Real Analysis

Modern Techniques and Their Applications

Second Edition

Gerald B. Folland

Pure and Applied Mathematics:
A Wiley-Interscience Series of Texts, Monographs, and Tracts

Real Analysis

Modern Techniques and Their Applications

Second Edition

Gerald B. Folland

UP DILIMAN
COLLEGE OF SCIENCE
CENTRAL LIBRARY

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

This book is printed on acid-free paper. ∞

Copyright © 1999 by John Wiley & Sons, Inc. All rights reserved.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4744. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008. E-Mail: PERMREQ@WILEY.COM.

Library of Congress Cataloging-in-Publication Data:

Folland, Gerald B.
Real analysis : modern techniques and their applications / Gerald B. Folland. — 2nd ed.
p. cm. — (Pure and applied mathematics)
"A Wiley-Interscience publication."
Includes bibliographical references and index.
ISBN 0-471-31716-0 (cloth : alk. paper)
1. Mathematical analysis. 2. Functions of real variables.
I. Title. II. Series: Pure and applied mathematics (John Wiley & Sons : Unnumbered)
QA300.F67 1999
515—dc21 98-37260

Printed in the United States of America

10 9 8 7 6 5 4 3 2

To my mother
Helen B. Folland
and to the memory of my father
Harold F. Folland

Preface

The name "real analysis" is something of an anachronism. Originally applied to the theory of functions of a real variable, it has come to encompass several subjects of a more general and abstract nature that underlie much of modern analysis. These general theories and their applications are the subject of this book, which is intended primarily as a text for a graduate-level analysis course. Chapters 1 through 7 are devoted to the core material from measure and integration theory, point set topology, and functional analysis that is a part of most graduate curricula in mathematics, together with a few related but less standard items with which I think all analysts should be acquainted. The last four chapters contain a variety of topics that are meant to introduce some of the other branches of analysis and to illustrate the uses of the preceding material. I believe these topics are all interesting and important, but their selection in preference to others is largely a matter of personal predilection.

The things one needs to know in order to read this book are as follows:

1. First and foremost, the classical theory of functions of a real variable: limits and continuity, differentiation and (Riemann) integration, infinite series, uniform convergence, and the notion of a metric space.

2. The arithmetic of complex numbers and the basic properties of the complex exponential function $e^{x+iy} = e^x(\cos y + i \sin y)$. (More advanced results from complex function theory are used only in the proof of the Riesz-Thorin theorem and in a few exercises and remarks.)

3. Some elementary set theory.

4. A bit of linear algebra — actually, not much beyond the definitions of vector spaces, linear mappings, and determinants.

All of the necessary material in (1) and (2) can be found in W. Rudin's classic *Principles of Mathematical Analysis* (3rd ed., McGraw-Hill, 1976) or its descendants such as R. S. Strichatrz's *The Way of Analysis* (Jones and Bartlett, 1995) or S. G. Krantz's *Real Analysis and Foundations* (CRC Press, 1991). A summary of the relevant facts about sets and metric spaces is provided here in Chapter 0. The reader should begin this book by examining §0.1 and §0.5 to become familiar with my notation and terminology; the rest of Chapter 0 can then be referred to as needed.

Each chapter concludes with a section entitled "Notes and References." These sections contain miscellaneous remarks, acknowledgments of sources, indications of results not discussed in the text, references for further reading, and historical notes. The latter are quite sketchy, although references to more detailed sources are provided; they are intended mainly to give an idea of how the subject grew out of its classical origins. I found it entertaining and instructive to read some of the original papers, and I hope to encourage others to do the same.

A sizable portion of this book is devoted to exercises. They are mostly in the form of assertions to be proved, and they range from trivial to difficult; hints and intermediate steps are provided for the more complicated ones. Every reader should peruse them, although only the most ambitious will try to work them all out. They serve several purposes: amplification of results and completion of proofs in the text, discussion of examples and counterexamples, applications of theorems, and development of further ideas. Instructors will probably wish to do some of the exercises in class; to maximize flexibility and minimize verbosity, I have followed the principle of "When in doubt, leave it as an exercise," especially with regard to examples. Exercises occur at the end of each section, but they are numbered consecutively within each chapter. In referring to them, "Exercise n" means the nth exercise in the present chapter unless another section is explicitly mentioned.

The topics in the book are arranged so as to allow some flexibility of presentation. For example, Chapters 4 and 5 do not depend on Chapters 1–3 except for a few examples and exercises. On the other hand, if one wishes to proceed quickly to L^p theory, one can skip from §3.3 to §§5.1–2 and thence to Chapter 6. Chapters 10 and 11 are independent of Chapters 8 and 9 except that the ideas in §8.6 are used in Chapter 10.

The new features of this edition are as follows:

- The material on the n-dimensional Lebesgue integral (§§2.6–7) has been rearranged and expanded.

- Tychonoff's theorem (§4.6) is proved by an elegant argument recently discovered by Paul Chernoff.

- The chapter on Fourier analysis has been split into two chapters (8 and 9). The material on Fourier series and integrals (§§8.3–5) has been rearranged and now contains the Dirichlet-Jordan theorem on convergence of Fourier series.

The material on distributions (§§9.1–2) has been extensively rewritten and expanded.

- A section on self-similarity and Hausdorff dimension (§11.3) has been added, replacing the outdated calculation of the Hausdorff dimension of Cantor sets in the old §10.2.

- Innumerable small changes have been made in the hope of improving the exposition.

The writer of a text on such a well-developed subject as real analysis must necessarily be indebted to his predecessors. I kept a large supply of books on hand while writing this one; they are too numerous to list here, but most of them can be found in the bibliography. I am also happy to acknowledge the influence of two of my teachers: the late Lynn Loomis, from whose lectures I first learned this subject, and Elias Stein, who has done much to shape my point of view. Finally, I am grateful to a number of people — especially Steven Krantz, Kenneth Ross, and William Faris — whose comments and corrigenda concerning the first edition have helped me to prepare the new one.

GERALD B. FOLLAND

Seattle, Washington

Contents

	Preface	*vii*
0	*Prologue*	*1*
	0.1 The Language of Set Theory	*1*
	0.2 Orderings	*4*
	0.3 Cardinality	*6*
	0.4 More about Well Ordered Sets	*9*
	0.5 The Extended Real Number System	*10*
	0.6 Metric Spaces	*13*
	0.7 Notes and References	*16*
1	*Measures*	*19*
	1.1 Introduction	*19*
	1.2 σ-algebras	*21*
	1.3 Measures	*24*
	1.4 Outer Measures	*28*
	1.5 Borel Measures on the Real Line	*33*
	1.6 Notes and References	*40*

2 *Integration* 43
- 2.1 *Measurable Functions* 43
- 2.2 *Integration of Nonnegative Functions* 49
- 2.3 *Integration of Complex Functions* 52
- 2.4 *Modes of Convergence* 60
- 2.5 *Product Measures* 64
- 2.6 *The n-dimensional Lebesgue Integral* 70
- 2.7 *Integration in Polar Coordinates* 77
- 2.8 *Notes and References* 81

3 *Signed Measures and Differentiation* 85
- 3.1 *Signed Measures* 85
- 3.2 *The Lebesgue-Radon-Nikodym Theorem* 88
- 3.3 *Complex Measures* 93
- 3.4 *Differentiation on Euclidean Space* 95
- 3.5 *Functions of Bounded Variation* 100
- 3.6 *Notes and References* 109

4 *Point Set Topology* 113
- 4.1 *Topological Spaces* 113
- 4.2 *Continuous Maps* 119
- 4.3 *Nets* 125
- 4.4 *Compact Spaces* 128
- 4.5 *Locally Compact Hausdorff Spaces* 131
- 4.6 *Two Compactness Theorems* 136
- 4.7 *The Stone-Weierstrass Theorem* 138
- 4.8 *Embeddings in Cubes* 143
- 4.9 *Notes and References* 146

5 *Elements of Functional Analysis* 151
- 5.1 *Normed Vector Spaces* 151
- 5.2 *Linear Functionals* 157
- 5.3 *The Baire Category Theorem and its Consequences* 161
- 5.4 *Topological Vector Spaces* 165
- 5.5 *Hilbert Spaces* 171
- 5.6 *Notes and References* 179

6 L^p Spaces 181
- 6.1 Basic Theory of L^p Spaces 181
- 6.2 The Dual of L^p 188
- 6.3 Some Useful Inequalities 193
- 6.4 Distribution Functions and Weak L^p 197
- 6.5 Interpolation of L^p Spaces 200
- 6.6 Notes and References 208

7 Radon Measures 211
- 7.1 Positive Linear Functionals on $C_c(X)$ 211
- 7.2 Regularity and Approximation Theorems 216
- 7.3 The Dual of $C_0(X)$ 221
- 7.4 Products of Radon Measures 226
- 7.5 Notes and References 231

8 Elements of Fourier Analysis 235
- 8.1 Preliminaries 235
- 8.2 Convolutions 239
- 8.3 The Fourier Transform 247
- 8.4 Summation of Fourier Integrals and Series 257
- 8.5 Pointwise Convergence of Fourier Series 263
- 8.6 Fourier Analysis of Measures 270
- 8.7 Applications to Partial Differential Equations 273
- 8.8 Notes and References 278

9 Elements of Distribution Theory 281
- 9.1 Distributions 281
- 9.2 Compactly Supported, Tempered, and Periodic Distributions 291
- 9.3 Sobolev Spaces 301
- 9.4 Notes and References 310

10 Topics in Probability Theory 313
- 10.1 Basic Concepts 313
- 10.2 The Law of Large Numbers 320
- 10.3 The Central Limit Theorem 325
- 10.4 Construction of Sample Spaces 328
- 10.5 The Wiener Process 330
- 10.6 Notes and References 336

11 More Measures and Integrals 339
11.1 Topological Groups and Haar Measure 339
11.2 Hausdorff Measure 348
11.3 Self-similarity and Hausdorff Dimension 355
11.4 Integration on Manifolds 361
11.5 Notes and References 363

Bibliography 365

Index of Notation 377

Index 379

Real Analysis

0

Prologue

The purpose of this introductory chapter is to establish the notation and terminology that will be used throughout the book and to present a few diverse results from set theory and analysis that will be needed later. The style here is deliberately terse, since this chapter is intended as a reference rather than a systematic exposition.

0.1 THE LANGUAGE OF SET THEORY

It is assumed that the reader is familiar with the basic concepts of set theory; the following discussion is meant mainly to fix our terminology.

Number Systems. Our notation for the fundamental number systems is as follows:

$\mathbb{N}$ = the set of positive integers (not including zero)

$\mathbb{Z}$ = the set of integers

$\mathbb{Q}$ = the set of rational numbers

$\mathbb{R}$ = the set of real numbers

$\mathbb{C}$ = the set of complex numbers

Logic. We shall avoid the use of special symbols from mathematical logic, preferring to remain reasonably close to standard English. We shall, however, use the abbreviation **iff** for "if and only if."

One point of elementary logic that is often insufficiently appreciated by students is the following: If A and B are mathematical assertions and $-A$, $-B$ are their

negations, the statement "A implies B" is logically equivalent to the contrapositive statement "$-B$ implies $-A$." Thus one may prove that A implies B by assuming $-B$ and deducing $-A$, and we shall frequently do so. This is not the same as *reductio ad absurdum,* which consists of assuming both A and $-B$ and deriving a contradiction.

Sets. The words "family" and "collection" will be used synonymously with "set," usually to avoid phrases like "set of sets." The empty set is denoted by $\varnothing$, and the family of all subsets of a set X is denoted by $\mathcal{P}(X)$:

$$\mathcal{P}(X) = \{E : E \subset X\}.$$

Here and elsewhere, the inclusion sign $\subset$ is interpreted in the weak sense; that is, the assertion "$E \subset X$" includes the possibility that $E = X$.

If $\mathcal{E}$ is a family of sets, we can form the union and intersection of its members:

$$\bigcup_{E\in\mathcal{E}} E = \{x : x \in E \text{ for some } E \in \mathcal{E}\},$$
$$\bigcap_{E\in\mathcal{E}} E = \{x : x \in E \text{ for all } E \in \mathcal{E}\}.$$

Usually it is more convenient to consider indexed families of sets:

$$\mathcal{E} = \{E_\alpha : \alpha \in A\} = \{E_\alpha\}_{\alpha\in A},$$

in which case the union and intersection are denoted by

$$\bigcup_{\alpha\in A} E_\alpha, \qquad \bigcap_{\alpha\in A} E_\alpha.$$

If $E_\alpha \cap E_\beta = \varnothing$ whenever $\alpha \neq \beta$, the sets E_α are called **disjoint**. The terms "disjoint collection of sets" and "collection of disjoint sets" are used interchangeably, as are "disjoint union of sets" and "union of disjoint sets."

When considering families of sets indexed by $\mathbb{N}$, our usual notation will be

$$\{E_n\}_{n=1}^\infty \quad \text{or} \quad \{E_n\}_1^\infty,$$

and likewise for unions and intersections. In this situation, the notions of **limit superior** and **limit inferior** are sometimes useful:

$$\limsup E_n = \bigcap_{k=1}^\infty \bigcup_{n=k}^\infty E_n, \qquad \liminf E_n = \bigcup_{k=1}^\infty \bigcap_{n=k}^\infty E_n.$$

The reader may verify that

$$\limsup E_n = \{x : x \in E_n \text{ for infinitely many } n\},$$
$$\liminf E_n = \{x : x \in E_n \text{ for all but finitely many } n\}.$$

If E and F are sets, we denote their **difference** by $E \setminus F$:

$$E \setminus F = \{x : x \in E \text{ and } x \notin F\},$$

and their **symmetric difference** by $E \triangle F$:

$$E \triangle F = (E \setminus F) \cup (F \setminus E).$$

When it is clearly understood that all sets in question are subsets of a fixed set X, we define the **complement** E^c of a set E (in X):

$$E^c = X \setminus E.$$

In this situation we have **deMorgan's laws**:

$$\Big(\bigcup_{\alpha \in A} E_\alpha\Big)^c = \bigcap_{\alpha \in A} E_\alpha^c, \qquad \Big(\bigcap_{\alpha \in A} E_\alpha\Big)^c = \bigcup_{\alpha \in A} E_\alpha^c.$$

If X and Y are sets, their **Cartesian product** $X \times Y$ is the set of all ordered pairs (x, y) such that $x \in X$ and $y \in Y$. A **relation** from X to Y is a subset of $X \times Y$. (If $Y = X$, we speak of a relation **on** X.) If R is a relation from X to Y, we shall sometimes write xRy to mean that $(x, y) \in R$. The most important types of relations are the following:

- Equivalence relations. An **equivalence relation** on X is a relation R on X such that

  xRx for all $x \in X$,

  xRy iff yRx,

  xRz whenever xRy and yRz for some y.

  The **equivalence class** of an element x is $\{y \in X : xRy\}$. X is the disjoint union of these equivalence classes.

- Orderings. See §0.2.

- Mappings. A **mapping** $f : X \to Y$ is a relation R from X to Y with the property that for every $x \in X$ there is a unique $y \in Y$ such that xRy, in which case we write $y = f(x)$. Mappings are sometimes called **maps** or **functions**; we shall generally reserve the latter name for the case when Y is $\mathbb{C}$ or some subset thereof.

If $f : X \to Y$ and $g : Y \to Z$ are mappings, we denote by $g \circ f$ their **composition**:

$$g \circ f : X \to Z, \qquad g \circ f(x) = g(f(x)).$$

If $D \subset X$ and $E \subset Y$, we define the **image** of D and the **inverse image** of E under a mapping $f : X \to Y$ by

$$f(D) = \{f(x) : x \in D\}, \qquad f^{-1}(E) = \{x : f(x) \in E\}.$$

It is easily verified that the map $f^{-1} : \mathcal{P}(Y) \to \mathcal{P}(X)$ defined by the second formula commutes with union, intersections, and complements:

$$f^{-1}\Big(\bigcup_{\alpha\in A} E_\alpha\Big) = \bigcup_{\alpha\in A} f^{-1}(E_\alpha), \qquad f^{-1}\Big(\bigcap_{\alpha\in A} E_\alpha\Big) = \bigcap_{\alpha\in A} f^{-1}(E_\alpha),$$
$$f^{-1}(E^c) = \big(f^{-1}(E)\big)^c.$$

(The direct image mapping $f : \mathcal{P}(X) \to \mathcal{P}(Y)$ commutes with unions, but in general not with intersections or complements.)

If $f : X \to Y$ is a mapping, X is called the **domain** of f and $f(X)$ is called the **range** of f. f is said to be **injective** if $f(x_1) = f(x_2)$ only when $x_1 = x_2$, **surjective** if $f(X) = Y$, and **bijective** if it is both injective and surjective. If f is bijective, it has an **inverse** $f^{-1} : Y \to X$ such that $f^{-1} \circ f$ and $f \circ f^{-1}$ are the identity mappings on X and Y, respectively. If $A \subset X$, we denote by $f|A$ the restriction of f to A:

$$(f|A) : A \to Y, \qquad (f|A)(x) = f(x) \text{ for } x \in A.$$

A **sequence** in a set X is a mapping from $\mathbb{N}$ into X. (We also use the term **finite sequence** to mean a map from $\{1, \ldots, n\}$ into X where $n \in \mathbb{N}$.) If $f : \mathbb{N} \to X$ is a sequence and $g : \mathbb{N} \to \mathbb{N}$ satisfies $g(n) < g(m)$ whenever $n < m$, the composition $f \circ g$ is called a **subsequence** of f. It is common, and often convenient, to be careless about distinguishing between sequences and their ranges, which are subsets of X indexed by $\mathbb{N}$. Thus, if $f(n) = x_n$, we speak of the sequence $\{x_n\}_1^\infty$; whether we mean a mapping from $\mathbb{N}$ to X or a subset of X will be clear from the context.

Earlier we defined the Cartesian product of two sets. Similarly one can define the Cartesian product of n sets in terms of ordered n-tuples. However, this definition becomes awkward for infinite families of sets, so the following approach is used instead. If $\{X_\alpha\}_{\alpha\in A}$ is an indexed family of sets, their **Cartesian product** $\prod_{\alpha\in A} X_\alpha$ is the set of all maps $f : A \to \bigcup_{\alpha\in A} X_\alpha$ such that $f(\alpha) \in X_\alpha$ for every $\alpha \in A$. (It should be noted, and then promptly forgotten, that when $A = \{1, 2\}$, the previous definition of $X_1 \times X_2$ is set-theoretically different from the present definition of $\prod_1^2 X_j$. Indeed, the latter concept depends on mappings, which are defined in terms of the former one.) If $X = \prod_{\alpha\in A} X_\alpha$ and $\alpha \in A$, we define the αth **projection** or **coordinate map** $\pi_\alpha : X \to X_\alpha$ by $\pi_\alpha(f) = f(\alpha)$. We also frequently write x and x_α instead of f and $f(\alpha)$ and call x_α the αth **coordinate** of x.

If the sets X_α are all equal to some fixed set Y, we denote $\prod_{\alpha\in A} X_\alpha$ by Y^A:

$$Y^A = \text{ the set of all mappings from } A \text{ to } Y.$$

If $A = \{1, \ldots, n\}$, Y^A is denoted by Y^n and may be identified with the set of ordered n-tuples of elements of Y.

0.2 ORDERINGS

A **partial ordering** on a nonempty set X is a relation R on X with the following properties:

- if xRy and yRz, then xRz;
- if xRy and yRx, then $x = y$;
- xRx for all x.

If R also satisfies

- if $x, y \in X$, then either xRy or yRx,

then R is called a **linear** (or **total**) ordering. For example, if E is any set, then $\mathcal{P}(E)$ is partially ordered by inclusion, and $\mathbb{R}$ is linearly ordered by its usual ordering. Taking this last example as a model, we shall usually denote partial orderings by $\leq$, and we write $x < y$ to mean that $x \leq y$ but $x \neq y$. We observe that a partial ordering on X naturally induces a partial ordering on every nonempty subset of X. Two partially ordered sets X and Y are said to be **order isomorphic** if there is a bijection $f : X \to Y$ such that $x_1 \leq x_2$ iff $f(x_1) \leq f(x_2)$.

If X is partially ordered by $\leq$, a **maximal** (resp. **minimal**) **element** of X is an element $x \in X$ such that the only $y \in X$ satisfying $x \leq y$ (resp. $x \geq y$) is x itself. Maximal and minimal elements may or may not exist, and they need not be unique unless the ordering is linear. If $E \subset X$, an **upper** (resp. **lower**) **bound** for E is an element $x \in X$ such that $y \leq x$ (resp. $x \leq y$) for all $y \in E$. An upper bound for E need not be an element of E, and unless E is linearly ordered, a maximal element of E need not be an upper bound for E. (The reader should think up some examples.)

If X is linearly ordered by $\leq$ and every nonempty subset of X has a (necessarily unique) minimal element, X is said to be **well ordered** by $\leq$, and (in defiance of the laws of grammar) $\leq$ is called a **well ordering** on X. For example, $\mathbb{N}$ is well ordered by its natural ordering.

We now state a fundamental principle of set theory and derive some consequences of it.

0.1 The Hausdorff Maximal Principle. *Every partially ordered set has a maximal linearly ordered subset.*

In more detail, this means that if X is partially ordered by $\leq$, there is a set $E \subset X$ that is linearly ordered by $\leq$, such that no subset of X that properly includes E is linearly ordered by $\leq$. Another version of this principle is the following:

0.2 Zorn's Lemma. *If X is a partially ordered set and every linearly ordered subset of X has an upper bound, then X has a maximal element.*

Clearly the Hausdorff maximal principle implies Zorn's lemma: An upper bound for a maximal linearly ordered subset of X is a maximal element of X. It is also not difficult to see that Zorn's lemma implies the Hausdorff maximal principle. (Apply Zorn's lemma to the collection of linearly ordered subsets of X, which is partially ordered by inclusion.)

0.3 The Well Ordering Principle. *Every nonempty set X can be well ordered.*

Proof. Let $\mathcal{W}$ be the collection of well orderings of subsets of X, and define a partial ordering on $\mathcal{W}$ as follows. If $\leq_1$ and $\leq_2$ are well orderings on the subsets E_1 and E_2, then $\leq_1$ precedes $\leq_2$ in the partial ordering if (i) $\leq_2$ extends $\leq_1$, i.e., $E_1 \subset E_2$ and $\leq_1$ and $\leq_2$ agree on E_1, and (ii) if $x \in E_2 \setminus E_1$ then $y \leq_2 x$ for all $y \in E_1$. The reader may verify that the hypotheses of Zorn's lemma are satisfied, so that $\mathcal{W}$ has a maximal element. This must be a well ordering on X itself, for if $\leq$ is a well ordering on a proper subset E of X and $x_0 \in X \setminus E$, then $\leq$ can be extended to a well ordering on $E \cup \{x_0\}$ by declaring that $x \leq x_0$ for all $x \in E$. ∎

0.4 The Axiom of Choice. *If $\{X_\alpha\}_{\alpha \in A}$ is a nonempty collection of nonempty sets, then $\prod_{\alpha \in A} X_\alpha$ is nonempty.*

Proof. Let $X = \bigcup_{\alpha \in A} X_\alpha$. Pick a well ordering on X and, for $\alpha \in A$, let $f(\alpha)$ be the minimal element of X_α. Then $f \in \prod_{\alpha \in A} X_\alpha$. ∎

0.5 Corollary. *If $\{X_\alpha\}_{\alpha \in A}$ is a disjoint collection of nonempty sets, there is a set $Y \subset \bigcup_{\alpha \in A} X_\alpha$ such that $Y \cap X_\alpha$ contains precisely one element for each $\alpha \in A$.*

Proof. Take $Y = f(A)$ where $f \in \prod_{\alpha \in A} X_\alpha$. ∎

We have deduced the axiom of choice from the Hausdorff maximal principle; in fact, it can be shown that the two are logically equivalent.

0.3 CARDINALITY

If X and Y are nonempty sets, we define the expressions

$$\operatorname{card}(X) \leq \operatorname{card}(Y), \qquad \operatorname{card}(X) = \operatorname{card}(Y), \qquad \operatorname{card}(X) \geq \operatorname{card}(Y)$$

to mean that there exists $f : X \to Y$ which is injective, bijective, or surjective, respectively. We also define

$$\operatorname{card}(X) < \operatorname{card}(Y), \qquad \operatorname{card}(X) > \operatorname{card}(Y)$$

to mean that there is an injection but no bijection, or a surjection but no bijection, from X to Y. Observe that we attach no meaning to the expression "$\operatorname{card}(X)$" when it stands alone; there are various ways of doing so, but they are irrelevant for our purposes (except when X is finite — see below). These relationships can be extended to the empty set by declaring that

$$\operatorname{card}(\varnothing) < \operatorname{card}(X) \text{ and } \operatorname{card}(X) > \operatorname{card}(\varnothing) \text{ for all } X \neq \varnothing.$$

For the remainder of this section we assume implicitly that all sets in question are nonempty in order to avoid special arguments for $\varnothing$. Our first task is to prove that the relationships defined above enjoy the properties that the notation suggests.

0.6 Proposition. $\operatorname{card}(X) \leq \operatorname{card}(Y)$ *iff* $\operatorname{card}(Y) \geq \operatorname{card}(X)$.

Proof. If $f : X \to Y$ is injective, pick $x_0 \in X$ and define $g : Y \to X$ by $g(y) = f^{-1}(y)$ if $y \in f(X)$, $g(y) = x_0$ otherwise. Then g is surjective. Conversely, if $g : Y \to X$ is surjective, the sets $g^{-1}(\{x\})$ $(x \in X)$ are nonempty and disjoint, so any $f \in \prod_{x \in X} g^{-1}(\{x\})$ is an injection from X to Y. ∎

0.7 Proposition. *For any sets X and Y, either* $\operatorname{card}(X) \leq \operatorname{card}(Y)$ *or* $\operatorname{card}(Y) \leq \operatorname{card}(X)$.

Proof. Consider the set $\mathcal{I}$ of all injections from subsets of X to Y. The members of $\mathcal{I}$ can be regarded as subsets of $X \times Y$, so $\mathcal{I}$ is partially ordered by inclusion. It is easily verified that Zorn's lemma applies, so $\mathcal{I}$ has a maximal element f, with (say) domain A and range B. If $x_0 \in X \setminus A$ and $y_0 \in Y \setminus B$, then f can be extended to an injection from $A \cup \{x_0\}$ to $Y \cup \{y_0\}$ by setting $f(x_0) = y_0$, contradicting maximality. Hence either $A = X$, in which case $\operatorname{card}(X) \leq \operatorname{card}(Y)$, or $B = Y$, in which case f^{-1} is an injection from Y to X and $\operatorname{card}(Y) \leq \operatorname{card}(X)$. ∎

0.8 The Schröder-Bernstein Theorem. *If* $\operatorname{card}(X) \leq \operatorname{card}(Y)$ *and* $\operatorname{card}(Y) \leq \operatorname{card}(X)$ *then* $\operatorname{card}(X) = \operatorname{card}(Y)$.

Proof. Let $f : X \to Y$ and $g : Y \to X$ be injections. Consider a point $x \in X$: If $x \in g(Y)$, we form $g^{-1}(x) \in Y$; if $g^{-1}(x) \in f(X)$, we form $f^{-1}(g^{-1}(x))$; and so forth. Either this process can be continued indefinitely, or it terminates with an element of $X \setminus g(Y)$ (perhaps x itself), or it terminates with an element of $Y \setminus f(X)$. In these three cases we say that x is in X_∞, X_X, or X_Y; thus X is the disjoint union of X_∞, X_X, and X_Y. In the same way, Y is the disjoint union of three sets Y_∞, Y_X, and Y_Y. Clearly f maps X_∞ onto Y_∞ and X_X onto Y_X, whereas g maps Y_Y onto X_Y. Therefore, if we define $h : X \to Y$ by $h(x) = f(x)$ if $X \in X_\infty \cup X_X$ and $h(x) = g^{-1}(x)$ if $x \in X_Y$, then h is bijective. ∎

0.9 Proposition. *For any set X,* $\operatorname{card}(X) < \operatorname{card}(\mathcal{P}(X))$.

Proof. On the one hand, the map $f(x) = \{x\}$ is an injection from X to $\mathcal{P}(X)$. On the other, if $g : X \to \mathcal{P}(X)$, let $Y = \{x \in X : x \notin g(x)\}$. Then $Y \notin g(X)$, for if $Y = g(x_0)$ for some $x_0 \in X$, any attempt to answer the question "Is $x_0 \in Y$?" quickly leads to an absurdity. Hence g cannot be surjective. ∎

A set X is called **countable** (or **denumerable**) if $\operatorname{card}(X) \leq \operatorname{card}(\mathbb{N})$. In particular, all finite sets are countable, and for these it is convenient to interpret "$\operatorname{card}(X)$" as the number of elements in X:

$$\operatorname{card}(X) = n \text{ iff } \operatorname{card}(X) = \operatorname{card}(\{1, \ldots, n\}).$$

If X is countable but not finite, we say that X is **countably infinite**.

0.10 Proposition.

a. *If X and Y are countable, so is $X \times Y$.*

b. *If A is countable and X_α is countable for every $\alpha \in A$, then $\bigcup_{\alpha \in A} X_\alpha$ is countable.*

c. *If X is countably infinite, then* $\operatorname{card}(X) = \operatorname{card}(\mathbb{N})$.

Proof. To prove (a) it suffices to prove that $\mathbb{N}^2$ is countable. But we can define a bijection from $\mathbb{N}$ to $\mathbb{N}^2$ by listing, for n successively equal to $2, 3, 4, \ldots$, those elements $(j, k) \in \mathbb{N}^2$ such that $j + k = n$ in order of increasing j, thus:

$$(1,1),\ (1,2),\ (2,1),\ (1,3),\ (2,2),\ (3,1),\ (1,4),\ (2,3),\ (3,2),\ (4,1), \ldots$$

As for (b), for each $\alpha \in A$ there is a surjective $f_\alpha : \mathbb{N} \to X_\alpha$, and then the map $f : \mathbb{N} \times A \to \bigcup_{\alpha \in A} X_\alpha$ defined by $f(n, \alpha) = f_\alpha(n)$ is surjective; the result therefore follows from (a). Finally, for (c) it suffices to assume that X is an infinite subset of $\mathbb{N}$. Let $f(1)$ be the smallest element of X, and define $f(n)$ inductively to be the smallest element of $E \setminus \{f(1), \ldots, f(n-1)\}$. Then f is easily seen to be a bijection from $\mathbb{N}$ to X. ∎

0.11 Corollary. *$\mathbb{Z}$ and $\mathbb{Q}$ are countable.*

Proof. $\mathbb{Z}$ is the union of the countable sets $\mathbb{N}$, $\{-n : n \in \mathbb{N}\}$, and $\{0\}$, and one can define a surjection $f : \mathbb{Z}^2 \to \mathbb{Q}$ by $f(m, n) = m/n$ if $n \neq 0$ and $f(m, 0) = 0$. ∎

A set X is said to have the **cardinality of the continuum** if $\operatorname{card}(X) = \operatorname{card}(\mathbb{R})$. We shall use the letter $\mathfrak{c}$ as an abbreviation for $\operatorname{card}(\mathbb{R})$:

$$\operatorname{card}(X) = \mathfrak{c} \text{ iff } \operatorname{card}(X) = \operatorname{card}(\mathbb{R}).$$

0.12 Proposition. $\operatorname{card}(\mathcal{P}(\mathbb{N})) = \mathfrak{c}$.

Proof. If $A \subset \mathbb{N}$, define $f(A) \in \mathbb{R}$ to be $\sum_{n \in A} 2^{-n}$ if $\mathbb{N} \setminus A$ is infinite and $1 + \sum_{n \in A} 2^{-n}$ if $\mathbb{N} \setminus A$ is finite. (In the two cases, $f(A)$ is the number whose base-2 decimal expansion is $0.a_1 a_2 \cdots$ or $1.a_1 a_2 \cdots$, where $a_n = 1$ if $n \in A$ and $a_n = 0$ otherwise.) Then $f : \mathcal{P}(\mathbb{N}) \to \mathbb{R}$ is injective. On the other hand, define $g : \mathcal{P}(\mathbb{Z}) \to \mathbb{R}$ by $g(A) = \log(\sum_{n \in A} 2^{-n})$ if A is bounded below and $g(A) = 0$ otherwise. Then g is surjective since every positive real number has a base-2 decimal expansion. Since $\operatorname{card}(\mathcal{P}(\mathbb{Z})) = \operatorname{card}(\mathcal{P}(\mathbb{N}))$, the result follows from the Schröder-Bernstein theorem. ∎

0.13 Corollary. *If* $\operatorname{card}(X) \geq \mathfrak{c}$, *then X is uncountable.*

Proof. Apply Proposition 0.9. ∎

The converse of this corollary is the so-called continuum hypothesis, whose validity is one of the famous undecidable problems of set theory; see §0.7.

