in \ell^1(\mathbb{Z})$. Use this to prove that

$$\sum_{n=1}^{\infty} \frac{\cos 2\pi n x}{2\pi^2 n^2} \;=\; x^2 - x + \frac{1}{6}, \qquad \text{for } x \in [0,1], \tag{9.46}$$

where the series converges uniformly on $[0,1]$.

(b) Prove Euler's Formula (see Problem 8.4.6).

(c) Find the value of $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^4}$.

9.3.26. Assume that α is a real number that is not an integer, and let

$$f(x) \;=\; \frac{\pi\, e^{\pi i \alpha}}{\sin \pi\alpha}\, e^{-2\pi i \alpha x}, \qquad \text{for } x \in [0,1].$$

Show that $\widehat{f}(n) = 1/(n+\alpha)$ for each $n \in \mathbb{Z}$, and use the Plancherel Equality to prove that

$$\sum_{n=-\infty}^{\infty} \frac{1}{(n+\alpha)^2} \;=\; \frac{\pi^2}{\sin^2 \pi\alpha}.$$

9.3.27. (a) Show that if $f \in L^1(\mathbb{T})$ and $g \in C(\mathbb{T})$ then $f * g \in C(\mathbb{T})$.

(b) Prove that convolution commutes with differentiation in the following sense: If $f \in L^1(\mathbb{T})$ and $g \in C^1(\mathbb{T})$, then $f * g \in C^1(\mathbb{T})$, and $(f*g)' = f * g'$.

9.3.28. Suppose that $f \in \mathrm{AC}(\mathbb{T})$, i.e., f is 1-periodic and is absolutely continuous on $[0,1]$.

(a) Prove that $\widehat{f'}(n) = 2\pi i n \widehat{f}(n)$ for $n \in \mathbb{Z}$, and use this to show that $n\widehat{f}(n) \to 0$ as $|n| \to \infty$.

(b) Show that if $\int_0^1 f(x)\,dx = 0$, then we have *Wirtinger's Inequality*:

$$\int_0^1 |f(x)|^2\,dx \;\le\; \frac{1}{4\pi^2} \int_0^1 |f'(x)|^2.$$

Further, equality holds if and only if $f(x) = ae^{2\pi i x} + be^{-2\pi i x}$ for some scalars $a, b \in \mathbb{C}$ (equivalently, $f(x) = c\cos(2\pi x) + id\sin(2\pi x)$, where $c = a+b$ and $d = a-b$).

9.3.29. Fix $0 < \alpha < 1$. Prove that if $f \in C(\mathbb{T})$ is Hölder continuous with exponent α, then

$$|\widehat{f}(n)| \;\le\; \frac{1}{2}\left(\frac{1}{2|n|}\right)^{\alpha}, \qquad \text{for all } n \neq 0.$$

9.3.30. Let $\{w_N\}_{N\in\mathbb{N}}$ be the Fejér kernel, and prove that the series $f = \sum_{k=1}^{\infty} 2^{-k}\, w_{2^k}$ converges in $L^1(\mathbb{T})$, but $\widehat{f} \notin \ell^1(\mathbb{Z})$.

9.3.31. Show that if a sequence $c = (c_n)_{n\in\mathbb{Z}}$ satisfies $\sum_{n\in\mathbb{Z}} |nc_n| < \infty$, then the function $\widehat{c}(\xi) = \sum_{n\in\mathbb{Z}} c_n e^{-2\pi i n\xi}$ is differentiable, and at every point $\xi \in \mathbb{R}$ we have

$$\widehat{c}\,'(\xi) \;=\; -2\pi i \sum_{n\in\mathbb{Z}} nc_n e^{-2\pi i n \xi} \;=\; \widehat{d}(\xi),$$

where $d = (-2\pi i n c_n)_{n\in\mathbb{Z}} \in \ell^1(\mathbb{Z})$.

9.3.32. Prove that $A(\mathbb{T}) = L^2(\mathbb{T}) * L^2(\mathbb{T})$. That is, show that $f \in A(\mathbb{T})$ if and only if $f = g * h$ for some $g, h \in L^2(\mathbb{T})$.

9.3.33. Given $f \in L^1(\mathbb{T})$ and $g \in L^\infty(\mathbb{T})$, prove *Fejér's Lemma*:

$$\lim_{m\to\infty} \int_0^1 f(x)\, g(mx)\, dx \;=\; \widehat{f}(0)\, \widehat{g}(0) \;=\; \biggl(\int_0^1 f(x)\, dx\biggr) \biggl(\int_0^1 g(x)\, dx\biggr).$$

9.3.34. Assume that $E \subseteq [0,1]$ is measurable and $|E| > 0$. Given $\delta \ge 0$, prove that there are at most finitely many positive integers n such that $\sin 2\pi n x \ge \delta$ for all $x \in E$.

9.3.35. Let $\{d_N\}_{N\in\mathbb{N}}$ be the Dirichlet kernel, where d_N is defined by equation (9.39). Prove that $\int_0^1 d_N = 1$ for each $N \in \mathbb{N}$, and for $N > 1$ we have

$$\frac{4}{\pi^2} \sum_{k=1}^{N} \frac{1}{k} \;\le\; \|d_N\|_1 \;\le\; 3 \,+\, \frac{4}{\pi^2} \sum_{k=1}^{N} \frac{1}{k}.$$

Conclude that the Dirichlet kernel is not an approximate identity for $\mathbb{T}$.

9.4 The Fourier Transform on $L^2(\mathbb{R})$

We defined the Fourier transform of functions in $L^1(\mathbb{R})$ in Section 9.2. Now we will consider functions that belong to $L^2(\mathbb{R})$.

For motivation, recall the analogous situation for Fourier series. Theorem 8.4.2 told us that the mapping $U\colon L^2(\mathbb{T}) \to \ell^2(\mathbb{Z})$ that sends a 1-periodic function $f \in L^2(\mathbb{T})$ to its sequence of Fourier coefficients $U(f) = \bigl(\widehat{f}(n)\bigr)_{n\in\mathbb{Z}}$ is a *unitary operator* in the sense of Definition 8.3.16. That is, U is linear, surjective, and isometric (i.e., it preserves the norms of vectors). The isometric nature of U is a direct consequence of the Plancherel Equality:

$$\|U(f)\|_2^2 \;=\; \sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2 \;=\; \int_0^1 |f(x)|^2\, dx \;=\; \|f\|_2^2.$$

Is there an analogue to U for the Fourier transform? That is, does the Fourier transform isometrically map functions in $L^2(\mathbb{R})$ to another Hilbert space? We will see that the answer is yes, but first we have to address a more basic issue, one that does not arise for Fourier series because $L^2[0,1] \subseteq L^1[0,1]$. In contrast, $L^2(\mathbb{R})$ is not contained in $L^1(\mathbb{R})$, so how do we even define the Fourier transform of a function in $L^2(\mathbb{R})$? Definition 9.2.1 told us that if f is *integrable* on $\mathbb{R}$ then its Fourier transform is

$$\widehat{f}(\xi) \;=\; \int_{-\infty}^{\infty} f(x)\, e^{-2\pi i \xi x}\, dx, \qquad \text{for } \xi \in \mathbb{R}. \tag{9.47}$$

However, there are functions in $L^2(\mathbb{R})$ that are not integrable, and for such functions the integral in equation (9.47) will not exist. On the other hand, $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ is dense in $L^2(\mathbb{R})$, and the Fourier transform is well-defined for all functions in this subspace, so perhaps there is a way to extend the definition of the Fourier transform from this dense subspace to all of $L^2(\mathbb{R})$.