0.14 Proposition.

a. If $\operatorname{card}(X) \le \mathfrak{c}$ *and* $\operatorname{card}(Y) \le \mathfrak{c}$, *then* $\operatorname{card}(X \times Y) \le \mathfrak{c}$.

b. If $\operatorname{card}(A) \le \mathfrak{c}$ *and* $\operatorname{card}(X_\alpha) \le \mathfrak{c}$ *for all* $\alpha \in A$, *then* $\operatorname{card}(\bigcup_{\alpha \in A} X_\alpha) \le \mathfrak{c}$.

Proof. For (a) it suffices to take $X = Y = \mathcal{P}(\mathbb{N})$. Define $\phi, \psi : \mathbb{N} \to \mathbb{N}$ by $\phi(n) = 2n$ and $\psi(n) = 2n - 1$. It is then easy to check that the map $f : \mathcal{P}(\mathbb{N})^2 \to \mathcal{P}(\mathbb{N})$ defined by $f(A, B) = \phi(A) \cup \psi(B)$ is bijective. (b) follows from (a) as in the proof of Proposition 0.10. ∎

0.4 MORE ABOUT WELL ORDERED SETS

The material in this section is optional; it is used only in a few exercises and in some notes at the ends of chapters.

Let X be a well ordered set. If $A \subset X$ is nonempty, A has a minimal element, which is its maximal lower bound or **infimum**; we shall denote it by $\inf A$. If A is bounded above, it also has a minimal upper bound or **supremum**, denoted by $\sup A$. If $x \in X$, we define the **initial segment** of x to be

$$I_x = \{y \in X : y < x\}.$$

The elements of I_x are called **predecessors** of x.

The principle of mathematical induction is equivalent to the fact that $\mathbb{N}$ is well ordered. It can be extended to arbitrary well ordered sets as follows:

0.15 The Principle of Transfinite Induction. *Let X be a well ordered set. If A is a subset of X such that $x \in A$ whenever $I_x \subset A$, then $A = X$.*

Proof. If $X \neq A$, let $x = \inf(X \setminus A)$. Then $I_x \subset A$ but $x \notin A$. ∎

0.16 Proposition. *If X is well ordered and $A \subset X$, then $\bigcup_{x \in A} I_x$ is either an initial segment or X itself.*

Proof. Let $J = \bigcup_{x \in A} I_x$. If $J \neq X$, let $b = \inf(X \setminus J)$. If there existed $y \in J$ with $y > b$, we would have $y \in I_x$ for some $x \in A$ and hence $b \in I_x$, contrary to construction. Hence $J \subset I_b$, and it is obvious that $I_b \subset J$. ∎

0.17 Proposition. *If X and Y are well ordered, then either X is order isomorphic to Y, or X is order isomorphic to an initial segment in Y, or Y is order isomorphic to an initial segment in X.*

Proof. Consider the set $\mathcal{F}$ of order isomorphisms whose domains are initial segments in X or X itself and whose ranges are initial segments in Y or Y itself. $\mathcal{F}$ is nonempty since the unique $f : \{\inf X\} \to \{\inf Y\}$ belongs to $\mathcal{F}$, and $\mathcal{F}$ is partially ordered by inclusion (its members being regarded as subsets of $X \times Y$).

An application of Zorn's lemma shows that $\mathcal{F}$ has a maximal element f, with (say) domain A and range B. If $A = I_x$ and $B = I_y$, then $A \cup \{x\}$ and $B \cup \{y\}$ are again initial segments of X and Y, and f could be extended by setting $f(x) = y$, contradicting maximality. Hence either $A = X$ or $B = Y$ (or both), and the result follows. ∎

0.18 Proposition. *There is an uncountable well ordered set Ω such that I_x is countable for each $x \in \Omega$. If Ω' is another set with the same properties, then Ω and Ω' are order isomorphic.*

Proof. Uncountable well ordered sets exist by the well ordering principle; let X be one. Either X has the desired property or there is a minimal element x_0 such that I_{x_0} is uncountable, in which case we can take $\Omega = I_{x_0}$. If Ω' is another such set, Ω' cannot be order isomorphic to an initial segment of Ω or vice versa, because Ω and Ω' are uncountable while their initial segments are countable, so Ω and Ω' are order isomorphic by Proposition 0.17. ∎

The set Ω in Proposition 0.18, which is essentially unique *qua* well ordered set, is called the **set of countable ordinals**. It has the following remarkable property:

0.19 Proposition. *Every countable subset of Ω has an upper bound.*

Proof. If $A \subset \Omega$ is countable, $\bigcup_{x \in A} I_x$ is countable and hence is not all of Ω. By Proposition 0.16, there exists $y \in \Omega$ such that $\bigcup_{x \in A} I_x = I_y$, and y is thus an upper bound for A. ∎

The set $\mathbb{N}$ of positive integers may be identified with a subset of Ω as follows. Set $f(1) = \inf \Omega$, and proceeding inductively, set $f(n) = \inf(\Omega \setminus \{f(1), \ldots, f(n-1)\})$. The reader may verify that f is an order isomorphism from $\mathbb{N}$ to I_ω, where ω is the minimal element of Ω such that I_ω is infinite.

It is sometimes convenient to add an extra element ω_1 to Ω to form a set $\Omega^* = \Omega \cup \{\omega_1\}$ and to extend the ordering on Ω to Ω^* by declaring that $x < \omega_1$ for all $x \in \Omega$. ω_1 is called the **first uncountable ordinal**. (The usual notation for ω_1 is Ω, since ω_1 is generally taken to be the set of countable ordinals itself.)

0.5 THE EXTENDED REAL NUMBER SYSTEM

It is frequently useful to adjoin two extra points ∞ $(= +\infty)$ and $-\infty$ to $\mathbb{R}$ to form the **extended real number system** $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, and to extend the usual ordering on $\mathbb{R}$ by declaring that $-\infty < x < \infty$ for all $x \in \mathbb{R}$. The completeness of $\mathbb{R}$ can then be stated as follows: Every subset A of $\overline{\mathbb{R}}$ has a least upper bound, or **supremum**, and a greatest lower bound, or **infimum**, which are denoted by $\sup A$ and $\inf A$. If $A = \{a_1, \ldots a_n\}$, we also write

$$\max(a_1, \ldots, a_n) = \sup A, \qquad \min(a_1, \ldots, a_n) = \inf A.$$

From completeness it follows that every sequence $\{x_n\}$ in $\overline{\mathbb{R}}$ has a **limit superior** and a **limit inferior**:

$$\limsup x_n = \inf_{k\geq 1}\Big(\sup_{n\geq k} x_n\Big), \qquad \liminf x_n = \sup_{k\geq 1}\Big(\inf_{n\geq k} x_n\Big).$$

The sequence $\{x_n\}$ converges (in $\mathbb{R}$) iff these two numbers are equal (and finite), in which case its limit is their common value. One can also define lim sup and lim inf for functions $f : \mathbb{R} \to \overline{\mathbb{R}}$, for instance:

$$\limsup_{x\to a} f(x) = \inf_{\delta>0}\Big(\sup_{0<|x-a|<\delta} f(x)\Big).$$

The arithmetical operations on $\mathbb{R}$ can be partially extended to $\overline{\mathbb{R}}$:

$$x \pm \infty = \pm\infty \ (x \in \mathbb{R}), \qquad \infty + \infty = \infty, \qquad -\infty - \infty = -\infty,$$
$$x \cdot (\pm\infty) = \pm\infty \ (x > 0), \qquad x \cdot (\pm\infty) = \mp\infty \ (x < 0).$$

We make no attempt to define $\infty - \infty$, but we abide by the convention that, unless otherwise stated,

$$0 \cdot (\pm\infty) = 0.$$

(The expression $0 \cdot \infty$ turns up now and then in measure theory, and for various reasons its proper interpretation is almost always 0.)

We employ the following notation for intervals in $\overline{\mathbb{R}}$: if $-\infty \leq a < b \leq \infty$,

$$(a,b) = \{x : a < x < b\}, \qquad [a,b] = \{x : a \leq x \leq b\},$$
$$(a,b] = \{x : a < x \leq b\}, \qquad [a,b) = \{x : a \leq x < b\}.$$

We shall occasionally encounter uncountable sums of nonnegative numbers. If X is an arbitrary set and $f : X \to [0,\infty]$, we define $\sum_{x\in X} f(x)$ to be the supremum of its finite partial sums:

$$\sum_{x\in X} f(x) = \sup\Big\{\sum_{x\in F} f(x) : F \subset X,\ F \text{ finite}\Big\}.$$

(Later we shall recognize this as the integral of f with respect to counting measure on X.)

0.20 Proposition. *Given $f : X \to [0,\infty]$, let $A = \{x : f(x) > 0\}$. If A is uncountable, then $\sum_{x\in X} f(x) = \infty$. If A is countably infinite, then $\sum_{x\in X} f(x) = \sum_1^\infty f(g(n))$ where $g : \mathbb{N} \to A$ is any bijection and the sum on the right is an ordinary infinite series.*

Proof. We have $A = \bigcup_1^\infty A_n$ where $A_n = \{x : f(x) > 1/n\}$. If A is uncountable, then some A_n must be uncountable, and $\sum_{x\in F} f(x) > \operatorname{card}(F)/n$ for F a finite subset of A_n; it follows that $\sum_{x\in X} f(x) = \infty$. If A is countably infinite,

$g : \mathbb{N} \to A$ is a bijection, and $B_N = g(\{1, \ldots, N\})$, then every finite subset F of A is contained in some B_N. Hence

$$\sum_{x \in F} f(x) \le \sum_1^N f(g(n)) \le \sum_{x \in X} f(x).$$

Taking the supremum over N, we find

$$\sum_{x \in F} f(x) \le \sum_1^\infty f(g(n)) \le \sum_{x \in X} f(x),$$

and then taking the supremum over F, we obtain the desired result. ∎

Some terminology concerning (extended) real-valued functions: A relation between numbers that is applied to functions is understood to hold pointwise. Thus $f \le g$ means that $f(x) \le g(x)$ for every x, and $\max(f, g)$ is the function whose value at x is $\max(f(x), g(x))$. If $X \subset \overline{\mathbb{R}}$ and $f : X \to \overline{\mathbb{R}}$, f is called **increasing** if $f(x) \le f(y)$ whenever $x \le y$ and **strictly increasing** if $f(x) < f(y)$ whenever $x < y$; similarly for **decreasing**. A function that is either increasing or decreasing is called **monotone**.

If $f : \mathbb{R} \to \mathbb{R}$ is an increasing function, then f has right- and left-hand limits at each point:

$$f(a+) = \lim_{x \searrow a} f(x) = \inf_{x > a} f(x), \qquad f(a-) = \lim_{x \nearrow a} f(x) = \sup_{x < a} f(x).$$

Moreover, the limiting values $f(\infty) = \sup_{a \in \mathbb{R}} f(x)$ and $f(-\infty) = \inf_{a \in \mathbb{R}} f(x)$ exist (possibly equal to $\pm\infty$). f is called **right continuous** if $f(a) = f(a+)$ for all $a \in \mathbb{R}$ and **left continuous** if $f(a) = f(a-)$ for all $a \in \mathbb{R}$.

For points x in $\mathbb{R}$ or $\mathbb{C}$, $|x|$ denotes the ordinary absolute value or modulus of x, $|a + ib| = \sqrt{a^2 + b^2}$. For points x in $\mathbb{R}^n$ or $\mathbb{C}^n$, $|x|$ denotes the Euclidean norm:

$$|x| = \left[\sum_1^n |x_j|^2\right]^{1/2}.$$

We recall that a set $U \subset \mathbb{R}$ is **open** if, for every $x \in U$, U includes an interval centered at x.

0.21 Proposition. *Every open set in $\mathbb{R}$ is a countable disjoint union of open intervals.*

Proof. If U is open, for each $x \in U$ consider the collection $\mathcal{I}_x$ of all open intervals I such that $x \in I \subset U$. It is easy to check that the union of any family of open intervals containing a point in common is again an open interval, and hence $J_x = \bigcup_{I \in \mathcal{I}_x} I$ is an open interval; it is the largest element of $\mathcal{I}_x$. If $x, y \in U$ then either $J_x = J_y$ or $J_x \cap J_y = \varnothing$, for otherwise $J_x \cup J_y$ would be a larger open interval than J_x in $\mathcal{I}_x$. Thus if $\mathcal{J} = \{J_x : x \in U\}$, the (distinct) members of $\mathcal{J}$ are disjoint, and $U = \bigcup_{J \in \mathcal{J}} J$. For each $J \in \mathcal{J}$, pick a rational number $f(J) \in J$. The map $f : \mathcal{J} \to \mathbb{Q}$ thus defined is injective, for if $J \ne J'$ then $J \cap J' = \varnothing$; therefore $\mathcal{J}$ is countable. ∎

0.6 METRIC SPACES

A **metric** on a set X is a function $\rho : X \times X \to [0, \infty)$ such that

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

(Intuitively, $\rho(x, y)$ is to be interpreted as the distance from x to y.) A set equipped with a metric is called a **metric space**. Some examples:

i. The Euclidean distance $\rho(x, y) = |x - y|$ is a metric on $\mathbb{R}^n$.

ii. $\rho_1(f, g) = \int_0^1 |f(x) - g(x)|\, dx$ and $\rho_\infty(f, g) = \sup_{0 \leq x \leq 1} |f(x) - g(x)|$ are metrics on the space of continuous functions on $[0, 1]$.

iii. If ρ is a metric on X and $A \subset X$, then $\rho|(A \times A)$ is a metric on A.

iv. If (X_1, ρ_1) and (X_2, ρ_2) are metric spaces, the **product metric** ρ on $X_1 \times X_2$ is given by

$$\rho\big((x_1, x_2),\ (y_1, y_2)\big) = \max\big(\rho_1(x_1, y_1),\ \rho_2(x_2, y_2)\big).$$

Other metrics are sometimes used on $X_1 \times X_2$, for instance,

$$\rho_1(x_1, y_1) + \rho_2(x_2, y_2) \quad \text{or} \quad \big[\rho_1(x_1, y_1)^2 + \rho_2(x_2, y_2)^2\big]^{1/2}.$$

These, however, are equivalent to the product metric in the sense that we shall define at the end of this section.

Let (X, ρ) be a metric space. If $x \in X$ and $r > 0$, the (open) **ball** of radius r about x is

$$B(r, x) = \big\{y \in X : \rho(x, y) < r\big\}.$$

A set $E \subset X$ is **open** if for every $x \in E$ there exists $r > 0$ such that $B(r, x) \subset E$, and **closed** if its complement is open. For example, every ball $B(r, x)$ is open, for if $y \in B(r, x)$ and $\rho(x, y) = s$ then $B(r - s, y) \subset B(r, x)$. Also, X and $\varnothing$ are both open and closed. Clearly the union of any family of open sets is open, and hence the intersection of any family of closed sets is closed. Also, the intersection (resp. union) of any finite family of open (resp. closed) sets is open (resp. closed). Indeed, if $U_1, \ldots U_n$ are open and $x \in \bigcap_1^n U_j$, for each j there exists $r_j > 0$ such that $B(r_j, x) \subset U_j$, and then $B(r, x) \subset \bigcap_1^n U_j$ where $r = \min(r_1, \ldots, r_n)$, so $\bigcap_1^n U_j$ is open.

If $E \subset X$, the union of all open sets $U \subset E$ is the largest open set contained in E; it is called the **interior** of E and is denoted by E^o. Likewise, the intersection of all closed sets $F \supset E$ is the smallest closed set containing E; it is called the **closure** of E and is denoted by $\overline{E}$. E is said to be **dense** in X if $\overline{E} = X$, and **nowhere dense** if

$\overline{E}$ has empty interior. X is called **separable** if it has a countable dense subset. (For example, $\mathbb{Q}^n$ is a countable dense subset of $\mathbb{R}^n$.) A sequence $\{x_n\}$ in X **converges** to $x \in X$ (symbolically: $x_n \to x$ or $\lim x_n = x$) if $\lim_{n\to\infty} \rho(x_n, x) = 0$.

0.22 Proposition. *If X is a metric space, $E \subset X$, and $x \in X$, the following are equivalent:*

a. $x \in \overline{E}$.

b. $B(r, x) \cap E \neq \varnothing$ *for all* $r > 0$.

c. There is a sequence $\{x_n\}$ in E that converges to x.

Proof. If $B(r, x) \cap E = \varnothing$, then $B(r, x)^c$ is a closed set containing E but not x, so $x \notin \overline{E}$. Conversely, if $x \notin \overline{E}$, since $(\overline{E})^c$ is open there exists $r > 0$ such that $B(r, x) \subset (\overline{E})^c \subset E^c$. Thus (a) is equivalent to (b). If (b) holds, for each $n \in \mathbb{N}$ there exists $x_n \in B(n^{-1}, x) \cap E$, so that $x_n \to x$. On the other hand, if $B(r, x) \cap E = \varnothing$, then $\rho(y, x) \geq r$ for all $y \in E$, so no sequence of E can converge to x. Thus (b) is equivalent to (c). ∎

If (X_1, ρ_1) and (X_2, ρ_2) are metric spaces, a map $f : X_1 \to X_2$ is called **continuous** at $x \in X$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that $\rho_2(f(y), f(x)) < \epsilon$ whenver $\rho_1(x, y) < \delta$ — in other words, such that $f^{-1}(B(\epsilon, f(x))) \supset B(\delta, x)$. The map f is called **continuous** if it is continuous at each $x \in X_1$ and **uniformly continuous** if, in addition, the δ in the definition of continuity can be chosen independent of x.

0.23 Proposition. *$f : X_1 \to X_2$ is continuous iff $f^{-1}(U)$ is open in X_1 for every open $U \subset X_2$.*

Proof. If the latter condition holds, then for every $x \in X_1$ and $\epsilon > 0$, the set $f^{-1}(B(\epsilon, f(x)))$ is open and contains x, so it contains some ball about x; this means that f is continuous at x. Conversely, suppose that f is continuous and U is open in X_2. For each $y \in U$ there exists $\epsilon_y > 0$ such that $B(\epsilon_y, y) \subset U$, and for each $x \in f^{-1}(\{y\})$ there exists $\delta_x > 0$ such that $B(\delta_x, x) \subset f^{-1}(B(\epsilon_y, y)) \subset f^{-1}(U)$. Thus $f^{-1}(U) = \bigcup_{x \in f^{-1}(U)} B(\delta_x, x)$ is open. ∎

A sequence $\{x_n\}$ in a metric space (X, ρ) is called **Cauchy** if $\rho(x_n, x_m) \to 0$ as $n, m \to \infty$. A subset E of X is called **complete** if every Cauchy sequence in E converges and its limit is in E. For example, $\mathbb{R}^n$ (with the Euclidean metric) is complete, whereas $\mathbb{Q}^n$ is not.

0.24 Proposition. *A closed subset of a complete metric space is complete, and a complete subset of an arbitrary metric space is closed.*

Proof. If X is complete, $E \subset X$ is closed, and $\{x_n\}$ is a Cauchy sequence in E, $\{x_n\}$ has a limit in X. By Proposition 0.22, $x \in \overline{E} = E$. If $E \subset X$ is complete and $x \in \overline{E}$, by Proposition (0.22) there is a sequence $\{x_n\}$ in E converging to x. $\{x_n\}$ is Cauchy, so its limit lies in E; thus $E = \overline{E}$. ∎

In a metric space (X, ρ) we can define the distance from a point to a set and the distance between two sets. Namely, if $x \in X$ and $E, F \subset X$,

$$\rho(x, E) = \inf\{\rho(x, y) : y \in E\},$$
$$\rho(E, F) = \inf\{\rho(x, y) : x \in E,\ y \in F\} = \inf\{\rho(x, F) : x \in E\}.$$

Observe that, by Proposition 0.22, $\rho(x, E) = 0$ iff $x \in \overline{E}$. We also define the **diameter** of $E \subset X$ to be

$$\operatorname{diam} E = \sup\{\rho(x, y) : x, y \in E\}.$$

E is called **bounded** if $\operatorname{diam} E < \infty$.

If $E \subset X$ and $\{V_\alpha\}_{\alpha \in A}$ is a family of sets such that $E \subset \bigcup_{\alpha \in A} V_\alpha$, $\{V_\alpha\}_{\alpha \in A}$ is called a **cover** of E, and E is said to be **covered** by the V_α's. E is called totally bounded if, for every $\epsilon > 0$, E can be covered by finitely many balls of radius ϵ. Every totally bounded set is bounded, for if $x, y \in \bigcup_1^n B(\epsilon, z_j)$, say $x \in B(\epsilon, z_1)$ and $y \in B(\epsilon, z_2)$, then

$$\rho(x, y) \le \rho(x, z_1) + \rho(z_1, z_2) + \rho(z_2, y) \le 2\epsilon + \max\{\rho(z_j, z_k) : 1 \le j, k \le n\}.$$

(The converse is false in general.) If E is totally bounded, so is $\overline{E}$, for it is easily seen that if $E \subset \bigcup_1^n B(\epsilon, z_j)$, then $\overline{E} \subset \bigcup_1^n B(2\epsilon, z_j)$.

0.25 Theorem. *If E is a subset of the metric space (X, ρ), the following are equivalent:*

a. E is complete and totally bounded.

b. **(The Bolzano-Weierstrass Property)** *Every sequence in E has a subsequence that converges to a point of E.*

c. **(The Heine-Borel Property)** *If $\{V_\alpha\}_{\alpha \in A}$ is a cover of E by open sets, there is a finite set $F \subset A$ such that $\{V_\alpha\}_{\alpha \in F}$ covers E.*

Proof. We shall show that (a) and (b) are equivalent, that (a) and (b) together imply (c), and finally that (c) implies (b).

(a) implies (b): Suppose that (a) holds and $\{x_n\}$ is a sequence in E. E can be covered by finitely many balls of radius 2^{-1}, and at least one of them must contain x_n for infinitely many n: say, $x_n \in B_1$ for $n \in N_1$. $E \cap B_1$ can be covered by finitely many balls of radius 2^{-2}, and at least one of them must contain x_n for infinitely many $n \in N_1$: say, $x_n \in B_2$ for $n \in N_2$. Continuing inductively, we obtain a sequence of balls B_j of radius 2^{-j} and a decreasing sequence of subsets N_j of $\mathbb{N}$ such that $x_n \in B_j$ for $n \in N_j$. Pick $n_1 \in N_1$, $n_2 \in N_2, \ldots$ such that $n_1 < n_2 < \cdots$. Then $\{x_{n_j}\}$ is a Cauchy sequence, for $\rho(x_{n_j}, x_{n_k}) < 2^{1-j}$ if $k > j$, and since E is complete, it has a limit in E.

(b) implies (a): We show that if either condition in (a) fails, then so does (b). If E is not complete, there is a Cauchy sequence $\{x_n\}$ in E with no limit in E. No subsequence of $\{x_n\}$ can converge in E, for otherwise the whole squence would converge to the same limit. On the other hand, if E is not totally bounded, let $\epsilon > 0$

be such that E cannot be covered by finitely many balls of radius ϵ. Choose $x_n \in E$ inductively as follows. Begin with any $x_1 \in E$, and having chosen $x_1, \ldots, x_n$, pick $x_{n+1} \in E \setminus \bigcup_1^n B(\epsilon, x_j)$. Then $\rho(x_n, x_m) > \epsilon$ for all n, m, so $\{x_n\}$ has no convergent subsequence.

(a) and (b) imply (c): It suffices to show that if (b) holds and $\{V_\alpha\}_{\alpha \in A}$ is a cover of E by open sets, there exists $\epsilon > 0$ such that every ball of radius ϵ that intersects E is contained in some V_α, for E can be covered by finitely many such balls by (a). Suppose to the contrary that for each $n \in \mathbb{N}$ there is a ball B_n of radius 2^{-n} such that $B_n \cap E \neq \varnothing$ and B_n is contained in no V_α. Pick $x_n \in B_n \cap E$; by passing to a subsequence we may assume that $\{x_n\}$ converges to some $x \in E$. We have $x \in V_\alpha$ for some α, and since V_α is open, there exists $\epsilon > 0$ such that $B(\epsilon, x) \subset V_\alpha$. But if n is large enough so that $\rho(x_n, x) < \epsilon/3$ and $2^{-n} < \epsilon/3$, then $B_n \subset B(\epsilon, x) \subset V_\alpha$, contradicting the assumption on B_n.

(c) implies (b): If $\{x_n\}$ is a sequence in E with no convergent subsequence, for each $x \in E$ there is a ball B_x centered at x that contains x_n for only finitely many n (otherwise some subsequence would converge to x). Then $\{B_x\}_{x \in E}$ is a cover of E by open sets with no finite subcover. ∎

A set E that possesses the properties (a)–(c) of Theorem 0.25 is called **compact**. Every compact set is closed (by Proposition 0.24) and bounded; the converse is false in general but true in $\mathbb{R}^n$.

0.26 Proposition. *Every closed and bounded subset of $\mathbb{R}^n$ is compact.*

Proof. Since closed subsets of $\mathbb{R}^n$ are complete, it suffices to show that bounded subsets of $\mathbb{R}^n$ are totally bounded. Since every bounded set is contained in some cube

$$Q = [-R, R]^n = \{x \in \mathbb{R}^n : \max(|x_1|, \ldots, |x_n|) \leq R\},$$

it is enough to show that Q is totally bounded. Given $\epsilon > 0$, pick an integer $k > R\sqrt{n}/\epsilon$, and express Q as the union of k^n congruent subcubes by dividing the interval $[-R, R]$ into k equal pieces. The side length of these subcubes is $2R/k$ and hence their diameter is $\sqrt{n}(2R/k) < 2\epsilon$, so they are contained in the balls of radius ϵ about their centers. ∎

Two metrics ρ_1 and ρ_2 on a set X are called **equivalent** if

$$C\rho_1 \leq \rho_2 \leq C'\rho_1 \text{ for some } C, C' > 0.$$

It is easily verified that equivalent metrics define the same open, closed, and compact sets, the same convergent and Cauchy sequences, and the same continuous and uniformly continuous mappings. Consequently, most results concerning metric spaces depend not on the particular metric chosen but only on its equivalence class.

0.7 NOTES AND REFERENCES

§§0.1–0.4: The best exposition of set theory for beginners is Halmos [63], and Smullyan and Fitting [135] is a good text on a more advanced level. Kelley [83]

also contains a concise account of of basic axiomatic set theory. All of these books present a deduction of the Hausdorff maximal principle from the axiom of choice, as does Hewitt and Stromberg [76].

The axiom of choice (or one of the propositions equivalent to it) is generally taken as one of the basic postulates in the axiomatic formulations of set theory. Some mathematicians of the intuitionist or constructivist persuasion reject it on the grounds that one has not proved the existence of a mathematical object until one has shown how to construct it in some reasonably explicit fashion, whereas the whole point of the axiom of choice is to provide existence theorems when constructive methods fail (or are too cumbersome for comfort). People who are seriously bothered by such objections belong to a minority that does not include the present writer; in this book the axiom of choice is used sparingly but freely.

The **continuum hypothesis** is the assertion that if $\mathrm{card}(X) < \mathfrak{c}$, then X is countable. (Since it follows easily from the construction of Ω, the set of countable ordinals, that $\mathrm{card}(\Omega) \leq \mathrm{card}(X)$ for any uncountable X, an equivalent assertion is that $\mathrm{card}(\Omega) = \mathfrak{c}$.) It is known, thanks to Gödel and Cohen, that the continuum hypothesis and its negation are both consistent with the standard axioms of set theory including the axiom of choice, assuming that those axioms are themselves consistent. (An exposition of the consistency and independence theorems for the axiom of choice and the continuum hypothesis can be found in Smullyan and Fitting [135].) Some mathematicians are willing to accept the continuum hypothesis as true, seemingly as a matter of convenience, but Gödel [56] and Cohen [26, p. 151] have both expressed suspicions that it should be false, and as of this writing no one has found any really compelling evidence on one side or the other. My own feeling, subject to revision in the event of a major breakthrough in set theory, is that if the answer to one's question turns out to depend on the continuum hypothesis, one should give up and ask a different question.

§0.6: A more detailed discussion of metric spaces can be found in Loomis and Sternberg [95] and DePree and Swartz [32].

1

Measures

In this chapter we set forth the basic concepts of measure theory, develop a general procedure for constructing nontrivial examples of measures, and apply this procedure to construct measures on the real line.

1.1 INTRODUCTION

One of the most venerable problems in geometry is to determine the area or volume of a region in the plane or in 3-space. The techniques of integral calculus provide a satisfactory solution to this problem for regions that are bounded by "nice" curves or surfaces but are inadequate to handle more complicated sets, even in dimension one. Ideally, for $n \in \mathbb{N}$ we would like to have a function μ that assigns to each $E \subset \mathbb{R}^n$ a number $\mu(E) \in [0, \infty]$, the n-dimensional measure of E, such that $\mu(E)$ is given by the usual integral formulas when the latter apply. Such a function μ should surely possess the following properties:

i. If $E_1, E_2, \ldots$ is a finite or infinite sequence of disjoint sets, then

$$\mu(E_1 \cup E_2 \cup \cdots) = \mu(E_1) + \mu(E_2) + \cdots.$$

ii. If E is congruent to F (that is, if E can be transformed into F by translations, rotations, and reflections), then $\mu(E) = \mu(F)$.

iii. $\mu(Q) = 1$, where Q is the unit cube

$$Q = \{x \in \mathbb{R}^n : 0 \leq x_j < 1 \text{ for } j = 1, \ldots, n\}.$$

Unfortunately, these conditions are mutually inconsistent. Let us see why this is true for $n = 1$. (The argument can easily be adapted to higher dimensions.) To begin with, we define an equivalence relation on $[0, 1)$ by declaring that $x \sim y$ iff $x - y$ is rational. Let N be a subset of $[0, 1)$ that contains precisely one member of each equivalence class. (To find such an N, one must invoke the axiom of choice.) Next, let $R = \mathbb{Q} \cap [0, 1)$, and for each $r \in R$ let

$$N_r = \{x + r : x \in N \cap [0, 1 - r)\} \cup \{x + r - 1 : x \in N \cap [1 - r, 1)\}.$$

That is, to obtain N_r, shift N to the right by r units and then shift the part that sticks out beyond $[0, 1)$ one unit to the left. Then $N_r \subset [0, 1)$, and every $x \in [0, 1)$ belongs to precisely one N_r. Indeed, if y is the element of N that belongs to the equivalence class of x, then $x \in N_r$ where $r = x - y$ if $x \geq y$ or $r = x - y + 1$ if $x < y$; on the other hand, if $x \in N_r \cap N_s$, then $x - r$ (or $x - r + 1$) and $x - s$ (or $x - s + 1$) would be distinct elements of N belonging to the same equivalence class, which is impossible.

Suppose now that $\mu : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ satisfies (i), (ii), and (iii). By (i) and (ii),

$$\mu(N) = \mu(N \cap [0, 1 - r)) + \mu(N \cap [1 - r, 1)) = \mu(N_r)$$

for any $r \in R$. Also, since R is countable and $[0, 1)$ is the disjoint union of the N_r's,

$$\mu([0, 1)) = \sum_{r \in R} \mu(N_r)$$

by (i) again. But $\mu([0, 1)) = 1$ by (iii), and since $\mu(N_r) = \mu(N)$, the sum on the right is either 0 (if $\mu(N) = 0$) or ∞ (if $\mu(N) > 0$). Hence no such μ can exist.