To investigate this, we first consider functions that are both integrable and square-integrable. In fact, we will restrict our attention to functions in $C_c^2(\mathbb{R})$. This space will be convenient for our purposes because it is dense in both $L^1(\mathbb{R})$ and in $L^2(\mathbb{R})$ and, as we show next, if $f \in C_c^2(\mathbb{R})$ then both f and $\widehat{f}$ are continuous and have good decay.

Lemma 9.4.1. *If $f \in C_c^2(\mathbb{R})$, then the following statements hold.*

(a) *There is a constant $C \geq 0$ such that $|\widehat{f}(\xi)| \leq C/|\xi|^2$ for all $\xi \neq 0$.*

(b) *f and $\widehat{f}$ both belong to $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.*

(c) *f and $\widehat{f}$ are continuous.*

Proof. Since $f \in C_c^2(\mathbb{R})$, we know that f is continuous, integrable, and square-integrable. Its Fourier transform $\widehat{f}$ exists and is defined by equation (9.47). The Riemann–Lebesgue Lemma implies that $\widehat{f} \in C_0(\mathbb{R})$, so $\widehat{f}$ is continuous and bounded. Away from the origin, equation (9.29) tells us that

$$|\widehat{f}(\xi)| \;\leq\; \frac{\|f''\|_1}{4\pi^2|\xi|^2}, \qquad \text{for } \xi \neq 0.$$

This is sufficient decay to ensure that $\widehat{f} \in L^1(\mathbb{R})$ and $\widehat{f} \in L^2(\mathbb{R})$. $\square$

Now we prove that the mapping that sends f to $\widehat{f}$ is isometric with respect to the L^2-norm *on the domain* $C_c^2(\mathbb{R})$.

Lemma 9.4.2. $\|\widehat{f}\|_2 = \|f\|_2$ *for all* $f \in C_c^2(\mathbb{R})$.

Proof. Fix $f \in C_c^2(\mathbb{R})$. Applying Lemma 9.4.1, we have that f and $\widehat{f}$ are each continuous, integrable, and square-integrable. We define the *involution* of f to be

$$\widetilde{f}(x) \;=\; \overline{f(-x)}.$$

This is an integrable function, and by making a change of variables we see that its Fourier transform is

$$\big(\widetilde{f}\big)^{\wedge}(\xi) \;=\; \int_{-\infty}^{\infty} \overline{f(-x)}\, e^{-2\pi i \xi x}\, dx \;=\; \overline{\widehat{f}(\xi)}.$$

We will also need the *autocorrelation* of f, which is

$$g(x) \;=\; \big(f * \widetilde{f}\big)(x) \;=\; \int_{-\infty}^{\infty} f(y)\, \overline{f(y-x)}\, dy. \tag{9.48}$$

Since $L^1(\mathbb{R})$ is closed under convolution, we have $g \in L^1(\mathbb{R})$. Additionally, g is continuous because both f and $\widetilde{f}$ are continuous and integrable. Since the Fourier transform converts convolution to multiplication, we compute that

$$\widehat{g}(\xi) \;=\; \bigl(f * \widetilde{f}\,\bigr)^{\wedge}(\xi) \;=\; \widehat{f}(\xi)\,\overline{\widehat{f}(\xi)} \;=\; |\widehat{f}(\xi)|^2 \;\in\; L^1(\mathbb{R}).$$

Thus g and $\widehat{g}$ are both integrable, so the Inversion Formula (Theorem 9.2.9) implies that $g(x) = \bigl(\widehat{g}\,\bigr)^{\vee}(x)$ for every x. Evaluating the continuous function g at $x = 0$ yields

$$g(0) \;=\; \bigl(\widehat{g}\,\bigr)^{\vee}(0) \;=\; \int_{-\infty}^{\infty} \widehat{g}(\xi)\,d\xi \;=\; \int_{-\infty}^{\infty} |\widehat{f}(\xi)|^2\,d\xi \;=\; \|\widehat{f}\,\|_2^2.$$

On the other hand, evaluating equation (9.48) at $x = 0$ gives

$$g(0) \;=\; \bigl(f * \widetilde{f}\,\bigr)(0) \;=\; \int_{-\infty}^{\infty} f(y)\,\overline{f(y)}\,dy \;=\; \|f\|_2^2.$$

Therefore $\|\widehat{f}\,\|_2 = \|f\|_2$. $\square$

Lemma 9.4.2 implies that the operator $\mathcal{F}\colon C_c^2(\mathbb{R}) \to L^2(\mathbb{R})$ defined by $\mathcal{F}(f) = \widehat{f}$ is *linear and isometric* (with respect to the L^2-norm). Now, $C_c^2(\mathbb{R})$ is not complete with respect to the L^2-norm, but it is dense in $L^2(\mathbb{R})$. Thus $\mathcal{F}$ is a "very nice" linear map whose domain is a dense subspace of the Hilbert space $L^2(\mathbb{R})$. We will show that we can extend $\mathcal{F}$ so that its domain is all of $L^2(\mathbb{R})$, and we can do so in such a way that the mapping $\mathcal{F}\colon L^2(\mathbb{R}) \to L^2(\mathbb{R})$ *is linear, bijective, and isometric.*

To do this, fix any function $f \in L^2(\mathbb{R})$. Since $C_c^2(\mathbb{R})$ is dense in $L^2(\mathbb{R})$, there exists a sequence $\{f_n\}_{n\in\mathbb{N}}$ in $C_c^2(\mathbb{R})$ such that $f_n \to f$ in L^2-norm. Consequently, $\{f_n\}_{n\in\mathbb{N}}$ is Cauchy in $L^2(\mathbb{R})$. We have $f_m - f_n \in C_c^2(\mathbb{R})$ for every m and n, so we can apply Lemma 9.4.2 to obtain

$$\bigl\|\widehat{f_m} - \widehat{f_n}\bigr\|_2 \;=\; \bigl\|(f_m - f_n)^{\wedge}\bigr\|_2 \;=\; \|f_m - f_n\|_2.$$

This implies that $\bigl\{\widehat{f_n}\bigr\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^2(\mathbb{R})$. Since $L^2(\mathbb{R})$ is complete, this sequence must converge. Therefore, there exists some function $\widehat{f} \in L^2(\mathbb{R})$ such that $\widehat{f_n} \to \widehat{f}$ in L^2-norm.

We would like to define $\widehat{f}$ to be the Fourier transform of f, but there is a complication. There could be many sequences in $C_c^2(\mathbb{R})$ that converge to f, and so we could obtain a different function $\widehat{f}$ if we chose a different sequence $\{f_n\}_{n\in\mathbb{N}}$. Therefore, we must show that $\widehat{f}$ is *well-defined.* That is, we must show that no matter which functions $f_n \in C_c^2(\mathbb{R})$ that we choose that satisfy $\|f - f_n\|_2 \to 0$, we obtain the same result for $\widehat{f}$.

To see this, suppose that $\{h_n\}_{n\in\mathbb{N}}$ is another sequence of functions in $C_c^2(\mathbb{R})$ such that $\|f - h_n\|_2 \to 0$. Then $\{h_n\}_{n\in\mathbb{N}}$ is Cauchy in L^2-norm, and since $\|\widehat{h}_m - \widehat{h}_n\|_2 = \|h_m - h_n\|_2$, we see that $\{\widehat{h}_n\}_{n\in\mathbb{N}}$ is Cauchy in $L^2(\mathbb{R})$ and therefore converges to some function $\widehat{h} \in L^2(\mathbb{R})$. Applying the continuity of the norm, it follows that

$$\|\widehat{f} - \widehat{h}\|_2 = \lim_{n\to\infty} \|\widehat{f}_n - \widehat{h}_n\|_2 = \lim_{n\to\infty} \|f_n - h_n\|_2 = \|f - f\|_2 = 0.$$

Thus $\widehat{h} = \widehat{f}$ a.e., so they are the same element of $L^2(\mathbb{R})$. We therefore can make the following definition.

Definition 9.4.3 (The Fourier Transform on $L^2(\mathbb{R})$). Given $f \in L^2(\mathbb{R})$, let $\{f_n\}_{n\in\mathbb{N}}$ be any sequence in $C_c^2(\mathbb{R})$ such that $f_n \to f$ in L^2-norm. Then the *Fourier transform* of f is the function $\widehat{f} \in L^2(\mathbb{R})$ such that $\widehat{f}_n \to \widehat{f}$ in L^2-norm. $\diamondsuit$

This defines the Fourier transform of every square-integrable function. However, we now have two Fourier transforms, one defined on $L^1(\mathbb{R})$ and one on $L^2(\mathbb{R})$. We show next that these two definitions coincide for any function that belongs to both spaces. Note that if $f \in L^1(\mathbb{R})$, then $\widehat{f}$ is a continuous function that is defined by the integral that appears in equation (9.47). In contrast, if $f \in L^2(\mathbb{R})$, then $\widehat{f}$ is only implicitly defined as the L^2-norm limit of $\widehat{f}_n$ where $f_n \in C_c^2(\mathbb{R})$ and $f_n \to f$ in L^2-norm. Hence, if $f \in L^2(\mathbb{R})$, then its Fourier transform $\widehat{f}$ is an element of $L^2(\mathbb{R})$, and therefore is only defined up to sets of measure zero.

Lemma 9.4.4. *If $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, then the function $\widehat{f}$ given by equation* (9.47) *is equal almost everywhere to the function $\widehat{f}$ given by Definition 9.4.3.*

Proof. Fix a function $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Let $\widehat{f}$ be the function defined by equation (9.47), and let F be the L^2-Fourier transform of f as given by Definition 9.4.3.

The proof of Theorem 9.1.12 shows how to explicitly construct functions $f_N \in C_c^\infty(\mathbb{R})$ that converge to f in L^1-norm. Specifically, if f_N is defined as in equation (9.5), then $\|f - f_N\|_1 \to 0$. Replacing the L^1-norm by the L^2-norm, exactly the same proof shows that we *also* have $\|f - f_N\|_2 \to 0$ (compare Problem 9.1.22).

Now, since $\|f - f_N\|_1 \to 0$, Lemma 9.2.3 implies that $\widehat{f_N} \to \widehat{f}$ uniformly, and hence pointwise. On the other hand, since $\|f - f_N\|_2 \to 0$, we have by definition that $\widehat{f_N} \to F$ in L^2-norm. Hence there is a subsequence of the $\widehat{f_N}$ that converges to F pointwise a.e. But this subsequence also converges to $\widehat{f}$ pointwise, so we conclude that $F = \widehat{f}$ a.e. $\square$

In summary, we have defined the Fourier transform of every function in $L^1(\mathbb{R}) \cup L^2(\mathbb{R})$. For functions in $L^1(\mathbb{R})$ the Fourier transform is given by equation (9.47), while for functions in $L^2(\mathbb{R})$ it is given by Definition 9.4.3.

For functions that belong to both $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$ these two definitions coincide in the usual almost everywhere sense.

We show next that the Fourier transform is isometric on all of $L^2(\mathbb{R})$.

Lemma 9.4.5. (a) $\|\widehat{f}\|_2 = \|f\|_2$ *for every* $f \in L^2(\mathbb{R})$.

(b) *If* $\{f_n\}_{n\in\mathbb{N}}$ *is any sequence in* $L^2(\mathbb{R})$ *such that* $f_n \to f$ *in* L^2*-norm, then* $\widehat{f_n} \to \widehat{f}$ *in* L^2*-norm.*

Proof. (a) Fix $f \in L^2(\mathbb{R})$, and choose any functions $f_n \in C_c^2(\mathbb{R})$ such that $f_n \to f$ in L^2-norm. Then $\widehat{f_n} \to \widehat{f}$ in L^2-norm by definition. Since $f_n \in C_c^2(\mathbb{R})$ we have $\|\widehat{f_n}\|_2 = \|f_n\|_2$ by Lemma 9.4.2. Therefore, by the continuity of the norm, we obtain

$$\|\widehat{f}\|_2 \;=\; \lim_{n\to\infty} \|\widehat{f_n}\|_2 \;=\; \lim_{n\to\infty} \|f_n\|_2 \;=\; \|f\|_2.$$

(b) Assume that $f_n, f \in L^2(\mathbb{R})$ are such that $\|f - f_n\|_2 \to 0$. Applying part (a), it follows that

$$\|\widehat{f} - \widehat{f_n}\|_2 \;=\; \|(f-f_n)^\wedge\|_2 \;=\; \|f - f_n\|_2 \;\to\; 0. \qquad \square$$

Now we show that the Fourier transform is a unitary operator on $L^2(\mathbb{R})$.

Theorem 9.4.6. *The Fourier transform* $\mathcal{F}\colon L^2(\mathbb{R}) \to L^2(\mathbb{R})$ *is a unitary operator, i.e.,* $\mathcal{F}$ *is linear, isometric, and surjective. In particular, we have the* Plancherel Equality,

$$\big\|\widehat{f}\big\|_2 \;=\; \|f\|_2, \qquad \text{for all } f \in L^2(\mathbb{R}), \tag{9.49}$$

and the Parseval Equality,

$$\big\langle \widehat{f}, \widehat{g} \big\rangle \;=\; \langle f, g\rangle, \qquad \text{for all } f, g \in L^2(\mathbb{R}). \tag{9.50}$$

Proof. If $f \in L^2(\mathbb{R})$, then $\widehat{f} \in L^2(\mathbb{R})$ by definition, so $\mathcal{F}$ maps $L^2(\mathbb{R})$ into itself. Lemma 9.4.5 shows that equation (9.49) holds, so $\mathcal{F}$ is isometric. The reader should verify that $\mathcal{F}$ is linear. Consequently, Lemma 8.3.15 implies that $\mathcal{F}$ preserves inner products, i.e., equation (9.50) holds. Hence, it only remains to show that $\mathcal{F}$ is surjective.

First we will prove that range$(\mathcal{F})$ is dense in $L^2(\mathbb{R})$. To do this, choose any function $f \in C_c^2(\mathbb{R})$. By Lemma 9.4.1, both f and $\widehat{f}$ belong to $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. The inverse Fourier transform of f is defined by $\check{f}(\xi) = \widehat{f}(-\xi)$, so we also have $\check{f} \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Since f and $\widehat{f}$ are both integrable, the Inversion Formula (Theorem 9.2.9) implies that

$$f \;=\; \big(\check{f}\big)^\wedge \;=\; \mathcal{F}(\check{f}).$$

Thus, f and $\check{f}$ both belong to $L^2(\mathbb{R})$ and $f = \mathcal{F}(\check{f})$, so we conclude that $f \in \operatorname{range}(\mathcal{F})$. This shows that $C_c^2(\mathbb{R}) \subseteq \operatorname{range}(\mathcal{F})$. But $C_c^2(\mathbb{R})$ is dense in $L^2(\mathbb{R})$, so $\operatorname{range}(\mathcal{F})$ must be dense in $L^2(\mathbb{R})$.