Faced with this discouraging situation, one might consider weakening (i) so that additivity is required to hold only for finite sequences. This is not a very good idea, as we shall see: The additivity for countable sequences is what makes all the limit and continuity results of the theory work smoothly. Moreover, in dimensions $n \geq 3$, even this weak form of (i) is inconsistent with (ii) and (iii). Indeed, in 1924 Banach and Tarski proved the following amazing result:

> Let U and V be arbitrary bounded open sets in $\mathbb{R}^n$, $n \geq 3$. There exist $k \in \mathbb{N}$ and subsets $E_1, \ldots, E_k, F_1, \ldots, F_k$ of $\mathbb{R}^n$ such that
>
> - the E_j's are disjoint and their union is U;
> - the F_j's are disjoint and their union is V;
> - E_j is congruent to F_j for $j = 1, \ldots, k$.

Thus one can cut up a ball the size of a pea into a finite number of pieces and rearrange them to form a ball the size of the earth! Needless to say, the sets E_j and F_j are *very* bizarre. They cannot be visualized accurately, and their construction depends on the axiom of choice. But their existence clearly precludes the construction of any $\mu : \mathcal{P}(\mathbb{R}^n) \to [0, \infty]$ that assigns positive, finite values to bounded open sets and satisfies (i) for finite sequences as well as (ii).

The moral of these examples is that $\mathbb{R}^n$ contains subsets which are so strangely put together that it is impossible to define a geometrically reasonable notion of measure for them, and the remedy for the situation is to discard the requirement that μ should be defined on *all* subsets of $\mathbb{R}^n$. Rather, we shall content ourselves with constructing μ on a class of subsets of $\mathbb{R}^n$ that includes all the sets one is likely to meet in practice unless one is deliberately searching for pathological examples. This construction will be carried out for $n = 1$ in §1.5 and for $n > 1$ in §2.6.

It is worthwhile, and not much extra work, to develop the theory in much greater generality. The conditions (ii) and (iii) are directly related to Euclidean geometry, but set functions satisfying (i), called *measures*, arise also in a great many other situations. For example, in a physics problem involving mass distributions, $\mu(E)$ could represent the total mass in the region E. For another example, in probability theory one considers a set X that represents the possible outcomes of an experiment, and for $E \subset X$, $\mu(E)$ is the probability that the outcome lies in E. We therefore begin by studying the theory of measures on abstract sets.

1.2 σ-ALGEBRAS

In this section we discuss the families of sets that serve as the domains of measures.

Let X be a nonempty set. An **algebra** of sets on X is a nonempty collection $\mathcal{A}$ of subsets of X that is closed under finite unions and complements; in other words, if $E_1, \ldots, E_n \in \mathcal{A}$, then $\bigcup_1^n E_j \in \mathcal{A}$; and if $E \in \mathcal{A}$, then $E^c \in \mathcal{A}$. A **σ-algebra** is an algebra that is closed under countable unions. (Some authors use the terms **field** and **σ-field** instead of algebra and σ-algebra.)

We observe that since $\bigcap_j E_j = (\bigcup_j E_j^c)^c$, algebras (resp. σ-algebras) are also closed under finite (resp. countable) intersections. Moreover, if $\mathcal{A}$ is an algebra, then $\varnothing \in \mathcal{A}$ and $X \in \mathcal{A}$, for if $E \in \mathcal{A}$ we have $\varnothing = E \cap E^c$ and $X = E \cup E^c$.

It is worth noting that an algebra $\mathcal{A}$ is a σ-algebra provided that it is closed under countable *disjoint* unions. Indeed, suppose $\{E_j\}_1^\infty \subset \mathcal{A}$. Set

$$F_k = E_k \setminus \left[\bigcup_1^{k-1} E_j\right] = E_k \cap \left[\bigcup_1^{k-1} E_j\right]^c.$$

Then the F_k's belong to $\mathcal{A}$ and are disjoint, and $\bigcup_1^\infty E_j = \bigcup_1^\infty F_k$. This device of replacing a sequence of sets by a disjoint sequence is worth remembering; it will be used a number of times below.

Some examples: If X is any set, $\mathcal{P}(X)$ and $\{\varnothing, X\}$ are σ-algebras. If X is uncountable, then

$$\mathcal{A} = \{E \subset X : E \text{ is countable or } E^c \text{ is countable}\}$$

is a σ-algebra, called the **σ-algebra of countable or co-countable sets**. (The point here is that if $\{E_j\}_1^\infty \subset \mathcal{A}$, then $\bigcup_1^\infty E_j$ is countable if all E_j are countable and is co-countable otherwise.)

It is trivial to verify that the intersection of any family of σ-algebras on X is again a σ-algebra. It follows that if $\mathcal{E}$ is any susbset of $\mathcal{P}(X)$, there is a unique smallest σ-algebra $\mathcal{M}(\mathcal{E})$ containing $\mathcal{E}$, namely, the intersection of all σ-algebras containing $\mathcal{E}$. (There is always at least one such, namely, $\mathcal{P}(X)$.) $\mathcal{M}(\mathcal{E})$ is called the σ-algebra **generated** by $\mathcal{E}$. The following observation is often useful:

1.1 Lemma. *If $\mathcal{E} \subset \mathcal{M}(\mathcal{F})$ then $\mathcal{M}(\mathcal{E}) \subset \mathcal{M}(\mathcal{F})$.*

Proof. $\mathcal{M}(\mathcal{F})$ is a σ-algebra containing $\mathcal{E}$; it therefore contains $\mathcal{M}(\mathcal{E})$. ∎

If X is any metric space, or more generally any topological space (see Chapter 4), the σ-algebra generated by the family of open sets in X (or, equivalently, by the family of closed sets in X) is called the **Borel σ-algebra** on X and is denoted by $\mathcal{B}_X$. Its members are called **Borel sets**. $\mathcal{B}_X$ thus includes open sets, closed sets, countable intersections of open sets, countable unions of closed sets, and so forth.

There is a standard terminology for the levels in this hierarchy. A countable intersection of open sets is called a $\boldsymbol{G_\delta}$ set; a countable union of closed sets is called an $\boldsymbol{F_\sigma}$ set; a countable union of G_δ sets is called a $\boldsymbol{G_{\delta\sigma}}$ set; a countable intersection of F_σ sets is called an $\boldsymbol{F_{\sigma\delta}}$ set; and so forth. (δ and σ stand for the German *Durchschnitt* and *Summe*, that is, intersection and union.)

The Borel σ-algebra on $\mathbb{R}$ will play a fundamental role in what follows. For future reference we note that it can be generated in a number of different ways:

1.2 Proposition. *$\mathcal{B}_\mathbb{R}$ is generated by each of the following:*

a. the open intervals: $\mathcal{E}_1 = \{(a,b) : a < b\}$,

b. the closed intervals: $\mathcal{E}_2 = \{[a,b] : a < b\}$,

c. the half-open intervals: $\mathcal{E}_3 = \{(a,b] : a < b\}$ or $\mathcal{E}_4 = \{[a,b) : a < b\}$,

d. the open rays: $\mathcal{E}_5 = \{(a,\infty) : a \in \mathbb{R}\}$ or $\mathcal{E}_6 = \{(-\infty,a) : a \in \mathbb{R}\}$,

e. the closed rays: $\mathcal{E}_7 = \{[a,\infty) : a \in \mathbb{R}\}$ or $\mathcal{E}_8 = \{(-\infty,a] : a \in \mathbb{R}\}$.

Proof. The elements of $\mathcal{E}_j$ for $j \neq 3, 4$ are open or closed, and the elements of $\mathcal{E}_3$ and $\mathcal{E}_4$ are G_δ sets — for example, $(a,b] = \bigcap_1^\infty (a, b+n^{-1})$. All of these are Borel sets, so by Lemma 1.1, $\mathcal{M}(\mathcal{E}_j) \subset \mathcal{B}_\mathbb{R}$ for all j. On the other hand, every open set in $\mathbb{R}$ is a countable union of open intervals, so by Lemma 1.1 again, $\mathcal{B}_\mathbb{R} \subset \mathcal{M}(\mathcal{E}_1)$. That $\mathcal{B}_\mathbb{R} \subset \mathcal{M}(\mathcal{E}_j)$ for $j \geq 2$ can now be established by showing that all open intervals lie in $\mathcal{M}(\mathcal{E}_j)$ and applying Lemma 1.1. For example, $(a,b) = \bigcup_1^\infty [a+n^{-1},\, b-n^{-1}] \in \mathcal{M}(\mathcal{E}_2)$. Verification of the other cases is left to the reader (Exercise 2). ∎

Let $\{X_\alpha\}_{\alpha\in A}$ be an indexed collection of nonempty sets, $X = \prod_{\alpha\in A} X_\alpha$, and $\pi_\alpha : X \to X_\alpha$ the coordinate maps. If $\mathcal{M}_\alpha$ is a σ-algebra on X_α for each α, the **product σ-algebra** on X is the σ-algebra generated by

$$\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathcal{M}_\alpha,\ \alpha \in A\}.$$

We denote this σ-algebra by $\bigotimes_{\alpha\in A} \mathcal{M}_\alpha$. (If $A = \{1,\ldots,n\}$ we also write $\bigotimes_1^n \mathcal{M}_j$ or $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$.) The significance of this definition will become clearer in §2.1;

for the moment we give an alternative, and perhaps more intuitive, characterization of product σ-algebras in the case of countably many factors.

1.3 Proposition. *If A is countable, then $\bigotimes_{\alpha\in A}\mathcal{M}_\alpha$ is the σ-algebra generated by $\{\prod_{\alpha\in A}E_\alpha : E_\alpha\in\mathcal{M}_\alpha\}$.*

Proof. If $E_\alpha\in\mathcal{M}_\alpha$, then $\pi_\alpha^{-1}(E_\alpha)=\prod_{\beta\in A}E_\beta$ where $E_\beta=X$ for $\beta\neq\alpha$; on the other hand, $\prod_{\alpha\in A}E_\alpha=\bigcap_{\alpha\in A}\pi_\alpha^{-1}(E_\alpha)$. The result therefore follows from Lemma 1.1. ∎

1.4 Proposition. *Suppose that $\mathcal{M}_\alpha$ is generated by $\mathcal{E}_\alpha$, $\alpha\in A$. Then $\bigotimes_{\alpha\in A}\mathcal{M}_\alpha$ is generated by $\mathcal{F}_1=\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha\in\mathcal{E}_\alpha,\ \alpha\in A\}$. If A is countable and $X_\alpha\in\mathcal{E}_\alpha$ for all α, $\bigotimes_{\alpha\in A}\mathcal{M}_\alpha$ is generated by $\mathcal{F}_2=\{\prod_{\alpha\in A}E_\alpha : E_\alpha\in\mathcal{E}_\alpha\}$.*

Proof. Obviously $\mathcal{M}(\mathcal{F}_1)\subset\bigotimes_{\alpha\in A}\mathcal{M}_\alpha$. On the other hand, for each α, the collection $\{E\subset X_\alpha : \pi_\alpha^{-1}(E)\in\mathcal{M}(\mathcal{F}_1)\}$ is easily seen to be a σ-algebra on X_α that contains $\mathcal{E}_\alpha$ and hence $\mathcal{M}_\alpha$. In other words, $\pi_\alpha^{-1}(E)\in\mathcal{M}(\mathcal{F}_1)$ for all $E\in\mathcal{M}_\alpha$, $\alpha\in A$, and hence $\bigotimes_{\alpha\in A}\mathcal{M}_\alpha\subset\mathcal{M}(\mathcal{F}_1)$. The second assertion follows from the first as in the proof of Proposition 1.3. ∎

1.5 Proposition. *Let $X_1,\dots,X_n$ be metric spaces and let $X=\prod_1^n X_j$, equipped with the product metric. Then $\bigotimes_1^n\mathcal{B}_{X_j}\subset\mathcal{B}_X$. If the X_j's are separable, then $\bigotimes_1^n\mathcal{B}_{X_j}=\mathcal{B}_X$.*

Proof. By Proposition 1.4, $\bigotimes_1^n\mathcal{B}_{X_j}$ is generated by the sets $\pi_j^{-1}(U_j)$, $1\le j\le n$, where U_j is open in X_j. Since these sets are open in X, Lemma 1.1 implies that $\bigotimes_1^n\mathcal{B}_{X_j}\subset\mathcal{B}_X$. Suppose now that C_j is a countable dense set in X_j, and let $\mathcal{E}_j$ be the collection of balls in X_j with rational radius and center in C_j. Then every open set in X_j is a union of members of $\mathcal{E}_j$ — in fact, a countable union since $\mathcal{E}_j$ itself is countable. Moreover, the set of points in X whose jth coordinate is in C_j for all j is a countable dense subset of X, and the balls of radius r in X are merely products of balls of radius r in the X_j's. It follows that $\mathcal{B}_{X_j}$ is generated by $\mathcal{E}_j$ and $\mathcal{B}_X$ is generated by $\{\prod_1^n E_j : E_j\in\mathcal{E}_j\}$. Therefore $\mathcal{B}_X=\bigotimes_1^n\mathcal{B}_{X_j}$ by Proposition 1.4. ∎

1.6 Corollary. $\mathcal{B}_{\mathbb{R}^n}=\bigotimes_1^n\mathcal{B}_{\mathbb{R}}$.

We conclude this section with a technical result that will be needed later. We define an **elementary family** to be a collection $\mathcal{E}$ of subsets of X such that

- $\varnothing\in\mathcal{E}$,
- if $E,F\in\mathcal{E}$ then $E\cap F\in\mathcal{E}$,
- if $E\in\mathcal{E}$ then E^c is a finite disjoint union of members of $\mathcal{E}$.

1.7 Proposition. *If $\mathcal{E}$ is an elementary family, the collection $\mathcal{A}$ of finite disjoint unions of members of $\mathcal{E}$ is an algebra.*

Proof. If $A, B \in \mathcal{E}$ and $B^c = \bigcup_1^J C_j$ ($C_j \in \mathcal{E}$, disjoint), then $A \setminus B = \bigcup_1^J (A \cap C_j)$ and $A \cup B = (A \setminus B) \cup B$, where these unions are disjoint, so $A \setminus B \in \mathcal{A}$ and $A \cup B \in \mathcal{A}$. It now follows by induction that if $A_1 \ldots, A_n \in \mathcal{E}$, then $\bigcup_1^n A_j \in \mathcal{A}$; indeed, by inductive hypothesis we may assume that $A_1, \ldots, A_{n-1}$ are disjoint, and then $\bigcup_1^n A_j = A_n \cup \bigcup_1^{n-1}(A_j \setminus A_n)$, which is a disjoint union. To see that $\mathcal{A}$ is closed under complements, suppose $A_1, \ldots A_n \in \mathcal{E}$ and $A_m^c = \bigcup_{j=1}^{J_m} B_m^j$ with $B_m^1, \ldots, B_m^{J_m}$ disjoint members of $\mathcal{E}$. Then

$$\Big(\bigcup_{m=1}^n A_m\Big)^c = \bigcap_{m=1}^n \Big(\bigcup_{j=1}^{J_m} B_m^j\Big) = \bigcup\{B_1^{j_1} \cap \cdots \cap B_n^{j_n} : 1 \le j_m \le J_m,\ 1 \le m \le n\},$$

which is in $\mathcal{A}$. ∎

Exercises

1. A family of sets $\mathcal{R} \subset \mathcal{P}(X)$ is called a **ring** if it is closed under finite unions and differences (i.e., if $E_1, \ldots, E_n \in \mathcal{R}$, then $\bigcup_1^n E_j \in \mathcal{R}$, and if $E, F \in \mathcal{R}$, then $E \setminus F \in \mathcal{R}$). A ring that is closed under countable unions is called a **σ-ring**.

a. Rings (resp. σ-rings) are closed under finite (resp. countable) intersections.

b. If $\mathcal{R}$ is a ring (resp. σ-ring), then $\mathcal{R}$ is an algebra (resp. σ-algebra) iff $X \in \mathcal{R}$.

c. If $\mathcal{R}$ is a σ-ring, then $\{E \subset X : E \in \mathcal{R} \text{ or } E^c \in \mathcal{R}\}$ is a σ-algebra.

d. If $\mathcal{R}$ is a σ-ring, then $\{E \subset X : E \cap F \in \mathcal{R} \text{ for all } F \in \mathcal{R}\}$ is a σ-algebra.

2. Complete the proof of Proposition 1.2.

3. Let $\mathcal{M}$ be an infinite σ-algebra.

a. $\mathcal{M}$ contains an infinite sequence of disjoint sets.

b. $\operatorname{card}(\mathcal{M}) \ge \mathfrak{c}$.

4. An algebra $\mathcal{A}$ is a σ-algebra iff $\mathcal{A}$ is closed under countable increasing unions (i.e., if $\{E_j\}_1^\infty \subset \mathcal{A}$ and $E_1 \subset E_2 \subset \cdots$, then $\bigcup_1^\infty E_j \in \mathcal{A}$).

5. If $\mathcal{M}$ is the σ-algebra generated by $\mathcal{E}$, then $\mathcal{M}$ is the union of the σ-algebras generated by $\mathcal{F}$ as $\mathcal{F}$ ranges over all countable subsets of $\mathcal{E}$. (Hint: Show that the latter object is a σ-algebra.)

1.3 MEASURES

Let X be a set equipped with a σ-algebra $\mathcal{M}$. A **measure** on $\mathcal{M}$ (or on $(X, \mathcal{M})$, or simply on X if $\mathcal{M}$ is understood) is a function $\mu : \mathcal{M} \to [0, \infty]$ such that

i. $\mu(\varnothing) = 0$,

ii. if $\{E_j\}_1^\infty$ is a sequence of disjoint sets in $\mathcal{M}$, then $\mu(\bigcup_1^\infty E_j) = \sum_1^\infty \mu(E_j)$.

Property (ii) is called **countable additivity**. It implies **finite additivity**:

ii′. if $E_1, \dots E_n$ are disjoint sets in $\mathcal{M}$, then $\mu(\bigcup_1^n E_j) = \sum_1^n \mu(E_j)$,

because one can take $E_j = \varnothing$ for $j > n$. A function μ that satisfies (i) and (ii′) but not necessarily (ii) is called a **finitely additive measure**.

If X is a set and $\mathcal{M} \subset \mathcal{P}(X)$ is a σ-algebra, $(X, \mathcal{M})$ is called a **measurable space** and the sets in $\mathcal{M}$ are called **measurable sets**. If μ is a measure on $(X, \mathcal{M})$, then $(X, \mathcal{M}, \mu)$ is called a **measure space**.

Let $(X, \mathcal{M}, \mu)$ be a measure space. Here is some standard terminology concerning the "size" of μ. If $\mu(X) < \infty$ (which implies that $\mu(E) < \infty$ for all $E \in \mathcal{M}$ since $\mu(X) = \mu(E) + \mu(E^c)$), μ is called **finite**. If $X = \bigcup_1^\infty E_j$ where $E_j \in \mathcal{M}$ and $\mu(E_j) < \infty$ for all j, μ is called **σ-finite**. More generally, if $E = \bigcup_1^\infty E_j$ where $E_j \in \mathcal{M}$ and $\mu(E_j) < \infty$ for all j, the set E is said to be **σ-finite** for μ. (It would be correct but more cumbersome to say that E is of σ-finite measure.) If for each $E \in \mathcal{M}$ with $\mu(E) = \infty$ there exists $F \in \mathcal{M}$ with $F \subset E$ and $0 < \mu(F) < \infty$, μ is called **semifinite**.

Every σ-finite measure is semifinite (Exercise 13), but not conversely. Most measures that arise in parctice are σ-finite, which is fortunate since non-σ-finite measures tend to exhibit pathological behavior. The properties of non-σ-finite measures will be explored from time to time in the exercises.

Let us examine a few examples of measures. These examples are of a rather trivial nature, although the first one is of practical importance. The construction of more interesting examples is a task to which we shall turn in the next two sections.

- Let X be any nonempty set, $\mathcal{M} = \mathcal{P}(X)$, and f any function from X to $[0, \infty]$. Then f determines a measure μ on $\mathcal{M}$ by the formula $\mu(E) = \sum_{x \in E} f(x)$. (For the definition of such possibly uncountable sums, see §0.5.) The reader may verify that μ is semifinite iff $f(x) < \infty$ for every $x \in X$, and μ is σ-finite iff μ is semifinite and $\{x : f(x) > 0\}$ is countable. Two special cases are of particular significance: If $f(x) = 1$ for all x, μ is called **counting measure**; and if, for some $x_0 \in X$, f is defined by $f(x_0) = 1$ and $f(x) = 0$ for $x \neq x_0$, μ is called the **point mass** or **Dirac measure** at x_0. (The same names are also applied to the restrictions of these measures to smaller σ-algebras on X.)

- Let X be an uncountable set, and let $\mathcal{M}$ be the σ-algebra of countable or co-countable sets. The function μ on $\mathcal{M}$ defined by $\mu(E) = 0$ if E is countable and $\mu(E) = 1$ if E is co-countable is easily seen to be a measure.

- Let X be an infinite set and $\mathcal{M} = \mathcal{P}(X)$. Define $\mu(E) = 0$ if E is finite, $\mu(E) = \infty$ if E is infinite. Then μ is a finitely additive measure but not a measure.

The basic properties of measures are summarized in the following theorem.

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

a. **(Monotonicity)** *If $E, F \in \mathcal{M}$ and $E \subset F$, then $\mu(E) \leq \mu(F)$.*

b. **(Subadditivity)** *If $\{E_j\}_1^\infty \subset \mathcal{M}$, then $\mu(\bigcup_1^\infty E_j) \leq \sum_1^\infty \mu(E_j)$.*

c. **(Continuity from below)** *If* $\{E_j\}_1^\infty \subset \mathcal{M}$ *and* $E_1 \subset E_2 \subset \cdots$, *then* $\mu(\bigcup_1^\infty E_j) = \lim_{j\to\infty} \mu(E_j)$.

d. **(Continuity from above)** *If* $\{E_j\}_1^\infty \subset \mathcal{M}$, $E_1 \supset E_2 \supset \cdots$, *and* $\mu(E_1) < \infty$, *then* $\mu(\bigcap_1^\infty E_j) = \lim_{j\to\infty} \mu(E_j)$.

Proof. (a) If $E \subset F$, then $\mu(F) = \mu(E) + \mu(F \setminus E) \geq \mu(E)$.

(b) Let $F_1 = E_1$ and $F_k = E_k \setminus (\bigcup_1^{k-1} E_j)$ for $k > 1$. Then the F_k's are disjoint and $\bigcup_1^n F_j = \bigcup_1^n E_j$ for all n. Therefore, by (a),

$$\mu\Big(\bigcup_1^\infty E_j\Big) = \mu\Big(\bigcup_1^\infty F_j\Big) = \sum_1^\infty \mu(F_j) \leq \sum_1^\infty \mu(E_j).$$

(c) Setting $E_0 = \varnothing$, we have

$$\mu\Big(\bigcup_1^\infty E_j\Big) = \sum_1^\infty \mu(E_j \setminus E_{j-1}) = \lim_{n\to\infty} \sum_1^n \mu(E_j \setminus E_{j-1}) = \lim_{n\to\infty} \mu(E_n).$$

(d) Let $F_j = E_1 \setminus E_j$; then $F_1 \subset F_2 \subset \cdots$, $\mu(E_1) = \mu(F_j) + \mu(E_j)$, and $\bigcup_1^\infty F_j = E_1 \setminus (\bigcap_1^\infty E_j)$. By (c), then,

$$\mu(E_1) = \mu\Big(\bigcap_1^\infty E_j\Big) + \lim_{j\to\infty} \mu(F_j) = \mu\Big(\bigcap_1^\infty E_j\Big) + \lim_{j\to\infty} \big[\mu(E_1) - \mu(E_j)\big].$$

Since $\mu(E_1) < \infty$, we may subtract it from both sides to yield the desired result. ∎

We remark that the condition $\mu(E_1) < \infty$ in part (d) could be replaced by $\mu(E_n) < \infty$ for some $n > 1$, as the first $n-1$ E_j's can be discarded from the sequence without affecting the intersection. However, some finiteness assumption is necessary, as it can happen that $\mu(E_j) = \infty$ for all j but $\mu(\bigcap_1^\infty E_j) < \infty$. (For example, let μ be counting measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ and let $E_j = \{n : n \geq j\}$; then $\bigcap_1^\infty E_j = \varnothing$.)

If $(X, \mathcal{M}, \mu)$ is a measure space, a set $E \in \mathcal{M}$ such that $\mu(E) = 0$ is called a **null** set. By subadditivity, any countable union of null sets is a null set, a fact which we shall use frequently. If a statement about points $x \in X$ is true except for x in some null set, we say that it is true **almost everywhere** (abbreviated **a.e.**), or for **almost every** x. (If more precision is needed, we shall speak of a **μ-null** set, or **μ-almost everywhere**).

If $\mu(E) = 0$ and $F \subset E$, then $\mu(F) = 0$ by monotonicity provided that $F \in \mathcal{M}$, but in general it need not be true that $F \in \mathcal{M}$. A measure whose domain includes all subsets of null sets is called **complete**. Completeness can sometimes obviate annoying technical points, and it can always be achieved by enlarging the domain of μ, as follows.

1.9 Theorem. *Suppose that* $(X, \mathcal{M}, \mu)$ *is a measure space. Let* $\mathcal{N} = \{N \in \mathcal{M} : \mu(N) = 0\}$ *and* $\overline{\mathcal{M}} = \{E \cup F : E \in \mathcal{M}$ *and* $F \subset N$ *for some* $N \in \mathcal{N}\}$. *Then* $\overline{\mathcal{M}}$ *is a* σ*-algebra, and there is a unique extension* $\overline{\mu}$ *of* μ *to a complete measure on* $\overline{\mathcal{M}}$.

Proof. Since $\mathcal{M}$ and $\mathcal{N}$ are closed under countable unions, so is $\overline{\mathcal{M}}$. If $E \cup F \in \overline{\mathcal{M}}$ where $E \in \mathcal{M}$ and $F \subset N \in \mathcal{N}$, we can assume that $E \cap N = \varnothing$ (otherwise, replace F and N by $F \setminus E$ and $N \setminus E$). Then $E \cup F = (E \cup N) \cap (N^c \cup F)$, so $(E \cup F)^c = (E \cup N)^c \cup (N \setminus F)$. But $(E \cup N)^c \in \mathcal{M}$ and $N \setminus F \subset N$, so that $(E \cup F)^c \in \overline{\mathcal{M}}$. Thus $\overline{\mathcal{M}}$ is a σ-algebra.

If $E \cup F \in \overline{\mathcal{M}}$ as above, we set $\overline{\mu}(E \cup F) = \mu(E)$. This is well defined, since if $E_1 \cup F_1 = E_2 \cup F_2$ where $F_j \subset N_j \in \mathcal{N}$, then $E_1 \subset E_2 \cup N_2$ and so $\mu(E_1) \leq \mu(E_2) + \mu(N_2) = \mu(E_2)$, and likewise $\mu(E_2) \leq \mu(E_1)$. It is easily verified that $\overline{\mu}$ is a complete measure on $\overline{\mathcal{M}}$, and that $\overline{\mu}$ is the only measure on $\overline{\mathcal{M}}$ that extends μ; details are left to the reader (Exercise 6). ∎

The measure $\overline{\mu}$ in Theorem 1.9 is called the **completion** of μ, and $\overline{\mathcal{M}}$ is called the **completion** of $\mathcal{M}$ with respect to μ.

Exercises

6. Complete the proof of Theorem 1.9.

7. If $\mu_1, \ldots, \mu_n$ are measures on $(X, \mathcal{M})$ and $a_1, \ldots, a_n \in [0, \infty)$, then $\sum_1^n a_j \mu_j$ is a measure on $(X, \mathcal{M})$.

8. If $(X, \mathcal{M}, \mu)$ is a measure space and $\{E_j\}_1^\infty \subset \mathcal{M}$, then $\mu(\liminf E_j) \leq \liminf \mu(E_j)$. Also, $\mu(\limsup E_j) \geq \limsup \mu(E_j)$ provided that $\mu(\bigcup_1^\infty E_j) < \infty$.

9. If $(X, \mathcal{M}, \mu)$ is a measure space and $E, F \in \mathcal{M}$, then $\mu(E) + \mu(F) = \mu(E \cup F) + \mu(E \cap F)$.

10. Given a measure space $(X, \mathcal{M}, \mu)$ and $E \in \mathcal{M}$, define $\mu_E(\mathcal{A}) = \mu(A \cap E)$ for $A \in \mathcal{M}$. Then μ_E is a measure.

11. A finitely additive measure μ is a measure iff it is continuous from below as in Theorem 1.8c. If $\mu(X) < \infty$, μ is a measure iff it is continuous from above as in Theorem 1.8d.

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

a. If $E, F \in \mathcal{M}$ and $\mu(E \triangle F) = 0$, then $\mu(E) = \mu(F)$.

b. Say that $E \sim F$ if $\mu(E \triangle F) = 0$; then $\sim$ is an equivalence relation on $\mathcal{M}$.

c. For $E, F \in \mathcal{M}$, define $\rho(E, F) = \mu(E \triangle F)$. Then $\rho(E, G) \leq \rho(E, F) + \rho(F, G)$, and hence ρ defines a metric on the space $\mathcal{M}/\sim$ of equivalence classes.

13. Every σ-finite measure is semifinite.

14. If μ is a semifinite measure and $\mu(E) = \infty$, for any $C > 0$ there exists $F \subset E$ with $C < \mu(F) < \infty$.

15. Given a measure μ on $(X, \mathcal{M})$, define μ_0 on $\mathcal{M}$ by $\mu_0(E) = \sup\{\mu(F) : F \subset E \text{ and } \mu(F) < \infty\}$.

a. μ_0 is a semifinite measure. It is called the **semifinite part** of μ.

b. If μ is semifinite, then $\mu = \mu_0$. (Use Exercise 14.)

c. There is a measure ν on $\mathcal{M}$ (in general, not unique) which assumes only the values 0 and ∞ such that $\mu = \mu_0 + \nu$.

16. Let $(X, \mathcal{M}, \mu)$ be a measure space. A set $E \subset X$ is called **locally measurable** if $E \cap A \in \mathcal{M}$ for all $A \in \mathcal{M}$ such that $\mu(A) < \infty$. Let $\widetilde{\mathcal{M}}$ be the collection of all locally measurable sets. Clearly $\mathcal{M} \subset \widetilde{\mathcal{M}}$; if $\mathcal{M} = \widetilde{\mathcal{M}}$, then μ is called **saturated**.

a. If μ is σ-finite, then μ is saturated.

b. $\widetilde{\mathcal{M}}$ is a σ-algebra.

c. Define $\widetilde{\mu}$ on $\widetilde{\mathcal{M}}$ by $\widetilde{\mu}(E) = \mu(E)$ if $E \in \mathcal{M}$ and $\widetilde{\mu}(E) = \infty$ otherwise. Then $\widetilde{\mu}$ is a saturated measure on $\widetilde{\mathcal{M}}$, called the **saturation** of μ.

d. If μ is complete, so is $\widetilde{\mu}$.

e. Suppose that μ is semifinite. For $E \in \widetilde{\mathcal{M}}$, define $\underline{\mu}(E) = \sup\{\mu(A) : A \in \mathcal{M} \text{ and } A \subset E\}$. Then $\underline{\mu}$ is a saturated measure on $\widetilde{\mathcal{M}}$ that extends μ.

f. Let X_1, X_2 be disjoint uncountable sets, $X = X_1 \cup X_2$, and $\mathcal{M}$ the σ-algebra of countable or co-countable sets in X. Let μ_0 be counting measure on $\mathcal{P}(X_1)$, and define μ on $\mathcal{M}$ by $\mu(E) = \mu_0(E \cap X_1)$. Then μ is a measure on $\mathcal{M}$, $\widetilde{\mathcal{M}} = \mathcal{P}(X)$, and in the notation of parts (c) and (e), $\widetilde{\mu} \neq \underline{\mu}$.

1.4 OUTER MEASURES

In this section we develop the tools we shall use to construct measures. To motivate the ideas, it may be useful to recall the procedure used in calculus to define the area of a bounded region E in the plane $\mathbb{R}^2$. One draws a grid of rectangles in the plane and approximates the area of E from below by the sum of the areas of the rectangles in the grid that are subsets of E, and from above by the sum of the areas of the rectangles in the grid that intersect E. The limits of these approximations as the grid is taken finer and finer give the "inner area" and "outer area" of E, and if they are equal, their common value is the "area" of E. (We shall discuss these matters in more detail in §2.6.) The key idea here is that of outer area, since if R is a large rectangle containing E, the inner area of E is just the area of R minus the outer area of $R \setminus E$.