Since the range is dense, its closure is all of $L^2(\mathbb{R})$. However, since $\mathcal{F}$ is isometric, Problem 9.4.9 implies that $\operatorname{range}(\mathcal{F})$ is a closed subset of $L^2(\mathbb{R})$. Therefore $\operatorname{range}(\mathcal{F})$ equals its closure, which is $L^2(\mathbb{R})$, so $\mathcal{F}$ is surjective. $\square$

Since the Fourier transform $\mathcal{F}\colon L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is unitary, it has an inverse $\mathcal{F}^{-1}\colon L^2(\mathbb{R}) \to L^2(\mathbb{R})$ that is also unitary. We call $\mathcal{F}^{-1}$ the *inverse Fourier transform*, and if $f \in L^2(\mathbb{R})$ then we say that $\check{f} = \mathcal{F}^{-1}(f)$ is the *inverse Fourier transform of* f. As functions in $L^2(\mathbb{R})$,

$$f = \big(\widehat{f}\,\big)^{\vee} = \big(\check{f}\,\big)^{\wedge},$$

i.e., these functions are equal almost everywhere. The Plancherel and Parseval Equalities hold for the inverse Fourier transform. That is, for all f and g in $L^2(\mathbb{R})$ we have

$$\big\|\check{f}\,\big\|_2 = \|f\|_2 \qquad \text{and} \qquad \big\langle \check{f}, \check{g} \big\rangle = \langle f, g\rangle.$$

Example 9.4.7. As an application, we will compute the Fourier transform of the *sinc function*

$$s(x) = \operatorname{sinc}(x) = \frac{\sin \pi x}{\pi x},$$

which is square-integrable but not integrable. Therefore its Fourier transform is not given by equation (9.47).

First consider the box function $\chi = \chi_{[-\frac{1}{2},\frac{1}{2}]}$. Since χ is integrable, we can use equation (9.47) to compute its Fourier transform:

$$\widehat{\chi}(\xi) = \int_{-\infty}^{\infty} \chi(x)\, e^{-2\pi i \xi x}\, dx = \int_{-1/2}^{1/2} e^{-2\pi i \xi x}\, dx = \frac{\sin \pi\xi}{\pi\xi} = s(\xi).$$

Because χ is even, a similar calculation shows that its inverse Fourier transform is $\check{\chi} = s$. Since χ belongs to $L^2(\mathbb{R})$, it satisfies the Inversion Formula for the L^2 Fourier transform. Therefore,

$$\chi = \big(\check{\chi}\big)^{\wedge} = \widehat{s}.$$

This is an equality of functions in $L^2(\mathbb{R})$, i.e., it holds a.e. Thus, even though we cannot use equation (9.47) to compute the Fourier transform of s, we have demonstrated that $\widehat{s} = \chi$. A similar computation shows that $\check{s} = \chi$. $\diamondsuit$

Many formulas that hold for the Fourier transform of functions in $L^1(\mathbb{R})$ have analogues that hold for functions in $L^2(\mathbb{R})$. For example, if $f \in L^1(\mathbb{R})$ then we know that $\check{f}(\xi) = \widehat{f}(-\xi)$ for every ξ. We show next that this implies that a similar formula holds for functions in $L^2(\mathbb{R})$.

Lemma 9.4.8. *If $f \in L^2(\mathbb{R})$ then $\check{f}(\xi) = \widehat{f}(-\xi)$ for almost every $\xi \in \mathbb{R}$.*

Proof. Fix $f \in L^2(\mathbb{R})$. There exist functions $f_n \in C_c^2(\mathbb{R})$ that converge to f in L^2-norm. By the Plancherel Equality, it follows that $\widehat{f_n} \to \widehat{f}$ in L^2-norm, and consequently there exists a subsequence $\{g_n\}_{n\in\mathbb{N}}$ of $\{f_n\}_{n\in\mathbb{N}}$ such that $\widehat{g_n} \to \widehat{f}$ pointwise a.e.

Now, since $\{g_n\}_{n\in\mathbb{N}}$ is a subsequence of $\{f_n\}_{n\in\mathbb{N}}$, we have that $g_n \to f$ in L^2-norm. Therefore, the Plancherel Equality for the inverse Fourier transform implies that $\check{g}_n \to \check{f}$ in L^2-norm. Consequently, there exists a subsequence $\{h_n\}_{n\in\mathbb{N}}$ of $\{g_n\}_{n\in\mathbb{N}}$ such that $\check{h}_n \to \check{f}$ a.e.

Since $\{h_n\}_{n\in\mathbb{N}}$ is a subsequence of $\{g_n\}_{n\in\mathbb{N}}$, we conclude that we have both $\widehat{h}_n \to \widehat{f}$ a.e. and $\check{h}_n \to \check{f}$ a.e. But h_n belongs to $L^1(\mathbb{R})$, so $\check{h}_n(\xi) = \widehat{h}_n(-\xi)$ for every ξ. Therefore, for a.e. ξ we have

$$\check{f}(\xi) \;=\; \lim_{n\to\infty} \check{h}_n(\xi) \;=\; \lim_{n\to\infty} \widehat{h}_n(-\xi) \;=\; \widehat{f}(-\xi). \qquad \square$$

It is possible to extend the Fourier transform beyond $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$. The process of *interpolation* allows us to define the Fourier transform of any function in $L^p(\mathbb{R})$ for indices in the range $1 \le p \le 2$. We can even go much further and define the Fourier transform of every *tempered distribution*. For details we refer to texts such as [DM72], [Ben97], [Kat04], or [Heil11].

Problems

9.4.9. Let X and Y be Banach spaces, and assume that $A\colon X \to Y$ is both linear and isometric (that is, $\|Ax\| = \|x\|$ for all $x \in X$). Prove that $\mathrm{range}(A) = \{Ax : x \in X\}$ is a closed subspace of Y.

9.4.10. Suppose that $f \in L^2(\mathbb{R})$ is such that $\widehat{f} \in L^1(\mathbb{R})$. Show that $f \in C_0(\mathbb{R})$ and $\|f\|_\infty \le \|\widehat{f}\|_1$. Exhibit a function $f \in L^2(\mathbb{R})\backslash L^1(\mathbb{R})$ such that $\widehat{f} \in L^1(\mathbb{R})$.

9.4.11. (a) Show that if $f \in L^1(\mathbb{R})$ and $\widehat{f} \in L^2(\mathbb{R})$, then $f \in L^2(\mathbb{R})$.

(b) Use part (a) to show that the Plancherel Equality holds for functions in $L^1(\mathbb{R})$, i.e., if $f \in L^1(\mathbb{R})$, then

$$\int_{-\infty}^{\infty} |f(x)|^2\,dx \;=\; \int_{-\infty}^{\infty} |\widehat{f}(\xi)|^2\,d\xi,$$

in the sense that one side is finite if and only if the other side is finite, and in this case they are equal, otherwise both are infinite.

(c) Exhibit a function $f \in L^1(\mathbb{R}) \setminus L^2(\mathbb{R})$ such that $\widehat{f} \notin L^1(\mathbb{R})$.

(d) Show that the Riemann–Lebesgue Lemma does not hold for all functions in $L^2(\mathbb{R})$. Specifically, show that there is a square-integrable f such that $\widehat{f}$ is continuous, yet $\widehat{f}(\xi)$ does not converge to zero as $|\xi| \to \infty$.

9.4.12. Prove the following facts about convolution of functions in $L^2(\mathbb{R})$.

(a) If $f, g \in L^2(\mathbb{R})$, then $(fg)^\wedge$ is continuous, $(fg)^\wedge = \widehat{f} * \widehat{g}$, and $f * g = (\widehat{f}\,\widehat{g})^\vee$.

(b) If $f, g \in L^2(\mathbb{R})$ and $f * g \in L^2(\mathbb{R})$, then $(f * g)^\wedge = \widehat{f}\,\widehat{g}$. In particular, this is the case if $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and $g \in L^2(\mathbb{R})$.

(c) $L^2(\mathbb{R}) * L^2(\mathbb{R}) = A(\mathbb{R})$ (defined in equation (9.19)). That is, $f \in A(\mathbb{R})$ if and only if $f = g * h$ for some $g, h \in L^2(\mathbb{R})$.

9.4.13. Show that $\|f * g\|_2^2 \le \|f * f\|_2 \, \|g * g\|_2$ for all $f, g \in L^2(\mathbb{R})$, but the inequality $\|f * g\|_1^2 \le \|f * f\|_1 \, \|g * g\|_1$ cannot hold for all $f, g \in L^1(\mathbb{R})$.

9.4.14. Exhibit a nontrivial function $f \in L^2(\mathbb{R})$ that satisfies $f = f * f$ a.e. Contrast this with Problem 9.2.22, which shows there are no such functions in $L^1(\mathbb{R})$.

9.4.15. Given $T > 0$, we define the *Dirichlet function* $d_{2\pi T}$ to be

$$d_{2\pi T}(\xi) \;=\; \frac{\sin 2\pi T\xi}{\pi\xi}.$$

Although $d_{2\pi T}$ is not integrable, it does belong to $L^2(\mathbb{R})$ and therefore has a Fourier transform in the sense of Definition 9.4.3.

(a) Prove that $\widehat{d_{2\pi T}} = \chi_{[-T,T]}$.

(b) Show that if $f \in L^2(\mathbb{R})$, then $f * d_{2\pi T} \in L^2(\mathbb{R})$ and

$$(f * d_{2\pi T})^\wedge \;=\; \widehat{f}\,\chi_{[-T,T]} \;\to\; \widehat{f} \quad \text{as } T \to \infty,$$

where the convergence is in L^2-norm.

(c) Show that if $f \in L^2(\mathbb{R})$, then $f * d_{2\pi T} \to f$ in L^2-norm as $T \to \infty$. Note that the *Dirichlet kernel* $\{d_{2\pi N}\}_{N \in \mathbb{N}}$ does not form an approximate identity.

9.4.16. (a) Show that there exist nontrivial functions $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ such that $f * d_{2\pi T} = 0$.

(b) Use the Plancherel Equality to show that

$$\int_0^\infty \frac{\sin^2 t}{t^2}\, dt \;=\; \frac{\pi}{2}.$$

As a consequence, $\int_0^\infty \frac{\sin^2 t}{t^2}\, dt = \int_0^\infty \frac{\sin t}{t}\, dt$, where the latter integral is an improper Riemann integral (see Problem 4.6.19).

(c) Generalizing part (b), use the Parseval Equality to show that if $j \in \mathbb{N}$ and r is any real number with $r \geq j$, then

$$\int_{-\infty}^{\infty} \Big(\frac{\sin t}{t}\Big)^j \, \frac{\sin rt}{t} \, dt \;=\; \pi.$$

9.4.17. Given $a, b > 0$, compute $\displaystyle\int_{-\infty}^{\infty} \frac{dx}{(x^2+a^2)\,(x^2+b^2)}$.

9.4.18. Fix $g \in L^2(\mathbb{R})$. Prove that $\{T_a g\}_{a \in \mathbb{R}}$ is complete in $L^2(\mathbb{R})$ if and only if $\widehat{g}(\xi) \neq 0$ a.e.

9.4.19. Given $g \in L^2(\mathbb{R})$, show that

$$\{T_k g\}_{k \in \mathbb{Z}} \text{ is an orthonormal sequence} \quad \Longleftrightarrow \quad \sum_{k \in \mathbb{Z}} |\widehat{g}(\xi - k)|^2 \;=\; 1 \text{ a.e.}$$

9.4.20. (a) Fix $a > 1$, $b > 0$, $c > 0$, and let $\psi \in L^2(\mathbb{R})$ be such that $\mathrm{supp}(\widehat{\psi}\,) \subseteq [c, c+b^{-1}]$ and

$$\sum_{n \in \mathbb{Z}} |\widehat{\psi}(a^n \xi)|^2 \;=\; b \qquad \text{for a.e. } \xi \geq 0.$$

For all $k, n \in \mathbb{Z}$, define

$$\psi_{kn}(x) \;=\; a^{n/2} \psi(a^n x - bk).$$

Prove that the *wavelet system* $\mathcal{W}(\psi) = \{\psi_{kn}\}_{k,n \in \mathbb{Z}}$ satisfies

$$\sum_{k,n \in \mathbb{Z}} |\langle f, \psi_{kn} \rangle|^2 \;=\; \frac{1}{b}\, \|f\|_2^2, \qquad \text{for all } f \in H^2_+(\mathbb{R}), \tag{9.51}$$

where $H^2_+(\mathbb{R}) \;=\; \big\{f \in L^2(\mathbb{R}) : \mathrm{supp}(\widehat{f}\,) \subseteq [0, \infty)\big\}$.

Remark: Using the language of frame theory, equation (9.51) says that $\mathcal{W}(\psi)$ is a *tight frame* for $H^2_+(\mathbb{R})$.

(b) Exhibit functions $\psi^1, \psi^2 \in L^2(\mathbb{R})$ such that $\widehat{\psi^1}, \widehat{\psi^2}$ are continuous, and $\mathcal{W}(\psi_1) \cup \mathcal{W}(\psi_2)$ is a *Parseval frame* for $L^2(\mathbb{R})$, i.e.,

$$\sum_{k,n \in \mathbb{Z}} |\langle f, \psi^1_{kn} \rangle|^2 \;+\; \sum_{k,n \in \mathbb{Z}} |\langle f, \psi^2_{kn} \rangle|^2 \;=\; \|f\|_2^2, \qquad \text{for all } f \in L^2(\mathbb{R}).$$

Hints for Selected Exercises and Problems

1.4.5 Theorem 1.2.8.

2.1.24 If $x \in (1/3, 2/3)$ then $c_1 = 1$.

2.1.37 We have *not yet shown* that Lebesgue measure is invariant under rotations. If U is an orthogonal matrix and Q is a cube in $\mathbb{R}^d$ with sides parallel to the coordinate axes, then $U(Q)$ is a cube but its sides need not be parallel to the coordinate axes, so we do not yet know whether $|Q|_e$ and $|U(Q)|_e$ are equal. On the other hand, every cube is contained in a ball, an orthogonal matrix maps balls to balls, and every ball is contained in a cube.

2.1.40 Consider $\cup_{r \in \mathbb{Q}} (r - Z)$.

2.2.36 No.

2.2.43 (b) Consider $E = (-\infty, 0) \cup N$.

2.3.21 Let $K \subseteq E$ be compact with $|K| > 0$.

3.1.18 (b) Consider $f^{-1}(U)$ where $U = \{x + iy : x \in (a,b),\, y \in \mathbb{R}\}$.

4.2.16 $\Leftarrow$. $\int_E (\psi - \phi)$. Do not try to integrate f if it has not been shown to be measurable.

4.2.17 (a) Consider $\{\varepsilon n \le f < \varepsilon(n+1)\} \times [\varepsilon n, \varepsilon(n+1))$.

4.4.23 $|f| + |f_n| - |f - f_n| \ge 0$.

4.5.31 Part (b) is not a consequence of part (a) since $f(t)/t$ need not be integrable.

4.6.18 By the MCT, $\int_0^\infty x\, e^{-x^2(1+y^2)}\, dx = \lim_{n\to\infty} \int_0^n \cdots$ (improper Riemann integral).

4.6.21 (d) What is $\chi_{\{g>t\}}(x)$ as a function of t? (f) Compare $n\omega(2n)$ to $\int_n^{2n} \omega$.