The abstract generalization of the notion of outer area is as follows. An **outer measure** on a nonempty set X is a function $\mu^* : \mathcal{P}(X) \to [0, \infty]$ that satisfies

- $\mu^*(\varnothing) = 0$,
- $\mu^*(A) \leq \mu^*(B)$ if $A \subset B$,
- $\mu^*(\bigcup_1^\infty A_j) \leq \sum_1^\infty \mu^*(A_j)$.

The most common way to obtain outer measures is to start with a family $\mathcal{E}$ of "elementary sets" on which a notion of measure is defined (such as rectangles in the plane) and then to approximate arbitrary sets "from the outside" by countable unions of members of $\mathcal{E}$. The precise construction is as follows.

1.10 Proposition. *Let $\mathcal{E} \subset \mathcal{P}(X)$ and $\rho : \mathcal{E} \to [0, \infty]$ be such that $\varnothing \in \mathcal{E}$, $X \in \mathcal{E}$, and $\rho(\varnothing) = 0$. For any $A \subset X$, define*

$$\mu^*(A) = \inf\Big\{\sum_1^\infty \mu(E_j) : E_j \in \mathcal{E} \text{ and } A \subset \bigcup_1^\infty E_j\Big\}.$$

Then μ^* is an outer measure.

Proof. For any $A \subset X$ there exists $\{E_j\}_1^\infty \subset \mathcal{E}$ such that $A \subset \bigcup_1^\infty E_j$ (take $E_j = X$ for all j) so the definition of μ^* makes sense. Obviously $\mu^*(\varnothing) = 0$ (take $E_j = \varnothing$ for all j), and $\mu^*(A) \le \mu^*(B)$ for $A \subset B$ because the set over which the infimum is taken in the definition of $\mu^*(A)$ includes the corresponding set in the definition of $\mu^*(B)$. To prove the countable subadditivity, suppose $\{A_j\}_1^\infty \subset \mathcal{P}(X)$ and $\epsilon > 0$. For each j there exists $\{E_j^k\}_{k=1}^\infty \subset \mathcal{E}$ such that $A_j \subset \bigcup_{k=1}^\infty E_j^k$ and $\sum_{k=1}^\infty \rho(E_j^k) \le \mu^*(A_j) + \epsilon 2^{-j}$. But then if $A = \bigcup_1^\infty A_j$, we have $A \subset \bigcup_{j,k=1}^\infty E_j^k$ and $\sum_{j,k} \rho(E_j^k) \le \sum_j \mu^*(A_j) + \epsilon$, whence $\mu^*(A) \le \sum_j \mu^*(A_j) + \epsilon$. Since ϵ is arbitrary, we are done. ∎

The fundamental step that leads from outer measures to measures is as follows. If μ^* is an outer measure on X, a set $A \subset X$ is called **μ^*-measurable** if

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \text{ for all } E \subset X.$$

Of course, the inequality $\mu^*(E) \le \mu^*(E \cap A) + \mu^*(E \cap A^c)$ holds for any A and E, so to prove that A is μ^*-measurable, it suffices to prove the reverse inequality. The latter is trivial if $\mu^*(E) = \infty$, so we see that A is μ^*-measurable iff

$$\mu^*(E) \ge \mu^*(E \cap A) + \mu^*(E \cap A^c) \text{ for all } E \subset X \text{ such that } \mu^*(E) < \infty.$$

Some motivation for the notion of μ^*-measurability can be obtained by referring to the discussion at the beginning of this section. If E is a "well-behaved" set such that $E \supset A$, the equation $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$ says that the outer measure of A, $\mu^*(A)$, is equal to the "inner measure" of A, $\mu^*(E) - \mu^*(E \cap A^c)$. The leap from "well-behaved" sets containing A to arbitrary subsets of X a large one, but it is justified by the following theorem.

1.11 Carathéodory's Theorem. *If μ^* is an outer measure on X, the collection $\mathcal{M}$ of μ^*-measurable sets is a σ-algebra, and the restriction of μ^* to $\mathcal{M}$ is a complete measure.*

Proof. First, we observe that $\mathcal{M}$ is closed under complements since the definition of μ^*-measurability of A is symmetric in A and A^c. Next, if $A, B \in \mathcal{M}$ and $E \subset X$,

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c)\\ &= \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c).\end{aligned}$$

But $(A \cup B) = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$, so by subadditivity,

$$\mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) \ge \mu^*(E \cap (A \cup B)),$$

and hence

$$\mu^*(E) \geq \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c).$$

It follows that $A \cup B \in \mathcal{M}$, so $\mathcal{M}$ is an algebra. Moreover, if $A, B \in \mathcal{M}$ and $A \cap B = \varnothing$,

$$\mu^*(A \cup B) = \mu^*((A \cup B) \cap A) + \mu^*((A \cup B) \cap A^c) = \mu^*(A) + \mu^*(B),$$

so μ^* is finitely additive on $\mathcal{M}$.

To show that $\mathcal{M}$ is a σ-algebra, it will suffice to show that $\mathcal{M}$ is closed under countable disjoint unions. If $\{A_j\}_1^\infty$ is a sequence of disjoint sets in $\mathcal{M}$, let $B_n = \bigcup_1^n A_j$ and $B = \bigcup_1^\infty A_j$. Then for any $E \subset X$,

$$\begin{aligned}\mu^*(E \cap B_n) &= \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^c)\\ &= \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1}),\end{aligned}$$

so a simple induction shows that $\mu^*(E \cap B_n) = \sum_1^n \mu^*(E \cap A_j)$. Therefore,

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \sum_1^n \mu^*(E \cap A_j) + \mu^*(E \cap B^c),$$

and letting $n \to \infty$ we obtain

$$\begin{aligned}\mu^*(E) &\geq \sum_1^\infty \mu^*(E \cap A_j) + \mu^*(E \cap B^c) \geq \mu^*\Big(\bigcup_1^\infty (E \cap A_j)\Big) + \mu^*(E \cap B^c)\\ &= \mu^*(E \cap B) + \mu^*(E \cap B^c) \geq \mu^*(E).\end{aligned}$$

All the inequalities in this last calculation are thus equalities. It follows that $B \in \mathcal{M}$ and — taking $E = B$ — that $\mu^*(B) = \sum_1^\infty \mu^*(A_j)$, so μ^* is countably additive on $\mathcal{M}$. Finally, if $\mu^*(A) = 0$, for any $E \subset X$ we have

$$\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A^c) \leq \mu^*(E),$$

so that $A \in \mathcal{M}$. Therefore $\mu^*|\mathcal{M}$ is a complete measure. ∎

Our first applications of Carathéodory's theorem will be in the context of extending measures from algebras to σ-algebras. More precisely, if $\mathcal{A} \subset \mathcal{P}(X)$ is an algebra, a function $\mu_0 : \mathcal{A} \to [0, \infty]$ will be called a **premeasure** if

- $\mu_0(\varnothing) = 0$,
- if $\{A_j\}_1^\infty$ is a sequence of disjoint sets in $\mathcal{A}$ such that $\bigcup_1^\infty A_j \in \mathcal{A}$, then $\mu_0(\bigcup_1^\infty A_j) = \sum_1^\infty \mu_0(A_j)$.

In particular, a premeasure is finitely additive since one can take $A_j = \varnothing$ for j large. The notions of finite and σ-finite premeasures are defined just as for measures. If μ_0

is a premeasure on $\mathcal{A} \subset \mathcal{P}(X)$, it induces an outer measure on X in accordance with Proposition 1.10, namely,

$$\mu^*(E) = \inf\Big\{\sum_1^\infty \mu_0(A_j) : A_j \in \mathcal{A},\ E \subset \bigcup_1^\infty A_j\Big\}. \tag{1.12}$$

1.13 Proposition. *If μ_0 is a premeasure on $\mathcal{A}$ and μ^* is defined by (1.12), then*

a. $\mu^*|\mathcal{A} = \mu_0$;

b. every set in $\mathcal{A}$ is μ^* measurable.

Proof. (a) Suppose $E \in \mathcal{A}$. If $E \subset \bigcup_1^\infty A_j$ with $A_j \in \mathcal{A}$, let $B_n = E \cap (A_n \setminus \bigcup_1^{n-1} A_j)$. Then the B_n's are disjoint members of $\mathcal{A}$ whose union is E, so $\mu_0(E) = \sum_1^\infty \mu_0(B_j) \le \sum_1^\infty \mu_0(A_j)$. It follows that $\mu_0(E) \le \mu^*(E)$, and the reverse inequality is obvious since $E \subset \bigcup_1^\infty A_j$ where $A_1 = E$ and $A_j = \varnothing$ for $j > 1$.

(b) If $A \in \mathcal{A}$, $E \subset X$, and $\epsilon > 0$, there is a sequence $\{B_j\}_1^\infty \subset \mathcal{A}$ with $E \subset \bigcup_1^\infty B_j$ and $\sum_1^\infty \mu_0(B_j) \le \mu^*(E) + \epsilon$. Since μ_0 is additive on $\mathcal{A}$,

$$\mu^*(E) + \epsilon \ge \sum_1^\infty \mu_0(B_j \cap A) + \sum_1^\infty \mu_0(B_j \cap A^c) \ge \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

Since ϵ is arbitrary, A is μ^*-measurable. ∎

1.14 Theorem. *Let $\mathcal{A} \subset \mathcal{P}(X)$ be an algebra, μ_0 a premeasure on $\mathcal{A}$, and $\mathcal{M}$ the σ-algebra generated by $\mathcal{A}$. There exists a measure μ on $\mathcal{M}$ whose restriction to $\mathcal{A}$ is μ_0 — namely, $\mu = \mu^*|\mathcal{M}$ where μ^* is given by (1.12). If ν is another measure on $\mathcal{M}$ that extends μ_0, then $\nu(E) \le \mu(E)$ for all $E \in \mathcal{M}$, with equality when $\mu(E) < \infty$. If μ_0 is σ-finite, then μ is the unique extension of μ_0 to a measure on $\mathcal{M}$.*

Proof. The first assertion follows from Carathéodory's theorem and Proposition 1.13 since the σ-algebra of μ^*-measurable sets includes $\mathcal{A}$ and hence $\mathcal{M}$. As for the second assertion, if $E \in \mathcal{M}$ and $E \subset \bigcup_1^\infty A_j$ where $A_j \in \mathcal{A}$, then $\nu(E) \le \sum_1^\infty \nu(A_j) = \sum_1^\infty \mu_0(A_j)$, whence $\nu(E) \le \mu(E)$. Also, if we set $A = \bigcup_1^\infty A_j$, we have

$$\nu(A) = \lim_{n\to\infty} \nu\Big(\bigcup_1^n A_j\Big) = \lim_{n\to\infty} \mu\Big(\bigcup_1^n A_j\Big) = \mu(A).$$

If $\mu(E) < \infty$, we can choose the A_j's so that $\mu(A) < \mu(E) + \epsilon$, hence $\mu(A \setminus E) < \epsilon$, and

$$\mu(E) \le \mu(A) = \nu(A) = \nu(E) + \nu(A \setminus E) \le \nu(E) + \mu(A \setminus E) \le \nu(E) + \epsilon.$$

Since ϵ is arbitrary, $\mu(E) = \nu(E)$. Finally, suppose $X = \bigcup_1^\infty A_j$ with $\mu_0(A_j) < \infty$, where we can assume that the A_j's are disjoint. Then for any $E \in \mathcal{M}$,

$$\mu(E) = \sum_1^\infty \mu(E \cap A_j) = \sum_1^\infty \nu(E \cap A_j) = \nu(E),$$

so $\nu = \mu$. ∎

The proof of this theorem yields more than the statement. Indeed, μ_0 may be extended to a measure on the algebra $\mathcal{M}^*$ of all μ^*-measurable sets. The relation between $\mathcal{M}$ and $\mathcal{M}^*$ is explored in Exercise 22 (along with Exercise 20b, which ensures that the outer measures induced by μ_0 and μ are the same).

Exercises

17. If μ^* is an outer measure on X and $\{A_j\}_1^\infty$ is a sequence of disjoint μ^*-measurable sets, then $\mu^*(E \cap (\bigcup_1^\infty A_j)) = \sum_1^\infty \mu^*(E \cap A_j)$ for any $E \subset X$.

18. Let $\mathcal{A} \subset \mathcal{P}(X)$ be an algebra, $\mathcal{A}_\sigma$ the collection of countable unions of sets in $\mathcal{A}$, and $\mathcal{A}_{\sigma\delta}$ the collection of countable intersections of sets in $\mathcal{A}_\sigma$. Let μ_0 be a premeasure on $\mathcal{A}$ and μ^* the induced outer measure.

a. For any $E \subset X$ and $\epsilon > 0$ there exists $A \in \mathcal{A}_\sigma$ with $E \subset A$ and $\mu^*(A) \leq \mu^*(E) + \epsilon$.

b. If $\mu^*(E) < \infty$, then E is μ^*-measurable iff there exists $B \in \mathcal{A}_{\sigma\delta}$ with $E \subset B$ and $\mu^*(B \setminus E) = 0$.

c. If μ_0 is σ-finite, the restriction $\mu^*(E) < \infty$ in (b) is superfluous.

19. Let μ^* be an outer measure on X induced from a finite premeasure μ_0. If $E \subset X$, define the **inner measure** of E to be $\mu_*(E) = \mu_0(X) - \mu^*(E^c)$. Then E is μ^*-measurable iff $\mu^*(E) = \mu_*(E)$. (Use Exercise 18.)

20. Let μ^* be an outer measure on X, $\mathcal{M}^*$ the σ-algebra of μ^*-measurable sets, $\overline{\mu} = \mu^*|\mathcal{M}^*$, and μ^+ the outer measure induced by $\overline{\mu}$ as in (1.12) (with $\overline{\mu}$ and $\mathcal{M}^*$ replacing μ_0 and $\mathcal{A}$).

a. If $E \subset X$, we have $\mu^*(E) \leq \mu^+(E)$, with equality iff there exists $A \in \mathcal{M}^*$ with $A \supset E$ and $\mu^*(A) = \mu^*(E)$.

b. If μ^* is induced from a premeasure, then $\mu^* = \mu^+$. (Use Exercise 18a.)

c. If $X = \{0, 1\}$, there exists an outer measure μ^* on X such that $\mu^* \neq \mu^+$.

21. Let μ^* be an outer measure induced from a premeasure and $\overline{\mu}$ the restriction of μ^* to the μ^*-measurable sets. Then $\overline{\mu}$ is saturated. (Use Exercise 18.)

22. Let $(X, \mathcal{M}, \mu)$ be a measure space, μ^* the outer measure induced by μ according to (1.12), $\mathcal{M}^*$ the σ-algebra of μ^*-measurable sets, and $\overline{\mu} = \mu^*|\mathcal{M}^*$.

a. If μ is σ-finite, then $\overline{\mu}$ is the completion of μ. (Use Exercise 18.)

b. In general, $\overline{\mu}$ is the saturation of the completion of μ. (See Exercises 16 and 21.)

23. Let $\mathcal{A}$ be the collection of finite unions of sets of the form $(a, b] \cap \mathbb{Q}$ where $-\infty \leq a < b \leq \infty$.

a. $\mathcal{A}$ is an algebra on $\mathbb{Q}$. (Use Proposition 1.7.)

b. The σ-algebra generated by $\mathcal{A}$ is $\mathcal{P}(\mathbb{Q})$.

c. Define μ_0 on $\mathcal{A}$ by $\mu_0(\varnothing) = 0$ and $\mu_0(A) = \infty$ for $A \neq \varnothing$. Then μ_0 is a premeasure on $\mathcal{A}$, and there is more than one measure on $\mathcal{P}(\mathbb{Q})$ whose restriction to $\mathcal{A}$ is μ_0.

24. Let μ be a finite measure on $(X, \mathcal{M})$, and let μ^* be the outer measure induced by μ. Suppose that $E \subset X$ satisfies $\mu^*(E) = \mu^*(X)$ (but not that $E \in \mathcal{M}$).

a. If $A, B \in \mathcal{M}$ and $A \cap E = B \cap E$, then $\mu(A) = \mu(B)$.

b. Let $\mathcal{M}_E = \{A \cap E : A \in \mathcal{M}\}$, and define the function ν on $\mathcal{M}_E$ defined by $\nu(A \cap E) = \mu(A)$ (which makes sense by (a)). Then $\mathcal{M}_E$ is a σ-algebra on E and ν is a measure on $\mathcal{M}_E$.

1.5 BOREL MEASURES ON THE REAL LINE

We are now in a position to construct a definitive theory for measuring subsets of $\mathbb{R}$ based on the idea that the measure of an interval is its length. We begin with a more general (but only slightly more complicated) construction that yields a large family of measures on $\mathbb{R}$ whose domain is the Borel σ-algebra $\mathcal{B}_{\mathbb{R}}$; such measures are called **Borel measures** on $\mathbb{R}$.

To motivate the ideas, suppose that μ is a finite Borel measure on $\mathbb{R}$, and let $F(x) = \mu((-\infty, x])$. (F is sometimes called the **distribution function** of μ.) Then F is increasing by Theorem 1.8a and right continuous by Theorem 1.8d since $(-\infty, x] = \bigcap_1^\infty (-\infty, x_n]$ whenever $x_n \searrow x$. (Recall the discussion of increasing functions in §0.5.) Moreover, if $b > a$, $(-\infty, b] = (-\infty, a] \cup (a, b]$, so $\mu((a, b]) = F(b) - F(a)$. Our procedure will be to turn this process around and construct a measure μ starting from an increasing, right-continuous function F. The special case $F(x) = x$ will yield the usual "length" measure.

The building blocks for our theory will be the left-open, right-closed intervals in $\mathbb{R}$ — that is, sets of the form $(a, b]$ or (a, ∞) or $\varnothing$, where $-\infty \le a < b < \infty$. In this section we shall refer to such sets as **h-intervals** (h for "half-open"). Clearly the intersection of two h-intervals is an h-interval, and the complement of an h-interval is an h-interval or the disjoint union of two h-intervals. By Proposition 1.7, the collection $\mathcal{A}$ of finite disjoint unions of h-intervals is an algebra, and by Proposition 1.2, the σ-algebra generated by $\mathcal{A}$ is $\mathcal{B}_{\mathbb{R}}$.

1.15 Proposition. *Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and right continuous. If $(a_j, b_j]$ ($j = 1, \ldots, n$) are disjoint h-intervals, let*

$$\mu_0\Big(\bigcup_1^n (a_j, b_j]\Big) = \sum_1^n [F(b_j) - F(a_j)],$$

and let $\mu_0(\varnothing) = 0$. Then μ_0 is a premeasure on the algebra $\mathcal{A}$.

Proof. First we must check that μ_0 is well defined, since elements of $\mathcal{A}$ can be represented in more than one way as disjoint unions of h-intervals. If $\{(a_j, b_j]\}_1^n$ are disjoint and $\bigcup_1^n (a_j, b_j] = (a, b]$, then, after perhaps relabeling the index j, we must have $a = a_1 < b_1 = a_2 < b_2 = \ldots < b_n = b$, so $\sum_1^n [F(b_j) - F(a_j)] = F(b) - F(a)$. More generally, if $\{I_i\}_1^n$ and $\{J_j\}_1^m$ are finite sequences of disjoint

h-intervals such that $\bigcup_1^n I_i = \bigcup_1^n J_j$, this reasoning shows that

$$\sum_i \mu_0(I_i) = \sum_{i,j} \mu_0(I_i \cap J_j) = \sum_j \mu_0(J_j).$$

Thus μ_0 is well defined, and it is finitely additive by construction.

It remains to show that if $\{I_j\}_1^\infty$ is a sequence of disjoint h-intervals with $\bigcup_1^\infty I_j \in \mathcal{A}$ then $\mu_0(\bigcup_1^\infty I_j) = \sum_1^\infty \mu_0(I_j)$. Since $\bigcup_1^\infty I_j$ is a finite union of h-intervals, the sequence $\{I_j\}_1^\infty$ can be partitioned into finitely many subsequences such that the union of the intervals in each subsequence is a single h-interval. By considering each subsequence separately and using the finite additivity of μ_0, we may assume that $\bigcup_1^\infty I_j$ is an h-interval $I = (a, b]$. In this case, we have

$$\mu_0(I) = \mu_0\Big(\bigcup_1^n I_j\Big) + \mu_0\Big(I \setminus \bigcup_1^n I_j\Big) \geq \mu_0\Big(\bigcup_1^n I_j\Big) = \sum_1^n \mu_0(I_j).$$

Letting $n \to \infty$, we obtain $\mu_0(I) \geq \sum_1^\infty \mu(I_j)$. To prove the reverse inequality, let us suppose first that a and b are finite, and let us fix $\epsilon > 0$. Since F is right continuous, there exists $\delta > 0$ such that $F(a+\delta) - F(a) < \epsilon$, and if $I_j = (a_j, b_j]$, for each j there exists $\delta_j > 0$ such that $F(b_j + \delta_j) - F(b_j) < \epsilon 2^{-j}$. The open intervals $(a_j,\, b_j + \delta_j)$ cover the compact set $[a+\delta,\, b]$, so there is a finite subcover. By discarding any $(a_j,\, b_j + \delta_j)$ that is contained in a larger one and relabeling the index j, we may assume that

- the intervals $(a_1,\, b_1 + \delta_1), \ldots, (a_N,\, b_N + \delta_N)$ cover $[a+\delta,\, b]$,
- $b_j + \delta_j \in (a_{j+1},\, b_{j+1} + \delta_{j+1})$ for $j = 1, \ldots, N-1$.

But then

$$\begin{aligned}
\mu_0(I) &< F(b) - F(a+\delta) + \epsilon \\
&\leq F(b_N + \delta_N) - F(a_1) + \epsilon \\
&= F(b_N + \delta_N) - F(a_N) + \sum_1^{N-1} \big[F(a_{j+1}) - F(a_j)\big] + \epsilon \\
&\leq F(b_N + \delta_N) - F(a_N) + \sum_1^{N-1} \big[F(b_j + \delta_j) - F(a_j)\big] + \epsilon \\
&< \sum_1^N \big[F(b_j) + \epsilon 2^{-j} - F(a_j)\big] + \epsilon \\
&< \sum_1^\infty \mu(I_j) + 2\epsilon.
\end{aligned}$$

Since ϵ is arbitrary, we are done when a and b are finite. If $a = -\infty$, for any $M < \infty$ the intervals $(a_j\, b_j + \delta_j)$ cover $[-M, b]$, so the same reasoning gives $F(b) - F(-M) \leq \sum_1^\infty \mu_0(I_j) + 2\epsilon$, whereas if $b = \infty$, for any $M < \infty$ we likewise obtain $F(M) - F(a) \leq \sum_1^\infty \mu_0(I_j) + 2\epsilon$. The desired result then follows by letting $\epsilon \to 0$ and $M \to \infty$. ∎

1.16 Theorem. *If $F : \mathbb{R} \to \mathbb{R}$ is any increasing, right continuous function, there is a unique Borel measure μ_F on $\mathbb{R}$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all a, b. If G is another such function, we have $\mu_F = \mu_G$ iff $F - G$ is constant. Conversely, if μ is a Borel measure on $\mathbb{R}$ that is finite on all bounded Borel sets and we define*

$$F(x) = \begin{cases} \mu((0,x]) & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -\mu((-x,0]) & \text{if } x < 0, \end{cases}$$

then F is increasing and right continuous, and $\mu = \mu_F$.

Proof. Each F induces a premeasure on $\mathcal{A}$ by Proposition 1.15. It is clear that F and G induce the same premeasure iff $F - G$ is constant, and that these premeasures are σ-finite (since $\mathbb{R} = \bigcup_{-\infty}^{\infty}(j, j+1]$). The first two assertions therefore follow from Theorem 1.14. As for the last one, the monotonicity of μ implies the monotonicity of F, and the continuity of μ from above and below implies the right continuity of F for $x \geq 0$ and $x < 0$. It is evident that $\mu = \mu_F$ on $\mathcal{A}$, and hence $\mu = \mu_F$ on $\mathcal{B}_{\mathbb{R}}$ by the uniqueness in Theorem 1.14. ∎

Several remarks are in order. First, this theory could equally well be developed by using intervals of the form $[a, b)$ and left continuous functions F. Second, if μ is a finite Borel measure on $\mathbb{R}$, then $\mu = \mu_F$ where $F(x) = \mu((-\infty, x])$ is the cumulative distribution function of μ; this differs from the F specified in Theorem 1.16 by the constant $\mu((-\infty, 0])$. Third, the theory of §1.4 gives, for each increasing and right continuous F, not only the Borel measure μ_F but a complete measure $\overline{\mu}_F$ whose domain includes $\mathcal{B}_{\mathbb{R}}$. In fact, $\overline{\mu}_F$ is just the completion of μ_F (Exercise 22a or Theorem 1.19 below), and one can show that its domain is always strictly larger than $\mathcal{B}_{\mathbb{R}}$. We shall usually denote this complete measure also by μ_F; it is called the **Lebesgue-Stieltjes measure** associated to F.

Lebesgue-Stieltjes measures enjoy some useful regularity properties that we now investigate. In this discussion we fix a complete Lebesgue-Stieltjes measure μ on $\mathbb{R}$ associated to the increasing, right continuous function F, and we denote by $\mathcal{M}_\mu$ the domain of μ. Thus, for any $E \in \mathcal{M}_\mu$,

$$\begin{aligned} \mu(E) &= \inf\Big\{\sum_1^\infty [F(b_j) - F(a_j)] : E \subset \bigcup_1^\infty (a_j, b_j]\Big\} \\ &= \inf\Big\{\sum_1^\infty \mu((a_j, b_j]) : E \subset \bigcup_1^\infty (a_j, b_j]\Big\}. \end{aligned}$$

We first observe that in the second formula for $\mu(E)$ we can replace h-intervals by open h-intervals:

1.17 Lemma. *For any $E \in \mathcal{M}_\mu$,*

$$\mu(E) = \inf\Big\{\sum_1^\infty \mu((a_j, b_j)) : E \subset \bigcup_1^\infty (a_j, b_j)\Big\}.$$

Proof. Let us call the quantity on the right $\nu(E)$. Suppose $E \subset \bigcup_1^\infty (a_j, b_j)$. Each (a_j, b_j) is a countable disjoint union of h-intervals I_j^k $(k = 1, 2, \ldots)$; specifically, $I_j^k = (c_j^k, c_j^{k+1}]$ where $\{c_j\}$ is any sequence such that $c_j^1 = a_j$ and c_j^k increases to b_j as $k \to \infty$. Thus $E \subset \bigcup_{j,k=1}^\infty I_j^k$, so

$$\sum_1^\infty \mu\big((a_j, b_j)\big) = \sum_{j,k=1}^\infty \mu(I_j^k) \geq \mu(E),$$

and hence $\nu(E) \geq \mu(E)$. On the other hand, given $\epsilon > 0$ there exists $\{(a_j, b_j]\}_1^\infty$ with $E \subset \bigcup_1^\infty (a_j, b_j]$ and $\sum_1^\infty \mu((a_j, b_j]) \leq \mu(E) + \epsilon$, and for each j there exists $\delta_j > 0$ such that $F(b_j + \delta_j) - F(b_j) < \epsilon 2^{-j}$. Then $E \subset \bigcup_1^\infty (a_j, b_j + \delta_j)$ and

$$\sum_1^\infty \mu\big((a_j, b_j + \delta_j)\big) \leq \sum_1^\infty \mu\big((a_j, b_j]\big) + \epsilon \leq \mu(E) + 2\epsilon,$$

so that $\nu(E) \leq \mu(E)$. ∎

1.18 Theorem. *If $E \in \mathcal{M}_\mu$, then*

$$\begin{aligned}\mu(E) &= \inf\big\{\mu(U) : U \supset E \text{ and } U \text{ is open}\big\}\\ &= \sup\big\{\mu(K) : K \subset E \text{ and } K \text{ is compact}\big\}.\end{aligned}$$

Proof. By Lemma 1.17, for any $\epsilon > 0$ there exist intervals (a_j, b_j) such that $E \subset \bigcup_1^\infty (a_j, b_j)$ and $\mu(E) \leq \sum_1^\infty \mu((a_j, b_j)) + \epsilon$. If $U = \bigcup_1^\infty (a_j, b_j)$ then U is open, $U \supset E$, and $\mu(U) \leq \mu(E) + \epsilon$. On the other hand, $\mu(U) \geq \mu(E)$ whenever $U \supset E$, so the first equality is valid. For the second one, suppose first that E is bounded. If E is closed, then E is compact and the equality is obvious. Otherwise, given $\epsilon > 0$ we can choose an open $U \supset \overline{E} \setminus E$ such that $\mu(U) \leq \mu(\overline{E} \setminus E) + \epsilon$. Let $K = \overline{E} \setminus U$. Then K is compact, $K \subset E$, and

$$\begin{aligned}\mu(K) &= \mu(E) - \mu(E \cap U) = \mu(E) - [\mu(U) - \mu(U \setminus E)]\\ &\geq \mu(E) - \mu(U) + \mu(\overline{E} \setminus E) \geq \mu(E) - \epsilon.\end{aligned}$$

If E is unbounded, let $E_j = E \cap (j, j+1]$. By the preceding argument, for any $\epsilon > 0$ there exist compact $K_j \subset E_j$ with $\mu(K_j) \geq \mu(E_j) - \epsilon 2^{-j}$. Let $H_n = \bigcup_{-n}^n K_j$. Then H_n is compact, $H_n \subset E$, and $\mu(H_n) \geq \mu(\bigcup_{-n}^n E_j) - \epsilon$. Since $\mu(E) = \lim_{n\to\infty} \mu(\bigcup_{-n}^n E_j)$, the result follows. ∎

1.19 Theorem. *If $E \subset \mathbb{R}$, the following are equivalent.*

a. $E \in \mathcal{M}_\mu$.

b. $E = V \setminus N_1$ *where V is a G_δ set and* $\mu(N_1) = 0$.

c. $E = H \cup N_2$ *where H is an F_σ set and* $\mu(N_2) = 0$.

Proof. Obviously (b) and (c) each imply (a) since μ is complete on $\mathcal{M}_\mu$. Suppose $E \in \mathcal{M}_\mu$ and $\mu(E) < \infty$. By Theorem 1.18, for $j \in \mathbb{N}$ we can choose an open $U_j \supset E$ and a compact $K_j \subset E$ such that

$$\mu(U_j) - 2^{-j} \leq \mu(E) \leq \mu(K_j) + 2^{-j}.$$

Let $V = \bigcap_1^\infty U_j$ and $H = \bigcup_1^\infty K_j$. Then $H \subset E \subset V$ and $\mu(V) = \mu(H) = \mu(E) < \infty$, so $\mu(V \setminus E) = \mu(E \setminus H) = 0$. The result is thus proved when $\mu(E) < \infty$; the extension to the general case is left to the reader (Exercise 25). ∎

The significance of Theorem 1.19 is that all Borel sets (or, more generally, all sets in $\mathcal{M}_\mu$) are of a reasonably simple form modulo sets of measure zero. This contrasts markedly with the machinations necessary to construct the Borel sets from the open sets when null sets are not excepted; see Proposition 1.23 below. Another version of the idea that general measurable sets can be approximated by "simple" sets is contained in the following proposition, whose proof is left to the reader (Exercise 26):

1.20 Proposition. *If $E \in \mathcal{M}_\mu$ and $\mu(E) < \infty$, then for every $\epsilon > 0$ there is a set A that is a finite union of open intervals such that $\mu(E \triangle A) < \epsilon$.*

We now examine the most important measure on $\mathbb{R}$, namely, **Lebesgue measure**: This is the complete measure μ_F associated to the function $F(x) = x$, for which the measure of an interval is simply its length. We shall denote it by m. The domain of m is called the class of **Lebesgue measurable** sets, and we shall denote it by $\mathcal{L}$. We shall also refer to the restriction of m to $\mathcal{B}_\mathbb{R}$ as Lebesgue measure.

Among the most significant properties of Lebesgue measure are its invariance under translations and simple behavior under dilations. If $E \subset \mathbb{R}$ and $s, r \in \mathbb{R}$, we define

$$E + s = \{x + s : x \in E\}, \qquad rE = \{rx : x \in E\}.$$

1.21 Theorem. *If $E \in \mathcal{L}$, then $E + s \in \mathcal{L}$ and $rE \in \mathcal{L}$ for all $s, r \in \mathbb{R}$. Moreover, $m(E + s) = m(E)$ and $m(rE) = |r|m(E)$.*

Proof. Since the collection of open intervals is invariant under translations and dilations, the same is true of $\mathcal{B}_\mathbb{R}$. For $E \in \mathcal{B}_\mathbb{R}$, let $m_s(E) = m(E + s)$ and $m^r(E) = m(rE)$. Then m_s and m^r clearly agree with m and $|r|m$ on finite unions of intervals, hence on $\mathcal{B}_\mathbb{R}$ by Theorem 1.14. In particular, if $E \in \mathcal{B}_\mathbb{R}$ and $m(E) = 0$, then $m(E + s) = m(rE) = 0$, from which it follows that the class of sets of Lebesgue measure zero is preserved by translations and dilations. It follows that $\mathcal{L}$ (the members of which are a union of a Borel set and a Lebesgue null set) is preserved by translation and dilations and that $m(E + s) = m(E)$ and $m(rE) = |r|m(E)$ for all $E \in \mathcal{L}$. ∎

The relation between the measure-theoretic and topological properties of subsets of $\mathbb{R}$ is delicate and contains some surprises. Consider the following facts. Every singleton set in $\mathbb{R}$ has Lebesgue measure zero, and hence so does every countable

set. In particular, $m(\mathbb{Q}) = 0$. Let $\{r_j\}_1^\infty$ be an enumeration of the rational numbers in $[0,1]$, and given $\epsilon > 0$, let I_j be the interval centered at r_j of length $\epsilon 2^{-j}$. Then the set $U = (0,1) \cap \bigcup_1^\infty I_j$ is open and dense in $[0,1]$, but $m(U) \le \sum_1^\infty \epsilon 2^{-j} = \epsilon$; its complement $K = [0,1] \setminus U$ is closed and nowhere dense, but $m(K) \ge 1 - \epsilon$. Thus a set that is open and dense, and hence topologically "large," can be measure-theoretically small, and a set that is nowhere dense, and hence topologically "small," can be measure-theoretically large. (A nonempty open set cannot have Lebesgue measure zero, however.)

The Lebesgue null sets include not only all countable sets but many sets having the cardinality of the continuum. We now present the standard example, the Cantor set, which is also of interest for other reasons.

Each $x \in [0,1]$ has a base-3 decimal expansion $x = \sum_1^\infty a_j 3^{-j}$ where $a_j = 0, 1,$ or 2. This expansion is unique unless x is of the form $p3^{-k}$ for some integers p, k, in which case x has two expansions: one with $a_j = 0$ for $j > k$ and one with $a_j = 2$ for $j > k$. Assuming p is not divisible by 3, one of these expansions will have $a_k = 1$ and the other will have $a_k = 0$ or 2. If we agree always to use the latter expansion, we see that

$$a_1 = 1 \text{ iff } \tfrac{1}{3} < x < \tfrac{2}{3},$$
$$a_1 \neq 1 \text{ and } a_2 = 1 \text{ iff } \tfrac{1}{9} < x < \tfrac{2}{9} \text{ or } \tfrac{7}{9} < x < \tfrac{8}{9},$$

and so forth. It will also be useful to observe that if $x = \sum a_j 3^{-j}$ and $y = \sum b_j 3^{-j}$, then $x < y$ iff there exists an n such that $a_n = b_n$ and $a_j = b_j$ for $j < n$.

The **Cantor set** C is the set of all $x \in [0,1]$ that have a base-3 expansion $x = \sum a_j 3^{-j}$ with $a_j \neq 1$ for all j. Thus C is obtained from [0,1] by removing the open middle third $(\frac{1}{3}, \frac{2}{3})$, then removing the open middle thirds $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ of the two remaining intervals, and so forth. The basic properties of C are summarized as follows:

1.22 Proposition. *Let C be the Cantor set.*

a. C is compact, nowhere dense, and totally disconnected (i.e., the only connected subsets of C are single points). Moreover, C has no isolated points.

b. $m(C) = 0$.

c. $\operatorname{card}(C) = \mathfrak{c}$.

Proof. We leave the proof of (a) to the reader (Exercise 27). As for (b), C is obtained from $[0,1]$ by removing one interval of length $\frac{1}{3}$, two intervals of length $\frac{1}{9}$, and so forth. Thus

$$m(C) = 1 - \sum_0^\infty \frac{2^j}{3^{j+1}} = 1 - \frac{1}{3} \cdot \frac{1}{1 - (2/3)} = 0.$$

Lastly, suppose $x \in C$, so that $x = \sum_0^\infty a_j 3^{-j}$ where $a_j = 0$ or 2 for all j. Let $f(x) = \sum_1^\infty b_j 2^{-j}$ where $b_j = a_j/2$. The series defining $f(x)$ is the base-2 expansion of a number in [0,1], and any number in [0,1] can be obtained in this way. Hence f maps C onto [0,1], and (c) follows. ∎

Let us examine the map f in the preceding proof more closely. One readily sees that if $x, y \in C$ and $x < y$, then $f(x) < f(y)$ unless x and y are the two endpoints of one of the intervals removed from [0,1] to obtain C. In this case $f(x) = p2^{-k}$ for some integers p, k, and $f(x)$ and $f(y)$ are the two base-2 expansions of this number. We can therefore extend f to a map from [0,1] to itself by declaring it to be constant on each interval missing from C. This extended f is still increasing, and since its range is all of [0,1] it cannot have any jump discontinuities; hence it is continuous. f is called the **Cantor function** or **Cantor-Lebesgue function**.

The construction of the Cantor set by starting with [0,1] and successively removing open middle thirds of intervals has an obvious generalization. If I is a bounded interval and $\alpha \in (0,1)$, let us call the open interval with the same midpoint as I and length equal to α times the length of I the "open middle αth" of I. If $\{\alpha_j\}_1^\infty$ is any sequence of numbers in $(0,1)$, then, we can define a decreasing sequence $\{K_j\}$ of closed sets as follows: $K_0 = [0,1]$, and K_j is obtained by removing the open middle α_jth from each of the intervals that make up K_{j-1}. The resulting limiting set $K = \bigcap_1^\infty K_j$ is called a **generalized Cantor set**. Generalized Cantor sets all share with the ordinary Cantor set the properties (a) and (c) in Proposition 1.22. As for their Lebesgue measure, clearly $m(K_j) = (1-\alpha_j)m(K_{j-1})$, so $m(K)$ is the infinite product $\prod_1^\infty (1-\alpha_j) = \lim_{n\to\infty} \prod_1^n (1-\alpha_j)$. If the α_j are all equal to a fixed $\alpha \in (0,1)$ (for example, $\alpha = \frac{1}{3}$ for the ordinary Cantor set), we have $m(K) = 0$. However, if $\alpha_j \to 0$ sufficiently rapidly as $j \to \infty$, $m(K)$ will be positive, and for any $\beta \in (0,1)$ one can choose α_j so that $m(K)$ will equal β; see Exercise 32. This gives another way of constructing nowhere dense sets of positive measure.

Not every Lebesgue measurable set is a Borel set. One can display examples of sets in $\mathcal{L} \setminus \mathcal{B}_\mathbb{R}$ by using the Cantor function; see Exercise 9 in Chapter 2. Alternatively, one can observe that since every subset of the Cantor set is Lebesgue measurable, we have $\text{card}(\mathcal{L}) = \text{card}(\mathcal{P}(\mathbb{R})) > \mathfrak{c}$, whereas $\text{card}(\mathcal{B}_\mathbb{R}) = \mathfrak{c}$. The latter fact follows from Proposition 1.23 below.

Exercises

25. Complete the proof of Theorem 1.19.

26. Prove Proposition 1.20. (Use Theorem 1.18.)

27. Prove Proposition 1.22a. (Show that if $x, y \in C$ and $x < y$, there exists $z \notin C$ such that $x < z < y$.)

28. Let F be increasing and right continuous, and let μ_F be the associated measure. Then $\mu_F(\{a\}) = F(a) - F(a-)$, $\mu_F([a,b)) = F(b-) - F(a-)$, $\mu_F([a,b]) = F(b) - F(a-)$, and $\mu_F((a,b)) = F(b-) - F(a)$.

29. Let E be a Lebesgue measurable set.

a. If $E \subset N$ where N is the nonmeasurable set described in §1.1, then $m(E) = 0$.

b. If $m(E) > 0$, then E contains a nonmeasurable set. (It suffices to assume $E \subset [0,1]$. In the notation of §1.1, $E = \bigcup_{r\in R} E \cap N_r$.)

30. If $E \in \mathcal{L}$ and $m(E) > 0$, for any $\alpha < 1$ there is an open interval I such that $m(E \cap I) > \alpha m(I)$.

31. If $E \in \mathcal{L}$ and $m(E) > 0$, the set $E - E = \{x - y : x, y \in E\}$ contains an interval centered at 0. (If I is as in Exercise 30 with $\alpha > \frac{3}{4}$, then $E - E$ contains $(-\frac{1}{2}m(I), \frac{1}{2}m(I))$.)

32. Suppose $\{\alpha_j\}_1^\infty \subset (0,1)$.
a. $\prod_1^\infty (1 - \alpha_j) > 0$ iff $\sum_1^\infty \alpha_j < \infty$. (Compare $\sum_1^\infty \log(1 - \alpha_j)$ to $\sum \alpha_j$.)
b. Given $\beta \in (0,1)$, exhibit a sequence $\{\alpha_j\}$ such that $\prod_1^\infty (1 - \alpha_j) = \beta$.

33. There exists a Borel set $A \subset [0,1]$ such that $0 < m(A \cap I) < m(I)$ for every subinterval I of $[0,1]$. (Hint: Every subinterval of $[0,1]$ contains Cantor-type sets of positive measure.)

1.6 NOTES AND REFERENCES

The history of measure theory is intimately connected with the history of integration theory, comments on which will be made in §2.7.

§1.1: The Banach-Tarski paradox appeared first in [11], but the following variant goes back to Hausdorff [68]:

> The unit sphere in $\mathbb{R}^3$, $\{x \in \mathbb{R}^3 : |x| = 1\}$, is the disjoint union of four sets $E_1, \ldots, E_4$ such that (a) E_1 is countable and (b) the sets E_2, E_3, E_4, and $E_3 \cup E_4$ are all images of each other under rotations.

An elementary exposition of the Banach-Tarski paradox and Hausdorff's result can be found in Stromberg [146].

§1.2: Our characterization of the σ-algebra $\mathcal{M}(\mathcal{E})$ generated by a family $\mathcal{E} \subset \mathcal{P}(X)$ is nonconstructive, and one might ask how to obtain $\mathcal{M}(\mathcal{E})$ explicitly from $\mathcal{E}$. The answer is rather complicated. One can begin as follows: Let $\mathcal{E}_1 = \mathcal{E} \cup \{E^c : E \in \mathcal{E}\}$, and for $j > 1$ define $\mathcal{E}_j$ to be the collection of all sets that are countable unions of sets in $\mathcal{E}_{j-1}$ or complements of such. Let $\mathcal{E}_\omega = \bigcup_1^\infty \mathcal{E}_j$: is $\mathcal{E}_\omega = \mathcal{M}(\mathcal{E})$? In general, no. $\mathcal{E}_\omega$ is closed under complements, but if $E_j \in \mathcal{E}_j \setminus \mathcal{E}_{j-1}$ for each j, there is no reason for $\bigcup_1^\infty E_j$ to be in $\mathcal{E}_\omega$. So one must start all over again. More precisely, one must define $\mathcal{E}_\alpha$ for every countable ordinal α by transfinite induction: If α has an immediate predecesor β, $\mathcal{E}_\alpha$ is the collection of sets that are countable unions of sets in $\mathcal{E}_\beta$ or complements of such; otherwise, $\mathcal{E}_\alpha = \bigcup_{\beta<\alpha} \mathcal{E}_\beta$. Then:

1.23 Proposition. $\mathcal{M}(E) = \bigcup_{\alpha\in\Omega} \mathcal{E}_\alpha$, *where Ω is the set of countable ordinals.*

Proof. Transfinite induction shows that $\mathcal{E}_\alpha \subset \mathcal{M}(E)$ for all $\alpha \in \Omega$, and hence $\bigcup_{\alpha\in\Omega} \mathcal{E}_\alpha \subset \mathcal{M}(E)$. The reverse inclusion follows from the fact that any sequence in Ω has a supremum in Ω (Proposition 0.19): If $E_j \in \mathcal{E}_{\alpha_j}$ for $j \in \mathbb{N}$ and $\beta = \sup\{\alpha_j\}$, then $E_j \in \mathcal{E}_\alpha$ for all j and hence $\bigcup_1^\infty E_j \in \mathcal{E}_\beta$ where β is the successor of α. ∎

Combining this with Proposition 0.14, we see that if $\operatorname{card}(\mathbb{N}) \leq \operatorname{card}(\mathcal{E}) \leq \mathfrak{c}$, then $\operatorname{card}(\mathcal{M}(\mathcal{E})) = \mathfrak{c}$. (Cf. Exercise 3.)

§1.3: Some authors prefer to take the domains of measures to be σ-rings rather than σ-algebras (see Exercise 1). The reason is that in dealing with "very large" spaces one can avoid certain pathologies by not attempting to measure "very large" sets. However, this point of view also has technical disadvantages, and it is no longer much in favor.

§1.4: Carathéodory's theorem appears in his treatise [22]. Theorem 1.14 has been attributed in the literature to Hahn, Carathéodory, and E. Hopf, but it is originally due to Fréchet [54]. The proof via Carathéodory's theorem was discovered independently by Hahn [60] and Kolmogorov [85].

See König [86] for a deeper study of the problem of constructing measures from more primitive data.

§1.5: Lebesgue originally defined the outer measure $m^*(E)$ of a set $E \subset \mathbb{R}$ in terms of countable coverings by intervals, as we have done. He then defined a bounded set E to be measurable if $m^*(E) + m^*((a,b) \setminus E) = b - a$, where (a,b) is an interval containing E, and an unbounded set to be measurable if its intersection with any bounded interval is measurable. Carathéodory's characterization of measurability, which is technically eaiser to work with, came later. For the equivalence of the two definitions, see Exercise 19.

One should convince oneself that the remarkably fussy proof of Proposition 1.15 is necessary by contemplating the complicated ways in which an h-interval can be decomposed into a disjoint union of h-subintervals. In any such decomposition the collection of right endpoints of the subintervals, when ordered from right to left, is a well ordered set, but it can be order isomorphic to any initial segment of the set of countable ordinals.

Lebesgue measure can be extended to a translation-invariant measure on σ-algebras that properly include $\mathcal{L}$; see Kakutani and Oxtoby [81]. Of course, such σ-algebras can never contain the nonmeasurable set discussed in §1. However, Lebesgue measure can be extended to a translation-invariant finitely additive measure on $\mathcal{P}(\mathbb{R})$, and its 2-dimensional analogue (see §2.6) can be extended to a finitely additive measure on $\mathcal{P}(\mathbb{R}^2)$ that is invariant under translations and rotations; see Banach [8]. The Banach-Tarski paradox prevents this result from being extended to higher dimensions.

In connection with the existence of nonmeasurable sets, Solovay [138] has proved a remarkable theorem which says in effect that it is impossible to prove the existence of Lebesgue nonmeasurable sets without using the axiom of choice. (The precise statement of the theorem involves to technical points of axiomatic set theory, which we shall not discuss here.) From the point of view of the working analyst, the effect of Solovay's theorem is to reaffirm the adequacy of the Lebesgue theory for all practical purposes.

See Rudin [124] for a terse solution of Exercise 33.

2

Integration

In the classical theory of integration on $\mathbb{R}$, $\int_a^b f(x)\,dx$ is defined as a limit of Riemann sums, which are integrals of functions that approximate f and are constant on subintervals of $[a, b]$. Similarly, on any measure space there is an obvious notion of integral for functions that are, in a suitable sense, locally constant, and it can be extended to an integral for more general functions. In this chapter, we develop the theory of integration on abstract measure spaces, paying particular attention to the Lebesgue integral on $\mathbb{R}$ and its generalization to $\mathbb{R}^n$.

2.1 MEASURABLE FUNCTIONS

We begin our study of integration theory with a discussion of measurable mappings, which are the morphisms in the category of measurable spaces.

We recall that any mapping $f : X \to Y$ between two sets induces a mapping $f^{-1} : \mathcal{P}(Y) \to \mathcal{P}(X)$, defined by $f^{-1}(E) = \{x \in X : f(x) \in E\}$, which preserves unions, intersections, and complements. Thus, if $\mathcal{N}$ is a σ-algebra on Y, $\{f^{-1}(E) : E \in \mathcal{N}\}$ is a σ-algebra on X. If $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ are measurable spaces, a mapping $f : X \to Y$ is called $(\mathcal{M}, \mathcal{N})$**-measurable**, or just **measurable** when $\mathcal{M}$ and $\mathcal{N}$ are understood, if $f^{-1}(E) \in \mathcal{M}$ for all $E \in \mathcal{N}$.

It is obvious that the composition of measurable mappings is measurable; that is, if $f : X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable and $g : Y \to Z$ is $(\mathcal{N}, \mathcal{O})$-measurable, then $g \circ f$ is $(\mathcal{M}, \mathcal{O})$-measurable.

2.1 Proposition. *If $\mathcal{N}$ is generated by $\mathcal{E}$, then $f : X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable iff $f^{-1}(E) \in \mathcal{M}$ for all $E \in \mathcal{E}$.*

Proof. The "only if" implication is trivial. For the converse, observe that $\{E \subset Y : f^{-1}(E) \in \mathcal{M}\}$ is a σ-algebra that contains $\mathcal{E}$; it therefore contains $\mathcal{N}$. ∎

2.2 Corollary. *If X and Y are metric (or topological) spaces, every continuous $f : X \to Y$ is $(\mathcal{B}_X, \mathcal{B}_Y)$-measurable.*

Proof. f is continuous iff $f^{-1}(U)$ is open in X for every open $U \subset Y$. ∎

If $(X, \mathcal{M})$ is a measurable space, a real- or complex-valued function f on X will be called **$\mathcal{M}$-measurable**, or just **measurable**, if it is $(\mathcal{M}, \mathcal{B}_{\mathbb{R}})$ or $(\mathcal{M}, \mathcal{B}_{\mathbb{C}})$ measurable. $\mathcal{B}_{\mathbb{R}}$ or $\mathcal{B}_{\mathbb{C}}$ is *always* understood as the σ-algebra on the range space unless otherwise specified. In particular, $f : \mathbb{R} \to \mathbb{C}$ is **Lebesgue** (resp. **Borel**) **measurable** if it is $(\mathcal{L}, \mathcal{B}_{\mathbb{C}})$ (resp. $(\mathcal{B}_{\mathbb{R}}, \mathcal{B}_{\mathbb{C}})$) measurable; likewise for $f : \mathbb{R} \to \mathbb{R}$.

Warning: If $f, g : \mathbb{R} \to \mathbb{R}$ are Lebesgue measurable, it does not follow that $f \circ g$ is Lebesgue measurable, even if g is assumed continuous. (If $E \in \mathcal{B}_{\mathbb{R}}$ we have $f^{-1}(E) \in \mathcal{L}$, but unless $f^{-1}(E) \in \mathcal{B}_{\mathbb{R}}$ there is no guarantee that $g^{-1}(f^{-1}(E))$ will be in $\mathcal{L}$. See Exercise 9.) However, if f is Borel measurable, then $f \circ g$ is Lebesgue or Borel measurable whenever g is.

2.3 Proposition. *If $(X, \mathcal{M})$ is a measurable space and $f : X \to \mathbb{R}$, the following are equivalent:*

a. *f is $\mathcal{M}$-measurable.*
b. *$f^{-1}((a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$.*
c. *$f^{-1}([a, \infty)) \in \mathcal{M}$ for all $a \in \mathbb{R}$.*
d. *$f^{-1}((-\infty, a)) \in \mathcal{M}$ for all $a \in \mathbb{R}$.*
e. *$f^{-1}((-\infty, a]) \in \mathcal{M}$ for all $a \in \mathbb{R}$.*

Proof. This follows from Propositions 1.2 and 2.1. ∎

Sometimes we wish to consider measurability on subsets of X. If $(X, \mathcal{M})$ is a measurable space, f is a function on X, and $E \in \mathcal{M}$, we say that f is **measurable on** E if $f^{-1}(B) \cap E \in \mathcal{M}$ for all Borel sets B. (Equivalently, $f|E$ is $\mathcal{M}_E$-measurable, where $\mathcal{M}_E = \{F \cap E : F \in \mathcal{M}\}$.)

Given a set X, if $\{(Y_\alpha, \mathcal{N}_\alpha)\}_{\alpha \in A}$ is a family of measurable spaces, and $f : X \to Y_\alpha$ is a map for each $\alpha \in A$, there is a unique smallest σ-algebra on X with respect to which the f_α's are all measurable, namely, the σ-algebra generated by the sets $f_\alpha^{-1}(E_\alpha)$ with $E_\alpha \in \mathcal{N}_\alpha$ and $\alpha \in A$. It is called the σ-algebra **generated** by $\{f_\alpha\}_{\alpha \in A}$. In particular, if $X = \prod_{\alpha \in A} Y_\alpha$, we see that the product σ-algebra on X, as defined in §1.2, is the σ-algebra generated by the coordinate maps $\pi_\alpha : X \to Y_\alpha$.

2.4 Proposition. *Let $(X, \mathcal{M})$ and $(Y_\alpha, \mathcal{N}_\alpha)$ $(\alpha \in A)$ be measurable spaces, $Y = \prod_{\alpha \in A} Y_\alpha$, $\mathcal{N} = \bigotimes_{\alpha \in A} \mathcal{N}_\alpha$, and $\pi_\alpha : Y \to Y_\alpha$ the coordinate maps. Then $f : X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable iff $f_\alpha = \pi_\alpha \circ f$ is $(\mathcal{M}, \mathcal{N}_\alpha)$-measurable for all α.*

Proof. If f is measurable, so is each f_α since the composition of measurable maps is measurable. Conversely, if each f_α is measurable, then for all $E_\alpha \in \mathcal{N}_\alpha$, $f^{-1}(\pi_\alpha^{-1}(E_\alpha)) = f_\alpha^{-1}(E_\alpha) \in \mathcal{M}$, whence f is measurable by Proposition 2.1. ∎

2.5 Corollary. *A function $f : X \to \mathbb{C}$ is $\mathcal{M}$-measurable iff* $\operatorname{Re} f$ *and* $\operatorname{Im} f$ *are $\mathcal{M}$-measurable.*

Proof. This follows since $\mathcal{B}_{\mathbb{C}} = \mathcal{B}_{\mathbb{R}^2} = \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}}$ by Proposition 1.5. ∎

It is sometimes convenient to consider functions with values in the extended real number system $\overline{\mathbb{R}} = [\infty, \infty]$. We define Borel sets in $\overline{\mathbb{R}}$ by $\mathcal{B}_{\overline{\mathbb{R}}} = \{E \subset \overline{\mathbb{R}} : E \cap \mathbb{R} \in \mathcal{B}_{\mathbb{R}}\}$. (This coincides with the usual definition of the Borel σ-algebra if we make $\overline{\mathbb{R}}$ into a metric space with metric $\rho(x, y) = |A(x) - A(y)|$, where $A(x) = \arctan x$.) It is easily verified as in Proposition 2.3 that $\mathcal{B}_{\overline{\mathbb{R}}}$ is generated by the rays $(a, \infty]$ or $[-\infty, a)$ $(a \in \mathbb{R})$, and we define $f : X \to \overline{\mathbb{R}}$ to be $\mathcal{M}$-measurable if it is $(\mathcal{M}, \mathcal{B}_{\overline{\mathbb{R}}})$-measurable. See Exercise 1.

We now establish that measurability is preserved under the familiar algebraic and limiting operations.

2.6 Proposition. *If $f, g : X \to \mathbb{C}$ are $\mathcal{M}$-measurable, then so are $f + g$ and fg.*

Proof. Define $F : X \to \mathbb{C} \times \mathbb{C}$, $\phi : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$, and $\psi : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$ by $F(x) = (f(x), g(x))$, $\phi(z, w) = z + w$, $\psi(z, w) = zw$. Since $\mathcal{B}_{\mathbb{C}\times\mathbb{C}} = \mathcal{B}_{\mathbb{C}} \otimes \mathcal{B}_{\mathbb{C}}$ by Proposition 1.5, F is $(\mathcal{M}, \mathcal{B}_{\mathbb{C}\times\mathbb{C}})$-measurable by Proposition 2.4, whereas ϕ and ψ are $(\mathbb{C}_{\mathbb{C}\times\mathbb{C}}, \mathcal{B}_{\mathbb{C}})$-measurable by Corollary 2.2. Thus $f + g = \phi \circ F$ and $fg = \psi \circ F$ are $\mathcal{M}$-measurable. ∎

Proposition 2.6 remains valid for $\overline{\mathbb{R}}$-valued functions provided one takes a little care with the indeterminate expressions $\infty - \infty$ and $0 \cdot \infty$. (Recall, however, that by convention we always define $0 \cdot \infty$ to be 0.) See Exercise 2.

2.7 Proposition. *If $\{f_j\}$ is a sequence of $\overline{\mathbb{R}}$-valued measurable functions on $(X, \mathcal{M})$, then the functions*

$$g_1(x) = \sup_j f_j(x), \qquad g_3(x) = \limsup_{j\to\infty} f_j(x),$$
$$g_2(x) = \inf_j f_j(x), \qquad g_4(x) = \liminf_{j\to\infty} f_j(x)$$

are all measurable. If $f(x) = \lim_{j\to\infty} f(x)$ exists for every $x \in X$, then f is measurable.

Proof. We have

$$g_1^{-1}((a, \infty]) = \bigcup_1^\infty f_j^{-1}((a, \infty]), \qquad g_2^{-1}([-\infty, a)) = \bigcup_1^\infty f_j^{-1}([-\infty, a)),$$

so g_1 and g_2 are measurable by Proposition 2.3. More generally, if $h_k(x) = \sup_{j>k} f_j(x)$ then h_k is measurable for each k, so $g_3 = \inf_k h_k$ is measurable, and likewise for g_4. Finally, if f exists then $f = g_3 = g_4$, so f is measurable. ∎

2.8 Corollary. *If $f, g : X \to \overline{\mathbb{R}}$ are measurable, then so are* $\max(f, g)$ *and* $\min(f, g)$.

2.9 Corollary. *If $\{f_j\}$ is a sequence of complex-valued measurable functions and $f(x) = \lim_{j\to\infty} f_j(x)$ exists for all x, then f is measurable.*

Proof. Apply Corollary 2.5. ∎

For future reference we present two useful decompositions of functions. First, if $f : X \to \overline{\mathbb{R}}$, we define the **positive** and **negative parts** of f to be

$$f^+(x) = \max\big(f(x), 0\big), \qquad f^-(x) = \max\big(-f(x), 0\big).$$

Then $f = f^+ - f^-$. If f is measurable, so are f^+ and f^-, by Corollary 2.8. Second, if $f : X \to \mathbb{C}$, we have its **polar decomposition**:

$$f = (\operatorname{sgn} f)|f|, \quad \text{where} \quad \operatorname{sgn} z = \begin{cases} z/|z| & \text{if } z \neq 0, \\ 0 & \text{if } z = 0. \end{cases}$$

Again, if f is measurable, so are $|f|$ and $\operatorname{sgn} f$. Indeed, $z \mapsto |z|$ is continuous on $\mathbb{C}$, and $z \mapsto \operatorname{sgn} z$ is continuous except at the origin. If $U \subset \mathbb{C}$ is open, $\operatorname{sgn}^{-1}(U)$ is either open or of the form $V \cup \{0\}$ where V is open, so sgn is Borel measurable. Therefore $|f| = |\cdot| \circ f$ and $\operatorname{sgn} f = \operatorname{sgn} \circ f$ are measurable.

We now discuss the functions that are the building blocks for the theory of integration. Suppose that $(X, \mathcal{M})$ is a measurable space. If $E \subset X$, the **characteristic function** χ_E of E (sometimes called the **indicator function** of E and denoted by 1_E) is defined by

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

It is easily checked that χ_E is measurable iff $E \in \mathcal{M}$. A **simple function** on X is a finite linear combination, with complex coefficients, of characteristic functions of sets in $\mathcal{M}$. (We do not allow simple functions to assume the values $\pm\infty$.) Equivalently, $f : X \to \mathbb{C}$ is simple iff f is measurable and the range of f is a finite subset of $\mathbb{C}$. Indeed, we have

$$f = \sum_1^n z_j \chi_{E_j}, \text{ where } E_j = f^{-1}(\{z_j\}) \text{ and } \operatorname{range}(f) = \{z_1, \ldots, z_n\}.$$

We call this the **standard representation** of f. It exhibits f as a linear combination, with distinct coefficients, of characteristic functions of disjoint sets whose union is X. *Note:* One of the coefficients z_j may well be 0, but the term $z_j\chi_{E_j}$ is still to be envisioned as part of the standard representation, as the set E_j may have a role to play when f interacts with other functions.

It is clear that if f and g are simple functions, then so are $f + g$ and fg. We now show that arbitrary measurable functions can be approximated in a nice way by simple functions.

2.10 Theorem. *Let $(X, \mathcal{M})$ be a measurable space.*

a. If $f : X \to [0, \infty]$ is measurable, there is a sequence $\{\phi_n\}$ of simple functions such that $0 \le \phi_1 \le \phi_2 \le \cdots \le f$, $\phi_n \to f$ pointwise, and $\phi_n \to f$ uniformly on any set on which f is bounded.

b. If $f : X \to \mathbb{C}$ is measurable, there is a sequence $\{\phi_n\}$ of simple functions such that $0 \le |\phi_1| \le |\phi_2| \le \cdots \le |f|$, $\phi_n \to f$ pointwise, and $\phi_n \to f$ uniformly on any set on which f is bounded.

Proof. (a) For $n = 0, 1, 2, \ldots$ and $0 \le k \le 2^{2n} - 1$, let

$$E_n^k = f^{-1}\big((k2^{-n}, (k+1)2^{-n}]\big) \quad \text{and} \quad F_n = f^{-1}\big((2^n, \infty]\big),$$

and define

$$\phi_n = \sum_{k=0}^{2^{2n}-1} k2^{-n}\chi_{E_n^k} + 2^n\chi_{F_n}.$$

(This formula is messy in print but easily understood graphically; see Figure 2.1.) It is easily checked that $\phi_n \le \phi_{n+1}$ for all n, and $0 \le f - \phi_n \le 2^{-n}$ on the set where $f \le 2^n$. The result therefore follows.

(b) If $f = g + ih$, we can apply part (a) to the positive and negative parts of g and h, obtaining sequences ψ_n^+, ψ_n^-, ζ_n^+, ζ_n^- of nonnegative simple functions that increase to g^+, g^-, h^+, h^-. Let $\phi_n = \psi_n^+ - \psi_n^- + i(\zeta_n^+ - \zeta_n^-)$; it is then a simple exercise to verify that ϕ_n has the desired properties. ∎

If μ is a measure on $(X, \mathcal{M})$, we may wish to except μ-null sets from consideration in studying measurable functions. In this respect, life is a bit simpler if μ is complete.

2.11 Proposition. *The following implications are valid iff the measure μ is complete:*

a. If f is measurable and $f = g$ μ-a.e., then g is measurable.

b. If f_n is measurable for $n \in \mathbb{N}$ and $f_n \to f$ μ-a.e., then f is measurable.

The proof is left to the reader (Exercise 10).

On the other hand, the following result shows that one is unlikely to commit any serious blunders by forgetting to worry about completeness of the measure.

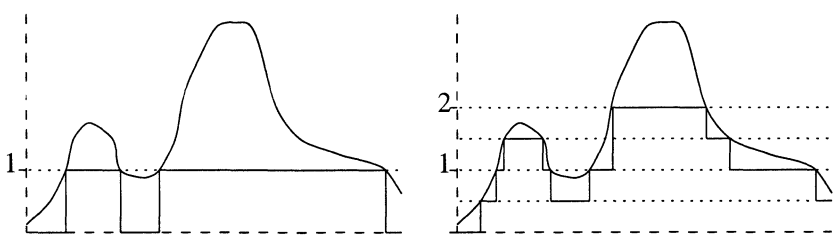

Fig. 2.1 The functions ϕ_0 (left) and ϕ_1 (right) in the proof of Theorem 2.10a.