4.6.27 $|(f*g)(x+h) - (f*g)(x)| = \bigl|\int f(x+h-y)\, g(y)\, dy \;-\; \int f(x-y)\, g(y)\, dy\bigr|$.

5.2.4 (a) Consider partitions that include $2/(n\pi)$. (b) Set $\alpha_n = (2/(4n\pi))^{1/2}$ and $\beta_n = (2/((4n-1)\pi))^{1/2}$. Show $\int_{\alpha_n}^{\beta_n} g'(x)\, dx = g(\beta_n) - g(\alpha_n)$. (c) Show h is Lipschitz.

5.2.11 (b) Consider $\Gamma' = \Gamma \cup \{x'\}$.

5.2.22 (a) Consider partitions that include $(2/(k\pi))^{1/b}$. (b) Consider $0 < x < y \le 1$ and set $h = y - x$. If $x^{b+1} < h$, then $|f(y) - f(x)| \le |f(y)| + |f(x)| \le y^b + x^b$; show $x^b \le h^\alpha$ and $y^b \le Ch^\alpha$. If $x^{b+1} \ge h$, use the MVT to show $|f(y) - f(x)| = h\,|f'(t)| \le \frac{2bh}{t} \le \cdots$.

5.4.6 First consider $D^+ f \ge \delta > 0$.

C. Heil, *Introduction to Real Analysis*, Graduate Texts in Mathematics 280,
https://doi.org/10.1007/978-3-030-26903-6

5.5.22 (a) Find a bounded E on which $|f| \geq \varepsilon$. For $|x|$ large, consider $B_{2|x|}(x)$.

6.1.10 Problem 6.1.9.

6.4.22 (c) Nonempty convex subsets of $[a,b]$ are intervals or points. (e) Lemma 6.2.4. (f) Intervals, then open sets, then measurable sets.

6.5.12 Corollary 6.5.8(b). Caution: $f_n \to f$ a.e. does not imply $f_n \circ g \to f \circ g$ a.e.

6.5.13 (b) Corollary 6.2.3. Do not assume $f \circ g$ must be measurable.

7.1.26 (d) Consider $\sum_{k=n+1}^{2n} |x_k|^p$.

7.2.16 q/p and $(q/p)'$.

7.2.20 Induction for $1 < p_1, \dots, p_n, r < \infty$. Alternative: Discrete Jensen.

7.2.22 Use Problem 7.2.21. First show $t\omega(t) \leq C\,\omega(t)^{1/p'}$.

7.2.23 (b) $\lim_{p\to 0^+} (x^p - 1)/p = \ln x$.

7.3.21 Fatou, Hölder, Egorov.

7.3.22 (a) First consider $g_N(x) = \sum_{n=1}^{N} |f_n(x)|$ and $g(x) = \sum_{n=1}^{\infty} |f_n(x)|$.

7.3.26 (a) Show a, b, $c \geq 0$ and $a \leq b + c$ implies $\frac{a}{1+a} \leq \frac{b}{1+b} + \frac{c}{1+c}$.

7.4.5 Converse to Hölder does not apply when $p = 1$. Consider $\widetilde{f}_h$ in equation (5.26).

8.1.12 Apply CBS to $\int_a^b f'(x)^{1/2}/f'(x)^{1/2}\,dx$.

8.3.30 Find a bounded function m such that $m(x) \neq 0$ a.e. and $f/m \notin L^2[a,b]$.

8.4.11 (a) $\{b^{1/2}e^{2\pi ibnx}\}_{n\in\mathbb{Z}}$ is an ONB for $L^2(I_k)$ where $I_k = [ak, ak + \frac{1}{b}]$. If $f \in C_c(\mathbb{R})$ then $f(x)\,\overline{g(x - ak)} \in L^2(I_k)$. $C_c(\mathbb{R})$ is dense.

9.1.33 Convolve with an approximate identity, and consider the Arzelá–Ascoli Theorem (for one statement of this theorem see [Heil18, Sec. 4.9]).

9.1.34 Let $J = \{j_1, j_2, \dots\}$ be a countable subset of I. Define $f(x_{j_n}) = n$ for $n \in \mathbb{N}$ and $f(x_i) = 0$ for $i \in I \backslash J_0$. Use the fact that $\{x_i\}_{i\in I}$ is a Hamel basis to extend $f(x)$ to $x \in \mathbb{R}$.

9.1.35 (a) Exercises 9.1.31 and 9.1.32 are helpful. (c) First show there is an *integer-valued* function $n(x)$ such that $f(x) = \alpha x + n(x)$ and $n(x + y) = n(x) + n(y)$.

9.2.29 Use the Inversion Formula to write $f(x+h) - f(x)$ in terms of $\widehat{f}$; break the integral into large $|\xi|$ and small $|\xi|$.

9.2.32 (c) Leibniz's rule: $(fg)^{(n)} = \sum_{k=0}^{n} \binom{n}{k} f^{(n-k)}\, g^{(k)}$.

9.3.25 (c) $\pi^4/90$.

9.3.33 Consider $f = e_n$ first.

9.3.34 Consider $\sum \frac{\sin 2\pi n_k x}{k}$.

9.3.35 For the lower estimate, $\frac{1}{2}\,\|d_N\|_1 \geq \int_0^{1/2} \frac{|\sin(2N+1)\pi x|}{\pi|x|}\,dx = \int_0^{N+\frac{1}{2}} \frac{|\sin \pi x|}{\pi|x|}\,dx \geq \sum_{k=0}^{N-1} \int_k^{k+1} \frac{|\sin \pi x|}{\pi|x|}\,dx$. For the upper estimate, show $\frac{1}{|\sin \pi x|} \leq \frac{1}{\pi|x|} + (1 - \frac{2}{\pi})$, $|x| \leq \frac{1}{2}$, and $\frac{1}{2}\,\|d_N\|_1 \leq \int_0^{1/2} \frac{|\sin(2N+1)\pi x|}{\pi|x|}\,dx + (1 - \frac{2}{\pi}) \int_0^{1/2} |\sin(2N+1)\pi x|\,dx$. Remark: *Euler's constant* is $\gamma = \lim_{N\to\infty} \Bigl(\sum_{k=1}^{N} \frac{1}{k} - \ln N\Bigr) \approx 0.57721566\dots$.

Index of Symbols

Sets

Symbol	Description	Reference
$\varnothing$	Empty set	Preliminaries
$B_r(x)$	Open ball of radius r centered at x	Definition 1.1.5
$\mathbb{C}$	Complex plane	Preliminaries
$\overline{\mathbf{F}}$	Choice of $\mathbb{C}$ or $[-\infty,\infty]$	Preliminaries
$\mathcal{L}=\mathcal{L}(\mathbb{R}^d)$	σ-algebra of Lebesgue measurable sets	Notation 2.2.2
$\mathbb{N}$	Natural numbers, $\{1,2,3,\dots\}$	Preliminaries
$\mathbb{Q}$	Rational numbers	Preliminaries
$\mathbb{R}$	Real line	Preliminaries
$\mathbb{T}$	Domain of 1-periodic functions	Section 9.3.1
$\mathbb{Z}$	Integers, $\{\dots,-1,0,1,\dots\}$	Preliminaries
$[-\infty,\infty]$	Extended real line	Preliminaries