2.12 Proposition. *Let $(X, \mathcal{M}, \mu)$ be a measure space and let $(X, \overline{\mathcal{M}}, \overline{\mu})$ be its completion. If f is an $\overline{\mathcal{M}}$-measurable function on X, there is an $\mathcal{M}$-measurable function g such that $f = g$ $\overline{\mu}$-almost everywhere.*

Proof. This is obvious from the definition of $\overline{\mu}$ if $f = \chi_E$ where $E \in \overline{\mathcal{M}}$, and hence if f is an $\overline{\mathcal{M}}$-measurable simple function. For the general case, choose a sequence $\{\phi_n\}$ of $\overline{\mathcal{M}}$-measurable simple functions that converge pointwise to f according to Theorem 2.10, and for each n let ψ_n be an $\mathcal{M}$-measurable simple function with $\psi_n = \phi_n$ except on a set $E_n \in \overline{\mathcal{M}}$ with $\overline{\mu}(E_n) = 0$. Choose $N \in \mathcal{M}$ such that $\mu(N) = 0$ and $N \supset \bigcup_1^\infty E_n$, and set $g = \lim \chi_{X \setminus N} \psi_n$. Then g is $\mathcal{M}$-measurable by Corollary 2.9, and $g = f$ on N^c. ∎

Exercises

In Exercises 1–7, $(X, \mathcal{M})$ is a measurable space.

1. Let $f : X \to \overline{\mathbb{R}}$ and $Y = f^{-1}(\mathbb{R})$. Then f is measurable iff $f^{-1}(\{-\infty\}) \in \mathcal{M}$, $f^{-1}(\{\infty\}) \in \mathcal{M}$, and f is measurable on Y.

2. Suppose $f, g : X \to \overline{\mathbb{R}}$ are measurable.
a. fg is measurable (where $0 \cdot (\pm\infty) = 0$).
b. Fix $a \in \overline{\mathbb{R}}$ and define $h(x) = a$ if $f(x) = -g(x) = \pm\infty$ and $h(x) = f(x) + g(x)$ otherwise. Then h is measurable.

3. If $\{f_n\}$ is a sequence of measurable functions on X, then $\{x : \lim f_n(x) \text{ exists}\}$ is a measurable set.

4. If $f : X \to \overline{\mathbb{R}}$ and $f^{-1}((r, \infty]) \in \mathcal{M}$ for each $r \in \mathbb{Q}$, then f is measurable.

5. If $X = A \cup B$ where $A, B \in \mathcal{M}$, a function f on X is measurable iff f is measurable on A and on B.

6. The supremum of an uncountable family of measurable $\overline{\mathbb{R}}$-valued functions on X can fail to be measurable (unless the σ-algebra $\mathcal{M}$ is very special).

7. Suppose that for each $\alpha \in \mathbb{R}$ we are given a set $E_\alpha \in \mathcal{M}$ such that $E_\alpha \subset E_\beta$ whenever $\alpha < \beta$, $\bigcup_{\alpha \in \mathbb{R}} E_\alpha = X$, and $\bigcap_{\alpha \in \mathbb{R}} E_\alpha = \varnothing$. Then there is a measurable function $f : X \to \mathbb{R}$ such that $f(x) \leq \alpha$ on E_α and $f(x) \geq \alpha$ on E_α^c for every α. (Use Exercise 4.)

8. If $f : \mathbb{R} \to \mathbb{R}$ is monotone, then f is Borel measurable.

9. Let $f : [0, 1] \to [0, 1]$ be the Cantor function (§1.5), and let $g(x) = f(x) + x$.
a. g is a bijection from $[0, 1]$ to $[0, 2]$, and $h = g^{-1}$ is continuous from [0,2] to [0,1].
b. If C is the Cantor set, $m(g(C)) = 1$.
c. By Exercise 29 of Chapter 1, $g(C)$ contains a Lebesgue nonmeasurable set A. Let $B = g^{-1}(A)$. Then B is Lebesgue measurable but not Borel.

d. There exist a Lebesgue measurable function F and a continuous function G on $\mathbb{R}$ such that $F \circ G$ is not Lebesgue measurable.

10. Prove Proposition 2.11.

11. Suppose that f is a function on $\mathbb{R} \times \mathbb{R}^k$ such that $f(x, \cdot)$ is Borel measurable for each $x \in \mathbb{R}$ and $f(\cdot, y)$ is continuous for each $y \in \mathbb{R}^k$. For $n \in \mathbb{N}$, define f_n as follows. For $i \in \mathbb{Z}$ let $a_i = i/n$, and for $a_i \leq x \leq a_{i+1}$ let

$$f_n(x, y) = \frac{f(a_{i+1}, y)(x - a_i) - f(a_i, y)(x - a_{i+1})}{a_{i+1} - a_i}.$$

Then f_n is Borel measurable on $\mathbb{R} \times \mathbb{R}^k$ and $f_n \to f$ pointwise; hence f is Borel measurable on $\mathbb{R} \times \mathbb{R}^k$. Conclude by induction that every function on $\mathbb{R}^n$ that is continuous in each variable separately is Borel measurable.

2.2 INTEGRATION OF NONNEGATIVE FUNCTIONS

In this section we fix a measure space $(X, \mathcal{M}, \mu)$, and we define

$$L^+ = \text{the space of all measurable functions from } X \text{ to } [0, \infty].$$

If ϕ is a simple function in L^+ with standard representation $\phi = \sum_1^n a_j \chi_{E_j}$, we define the **integral** of ϕ with respect to μ by

$$\int \phi \, d\mu = \sum_1^n a_j \mu(E_j)$$

(with the convention, as always, that $0 \cdot \infty = 0$). We note that $\int \phi \, d\mu$ may equal ∞. When there is no danger of confusion, we shall also write $\int \phi$ for $\int \phi \, d\mu$. Also, it is sometimes convenient to display the argument of ϕ explicitly, especially when $\phi(x)$ is given by a formula in terms of x or when there are other variables involved; in this case we shall use the notation $\int \phi(x) \, d\mu(x)$. (Some authors prefer to write $\int \phi(x) \, \mu(dx)$ instead.) Finally, if $A \in \mathcal{M}$, then $\phi\chi_A$ is also simple (viz., $\phi\chi_A = \sum a_j \chi_{A \cap E_j}$), and we define $\int_A \phi \, d\mu$ (or $\int_A \phi$ or $\int_A \phi(x) \, d\mu(x)$) to be $\int \phi\chi_A \, d\mu$. The same notational conventions will also apply to the inegrals of more general functions to be defined below. To summarize:

$$\int_A \phi \, d\mu = \int_A \phi = \int_A \phi(x) \, d\mu(x) = \int \phi\chi_A \, d\mu, \qquad \int = \int_X.$$

2.13 Proposition. *Let ϕ and ψ be simple functions in L^+.*

a. If $c \geq 0$, $\int c\phi = c \int \phi$.

b. $\int (\phi + \psi) = \int \phi + \int \psi$.

c. If $\phi \leq \psi$, then $\int \phi \leq \int \psi$.

d. The map $A \mapsto \int_A d\mu$ is a measure on $\mathcal{M}$.

Proof. (a) is trivial. For (b), let $\sum_1^n a_j \chi_{E_j}$ and $\sum_1^m b_k \chi_{F_k}$ be the standard representations of ϕ and ψ. Then $E_j = \bigcup_{k=1}^m (E_j \cap F_k)$ and $F_k = \bigcup_1^n (E_j \cap F_k)$ since $\bigcup_1^n E_j = \bigcup_1^n F_k = X$, and these unions are disjoint. Hence the finite additivity of μ implies that

$$\int \phi + \int \psi = \sum_{j,k} (a_j + b_k)\mu(E_j \cap F_k),$$

and the same reasoning show that the sum on the right equals $\int(\phi + \psi)$. Moreover, if $\phi \le \psi$, then $a_j \le b_k$ whenever $E_j \cap F_k \neq \varnothing$, so

$$\int \phi = \sum_{j,k} a_j \mu(E_j \cap F_k) \le \sum_{j,k} b_k \mu(E_j \cap F_k) = \int \psi,$$

which proves (c). Finally, if $\{A_k\}$ is a disjoint sequence in $\mathcal{M}$ and $A = \bigcup_1^\infty A_k$,

$$\int_A \phi = \sum_j a_j \mu(A \cap E_j) = \sum_{j,k} a_j \mu(A_k \cap E_j) = \sum_k \int_{A_k} \phi,$$

which establishes (d). ∎

We now extend the integral to all functions $f \in L^+$ by defining

$$\int f\, d\mu = \sup\left\{ \int \phi\, d\mu : 0 \le \phi \le f,\ \phi \text{ simple} \right\}.$$

By Proposition 2.13c, the two definitions of $\int f$ agree when f is simple, as the family of simple functions over which the supremum is taken includes f itself. Moreover, it is obvious from the definition that

$$\int f \le \int g \text{ whenever } f \le g, \text{ and } \int cf = c \int f \text{ for all } c \in [0, \infty).$$

The next step is to establish one of the fundamental convergence theorems.

2.14 The Monotone Convergence Theorem. *If $\{f_n\}$ is a sequence in L^+ such that $f_j \le f_{j+1}$ for all j, and $f = \lim_{n\to\infty} f_n$ $(= \sup_n f_n)$, then $\int f = \lim_{n\to\infty} \int f_n$.*

Proof. $\{\int f_n\}$ is an increasing sequence of numbers, so its limit exists (possibly equal to ∞). Moreover, $\int f_n \le \int f$ for all n, so $\lim \int f_n \le \int f$. To establish the reverse inequality, fix $\alpha \in (0, 1)$, let ϕ be a simple function with $0 \le \phi \le f$, and let $E_n = \{x : f_n(x) \ge \alpha\phi(x)\}$. Then $\{E_n\}$ is an increasing sequence of measurable sets whose union is X, and we have $\int f_n \ge \int_{E_n} f_n \ge \alpha \int_{E_n} \phi$. By Proposition 2.13d and Theorem 1.8c, $\lim \int_{E_n} \phi = \int \phi$, and hence $\lim \int f_n \ge \alpha \int \phi$. Since this is true for all $\alpha < 1$, it remains true for $\alpha = 1$, and taking the supremum over all simple $\phi \le f$, we obtain $\lim \int f_n \ge \int f$. ∎

The monotone convergence theorem is an essential tool in many situations, but its immediate significance for us is as follows. The definition of $\int f$ involves the supremum over a huge (usually uncountable) family of simple functions, so it may be difficult to evaluate $\int f$ directly from the definition. The monotone convergence theorem, however, assures us that to compute $\int f$ it is enough to compute $\lim \int \phi_n$ where $\{\phi_n\}$ is any sequence of simple functions that increase to f, and Theorem 2.10 guarantees that such sequences exist. As a first application, we establish the additivity of the integral.

2.15 Theorem. *If $\{f_n\}$ is a finite or infinite sequence in L^+ and $f = \sum_n f_n$, then $\int f = \sum_n \int f_n$.*

Proof. First consider two functions f_1 and f_2. By Theorem 2.10 we can find sequences $\{\phi_j\}$ and $\{\psi_j\}$ of nonnegative simple functions that increase to f_1 and f_2. Then $\{\phi_j + \psi_j\}$ increases to $f_1 + f_2$, so by the monotone convergence theorem and Theorem 2.13b,

$$\int (f_1 + f_2) = \lim \int (\phi_j + \psi_j) = \lim \int \phi_j + \lim \int \psi_j = \int f_1 + \int f_2.$$

Hence, by induction, $\int \sum_1^N f_n = \sum_1^N \int f_n$ for any finite N. Letting $N \to \infty$ and applying the monotone convergence theorem again, we obtain $\int \sum_1^\infty f_n = \sum_1^\infty \int f_n$. ∎

2.16 Proposition. *If $f \in L^+$, then $\int f = 0$ iff $f = 0$ a.e.*

Proof. This is obvious if f is simple: if $f = \sum_1^n a_j \chi_{E_j}$ with $a_j \geq 0$, then $\int f = 0$ iff for each j either $a_j = 0$ or $\mu(E_j) = 0$. In general, if $f = 0$ a.e. and ϕ is simple with $0 \leq \phi \leq f$, then $\phi = 0$ a.e., hence $\int f = \sup_{\phi \leq f} \int \phi = 0$. On the other hand, $\{x : f(x) > 0\} = \bigcup_1^\infty E_n$ where $E_n = \{x : f(x) > n^{-1}\}$, so if it is false that $f = 0$ a.e., we must have $\mu(E_n) > 0$ for some n. But then $f > n^{-1}\chi_{E_n}$, so $\int f \geq n^{-1}\mu(E_n) > 0$. ∎

2.17 Corollary. *If $\{f_n\} \subset L^+$, $f \in L^+$, and $f_n(x)$ increases to $f(x)$ for a.e. x, then $\int f = \lim \int f_n$.*

Proof. If $f_n(x)$ increases to $f(x)$ for $x \in E$ where $\mu(E^c) = 0$, then $f - f\chi_E = 0$ a.e. and $f_n - f_n\chi_E = 0$ a.e., so by the monotone convergence theorem, $\int f = \int f\chi_E = \lim \int f_n\chi_E = \lim \int f_n$. ∎

The hypothesis that the sequence $\{f_n\}$ be increasing, at least a.e., is essential for the monotone convergence theorem. For example, if X is $\mathbb{R}$ and μ is Lebesgue measure, we have $\chi_{(n,n+1)} \to 0$ and $n\chi_{(0,1/n)} \to 0$ pointwise, but $\int \chi_{(n,n+1)} = \int n\chi_{(0,1/n)} = 1$ for all n. As one sees by sketching the graphs, the trouble in these examples is that the area under the graph "escapes to infinity" as $n \to \infty$, so the area in the limit is less than one would expect. This is typical of the cases when the

integral of the limit is not the limit of the integrals, but in this situation there is still an inequality that remains valid. We deduce it from the following general result.

2.18 Fatou's Lemma. *If $\{f_n\}$ is any sequence in L^+, then*

$$\int (\liminf f_n) \leq \liminf \int f_n.$$

Proof. For each $k \geq 1$ we have $\inf_{n\geq k} f_n \leq f_j$ for $j \geq k$, hence $\int \inf_{n\geq k} f_n \leq \int f_j$ for $j \geq k$, hence $\int \inf_{n\geq k} f_n \leq \inf_{j\geq k} \int f_j$. Now let $k \to \infty$ and apply the monotone convergence theorem:

$$\int (\liminf f_n) = \lim_{k\to\infty} \int \left(\inf_{n\geq k} f_n\right) \leq \liminf \int f_n.$$

∎

2.19 Corollary. *If $\{f_n\} \subset L^+$, $f \in L^+$, and $f_n \to f$ a.e., then $\int f \leq \liminf \int f_n$.*

Proof. If $f_n \to f$ everywhere, the result is immediate from Fatou's lemma, and this can be achieved by modifying f_n and f on a null set without affecting the integrals, by Proposition 2.16. ∎

2.20 Proposition. *If $f \in L^+$ and $\int f < \infty$, then $\{x : f(x) = \infty\}$ is a null set and $\{x : f(x) > 0\}$ is σ-finite.*

The proof is left to the reader (Exercise 12).

Exercises

12. Prove Proposition 2.20. (See Proposition 0.20, where a special case is proved.)

13. Suppose $\{f_n\} \subset L^+$, $f_n \to f$ pointwise, and $\int f = \lim \int f_n < \infty$. Then $\int_E f = \lim \int_E f_n$ for all $E \in \mathcal{M}$. However, this need not be true if $\int f = \lim \int f_n = \infty$.

14. If $f \in L^+$, let $\lambda(E) = \int_E f\, d\mu$ for $E \in \mathcal{M}$. Then λ is a measure on $\mathcal{M}$, and for any $g \in L^+$, $\int g\, d\lambda = \int fg\, d\mu$. (First suppose that g is simple.)

15. If $\{f_n\} \subset L^+$, f_n decreases pointwise to f, and $\int f_1 < \infty$, then $\int f = \lim \int f_n$.

16. If $f \in L^+$ and $\int f < \infty$, for every $\epsilon > 0$ there exists $E \in \mathcal{M}$ such that $\mu(E) < \infty$ and $\int_E f > (\int f) - \epsilon$.

17. Assume Fatou's lemma and deduce the monotone convergence theorem from it.

2.3 INTEGRATION OF COMPLEX FUNCTIONS

We continue to work on a fixed measure space $(X, \mathcal{M}, \mu)$. The integral defined in the previous section can be extended to real-valued measurable functions f in an obvious

way; namely, if f^+ and f^- are the positive and negative parts of f and at least one of $\int f^+$ and $\int f^-$ is finite, we define

$$\int f = \int f^+ - \int f^-.$$

We shall be mainly concerned with the case where $\int f^+$ and $\int f^-$ are both finite; we then say that f is **integrable**. Since $|f| = f^+ + f^-$, it is clear that f is integrable iff $\int |f| < \infty$.

2.21 Proposition. *The set of integrable real-valued functions on X is a real vector space, and the integral is a linear functional on it.*

Proof. The first assertion follows from the fact that $|af + bg| \leq |a||f| + |b||g|$, and it is easy to check that $\int af = a \int f$ for any $a \in \mathbb{R}$. To show additivity, suppose that f and g are integrable and let $h = f + g$. Then $h^+ - h^- = f^+ - f^- + g^+ - g^-$, so $h^+ + f^- + g^- = h^- + f^+ + g^+$. By Theorem 2.15,

$$\int h^+ + \int f^- + \int g^- = \int h^- + \int f^+ + \int g^+,$$

and regrouping then yields the desired result:

$$\int h = \int h^+ - \int h^- = \int f^+ - \int f^- + \int g^+ - \int g^- = \int f + \int g.$$

∎

Next, if f is a complex-valued measurable function, we say that f is **integrable** if $\int |f| < \infty$. More generally, if $E \in \mathcal{M}$, f is **integrable on** E if $\int_E |f| < \infty$. Since $|f| \leq |\operatorname{Re} f| + |\operatorname{Im} f| \leq 2|f|$, f is integrable iff $\operatorname{Re} f$ and $\operatorname{Im} f$ are both integrable, and in this case we define

$$\int f = \int \operatorname{Re} f + i \int \operatorname{Im} f.$$

It follows easily that the space of complex-valued integrable functions is a complex vector space and that the integral is a complex-linear functional on it. We denote this space — provisionally — by $L^1(\mu)$ (or $L^1(X, \mu)$, or $L^1(X)$, or simply L^1, depending on the context). The superscript 1 is standard notation, but it will not assume any significance for us until Chapter 6.

2.22 Proposition. *If $f \in L^1$, then $|\int f| \leq \int |f|$.*

Proof. This is trivial if $\int f = 0$ and almost trivial if f is real, since

$$\left|\int f\right| = \left|\int f^+ - \int f^-\right| \leq \int f^+ + \int f^- = \int |f|.$$

If f is complex-valued and $\int f \neq 0$, let $\alpha = \overline{\operatorname{sgn}(\int f)}$. Then $|\int f| = \alpha \int f = \int \alpha f$. In particular, $\int \alpha f$ is real, so

$$\left|\int f\right| = \operatorname{Re} \int \alpha f = \int \operatorname{Re}(\alpha f) \leq \int |\operatorname{Re}(\alpha f)| \leq \int |\alpha f| = \int |f|.$$

∎

2.23 Proposition.

a. If $f \in L^1$, then $\{x : f(x) \neq 0\}$ is σ-finite.

b. If $f, g \in L^1$, then $\int_E f = \int_E g$ for all $E \in \mathcal{M}$ iff $\int |f - g| = 0$ iff $f = g$ a.e.

Proof. (a) and the second equivalence in (b) follow from Propositions 2.20 and 2.16. If $\int |f - g| = 0$, then by Proposition 2.22, for any $E \in \mathcal{M}$,

$$\left| \int_E f - \int_E g \right| \leq \int \chi_E |f - g| \leq \int |f - g| = 0,$$

so that $\int_E f = \int_E g$. On the other hand, if $u = \mathrm{Re}(f - g)$, $v = \mathrm{Im}(f - g)$, and it is false that $f = g$ a.e., then at least one of u^+, u^-, v^+, and v^- must be nonzero on a set of positive measure. If, say, $E = \{x : u^+(x) > 0\}$ has positive measure, then $\mathrm{Re}(\int_E f - \int_E g) = \int_E u^+ > 0$ since $u^- = 0$ on E; likewise in the other cases. ∎

This proposition shows that for the purposes of integration it makes no difference if we alter functions on null sets. Indeed, one can integrate functions f that are only *defined* on a measurable set E whose complement is null simply by defining f to be zero (or anything else) on E^c. In this fashion we can treat $\overline{\mathbb{R}}$-valued functions that are finite a.e. as real-valued functions for the purposes of integration.

With this in mind, we shall find it more convenient to redefine $L^1(\mu)$ to be the set of equivalence classes of a.e.-defined integrable functions on X, where f and g are considered equivalent iff $f = g$ a.e. This new $L^1(\mu)$ is still a complex vector space (under pointwise a.e. addition and scalar multiplication). Although we shall henceforth view $L^1(\mu)$ as a space of equivalence classes, we shall still employ the notation "$f \in L^1(\mu)$" to mean that f is an a.e.-defined integrable function. This minor abuse of notation is commonly accepted and rarely causes any confusion.

The new definition of $L^1(\mu)$ has two further advantages. First, if $\overline{\mu}$ is the completion of μ, Proposition 2.12 yields a natural one-to-one correspondence between $L^1(\overline{\mu})$ and $L^1(\mu)$, so we can (and shall) identify these spaces. Second, L^1 is a metric space with distance function $\rho(f, g) = \int |f - g|$. (The triangle inequality is easily verified, and obviously $\rho(f, g) = \rho(g, f)$; but to obtain the condition that $\rho(f, g) = 0$ only when $f = g$, one must identify functions that are equal a.e., according to Proposition 2.23b.) We shall refer to convergence with respect to this metric as **convergence in** L^1; thus $f_n \to f$ in L^1 iff $\int |f_n - f| \to 0$.

We now present the last of the three basic convergence theorems (the other two being the monotone convergence theorem and Fatou's lemma) and derive some useful consequences from it. In the context of integration on $\mathbb{R}$ with Lebesgue measure as in the discussion preceding Fatou's lemma, the idea behind this theorem is that if $f_n \to f$ a.e. and the graph of $|f_n|$ is confined to a region of the plane with finite area so that the area beneath it cannot escape to infinity, then $\int f_n \to \int f$.

2.24 The Dominated Convergence Theorem. *Let $\{f_n\}$ be a sequence in L^1 such that (a) $f_n \to f$ a.e., and (b) there exists a nonnegative $g \in L^1$ such that $|f_n| \leq g$ a.e. for all n. Then $f \in L^1$ and $\int f = \lim_{n\to\infty} \int f_n$.*

Proof. f is measurable (perhaps after redefinition on a null set) by Propositions 2.11 and 2.12, and since $|f| \leq g$ a.e., we have $f \in L^1$. By taking real and imaginary parts it suffices to assume that f_n and f are real-valued, in which case we have $g + f_n \geq 0$ a.e. and $g - f_n \geq 0$ a.e. Thus by Fatou's lemma,

$$\int g + \int f \leq \liminf \int (g + f_n) = \int g + \liminf \int f_n,$$
$$\int g - \int f \leq \liminf \int (g - f_n) = \int g - \limsup \int f_n.$$

Therefore, $\liminf \int f_n \geq \int f \geq \limsup \int f_n$, and the result follows. ∎

2.25 Theorem. *Suppose that $\{f_j\}$ is a sequence in L^1 such that $\sum_1^\infty \int |f_j| < \infty$. Then $\sum_1^\infty f_j$ converges a.e. to a function in L^1, and $\int \sum_1^\infty f_j = \sum_1^\infty \int f_j$.*

Proof. By Theorem 2.15, $\int \sum_1^\infty |f_j| = \sum_1^\infty \int |f_j| < \infty$, so the function $g = \sum_1^\infty |f_j|$ is in L^1. In particular, by Proposition 2.20, $\sum_1^\infty |f_j(x)|$ is finite for a.e. x, and for each such x the series $\sum_1^\infty f_j(x)$ converges. Moreover, $|\sum_1^n f_j| \leq g$ for all n, so we can apply the dominated convergence theorem to the sequence of partial sums to obtain $\int \sum_1^\infty f_j = \sum_1^\infty \int f_j$. ∎

2.26 Theorem. *If $f \in L^1(\mu)$ and $\epsilon > 0$, there is an integrable simple function $\phi = \sum a_j \chi_{E_j}$ such that $\int |f - \phi| \, d\mu < \epsilon$. (That is, the integrable simple functions are dense in L^1 in the L^1 metric.) If μ is a Lebesgue-Stieltjes measure on $\mathbb{R}$, the sets E_j in the definition of ϕ can be taken to be finite unions of open intervals; moreover, there is a continuous function g that vanishes outside a bounded interval such that $\int |f - g| \, d\mu < \epsilon$.*

Proof. Let $\{\phi_n\}$ be as in Theorem 2.10b; then $\int |\phi_n - f| < \epsilon$ for n sufficiently large by the dominated convergence theorem, since $|\phi_n - f| \leq 2|f|$. If $\phi_n = \sum a_j \chi_{E_j}$ where the E_j are disjoint and the a_j are nonzero, we observe that $\mu(E_j) = |a_j|^{-1} \int_{E_j} |\phi_n| \leq |a_j|^{-1} \int |f| < \infty$. Moreover, if E and F are measurable sets, we have $\mu(E \triangle F) = \int |\chi_E - \chi_F|$. Thus if μ is a Lebesgue-Stieltjes measure on $\mathbb{R}$, by Proposition 1.20 we can approximate χ_{E_j} arbitrarily closely in the L^1 metric by finite sums of functions χ_{I_k} where the I_k's are open intervals. Finally, if $I_k = (a, b)$ we can approximate χ_{I_k} in the L^1 metric by continuous functions that vanish outside (a, b). (For example, given $\epsilon > 0$, take g to be the continuous function that equals 0 on $(-\infty, a]$ and $[b, \infty)$, equals 1 on $[a + \epsilon, b - \epsilon]$, and is linear on $[a, a + \epsilon]$ and $[b - \epsilon, b]$.) Putting these facts together, we obtain the desired assertions. ∎

The next theorem gives a criterion, less restrictive than those found in most advanced calculus books, for the validity of interchanging a limit or a derivative with an integral.

2.27 Theorem. *Suppose that $f : X \times [a,b] \to \mathbb{C}$ $(-\infty < a < b < \infty)$ and that $f(\cdot,t) : X \to \mathbb{C}$ is integrable for each $t \in [a,b]$. Let $F(t) = \int_X f(x,t)\,d\mu(x)$.*

a. Suppose that there exists $g \in L^1(\mu)$ such that $|f(x,t)| \le g(x)$ for all x,t. If $\lim_{t\to t_0} f(x,t) = f(x,t_0)$ for every x, then $\lim_{t\to t_0} F(t) = F(t_0)$; in particular, if $f(x,\cdot)$ is continuous for each x, then F is continuous.

b. Suppose that $\partial f/\partial t$ exists and there is a $g \in L^1(\mu)$ such that $|(\partial f/\partial t)(x,t)| \le g(x)$ for all x,t. Then F is differentiable and $F'(x) = \int(\partial f/\partial t)(x,t)\,d\mu(x)$.

Proof. For (a), apply the dominated convergence theorem to $f_n(x) = f(x,t_n)$ where $\{t_n\}$ is any sequence in $[a,b]$ converging to t_0. For (b), observe that

$$\frac{\partial f}{\partial t}(x,t_0) = \lim h_n(x) \text{ where } h_n(x) = \frac{f(x,t_n) - f(x,t_0)}{t_n - t_0},$$

$\{t_n\}$ again being any sequence converging to t_0. It follows that $\partial f/\partial t$ is measurable, and by the mean value theorem,

$$|h_n(x)| \le \sup_{t\in[a,b]} \left|\frac{\partial f}{\partial t}(x,t)\right| \le g(x),$$

so the dominated convergence theorem can be invoked again to give

$$F'(t_0) = \lim \frac{F(t_n) - F(t_0)}{t_n - t_0} = \lim \int h_n(x)\,d\mu(x) = \int \frac{\partial f}{\partial t}(x,t)\,d\mu(x).$$

∎

The device of using sequences converging to t_0 in the preceding proof is technically necessary because the dominated convergence theorem deals only with sequences of functions. However, in such situations we shall usually just say "let $t \to t_0$" with the understanding that sequential convergence is underlying the argument.

It is important to note that in Theorem 2.27 the interval $[a,b]$ on which the estimates on f or $\partial f/\partial t$ hold might be a proper subinterval of an open interval I (perhaps $\mathbb{R}$ itself) on which $f(x,\cdot)$ is defined. If the hypotheses of (a) or (b) hold for all $[a,b] \subset I$, perhaps with the dominating function g depending on a and b, one obtains the continuity or differentiability of the integrated function F on all of I, as these properties are local in nature.

In the special case where the measure μ is Lebesgue measure on $\mathbb{R}$, the integral we have developed is called the **Lebesgue integral**. At this point it is appropriate to study the relation between the Lebesgue and Riemann integrals on $\mathbb{R}$. We shall use Darboux's characterization of the Riemann integral in terms of upper and lower sums, which we now recall.

Let $[a,b]$ be a compact interval. By a **partition** of $[a,b]$ we shall mean a finite sequence $P = \{t_j\}_0^n$ such that $a = t_0 < t_1 < \cdots < t_n = b$. Let f be an arbitrary bounded real-valued function on $[a,b]$. For each partition P we define

$$S_P f = \sum_1^n M_j(t_j - t_{j-1}), \qquad s_P f = \sum_1^n m_j(t_j - t_{j-1}),$$

where M_j and m_j are the supremum and infimum of f on $[t_{j-1}, t_j]$. Then we define

$$\overline{I}_a^b(f) = \inf_P S_P f, \qquad \underline{I}_a^b(f) = \sup_P s_P f,$$

where the infimum and supremum are taken over all partitions P. If $\overline{I}_a^b(f) = \underline{I}_a^b(f)$, their common value is the **Riemann integral** $\int_a^b f(x)\,dx$, and f is called **Riemann integrable**.

2.28 Theorem. *Let f be a bounded real-valued function on $[a, b]$.*

a. If f is Riemann integrable, then f is Lebesgue measurable (and hence integrable on $[a, b]$ since it is bounded), and $\int_a^b f(x)\,dx = \int_{[a,b]} f\,dm$.

b. f is Riemann integrable iff $\{x \in [a, b] : f$ is discontinuous at $x\}$ has Lebesgue measure zero.

Proof. Suppose that f is Riemann integrable. For each partition P let

$$G_P = \sum_1^n M_j \chi_{(t_{j-1}, t_j]}, \qquad g_P = \sum_1^n m_j \chi_{(t_{j-1}, t_j]}$$

(with the same notation as above), so that $S_P f = \int G_P\,dm$ and $s_P f = \int g_P\,dm$. There is a sequence $\{P_k\}$ of partitions whose mesh (i.e., $\max_j(t_j - t_{j-1})$) tends to zero, each of which includes the preceding one (so that g_{P_k} increases with k while G_{P_k} decreases), such that $S_{P_k} f$ and $s_{P_k} f$ converge to $\int_a^b f(x)\,dx$. Let $G = \lim G_{P_k}$ and $g = \lim g_{P_k}$. Then $g \le f \le G$, and by the dominated convergence theorem, $\int G\,dm = \int g\,dm = \int_a^b f(x)\,dx$. Hence $\int (G - g)\,dm = 0$, so $G = g$ a.e. by Proposition 2.16, and thus $G = f$ a.e. Since G is measurable (being the limit of a sequence of simple functions) and m is complete, f is measurable and $\int_{[a,b]} f\,dm = \int G\,dm = \int_a^b f(x)\,dx$. This proves (a), and the proof of (b) is outlined in Exercise 23. ∎

The (proper) Riemann integral is thus subsumed in the Lebesgue integral. Some improper Riemann integrals (the absolutely convergent ones) can be interpreted directly as Lebesgue integrals, but others still require a limiting procedure. For example, if f is Riemann integrable on $[0, b]$ for all $b > 0$ and Lebesgue integrable on $[0, \infty)$, then $\int_{[0,\infty)} f\,dm = \lim_{b\to\infty} \int_0^b f(x)\,dx$ (by the dominated convergence theorem), but the limit on the right can exist even when f is not integrable. (Example: $f = \sum_1^\infty n^{-1}(-1)^n \chi_{(n,n+1]}$.) Henceforth we shall generally use the notation $\int_a^b f(x)\,dx$ for Lebesgue integrals.

A few remarks comparing the construction of the Lebesgue and Riemann integrals may be helpful. Let f be a bounded measurable function on $[a, b]$, and for simplicity let us assume that $f \ge 0$. To compute the Riemann integral of f, one partitions the interval $[a, b]$ into subintervals and approximates f from above and below by functions that are constant on each subinterval. To compute the Lebesgue integral of

f, one picks a sequence of simple functions that increase to f. In particular, if one picks the sequence constructed in the proof of Theorem 2.10a (see Figure 2.1), one is in effect partitioning the *range* of f into subintervals I_j and approximating f by a constant on each of the sets $f^{-1}(I_j)$. This procedure requires a more sophisticated theory of measure to begin with since the sets $f^{-1}(I_j)$ can be complicated, even when f is continuous; but it is better adapted to the particular f under consideration and therefore more flexible — and more susceptible to generalization. (In the Lebesgue theory, the assumption that f is measurable removes the necessity of considering both upper and lower approximations; however, the latter point of view can also be made to work in the abstract setting. See Exercise 24.)

The Lebesgue theory offers two real advantages over the Riemann theory. First, much more powerful convergence theorems, such as the monotone and dominated convergence theorems, are available. These not only yield results previously unobtainable but also reduce the labor in proving classical theorems. Second, a wider class of functions can be integrated. For example, if R is the set of rational numbers in $[0,1]$, χ_R is not Riemann integrable, being everywhere discontinuous on $[0,1]$, but it is Lebesgue integrable, and $\int \chi_R \, dm = 0$. (Actually, this is in some sense a trivial example since χ_R agrees a.e. with the constant function 0. For a more interesting example, see Exercise 25.) Of course, virtually all functions that one meets in classical analysis are (locally) Riemann integrable, so this added generality is rarely used in computing specific integrals. However, it has the crucial effect that various metric spaces of functions whose metrics are defined in terms of integrals are *complete* when Lebesgue integrable functions are used but not when one considers only Riemann integrable functions. We shall investigate this situation more thoroughly later, especially in Chapter 6. (We have already proved the completeness of $L^1(\mu)$, disguised as Theorem 2.25. To remove the disguise, see Theorem 5.1.)

We conclude this section by introducing the most ubiquitous of the higher transcendental functions, the **gamma function** Γ, which will play a role in a number of places later on. If $z \in \mathbb{C}$ and $\operatorname{Re} z > 0$, define $f_z : (0,\infty) \to \mathbb{C}$ by $f_z(t) = t^{z-1}e^{-t}$. (Here $t^{z-1} = \exp[(z-1)\log t]$.) Since $|t^{z-1}| = t^{\operatorname{Re} z-1}$, we have $|f_z(t)| \le t^{\operatorname{Re} z-1}$, and also $|f_z(t)| \le C_z e^{-t/2}$ for $t \ge 1$. (The precise value of C_z can easily be found by maximizing $t^{\operatorname{Re} z-1}e^{-t/2}$, but it is of no importance here.) Since $\int_0^1 t^a \, dt < \infty$ for $a > -1$ and $\int_1^\infty e^{-t/2} \, dt < \infty$, we see that $f_z \in L^1((0,\infty))$ for $\operatorname{Re} z > 0$, and we define

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t} \, dt \qquad (\operatorname{Re} z > 0).$$

Since

$$\int_\epsilon^N t^z e^{-t} \, dt = -t^z e^{-t}\Big|_\epsilon^N + z\int_\epsilon^N t^{z-1}e^{-t} \, dt$$

by integration by parts, by letting $\epsilon \to 0$ and $N \to \infty$ we see that for $\operatorname{Re} z > 0$, Γ satisfies the functional equation

$$\Gamma(z+1) = z\Gamma(z).$$

This equation can then be used to extend Γ to (almost) the entire complex plane. Namely, for $-1 < \operatorname{Re} z \le 0$ we can *define* $\Gamma(z)$ to be $\Gamma(z+1)/z$, and by induction,

having defined $\Gamma(z)$ for $\operatorname{Re} z > -n$, we *define* $\Gamma(z)$ for $\operatorname{Re} z > -n-1$ to be $\Gamma(z+1)/z$. The result is a function defined on all of $\mathbb{C}$ except for singularities at the nonpositive integers where the algorithm just described involves division by zero.

We have $\Gamma(1) = \int_0^\infty e^{-t}\,dt = -e^{-t}|_0^\infty = 1$, so an n-fold application of the functional equation shows that $\Gamma(n+1) = n!$. (Another proof of this fact is outlined in Exercise 29.) Many of the applications of the gamma function involve the fact that it provides an extension of the factorial function to nonintegers.

Exercises

18. Fatou's lemma remains valid if the hypothesis that $f_n \in L^+$ is replaced by the hypothesis that f_n is measurable and $f_n \geq -g$ where $g \in L^+ \cap L^1$. What is the analogue of Fatou's lemma for nonpositive functions?

19. Suppose $\{f_n\} \subset L^1(\mu)$ and $f_n \to f$ uniformly.

a. If $\mu(X) < \infty$, then $f \in L^1(\mu)$ and $\int f_n \to \int f$.

b. If $\mu(X) = \infty$, the conclusions of (a) can fail. (Find examples on $\mathbb{R}$ with Lebesgue measure.)

20. (A generalized Dominated Convergence Theorem) If $f_n, g_n, f, g \in L^1$, $f_n \to f$ and $g_n \to g$ a.e., $|f_n| \leq g_n$, and $\int g_n \to \int g$, then $\int f_n \to \int f$. (Rework the proof of the dominated convergence theorem.)

21. Suppose $f_n, f \in L^1$ and $f_n \to f$ a.e. Then $\int |f_n - f| \to 0$ iff $\int |f_n| \to \int |f|$. (Use Exercise 20.)

22. Let μ be counting measure on $\mathbb{N}$. Interpret Fatou's lemma and the monotone and dominated convergence theorems as statements about infinite series.

23. Given a bounded function $f : [a,b] \to \mathbb{R}$, let

$$H(x) = \lim_{\delta \to 0} \sup_{|y-x| \leq \delta} f(y), \qquad h(x) = \lim_{\delta \to 0} \inf_{|y-x| \leq \delta} f(y).$$

Prove Theorem 2.28b by establishing the following lemmas:

a. $H(x) = h(x)$ iff f is continuous at x.

b. In the notation of the proof of Theorem 2.28a, $H = G$ a.e. and $h = g$ a.e. Hence H and h are Lebesgue measurable, and $\int_{[a,b]} H\,dm = \overline{I}_a^b(f)$ and $\int_{[a,b]} h\,dm = \underline{I}_a^b(f)$.

24. Let $(X, \mathcal{M}, \mu)$ be a measure space with $\mu(X) < \infty$, and let $(X, \overline{\mathcal{M}}, \overline{\mu})$ be its completion. Suppose $f : X \to \mathbb{R}$ is bounded. Then f is $\overline{\mathcal{M}}$-measurable (and hence in $L^1(\overline{\mu})$) iff there exist sequences $\{\phi_n\}$ and $\{\psi_n\}$ of $\mathcal{M}$-measurable simple functions such that $\phi_n \leq f \leq \psi_n$ and $\int (\psi_n - \phi_n)\,d\mu < n^{-1}$. In this case, $\lim \int \phi_n\,d\mu = \lim \int \psi_n\,d\mu = \int f\,d\overline{\mu}$.

25. Let $f(x) = x^{-1/2}$ if $0 < x < 1$, $f(x) = 0$ otherwise. Let $\{r_n\}_1^\infty$ be an enumeration of the rationals, and set $g(x) = \sum_1^\infty 2^{-n} f(x - r_n)$.

a. $g \in L^1(m)$, and in particular $g < \infty$ a.e.

b. g is discontinuous at every point and unbounded on every interval, and it remains so after any modification on a Lebesgue null set.

c. $g^2 < \infty$ a.e., but g^2 is not integrable on any interval.

26. If $f \in L^1(m)$ and $F(x) = \int_{-\infty}^{x} f(t)\,dt$, then F is continuous on $\mathbb{R}$.

27. Let $f_n(x) = ae^{-nax} - be^{-nbx}$ where $0 < a < b$.

a. $\sum_1^\infty \int_0^\infty |f_n(x)|\,dx = \infty$.

b. $\sum_1^\infty \int_0^\infty f_n(x)\,dx = 0$.

c. $\sum_1^\infty f_n \in L^1([0,\infty), m)$, and $\int_0^\infty \sum_1^\infty f_n(x)\,dx = \log(b/a)$.

28. Compute the following limits and justify the calculations:

a. $\lim_{n\to\infty} \int_0^\infty (1 + (x/n))^{-n} \sin(x/n)\,dx$.

b. $\lim_{n\to\infty} \int_0^1 (1 + nx^2)(1 + x^2)^{-n}\,dx$.

c. $\lim_{n\to\infty} \int_0^\infty n \sin(x/n)[x(1 + x^2)]^{-1}\,dx$.

d. $\lim_{n\to\infty} \int_a^\infty n(1 + n^2x^2)^{-1}\,dx$. (The answer depends on whether $a > 0$, $a = 0$, or $a < 0$. How does this accord with the various convergence theorems?)

29. Show that $\int_0^\infty x^n e^{-x}\,dx = n!$ by differentiating the equation $\int_0^\infty e^{-tx}\,dx = 1/t$. Similarly, show that $\int_{-\infty}^\infty x^{2n} e^{-x^2}\,dx = (2n)!\sqrt{\pi}/4^n n!$ by differentiating the equation $\int_{-\infty}^\infty e^{-tx^2}\,dx = \sqrt{\pi/t}$ (see Proposition 2.53).

30. Show that $\lim_{k\to\infty} \int_0^k x^n(1 - k^{-1}x)^k\,dx = n!$.

31. Derive the following formulas by expanding part of the integrand into an infinite series and justifying the term-by-term integration. Exercise 29 may be useful. (*Note:* In (d) and (e), term-by-term integration works, and the resulting series converges, only for $a > 1$, but the formulas as stated are actually valid for all $a > 0$.)

a. For $a > 0$, $\int_{-\infty}^\infty e^{-x^2} \cos ax\,dx = \sqrt{\pi} e^{-a^2/4}$.

b. For $a > -1$, $\int_0^1 x^a(1 - x)^{-1} \log x\,dx = \sum_1^\infty (a + k)^{-2}$.

c. For $a > 1$, $\int_0^\infty x^{a-1}(e^x - 1)^{-1}\,dx = \Gamma(a)\zeta(a)$, where $\zeta(a) = \sum_1^\infty n^{-a}$.

d. For $a > 1$, $\int_0^\infty e^{-ax} x^{-1} \sin x\,dx = \arctan(a^{-1})$.

e. For $a > 1$, $\int_0^\infty e^{-ax} J_0(x)\,dx = (s^2 + 1)^{-1/2}$, where $J_0(x) = \sum_0^\infty (-1)^n x^{2n}/4^n(n!)^2$ is the Bessel function of order zero.

2.4 MODES OF CONVERGENCE

If $\{f_n\}$ is a sequence of complex-valued functions on a set X, the statement "$f_n \to f$ as $n \to \infty$" can be taken in many different senses, for example, pointwise or uniform convergence. If X is a measure space, one can also speak of a.e. convergence or convergence in L^1. Of course, uniform convergence implies pointwise convergence, which in turn implies a.e. convergence (and not conversely, in general), but these

modes of convergence do not imply L^1 convergence or vice versa. It will be useful to keep in mind the following examples on $\mathbb{R}$ (with Lebesgue measure):

i. $f_n = n^{-1}\chi_{(0,n)}$.

ii. $f_n = \chi_{(n,n+1)}$.

iii. $f_n = n\chi_{[0,1/n]}$.

iv. $f_1 = \chi_{[0,1]}$, $f_2 = \chi_{[0,1/2]}$, $f_3 = \chi_{[1/2,1]}$, $f_4 = \chi_{[0,1/4]}$, $f_5 = \chi_{[1/4,1/2]}$, $f_6 = \chi_{[1/2,3/4]}$, $f_7 = \chi_{[3/4,1]}$, and in general, $f_n = \chi_{[j/2^k,(j+1)/2^k]}$ where $n = 2^k + j$ with $0 \le j < 2^k$.

In (i), (ii), and (iii), $f_n \to 0$ uniformly, pointwise, and a.e., repectively, but $f_n \not\to 0 \in L^1$ (in fact $\int |f_n| = \int f_n = 1$ for all n). In (iv), $f_n \to 0$ in L^1 since $\int |f_n| = 2^{-k}$ for $2^k \le n < 2^{k+1}$, but $f_n(x)$ does not converge for any $x \in [0,1]$ since there are infinitely many n for which $f_n(x) = 0$ and infinitely many for which $f_n(x) = 1$.

On the other hand, if $f_n \to f$ a.e. and $|f_n| \le g \in L^1$ for all n, then $f_n \to f$ in L^1. (This is clear from the dominated convergence theorem since $|f_n - f| \le 2g$.) Also, we shall see below that if $f_n \to f$ in L^1 then some subsequence converges to f a.e.

Another mode of convergence that is frequently useful is convergence in measure. We say that a sequence $\{f_n\}$ of measurable complex-valued functions on $(X, \mathcal{M}, \mu)$ is **Cauchy in measure** if for every $\epsilon > 0$,

$$\mu\big(\{x : |f_n(x) - f_m(x)| \ge \epsilon\}\big) \to 0 \text{ as } m, n \to \infty,$$

and that $\{f_n\}$ **converges in measure** to f if for every $\epsilon > 0$,

$$\mu\big(\{x : |f_n(x) - f(x)| \ge \epsilon\}\big) \to 0 \text{ as } n \to \infty.$$

For example, the sequences (i), (iii), and (iv) above converge to zero in measure, but (ii) is not Cauchy in measure.

2.29 Proposition. *If $f_n \to f$ in L^1, then $f_n \to f$ in measure.*

Proof. Let $E_{n,\epsilon} = \{x : |f_n(x) - f(x)| \ge \epsilon\}$. Then $\int |f_n - f| \ge \int_{E_{n,\epsilon}} |f_n - f| \ge \epsilon\mu(E_{n,\epsilon})$, so $\mu(E_{n,\epsilon}) \le \epsilon^{-1} \int |f_n - f| \to 0$. ∎

The converse of Proposition 2.29 is false, as examples (i) and (iii) show.

2.30 Theorem. *Suppose that $\{f_n\}$ is Cauchy in measure. Then there is a measurable function f such that $f_n \to f$ in measure, and there is a subsequence $\{f_{n_j}\}$ that converges to f a.e. Moreover, if also $f_n \to g$ in measure, then $g = f$ a.e.*

Proof. We can choose a subsequence $\{g_j\} = \{f_{n_j}\}$ of $\{f_n\}$ such that if $E_j = \{x : |g_j(x) - g_{j+1}(x)| \ge 2^{-j}\}$, then $\mu(E_j) \le 2^{-j}$. If $F_k = \bigcup_{j=k}^{\infty} E_j$, then $\mu(F_k) \le \sum_k^{\infty} 2^{-j} = 2^{1-k}$, and if $x \notin F_k$, for $i \ge j \ge k$ we have

$$|g_j(x) - g_i(x)| \le \sum_{l=j}^{i-1} |g_{l+1}(x) - g_l(x)| \le \sum_{l=j}^{i-1} 2^{-l} \le 2^{1-j}. \tag{2.31}$$

Thus $\{g_j\}$ is pointwise Cauchy on F_k^c. Let $F = \bigcap_1^\infty F_k = \limsup E_j$. Then $\mu(F) = 0$, and if we set $f(x) = \lim g_j(x)$ for $x \notin F$ and $f(x) = 0$ for $x \in F$, then f is measurable (see Exercises 3 and 5) and $g_j \to f$ a.e. Also, (2.31) shows that $|g_j(x) - f(x)| \le 2^{1-j}$ for $x \notin F_k$ and $j \ge k$. Since $\mu(F_k) \to 0$ as $k \to \infty$, it follows that $g_j \to f$ in measure. But then $f_n \to f$ in measure, because

$$\{x : |f_n(x) - f(x)| \ge \epsilon\} \subset \{x : |f_n(x) - g_j(x)| \ge \tfrac{1}{2}\epsilon\} \cup \{x : |g_j(x) - f(x)| \ge \tfrac{1}{2}\epsilon\},$$

and the sets on the right both have small measure when n and j are large. Likewise, if $f_n \to g$ in measure,

$$\{x : |f(x) - g(x)| \ge \epsilon\} \subset \{x : |f(x) - f_n(x)| \ge \tfrac{1}{2}\epsilon\} \cup \{x : |f_n(x) - g(x)| \ge \tfrac{1}{2}\epsilon\}$$

for all n, hence $\mu(\{x : |f(x) - g(x)| \ge \epsilon\}) = 0$ for all ϵ. Letting ϵ tend to zero through some sequence of values, we conclude that $f = g$ a.e. ∎

2.32 Corollary. *If $f_n \to f$ in L^1, there is a subsequence $\{f_{n_j}\}$ such that $f_{n_j} \to f$ a.e.*

Proof. Combine Proposition 2.29 and Theorem 2.30. ∎

If $f_n \to f$ a.e., it does not follow that $f_n \to f$ in measure, as example (ii) shows. However, this conclusion does hold on a finite measure space, where something considerably stronger is true.

2.33 Egoroff's Theorem. *Suppose that $\mu(X) < \infty$, and $f_1, f_2, \ldots$ and f are measurable complex-valued functions on X such that $f_n \to f$ a.e. Then for every $\epsilon > 0$ there exists $E \subset X$ such that $\mu(E) < \epsilon$ and $f_n \to f$ uniformly on E^c.*

Proof. Without loss of generality we may assume that $f_n \to f$ everywhere on X. For $k, n \in \mathbb{N}$ let

$$E_n(k) = \bigcup_{m=n}^{\infty} \{x : |f_m(x) - f(x)| \ge k^{-1}\}.$$

Then, for fixed k, $E_n(k)$ decreases as n increases, and $\bigcap_{n=1}^\infty E_n(k) = \varnothing$, so since $\mu(X) < \infty$ we conclude that $\mu(E_n(k)) \to 0$ as $n \to \infty$. Given $\epsilon > 0$ and $k \in \mathbb{N}$, choose n_k so large that $\mu(E_{n_k}(k)) < \epsilon 2^{-k}$ and let $E = \bigcup_{k=1}^\infty E_{n_k}(k)$. Then $\mu(E) < \epsilon$, and we have $|f_n(x) - f(x)| < k^{-1}$ for $n > n_k$ and $x \notin E$. Thus $f_n \to f$ uniformly on E^c. ∎

The type of convergence involved in the conclusion of Egoroff's theorem is sometimes called **almost uniform convergence**. It is not hard to see that almost uniform convergence implies a.e. convergence and convergence in measure (Exercise 39).

Exercises

32. Suppose $\mu(X) < \infty$. If f and g are complex-valued measurable functions on X, define

$$\rho(f,g) = \int \frac{|f-g|}{1+|f-g|}\, d\mu.$$

Then ρ is a metric on the space of measurable functions if we identify functions that are equal a.e., and $f_n \to f$ with respect to this metric iff $f_n \to f$ in measure.

33. If $f_n \geq 0$ and $f_n \to f$ in measure, then $\int f \leq \liminf \int f_n$.

34. Suppose $|f_n| \leq g \in L^1$ and $f_n \to f$ in measure.
a. $\int f = \lim \int f_n$.
b. $f_n \to f$ in L^1.

35. $f_n \to f$ in measure iff for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $\mu(\{x : |f_n(x) - f(x)| \geq \epsilon\}) < \epsilon$ for $n \geq N$.

36. If $\mu(E_n) < \infty$ for $n \in \mathbb{N}$ and $\chi_{E_n} \to f$ in L^1, then f is (a.e. equal to) the characteristic function of a measurable set.

37. Suppose that f_n and f are measurable complex-valued functions and $\phi : \mathbb{C} \to \mathbb{C}$.
a. If ϕ is continuous and $f_n \to f$ a.e., then $\phi \circ f_n \to \phi \circ f$ a.e.
b. If ϕ is uniformly continuous and $f_n \to f$ uniformly, almost uniformly, or in measure, then $\phi \circ f_n \to \phi \circ f$ uniformly, almost uniformly, or in measure, respectively.
c. There are counterexamples when the continuity assumptions on ϕ are not satisfied.

38. Suppose $f_n \to f$ in measure and $g_n \to g$ in measure.
a. $f_n + g_n \to f + g$ in measure.
b. $f_n g_n \to fg$ in measure if $\mu(X) < \infty$, but not necessarily if $\mu(X) = \infty$.

39. If $f_n \to f$ almost uniformly, then $f_n \to f$ a.e. and in measure.

40. In Egoroff's theorem, the hypothesis "$\mu(X) < \infty$" can be replaced by "$|f_n| \leq g$ for all n, where $g \in L^1(\mu)$."

41. If μ is σ-finite and $f_n \to f$ a.e., there exist measurable $E_1, E_2, \ldots \subset X$ such that $\mu((\bigcup_1^\infty E_j)^c) = 0$ and $f_n \to f$ uniformly on each E_j.

42. Let μ be counting measure on $\mathbb{N}$. Then $f_n \to f$ in measure iff $f_n \to f$ uniformly.

43. Suppose that $\mu(X) < \infty$ and $f : X \times [0,1] \to \mathbb{C}$ is a function such that $f(\cdot, y)$ is measurable for each $y \in [0,1]$ and $f(x, \cdot)$ is continuous for each $x \in X$.
a. If $0 < \epsilon, \delta < 1$ then $E_{\epsilon,\delta} = \{x : |f(x,y) - f(x,0)| \leq \epsilon \text{ for all } y < \delta\}$ is measurable.
b. For any $\epsilon > 0$ there is a set $E \subset X$ such that $\mu(E) < \epsilon$ and $f(\cdot, y) \to f(\cdot, 0)$ uniformly on E^c as $y \to 0$.

44. (**Lusin's Theorem**) If $f : [a, b] \to \mathbb{C}$ is Lebesgue measurable and $\epsilon > 0$, there is a compact set $E \subset [a, b]$ such that $\mu(E^c) < \epsilon$ and $f|E$ is continuous. (Use Egoroff's theorem and Theorem 2.26.)

2.5 PRODUCT MEASURES

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be measure spaces. We have already discussed the product σ-algebra $\mathcal{M} \otimes \mathcal{N}$ on $X \times Y$; we now construct a measure on $\mathcal{M} \otimes \mathcal{N}$ that is, in an obvious sense, the product of μ and ν.

To begin with, we define a (measurable) **rectangle** to be a set of the form $A \times B$ where $A \in \mathcal{M}$ and $B \in \mathcal{N}$. Clearly

$$(A \times B) \cap (E \times F) = (A \cap E) \times (B \cap F), \qquad (A \times B)^c = (X \times B^c) \cup (A^c \times B).$$

Therefore, by Proposition 1.7, the collection $\mathcal{A}$ of finite disjoint unions of rectangles is an algebra, and of course the σ-algebra it generates is $\mathcal{M} \otimes \mathcal{N}$.

Suppose $A \times B$ is a rectangle that is a (finite or countable) disjoint union of rectangles $A_j \times B_j$. Then for $x \in X$ and $y \in Y$,

$$\chi_A(x)\chi_B(y) = \chi_{A \times B}(x, y) = \sum \chi_{A_j \times B_j}(x, y) = \sum \chi_{A_j}(x)\chi_{B_j}(y).$$

If we integrate with respect to x and use Theorem 2.15, we obtain

$$\begin{aligned}\mu(A)\chi_B(y) = \int \chi_A(x)\chi_B(y)\, d\mu(x) &= \sum \int \chi_{A_j}(x)\chi_{B_j}(y)\, d\mu(x) \\ &= \sum \mu(A_j)\chi_{B_j}(y).\end{aligned}$$

In the same way, integration in y then yields

$$\mu(A)\nu(B) = \sum \mu(A_j)\nu(B_j).$$

It follows that if $E \in \mathcal{A}$ is the disjoint union of rectangles $A_1 \times B_1, \ldots, A_n \times B_n$, and we set

$$\pi(E) = \sum_1^n \mu(A_j)\nu(E_j)$$

(with the usual convention that $0 \cdot \infty = 0$), then π is well defined on $\mathcal{A}$ (since any two representations of E as a finite disjoint union of rectangles have a common refinement), and π is a premeasure on $\mathcal{A}$. According to Theorem 1.14, therefore, π generates an outer measure on $X \times Y$ whose restriction to $\mathcal{M} \times \mathcal{N}$ is a measure that extends π. We call this measure the **product** of μ and ν and denote it by $\mu \times \nu$. Moreover, if μ and ν are σ-finite — say, $X = \bigcup_1^\infty A_j$ and $Y = \bigcup_1^\infty B_k$ with $\mu(A_j) < \infty$ and $\nu(B_k) < \infty$ — then $X \times Y = \bigcup_{j,k} A_j \times B_k$, and $\mu \times \nu(A_j \times B_k) < \infty$, so $\mu \times \nu$ is also σ-finite. In this case, by Theorem 1.14, $\mu \times \nu$ is the unique measure on $\mathcal{M} \otimes \mathcal{N}$ such that $\mu \times \nu(A \times B) = \mu(A)\nu(B)$ for all rectangles $A \times B$.

The same construction works for any finite number of factors. That is, suppose $(X_j, \mathcal{M}_j, \mu_j)$ are measure spaces for $j = 1, \ldots, n$. If we define a rectangle to be a set of the form $A_1 \times \cdots \times A_n$ with $A_j \in \mathcal{M}_j$, then the collection $\mathcal{A}$ of finite disjoint unions of rectangles is an algebra, and the same procedure as above produces a measure $\mu_1 \times \cdots \times \mu_n$ on $\mathcal{M}_1 \otimes \cdots \otimes \mathcal{M}_n$ such that

$$\mu_1 \times \cdots \times \mu_n(A_1 \times \cdots \times A_n) = \prod_1^n \mu_j(A_j).$$

Moreover, if the μ_j's are σ-finite so that the extension from $\mathcal{A}$ to $\bigotimes_1^n \mathcal{M}_j$ is uniquely determined, the obvious associativity properties hold. For example, if we identify $X_1 \times X_2 \times X_3$ with $(X_1 \times X_2) \times X_3$, we have $\mathcal{M}_1 \otimes \mathcal{M}_2 \otimes \mathcal{M}_3 = (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3$ (the former being generated by sets of the form $A_1 \times A_2 \times A_3$ with $A_j \in \mathcal{M}_j$, and the latter by sets of the form $B \times A_3$ with $B \in \mathcal{M}_1 \otimes \mathcal{M}_2$ and $A_3 \in \mathcal{M}_3$), and $\mu_1 \times \mu_2 \times \mu_3 = (\mu_1 \times \mu_2) \times \mu_3$ (since they agree on sets of the form $A_1 \times A_2 \times A_3$, and hence in general by uniqueness). Details are left to the reader (Exercise 45). All of our results below have obvious extensions to products with n factors, but we shall stick to the case $n = 2$ for simplicity.

We return to the case of two measure spaces $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$. If $E \subset X \times Y$, for $x \in X$ and $y \in Y$ we define the **x-section** E_x and the **y-section** E^y of E by

$$E_x = \{y \in Y : (x, y) \in E\}, \qquad E^y = \{x \in X : (x, y) \in E\}.$$

Also, if f is a function on $X \times Y$ we define the **x-section** f_x and the **y-section** f^y of f by

$$f_x(y) = f^y(x) = f(x, y).$$

Thus, for example, $(\chi_E)_x = \chi_{E_x}$ and $(\chi_E)^y = \chi_{E^y}$.

2.34 Proposition.

a. If $E \in \mathcal{M} \times \mathcal{N}$, then $E_x \in \mathcal{N}$ for all $x \in X$ and $E^y \in \mathcal{M}$ for all $y \in Y$.

b. If f is $\mathcal{M} \otimes \mathcal{N}$-measurable, then f_x is $\mathcal{N}$-measurable for all $x \in X$ and f^y is $\mathcal{M}$-measurable for all $y \in Y$.

Proof. Let $\mathcal{R}$ be the collection of all subsets E of $X \times Y$ such that $E_x \in \mathcal{N}$ for all x and $E^y \in \mathcal{M}$ for all y. Then $\mathcal{R}$ obviously contains all rectangles (e.g., $(A \times B)_x = B$ if $x \in A$, $= \varnothing$ otherwise). Since $(\bigcup_1^\infty E_j)_x = \bigcup_1^\infty (E_j)_x$ and $(E^c)_x = (E_x)^c$, and likewise for y-sections, $\mathcal{R}$ is a σ-algebra. Therefore $\mathcal{R} \supset \mathcal{M} \otimes \mathcal{N}$, which proves (a). (b) follows from (a) because $(f_x)^{-1}(B) = (f^{-1}(B))_x$ and $(f^y)^{-1}(B) = (f^{-1}(B))^y$. ∎

Before proceeding further we need a technical lemma. We define a **monotone class** on a space X to be a subset $\mathcal{C}$ of $\mathcal{P}(X)$ that is closed under countable increasing unions and countable decreasing intersections (that is, if $E_j \in \mathcal{C}$ and $E_1 \subset E_2 \subset \cdots$, then $\bigcup E_j \in \mathcal{C}$, and likewise for intersections). Clearly every σ-algebra is a monotone class. Also, the intersection of any family of monotone classes is a monotone class,

so for any $\mathcal{E} \subset \mathcal{P}(X)$ there is a unique smallest monotone class containing $\mathcal{E}$, called the monotone class **generated by** $\mathcal{E}$.

2.35 The Monotone Class Lemma. *If $\mathcal{A}$ is an algebra of subsets of X, then the monotone class $\mathcal{C}$ generated by $\mathcal{A}$ coincides with the σ-algebra $\mathcal{M}$ generated by $\mathcal{A}$.*

Proof. Since $\mathcal{M}$ is a monotone class, we have $\mathcal{C} \subset \mathcal{M}$; and if we can show that $\mathcal{C}$ is a σ-algebra, we will have $\mathcal{M} \subset \mathcal{C}$. To this end, for $E \in \mathcal{C}$ let us define

$$\mathcal{C}(E) = \{F \in \mathcal{C} : E \setminus F,\ F \setminus E, \text{ and } E \cap F \text{ are in } \mathcal{C}\}.$$

Clearly $\varnothing$ and E are in $\mathcal{C}(E)$, and $E \in \mathcal{C}(F)$ iff $F \in \mathcal{C}(E)$. Also, it is easy to check that $\mathcal{C}(E)$ is a monotone class. If $E \in \mathcal{A}$, then $F \in \mathcal{C}(E)$ for all $F \in \mathcal{A}$ because $\mathcal{A}$ is an algebra; that is, $\mathcal{A} \subset \mathcal{C}(E)$, and hence $\mathcal{C} \subset \mathcal{C}(E)$. Therefore, if $F \in \mathcal{C}$, then $F \in \mathcal{C}(E)$ for all $E \in \mathcal{A}$. But this means that $E \in \mathcal{C}(F)$ for all $E \in \mathcal{A}$, so that $\mathcal{A} \subset \mathcal{C}(F)$ and hence $\mathcal{C} \subset \mathcal{C}(F)$. Conclusion: If $E, F \in \mathcal{C}$, then $E \setminus F$ and $E \cap F$ are in $\mathcal{C}$. Since $X \in \mathcal{A} \subset \mathcal{C}$, $\mathcal{C}$ is therefore an algebra. But then if $\{E_j\}_1^\infty \subset \mathcal{C}$, we have $\bigcup_1^n E_j \in \mathcal{C}$ for all n, and since $\mathcal{C}$ is closed under countable increasing unions it follows that $\bigcup_1^\infty E_j \in \mathcal{C}$. In short, $\mathcal{C}$ is a σ-algebra, and we are done. ∎

We now come to the main results of this section, which relate integrals on $X \times Y$ to integrals on X and Y.