Operations on Sets

Symbol	Description	Reference
$A^{\mathrm{C}}=X\backslash A$	Complement of a set $A\subseteq X$	Preliminaries
A°	Interior of a set A	Definition 1.1.5
$\overline{A}$	Closure of a set A	Definition 1.1.5
∂A	Boundary of a set A	Definition 1.1.5
$A\times B$	Cartesian product of A and B	Preliminaries
$\operatorname{dist}(A,B)$	Distance between two sets	Equation (2.11)
$E+h$	Translation of a set $E\subseteq\mathbb{R}^d$	Preliminaries
$\lvert E\rvert_e$	Exterior Lebesgue measure of $E\subseteq\mathbb{R}^d$	Definition 2.1.8
$\lvert E\rvert_i$	Inner Lebesgue measure of $E\subseteq\mathbb{R}^d$	Problem 2.2.43
$\lvert E\rvert$	Lebesgue measure of $E\subseteq\mathbb{R}^d$	Definition 2.2.1

C. Heil, *Introduction to Real Analysis*, Graduate Texts in Mathematics 280,
https://doi.org/10.1007/978-3-030-26903-6

$\liminf E_k$	Liminf of sets	Definition 2.1.14
$\limsup E_k$	Limsup of sets	Definition 2.1.14
$\inf(S)$	Infimum of a set of real numbers	Preliminaries
$\mathcal{P}(X)$	Power set of X	Preliminaries
$\mathrm{span}(\mathcal{F})$	Finite linear span of a set $\mathcal{F}$	Definition 1.2.1
$\overline{\mathrm{span}}(\mathcal{F})$	Closed linear span of $\mathcal{F}$	Notation 8.2.12
$\sup(S)$	Supremum of a set of real numbers	Preliminaries
$\mathrm{vol}(Q)$	Volume of a box Q	Definition 2.1.1

Sequences

Symbol	Description	Reference
$\{Q_k\}$	*Countable* sequence of boxes	Notation 2.1.3
$\{x_i\}_{i\in I}$	Sequence indexed by I	Preliminaries
$(x_i)_{i\in I}$	Sequence of scalars indexed by I	Preliminaries
$\liminf x_n$	liminf of a sequence of real numbers	Preliminaries
$\limsup x_n$	limsup of a sequence of real numbers	Preliminaries
δ_n	nth standard basis vector	Equation (7.14)

Functions

Symbol	Description	Reference
χ_A	Characteristic function of A	Preliminaries
$\mathrm{sinc}(x)$	sinc function	Exercise 4.3.2
$w(x)$	Fejér function	Exercise 9.1.10
$W(x)$	Hat function	Exercise 9.1.2

Operations on Functions

Symbol	Description	Reference
$\operatorname{ess\,sup} f$	Essential supremum of f	Definition 2.2.26
$\overline{f}$	Complex conjugate of f	Preliminaries
$\lvert f\rvert$	Absolute value of f	Preliminaries
f'	Derivative of f	Preliminaries
f^-	Negative part of f	Preliminaries
f^+	Positive part of f	Preliminaries
$\widetilde{f}_h$	Average of f over a ball of radius h	Section 5.5
$\widehat{f}(n)$	nth Fourier coefficient of f	Section 8.4
$\widehat{f}$	Fourier transform of f	Definition 9.2.1
$\check{f}$	Inverse Fourier transform of f	Definition 9.2.8
$f\vert_S$	Restriction of f to S	Preliminaries

$f(A)$	Direct image of A under f	Preliminaries
$f^{-1}(B)$	Inverse image of B under f	Preliminaries
$\{f > a\}$	Shorthand for $\{x : f(x) > a\}$	Preliminaries
$f_n \to f$ a.e.	Pointwise a.e. convergence	Notation 3.2.8
$f_n \nearrow f$	Monotone increasing sequence	Preliminaries
$f_n \xrightarrow{\mathrm{m}} f$	Convergence in measure	Definition 3.5.1
$f * g$	Convolution of f and g	Section 4.6.3
Mf	Maximal function of f	Definition 5.5.5
$\mathrm{range}(f)$	Range of f	Preliminaries
$\mathrm{supp}(f)$	Support of f	Section 1.3.1
$T_a f(x)$	Translation of f $(= f(x-a))$	Preliminaries
$V[f; a, b]$	Total variation of f on $[a, b]$	Definition 5.2.1
$V^+[f; a, b]$	Positive variation of f on $[a, b]$	Definition 5.2.13
$V^-[f; a, b]$	Negative variation of f on $[a, b]$	Definition 5.2.13

Some Vector Spaces

Symbol	Description	Reference
$A(\mathbb{R})$	Range of the Fourier transform	Section 9.2
$\mathrm{AC}[a, b]$	Absolutely continuous functions on $[a, b]$	Definition 6.1.1
$\mathrm{BV}[a, b]$	Functions of bounded variation on $[a, b]$	Definition 5.2.1
c_{00}	Finite sequences	Section 7.1.6
c_0	Sequences vanishing at infinity	Section 7.1.6
$C(X)$	Continuous functions on X	Section 1.3
$C_b(X)$	Bounded continuous functions on X	Section 1.3
$C_0(\mathbb{R}^d)$	Continuous functions vanishing at infinity	Section 1.3.1
$C_c(\mathbb{R}^d)$	Continuous, compactly supported functions	Section 1.3.1
$C^\alpha(I)$	Hölder continuous functions on an interval	Problem 1.4.5
$C^m(\mathbb{R})$	m-times differentiable functions	Section 1.3.1
$C^\infty(\mathbb{R})$	Infinitely differentiable functions	Section 1.3.1
ℓ^p	p-summable sequences	Definition 7.1.2
$L^1(E)$	Lebesgue space of integrable functions	Definition 4.4.3
$L^1_{\mathrm{loc}}(\mathbb{R}^d)$	Locally integrable functions	Definition 5.5.4
$L^p(E)$	Lebesgue space of p-integrable functions	Definition 7.2.1
$L^p(\mathbb{T})$	Space of 1-periodic L^p functions	Definition 9.3.1
$L^\infty(E)$	Space of essentially bounded functions	Definition 3.3.3
$\mathrm{Lip}(I)$	Lipschitz functions on an interval	Section 5.2.2
$\mathcal{S}(\mathbb{R})$	Schwartz space	Problem 9.2.32

Hilbert Space Notations

Symbol	Description	Reference
$\langle \cdot, \cdot \rangle$	Generic inner product	Definition 8.1.1
$\langle x, y \rangle$	Inner product of vectors in ℓ^2	Example 8.1.7
$\langle f, g \rangle$	Inner product of functions in $L^2(E)$	Example 8.1.8
$A^\perp$	Orthogonal complement of a set A	Definition 8.2.4
$f \perp g$	Orthogonal vectors	Definition 8.2.1
$x \cdot y$	Dot product of vectors x and y	Preliminaries

Some Norms

Symbol	Description	Reference
$\|\cdot\|$	Generic norm	Definition 1.2.3
$\|x\|$	Euclidean norm of a vector x	Preliminaries
$\|f\|_{\mathrm{u}}$	Uniform norm of a function f	Definition 1.3.1
$\|f\|_1$	L^1-norm of a function f	Definition 4.4.1
$\|f\|_p$	L^p-norm of a function f	Definition 7.2.1
$\|f\|_\infty$	L^∞-norm of a function f	Section 3.3
$\|f\|_{\mathrm{BV}}$	Bounded variation norm of a function f	Section 5.2.1
$\|x\|_p$	ℓ^p-norm of a sequence x	Definition 7.1.1
$\|x\|_\infty$	sup-norm of a sequence x	Definition 7.1.1

Miscellaneous Symbols

Symbol	Description	Reference
$\forall$	= "for all"	Math symbol
$\exists$	= "there exists"	Math symbol
a.e.	Almost everywhere	Notation 2.2.24
$\mathrm{d}(\cdot, \cdot)$	Generic metric	Definition 1.1.1
$\det(L)$	Determinant of a matrix L	Section 2.3.3
p'	Dual index to p	Preliminaries
δ_{ij}	Kronecker δ	Preliminaries
$\lvert\Gamma\rvert$	Mesh size of a partition Γ	Preliminaries
$\square$	End of proof	Preliminaries
$\diamondsuit$	End of Remark, Example, or Exercise	Preliminaries
$\diamondsuit$	End of Theorem whose proof is omitted	Preliminaries
$*$	Challenging Problem	Preliminaries

References

[Ask75] R. Askey, *Orthogonal Polynomials and Special Functions*, SIAM, Philadelphia, 1975.

[BS11] R. G. Bartle and D. R. Serbert, *Introduction to Real Analysis*, Fourth Edition, Wiley, New York, 2011.

[Ben97] J. J. Benedetto, *Harmonic Analysis and Applications*, CRC Press, Boca Raton, FL, 1997.

[BC09] J. J. Benedetto and W. Czaja, *Integration and Modern Analysis*, Birkhäuser, Boston, 2009.

[BP48] R. P. Boas and H. Pollard, The multiplicative completion of sets of functions, *Bull. Amer. Math. Soc.*, **54** (1948), pp. 518–522.

[Bro69] A. Brown, An elementary example of a continuous singular function, *Amer. Math. Monthly*, **76** (1969), pp. 295–297.

[BBT97] A. M. Bruckner, J. B. Bruckner, and B. S. Thomson, *Real Analysis*, Prentice-Hall, Upper Saddle River, NJ, 1997.

[Car66] L. Carleson, On convergence and growth of partial sums of Fourier series, *Acta Math.*, **116** (1966), pp. 135–157.

[Chr16] O. Christensen, *An Introduction to Frames and Riesz Bases*, Second Edition, Birkhäuser, Boston, 2016.

[Con90] J. B. Conway, *A Course in Functional Analysis*, Second Edition, Springer-Verlag, New York, 1990.

[Dal45] D. P. Dalzell, On the completeness of a series of normal orthogonal functions, J. London Math. Soc., **20** (1945), pp. 87–93.

[Dau92] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

[DM72] H. Dym and H. P. McKean, *Fourier Series and Integrals*, Academic Press, New York–London, 1972.

[Fol99] G. B. Folland, *Real Analysis*, Second Edition, Wiley, New York, 1999.

[Grö01] K. Gröchenig, *Foundations of Time-Frequency Analysis*, Birkhäuser, Boston, 2001.

[Haar10] A. Haar, Zür Theorie der Orthogonalen Funktionensystemen, *Math. Ann.*, **69** (1910), pp. 331–371; English translation in [HW06], pp. 155–188.

[HHW18] J. Haas, C. Heil, and M. D. Weir, *Thomas' Calculus*, 14th Edition, Pearson, Boston, 2018.

[Heil11] C. Heil, *A Basis Theory Primer*, Expanded Edition, Birkhäuser, Boston, 2011.

[Heil18] C. Heil, *Metrics, Norms, Inner Products, and Operator Theory*, Birkhäuser/Springer, Cham, 2018.

[HW06] C. Heil and D. F. Walnut, Eds., *Fundamental Papers in Wavelet Theory*, Princeton University Press, Princeton, NJ, 2006.

[HW96] E. Hernández and G. Weiss, *A First Course on Wavelets*, CRC Press, Boca Raton, FL, 1996.

C. Heil, *Introduction to Real Analysis*, Graduate Texts in Mathematics 280,
https://doi.org/10.1007/978-3-030-26903-6

[Her85] D. A. Herrero, From my chest of examples of Fourier transforms, *Rev. Un. Mat. Argentina*, **32** (1985), pp. 41–47.

[HR79] E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis, Vol. I*, Structure of topological groups, integration theory, group representations, Second Edition, Springer-Verlag, Berlin–New York, 1979.

[HJ90] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Corrected Reprint of the 1985 Original, Cambridge University Press, Cambridge, 1990.

[Hunt68] R. A. Hunt, On the convergence of Fourier series, in: *Orthogonal Expansions and their Continuous Analogues* (Proc. Conf., Edwardsville, Ill., 1967), D. T. Haimo, Ed., Southern Illinois Univ. Press, Carbondale, IL, 1968, pp. 235–255.

[Kat04] Y. Katznelson, *An Introduction to Harmonic Analysis*, Third Edition, Cambridge University Press, Cambridge, UK, 2004.

[KV95] J. Kovačević and M. Vetterli, *Wavelets and Subband Coding*, Prentice Hall, Englewood Cliffs, NJ, 1995.

[LL01] E. H. Lieb and M. Loss, *Analysis*, Second Edition, American Mathematical Society, Providence, RI, 2001.

[Mun75] J. R. Munkres, *Topology: A First Course*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

[Nat55] I. P. Natanson, *Theory of Functions of a Real Variable*, Translated from the Russian by Leo F. Boron with the collaboration of Edwin Hewitt, Frederick Ungar Publishing Co., New York, 1955.

[Rot02] J. J. Rotman, *Advanced Modern Algebra*, Prentice Hall, Upper Saddle River, NJ, 2002.

[Rud76] W. Rudin, *Principles of Mathematical Analysis*, Third Edition. McGraw-Hill, New York, 1976.

[Rud87] W. Rudin, *Real and Complex Analysis*, Third Edition. McGraw-Hill, New York, 1987.

[Rud90] W. Rudin, *Fourier Analysis on Groups*, Wiley, New York, 1990. Reprint of the 1962 original.

[SV69] J. Serrin and D. E. Varberg, A general chain rule for derivatives and the change of variables formula for the Lebesgue integral, *Amer. Math. Monthly*, **76** (1969), pp. 514–520.

[ST76] I. M. Singer and J. A. Thorpe, *Lecture Notes on Elementary Topology and Geometry*, Springer-Verlag, New York-Heidelberg, 1976 (reprint of the 1967 edition).

[SS03] E. M. Stein and R. Shakarchi, *Fourier Analysis, An Introduction*, Princeton University Press, Princeton, NJ, 2003.

[SS05] E. M. Stein and R. Shakarchi, *Real Analysis*, Princeton University Press, Princeton, NJ, 2005.

[Str06] G. Strang, *Linear Algebra and Its Applications*, Fourth Edition, Thomson Brooks/Cole, Belmont, CA, 2006.

[SN96] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley–Cambridge Press, Wellesley, MA, 1996.

[Str11] D. W. Stroock, *Essentials of Integration Theory for Analysis*, Springer, New York, 2011.

[Sze75] G. Szegö, *Orthogonal Polynomials*, Fourth edition, American Mathematical Society, Providence, R.I., 1975.

[Var65] D. E. Varberg, On absolutely continuous functions, *Amer. Math. Monthly*, **72** (1965), 831–841.

[Vit21] G. Vitali, Sulla condizione di chiusura di un sistema di funzioni ortogonali, *Att. R. Acad. Naz. Lincei, Rend. Cl. Sci. Fis., Mat. Natur.*, **30** (1921), 498–501.

[Wal02] D. F. Walnut, *An Introduction to Wavelet Analysis*, Birkhäuser Boston, Inc., Boston, MA, 2002.

[WZ77] R. L. Wheeden and A. Zygmund, *Measure and Integral*, Marcel Dekker, New York-Basel, 1977.

[You01] R. Young *An Introduction to Nonharmonic Fourier Series*, Revised First Edition, Academic Press, San Diego, 2001.

Index

C. Heil, *Introduction to Real Analysis*, Graduate Texts in Mathematics 280,
https://doi.org/10.1007/978-3-030-26903-6