2.36 Theorem. *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces. If $E \in \mathcal{M} \otimes \mathcal{N}$, then the functions $x \mapsto \nu(E_x)$ and $y \mapsto \mu(E^y)$ are measurable on X and Y, respectively, and*

$$\mu \times \nu(E) = \int \nu(E_x)\, d\mu(x) = \int \mu(E^y)\, d\nu(y).$$

Proof. First suppose that μ and ν are finite, and let $\mathcal{C}$ be the set of all $E \in \mathcal{M} \otimes \mathcal{N}$ for which the conclusions of the theorem are true. If $E = A \times B$, then $\nu(E_x) = \chi_A(x)\nu(B)$ and $\mu(E^y) = \mu(A)\chi_B(y)$, so clearly $E \in \mathcal{C}$. By additivity it follows that finite disjoint unions of rectangles are in $\mathcal{C}$, so by Lemma 2.35 it will suffice to show that $\mathcal{C}$ is a monotone class. If $\{E_n\}$ is an increasing sequence in $\mathcal{C}$ and $E = \bigcup_1^\infty E_n$, then the functions $f_n(y) = \mu((E_n)^y)$ are measurable and increase pointwise to $f(y) = \mu(E^y)$. Hence f is measurable, and by the monotone convergence theorem,

$$\int \mu(E^y)\, d\nu(y) = \lim \int \mu((E_n)^y)\, d\nu(y) = \lim \mu \times \nu(E_n) = \mu \times \nu(E).$$

Likewise $\mu \times \nu(E) = \int \nu(E_x)\, d\mu(x)$, so $E \in \mathcal{C}$. Similarly, if $\{E_n\}$ is a decreasing sequence in $\mathcal{C}$ and $\bigcap_1^\infty E_n$, the function $y \mapsto \mu((E_1)^y)$ is in $L^1(\nu)$ because $\mu((E_1)^y) \le \mu(X) < \infty$ and $\nu(Y) < \infty$, so the dominated convergence theorem can be applied to show that $E \in \mathcal{C}$. Thus $\mathcal{C}$ is a monotone class, and the proof is complete for the case of finite measure spaces.

Finally, if μ and ν are σ-finite, we can write $X \times Y$ as the union of an increasing sequence $\{X_j \times Y_j\}$ of rectangles of finite measure. If $E \in \mathcal{M} \otimes \mathcal{N}$, the preceding argument applies to $E \cap (X_j \times Y_j)$ for each j to give

$$\mu \times \nu(E \cap (X_j \times Y_j)) = \int \chi_{X_j}(x)\nu(E_x \cap Y_j)\, d\mu(x) = \int \chi_{Y_j}(y)\mu(E^y \cap X_j)\, d\nu(y),$$

and a final application of the monotone convergence theorem then yields the desired result. ∎

2.37 The Fubini-Tonelli Theorem. *Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces.*

a. (Tonelli) *If $f \in L^+(X \times Y)$, then the functions $g(x) = \int f_x\, d\nu$ and $h(y) = \int f^y\, d\mu$ are in $L^+(X)$ and $L^+(Y)$, respectively, and*

$$\begin{aligned} \int f\, d(\mu \times \nu) &= \int \left[\int f(x,y)\, d\nu(y)\right] d\mu(x) \\ &= \int \left[\int f(x,y)\, d\mu(x)\right] d\nu(y). \end{aligned} \tag{2.38}$$

b. (Fubini) *If $f \in L^1(\mu \times \nu)$, then $f_x \in L^1(\nu)$ for a.e. $x \in X$, $f^y \in L^1(\mu)$ for a.e. $y \in Y$, the a.e.-defined functions $g(x) = \int f_x\, d\nu$ and $h(x) = \int f^y\, d\nu$ are in $L^1(\mu)$ and $L^1(\nu)$, respectively, and (2.38) holds.*

Proof. Tonelli's theorem reduces to Theorem 2.36 in case f is a characteristic function, and it therefore holds for nonnegative simple functions by linearity. If $f \in L^+(X \times Y)$, let $\{f_n\}$ be a sequence of simple functions that increase pointwise to f as in Theorem 2.10. The monotone convergence theorem implies, first, that the corresponding g_n and h_n increase to g and h (so that g and h are measurable), and, second that

$$\int g\, d\mu = \lim \int g_n\, d\mu = \lim \int f_n\, d(\mu \times \nu) = \int f\, d(\mu \times \nu),$$
$$\int h\, d\nu = \lim \int h_n\, d\nu = \lim \int f_n\, d(\mu \times \nu) = \int f\, d(\mu \times \nu),$$

which is (2.38). This establishes Tonelli's theorem and also shows that if $f \in L^+(X \times Y)$ and $\int f\, d(\mu \times \nu) < \infty$, then $g < \infty$ a.e. and $h < \infty$ a.e., that is, $f_x \in L^1(\nu)$ for a.e. x and $f^y \in L^1(\mu)$ for a.e. y. If $f \in L^1(\mu \times \nu)$, then, the conclusion of Fubini's theorem follows by applying these results to the positive and negative parts of the real and imaginary parts of f. ∎

A few remarks are in order:

- We shall usually omit the brackets in the iterated integrals in (2.38), thus:

$$\int \left[\int f(x,y)\, d\mu(x)\right] d\nu(y) = \iint f(x,y)\, d\mu(x)\, d\nu(y) = \iint f\, d\mu\, d\nu.$$

- The hypothesis of σ-finiteness is necessary; see Exercise 46.

- The hypothesis $f \in L^+(X \times Y)$ or $f \in L^1(\mu \times \nu)$ is necessary, in two respects. First, it is possible for f_x and f^y to be measurable for all x, y and for the iterated integrals $\iint f\, d\mu\, d\nu$ and $\iint f\, d\nu\, d\mu$ to exist even if f is not $\mathcal{M} \otimes \mathcal{N}$-measurable. However, the iterated integrals need not then be equal; see Exercise 47. Second, if f is not nonnegative, it is possible for f_x and f^y to be integrable for all x, y and for the iterated integrals $\iint f\, d\mu\, d\nu$ and $\iint f\, d\nu\, d\mu$ to exist even if $\int |f|\, d(\mu \times \nu) = \infty$. But again, the iterated integrals need not be equal; see Exercise 48.

- The Fubini and Tonelli theorems are frequently used in tandem. Typically one wishes to reverse the order of integration in a double integral $\iint f\, d\mu\, d\nu$. *First* one verifies that $\int |f|\, d(\mu \times \nu) < \infty$ by using Tonelli's theorem to evaluate this integral as an iterated integral; *then* one applies Fubini's theorem to conclude that $\iint f\, d\mu\, d\nu = \iint f\, d\nu\, d\mu$. For examples, see the exercises in §2.6.

Even if μ and ν are complete, $\mu \times \nu$ is almost never complete. Indeed, suppose that there is a nonempty $A \in \mathcal{M}$ with $\mu(A) = 0$ and that $\mathcal{N} \neq \mathcal{P}(Y)$. (This is the case with $\mu = \nu =$ Lebesgue measure on $\mathbb{R}$, for example.) If $E \in \mathcal{P}(Y) \setminus \mathcal{N}$, then $A \times E \notin \mathcal{M} \otimes \mathcal{N}$ by Proposition 2.34, but $A \times E \subset A \times Y$, and $\mu \times \nu(A \times Y) = 0$.

If one wishes to work with complete measures, of course, one can consider the completion of $\mu \times \nu$. In this setting the relationship between the measurability of a function on $X \times Y$ and the measurability of its x-sections and y-sections is not so simple. However, the Fubini-Tonelli theorem is still valid when suitably reformulated:

2.39 The Fubini-Tonelli Theorem for Complete Measures. *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be complete, σ-finite measure spaces, and let $(X \times Y, \mathcal{L}, \lambda)$ be the completion of $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$. If f is $\mathcal{L}$-measurable and either (a) $f \geq 0$ or (b) $f \in L^1(\lambda)$, then f_x is $\mathcal{N}$-measurable for a.e. x and f^y is $\mathcal{M}$-measurable for a.e. y, and in case (b) f_x and f^y are also integrable for a.e. x and y. Moreover, $x \mapsto \int f_x\, d\nu$ and $y \mapsto \int f^y\, d\mu$ are measurable, and in case (b) also integrable, and*

$$\int f\, d\lambda = \iint f(x,y)\, d\mu(x)\, d\nu(y) = \iint f(x,y)\, d\nu(y)\, d\mu(x).$$

This theorem is a fairly easy corollary of Theorem 2.37; the proof is outlined in Exercise 49.

Exercises

45. If $(X_j, \mathcal{M}_j)$ is a measurable space for $j = 1, 2, 3$, then $\bigotimes_1^3 \mathcal{M}_j = (\mathcal{M}_1 \otimes \mathcal{M}_2) \otimes \mathcal{M}_3$. Moreover, if μ_j is a σ-finite measure on $(X_j, \mathcal{M}_j)$, then $\mu_1 \times \mu_2 \times \mu_3 = (\mu_1 \times \mu_2) \times \mu_3$.

46. Let $X = Y = [0,1]$, $\mathcal{M} = \mathcal{N} = \mathcal{B}_{[0,1]}$, $\mu =$ Lebesgue measure, and $\nu =$ counting measure. If $D = \{(x,x) : x \in [0,1]\}$ is the diagonal in $X \times Y$, then $\iint \chi_D\, d\mu\, d\nu$,

$\iint \chi_D \, d\nu \, d\mu$, and $\int \chi_D \, d(\mu \times \nu)$ are all unequal. (To compute $\int \chi_D \, d(\mu \times \nu) = \mu \times \nu(D)$, go back to the definition of $\mu \times \nu$.)

47. Let $X = Y$ be an uncountable linearly ordered set such that for each $x \in X$, $\{y \in X : y < x\}$ is countable. (Example: the set of countable ordinals.) Let $\mathcal{M} = \mathcal{N}$ be the σ-algebra of countable or co-countable sets, and let $\mu = \nu$ be defined on $\mathcal{M}$ by $\mu(A) = 0$ if A is countable and $\mu(A) = 1$ if A is co-countable. Let $E = \{(x, y) \in X \times X : y < x\}$. Then E_x and E^y are measurable for all x, y, and $\iint \chi_E \, d\mu \, d\nu$ and $\iint \chi_E \, d\nu \, d\mu$ exist but are not equal. (If one believes in the continuum hypothesis, one can take $X = [0, 1]$ [with a nonstandard ordering] and thus obtain a set $E \subset [0, 1]^2$ such that E_x is countable and E^y is co-countable [in particular, Borel] for all x, y, but E is not Lebesgue measurable.)

48. Let $X = Y = \mathbb{N}$, $\mathcal{M} = \mathcal{N} = \mathcal{P}(\mathbb{N})$, $\mu = \nu =$ counting measure. Define $f(m, n) = 1$ if $m = n$, $f(m, n) = -1$ if $m = n + 1$, and $f(m, n) = 0$ otherwise. Then $\int |f| \, d(\mu \times \nu) = \infty$, and $\iint f \, d\mu \, d\nu$ and $\iint f \, d\nu \, d\mu$ exist and are unequal.

49. Prove Theorem 2.39 by using Theorem 2.37 and Proposition 2.12 together with the following lemmas.

a. If $E \in \mathcal{M} \times \mathcal{N}$ and $\mu \times \nu(E) = 0$, then $\nu(E_x) = \mu(E^y) = 0$ for a.e. x and y.

b. If f is $\mathcal{L}$-measurable and $f = 0$ λ-a.e., then f_x and f^y are integrable for a.e. x and y, and $\int f_x \, d\nu = \int f^y \, d\mu = 0$ for a.e. x and y. (Here the completeness of μ and ν is needed.)

50. Suppose $(X, \mathcal{M}, \mu)$ is a σ-finite measure space and $f \in L^+(X)$. Let

$$G_f = \{(x, y) \in X \times [0, \infty] : y \leq f(x)\}.$$

Then G_f is $\mathcal{M} \times \mathcal{B}_{\mathbb{R}}$-measurable and $\mu \times m(G_f) = \int f \, d\mu$; the same is also true if the inequality $y \leq f(x)$ in the definition of G_f is replaced by $y < f(x)$. (To show measurability of G_f, note that the map $(x, y) \mapsto f(x) - y$ is the composition of $(x, y) \mapsto (f(x), y)$ and $(z, y) \mapsto z - y$.) This is the definitive statement of the familiar theorem from calculus, "the integral of a function is the area under its graph."

51. Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be arbitrary measure spaces (not necessarily σ-finite).

a. If $f : X \to \mathbb{C}$ is $\mathcal{M}$-measurable, $g : Y \to \mathbb{C}$ is $\mathcal{N}$-measurable, and $h(x, y) = f(x)g(y)$, then h is $\mathcal{M} \otimes \mathcal{N}$-measurable.

b. If $f \in L^1(\mu)$ and $g \in L^1(\nu)$, then $h \in L^1(\mu \times \nu)$ and $\int h \, d(\mu \times \nu) = [\int f \, d\mu][\int g \, d\nu]$.

52. The Fubini-Tonelli theorem is valid when $(X, \mathcal{M}, \mu)$ is an arbitrary measure space and Y is a countable set, $\mathcal{N} = \mathcal{P}(Y)$, and ν is counting measure on Y. (Cf. Theorems 2.15 and 2.25.)

2.6 THE n-DIMENSIONAL LEBESGUE INTEGRAL

Lebesgue measure m^n on $\mathbb{R}^n$ is the completion of the n-fold product of Lebesgue measure on $\mathbb{R}$ with itself, that is, the completion of $m \times \cdots \times m$ on $\mathcal{B}_{\mathbb{R}} \otimes \cdots \otimes \mathcal{B}_{\mathbb{R}} = \mathcal{B}_{\mathbb{R}^n}$, or equivalently the completion of $m \times \cdots \times m$ on $\mathcal{L} \otimes \cdots \otimes \mathcal{L}$. The domain $\mathcal{L}^n$ of m^n is the class of **Lebesgue measurable** sets in $\mathbb{R}^n$; sometimes we shall also consider m^n as a measure on the smaller domain $\mathcal{B}_{\mathbb{R}^n}$. When there is no danger of confusion, we shall usually omit the superscript n and write m for m^n, and as in the case $n = 1$, we shall usually write $\int f(x)\,dx$ for $\int f\,dm$.

We begin by establishing the extensions of some of the results in §1.5 to the n-dimensional case. In what follows, if $E = \prod_1^n E_j$ is a rectangle in $\mathbb{R}^n$, we shall refer to the sets $E_j \subset \mathbb{R}$ as the **sides** of E.

2.40 Theorem. *Suppose $E \in \mathcal{L}^n$.*

a. $m(E) = \inf\{m(U) : U \supset E,\ U \text{ open}\} = \sup\{m(K) : K \subset E,\ K \text{ compact}\}$.

b. $E = A_1 \cup N_1 = A_2 \setminus N_2$ *where A_1 is an F_σ set, A_2 is a G_δ set, and* $m(N_1) = m(N_2) = 0$.

c. If $m(E) < \infty$, for any $\epsilon > 0$ there is a finite collection $\{R_j\}_1^N$ of disjoint rectangles whose sides are intervals such that $m(E \triangle \bigcup_1^N R_j) < \epsilon$.

Proof. By the definition of product measures, if $E \in \mathcal{L}^n$ and $\epsilon > 0$ there is a countable family $\{T_j\}$ of rectangles such that $E \subset \bigcup_1^\infty T_j$ and $\sum_1^\infty m(T_j) \le m(E) + \epsilon$. For each j, by applying Theorem 1.18 to the sides of R_j we can find a rectangle $U_j \supset F_j$ whose sides are open sets such that $m(U_j) < m(T_j) + \epsilon 2^{-j}$. If $U = \bigcup_1^\infty U_j$, then U is open and $m(U) \le \sum_1^\infty m(U_j) \le m(E) + 2\epsilon$. This proves the first equation in part (a); the second one, and part (b), then follow as in the proofs of Theorems 1.18 and 1.19. Next, if $m(E) < \infty$, then $m(U_j) < \infty$ for all j. Since the sides of U_j are countable unions of open intervals, by taking suitable finite subunions we obtain rectangles $V_j \subset U_j$ whose sides are finite unions of intervals such that $m(V_j) \ge m(U_j) - \epsilon 2^{-j}$. If N is sufficiently large, then, we have

$$m\Big(E \setminus \bigcup_1^N V_j\Big) \le m\Big(\bigcup_1^N U_j \setminus V_j\Big) + m\Big(\bigcup_{N+1}^\infty U_j\Big) < 2\epsilon$$

and

$$m\Big(\bigcup_1^N V_j \setminus E\Big) \le m\Big(\bigcup_1^\infty U_j \setminus E\Big) < \epsilon,$$

so that $m(E \triangle \bigcup_1^N V_j) < 3\epsilon$. Since $\bigcup_1^N V_j$ can be expressed as a finite disjoint union of rectangles whose sides are intervals, we have proved (c). ∎

2.41 Theorem. *If $f \in L^1(m)$ and $\epsilon > 0$, there is a simple function $\phi = \sum_1^N a_j \chi_{R_j}$, where each R_j is a product of intervals, such that $\int |f - \phi| < \epsilon$, and there is a continuous function g that vanishes outside a bounded set such that $\int |f - g| < \epsilon$.*

Proof. As in the proof of Theorem 2.26, approximate f by simple functions, then use Theorem 2.40c to approximate the latter by functions ϕ of the desired form. Finally, approximate such ϕ's by continuous functions by applying an obvious generalization of the argument in the proof of Theorem 2.26. ∎

2.42 Theorem. *Lebesgue measure is translation-invariant. More precisely, for $a \in \mathbb{R}^n$ define $\tau_a : \mathbb{R}^n \to \mathbb{R}^n$ by $\tau_a(x) = x + a$.*

a. If $E \in \mathcal{L}^n$, then $\tau_a(E) \in \mathcal{L}^n$ and $m(\tau_a(E)) = m(E)$.

b. If $f : \mathbb{R}^n \to \mathbb{C}$ is Lebesgue measurable, then so is $f \circ \tau_a$. Moreover, if either $f \geq 0$ or $f \in L^1(m)$, then $\int (f \circ \tau_a)\, dm = \int f\, dm$.

Proof. Since τ_a and its inverse τ_{-a} are continuous, they preserve the class of Borel sets. The formula $m(\tau_a(E)) = m(E)$ follows easily from the one-dimensional result (Theorem 1.21) if E is a rectangle, and it then follows for general Borel sets since m is determined by its action on rectangles (the uniqueness in Theorem 1.14). In particular, the collection of Borel sets E such that $m(E) = 0$ is invariant under τ_a. Assertion (a) now follows immediately.

If f is Lebesgue measurable and B is a Borel set in $\mathbb{C}$, we have $f^{-1}(B) = E \cup N$ where E is Borel and $m(N) = 0$. But $\tau_a^{-1}(E)$ is Borel and $m(\tau_a^{-1}(N)) = 0$, so $(f \circ \tau_a)^{-1}(B) \in \mathcal{L}^n$ and f is Lebesgue measurable. The equality $\int (f \circ \tau_a)\, d\mu = \int f\, d\mu$ reduces to the equality $m(\tau_{-a}(E)) = m(E)$ when $f = \chi_E$. It is then true for simple functions by linearity, and hence for nonnegative measurable functions by the definition of the integral. Taking positive and negative parts of real and imaginary parts then yields the result for $f \in L^1(m)$. ∎

Let us now compare Lebesgue measure on $\mathbb{R}^n$ to the more naive theory of n-dimensional measure usually found in advanced calculus books. In this discussion, a **cube** in $\mathbb{R}^n$ is a Cartesian product of n closed intervals whose side lengths are all equal.

For $k \in \mathbb{Z}$, let $\mathcal{Q}_k$ be the collection of cubes whose side length is 2^{-k} and whose vertices are in the lattice $(2^{-k}\mathbb{Z})^n$. (That is, $\prod_1^n [a_j, b_j] \in \mathcal{Q}_k$ iff $2^k a_j$ and $2^k b_j$ are integers and $b_j - a_j = 2^{-k}$ for all j.) Note that any two cubes in $\mathcal{Q}_k$ have disjoint interiors, and that the cubes in $\mathcal{Q}_{k+1}$ are obtained from the cubes in $\mathcal{Q}_k$ by bisecting the sides.

If $E \subset \mathbb{R}^n$, we define the inner and outer approximations to E by the grid of cubes $\mathcal{Q}_k$ to be

$$\underline{A}(E,k) = \bigcup \{Q \in \mathcal{Q}_k : Q \subset E\}, \qquad \overline{A}(E,k) = \bigcup \{Q \in \mathcal{Q}_k : Q \cap E \neq \varnothing\}.$$

(See Figure 2.2.) The measure of $\underline{A}(E,k)$ (in either the naive geometric sense or the Lebesgue sense) is just 2^{-nk} times the number of cubes in $\mathcal{Q}_k$ that lie in $\underline{A}(E,k)$, and we denote it by $m(\underline{A}(E,k))$; likewise for $m(\overline{A}(E,k))$. Also, the sets $\underline{A}(E,k)$ increase with k while the sets $\overline{A}(E,k)$ decrease, because each cube in $\mathcal{Q}_k$ is a union of cubes in $\mathcal{Q}_{k+1}$. Hence the limits

$$\underline{\kappa}(E) = \lim_{k\to\infty} m(\underline{A}(E,k)), \qquad \overline{\kappa}(E) = \lim_{k\to\infty} m(\overline{A}(E,k))$$

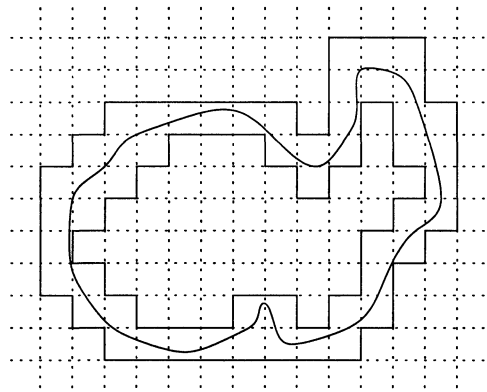

Fig. 2.2 Approximations to the inner and outer content of a set.

exist. They are called the **inner** and **outer content** of E, and if they are equal, their common value $\kappa(E)$ is the **Jordan content** of E.

Two comments: First, Jordan content is usually defined using general rectangles whose sides are intervals rather than our dyadic cubes, but the result is the same. Second, although all the definitons above make sense for arbitrary $E \subset \mathbb{R}^n$, the theory of Jordan content is meaningful only if E is bounded, for otherwise $\overline{\kappa}(E)$ always equals ∞.

Let

$$\underline{A}(E) = \bigcup_1^\infty \underline{A}(E,k), \qquad \overline{A}(E) = \bigcap_1^\infty \overline{A}(E,k).$$

Then $\underline{A}(E) \subset E \subset \overline{A}(E)$, $\underline{A}(E)$ and $\overline{A}(E)$ are Borel sets, and $\underline{\kappa}(E) = m(\underline{A}(E))$ and $\overline{\kappa}(E) = m(\overline{A}(E))$. Thus the Jordan content of E exists iff $m(\overline{A}(E) \setminus \underline{A}(E)) = 0$, which implies that E is Lebesgue measurable and $m(E) = \kappa(E)$.

To clarify further the relationship between Lebesgue measure and the approximation process leading to Jordan content, we establish the following lemma. (The second part of the lemma will be used later.)

2.43 Lemma. *If $U \subset \mathbb{R}^n$ is open, then $U = \underline{A}(U)$. Moreover, U is a countable union of cubes with disjoint interiors.*

Proof. If $x \in U$, let $\delta = \inf\{|y-x| : y \notin U\}$, which is positive since U is open. If Q is a cube in $\mathcal{Q}_k$ that contains x, then every $y \in Q$ is at a distance at most $2^{-k}\sqrt{n}$ from x (the worst case being when $|x_j - y_j| = 2^{-k}$ for all j), so we will have $Q \subset U$ provided k is large enough so that $2^{-k}\sqrt{n} < \delta$. But then $x \in \underline{A}(U,k) \subset \underline{A}(U)$.

This shows that $\underline{A}(U) = U$, and the second assertion follows by writing $\underline{A}(U) = \underline{A}(U,0) \cup \bigcup_1^\infty [\underline{A}(U,k) \setminus \underline{A}(U,k-1)]$. $\underline{A}(U,0)$ is a (countable) union of cubes in $\mathcal{Q}_0$, and for $k \geq 1$, the closure of $\underline{A}(U,k) \setminus \underline{A}(U,k-1)$ is a (countable) union of cubes in $\mathcal{Q}_k$. These cubes all have disjoint interiors, and the result follows. ∎

Lemma 2.43 immediately implies that *the Lebesgue measure of any open set is equal to its inner content.* On the other hand, suppose that $F \subset \mathbb{R}^n$ is compact. We can find a large cube, say $Q_0 = \{x : \max |x_j| \leq 2^M\}$, whose interior $\operatorname{int}(Q_0)$ contains F. If $Q \in \mathcal{Q}_k$ and $Q \subset Q_0$ then either $Q \cap F \neq \varnothing$ or $Q \subset (Q_0 \setminus F)$, so $m(\overline{A}(F,k)) + m(\underline{A}(Q_0 \setminus F, k)) = m(Q_0)$. Letting $k \to \infty$, we see that $\overline{\kappa}(F) + \underline{\kappa}(Q_0 \setminus F) = m(Q_0)$. But $Q_0 \setminus F$ is the union of the open set $\operatorname{int}(Q_0) \setminus F$ and the boundary of Q_0, which has content zero, so that $\underline{\kappa}(Q_0 \setminus F) = \underline{\kappa}(\operatorname{int}(Q_0) \setminus F) = m(Q_0 \setminus F)$. It follows that *the Lebesgue measure of any compact set is equal to its outer content.*

Combining these results with Theorem 2.40a, we can see exactly how Lebesgue measure compares to Jordan content. The Jordan content of E is defined by approximating E from the inside and the outside by finite unions of cubes. The Lebesgue measure of E, on the other hand, is given by a two-step approximation process: First one approximates E from the outside by open sets and from the inside by compact sets, and then approximates the open sets from the inside and the compact sets from the outside by finite unions of cubes. The Lebesgue measurable sets are precisely those for which these outer-inner and inner-outer approximations give the same answer in the limit. (Cf. Exercise 19 in §1.4.)

We now investigate the behavior of the Lebesgue integral under linear transformations. We identify a linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ with the matrix $(T_{ij}) = (e_i \cdot Te_j)$ where $\{e_j\}$ is the standard basis for $\mathbb{R}^n$. We denote the determinant of this matrix by $\det T$ and recall that $\det(T \circ S) = (\det T)(\det S)$. Furthermore, we employ the standard notation $GL(n, \mathbb{R})$ (the "general linear" group) for the group of invertible linear transformations of $\mathbb{R}^n$. We shall need the fact from elementary linear algebra that every $T \in GL(n, \mathbb{R})$ can be written as the product of finitely many transformations of three "elementary" types. The first type multiplies one coordinate by a nonzero constant c and leaves the others fixed; the second type adds a multiple of one coordinate to some other coordinate and leaves all but the latter fixed; the third type interchanges two coordinates and leaves the others fixed. In symbols:

$$\begin{aligned} T_1(x_1, \ldots, x_j, \ldots, x_n) &= (x_1, \ldots, cx_j, \ldots, x_n) \qquad (c \neq 0), \\ T_2(x_1, \ldots, x_j, \ldots, x_n) &= (x_1, \ldots, x_j + cx_k, \ldots, x_n) \qquad (k \neq j), \\ T_3(x_1, \ldots, x_j, \ldots, x_k, \ldots, x_n) &= (x_1, \ldots, x_k, \ldots, x_j, \ldots, x_n). \end{aligned}$$

That every invertible transformation is a product of transformations of these three types is simply the fact that every nonsingular matrix can be row-reduced to the identity matrix.

2.44 Theorem. *Suppose $T \in GL(n, \mathbb{R})$.*

a. If f is a Lebesgue measurable function on $\mathbb{R}^n$, so is $f \circ T$. If $f \geq 0$ or $f \in L^1(m)$, then

$$\int f(x)\, dx = |\det T| \int f \circ T(x)\, dx. \tag{2.45}$$

b. If $E \in \mathcal{L}^n$, then $T(E) \in \mathcal{L}^n$ and $m(T(E)) = |\det T| m(E)$.

Proof. First suppose that f is Borel measurable. Then $f \circ T$ is Borel measurable since T is continuous. If (2.45) is true for the transformations T and S, it is also true for $T \circ S$, since

$$\int f(x)\,dx = |\det T| \int f \circ T(x)\,dx = |\det T|\,|\det S| \int (f \circ T) \circ S(x)\,dx$$
$$= |\det(T \circ S)| \int f \circ (T \circ S)(x)\,dx.$$

Hence is suffices to prove (2.45) when T is of the types T_1, T_2, T_3 described above. But this is a simple consequence of the Fubini-Tonelli theorem. For T_3 we interchange the order of integration in the variables x_j and x_k, and for T_1 and T_2 we integrate first with respect to x_j and use the one-dimensional formulas

$$\int f(t)\,dt = |c| \int f(ct)\,dt, \qquad \int f(t+a)\,dt = \int f(t)\,dt,$$

which follow from Theorem 1.21. Since it is easily verified that $\det T_1 = c$, $\det T_2 = 1$, and $\det T_3 = -1$, (2.45) is proved. Moreover, if E is a Borel set, so is $T(E)$ (since T^{-1} is continuous), and by taking $f = \chi_{T(E)}$, we obtain $m(T(E)) = |\det T| m(E)$. In particular, the class of Borel null sets is invariant under T and T^{-1}, and hence so is $\mathcal{L}^n$. The result for Lebesgue measurable functions and sets now follows as in the proof of Theorem 2.42. ∎

2.46 Corollary. *Lebesgue measure is invariant under rotations.*

Proof. Rotations are linear maps satisfying $TT^* = I$ where T^* is the transpose of T. Since $\det T = \det T^*$, this condition implies that $|\det T| = 1$. ∎

Next we shall generalize Theorem 2.44 to differentiable maps. This result will not be used elsewhere in this book and may be omitted on a first reading. We shall prove a generalization of it, by somewhat different methods, in §11.2.

Let $G = (g_1, \ldots, g_n)$ be a map from an open set $\Omega \subset \mathbb{R}^n$ into $\mathbb{R}^n$ whose components g_j are of class C^1, i.e., have continuous first-order partial derivatives. We denote by $D_x G$ the linear map defined by the matrix $\big((\partial g_i/\partial x_j)(x)\big)$ of partial derivatives at x. (Observe that if G is linear, then $D_x G = G$ for all x.) G is called a C^1 **diffeomorphism** if G is injective and $D_x G$ is invertible for all $x \in \Omega$. In this case, the inverse function theorem guarantees that $G^{-1} : G(\Omega) \to \Omega$ is also a C^1 diffeomorphism and that $D_x(G^{-1}) = [D_{G^{-1}(x)} G]^{-1}$ for all $x \in G(\Omega)$.

2.47 Theorem. *Suppose that Ω is an open set in $\mathbb{R}^n$ and $G : \Omega \to \mathbb{R}^n$ is a C^1 diffeomorphism.*

a. If f is a Lebesgue measurable function on $G(\Omega)$, then $f \circ G$ is Lebesgue measurable on Ω. If $f \geq 0$ or $f \in L^1(G(\Omega), m)$, then

$$\int_{G(\Omega)} f(x)\,dx = \int_\Omega f \circ G(x) |\det D_x G|\,dx.$$

b. If $E \subset \Omega$ and $E \in \mathcal{L}^n$, then $G(E) \in \mathcal{L}^n$ and $m(G(E)) = \int_E |\det D_x G|\,dx$.

Proof. It suffices to consider Borel measurable functions and sets. Since G and G^{-1} are both continuous, there are no measurability problems in this case, and the general case follows as in the proof of Theorem 2.42.

A bit of notation: For $x \in \mathbb{R}^n$ and $T = (T_{ij}) \in GL(n, \mathbb{R})$, we set

$$\|x\| = \max_{1 \le j \le n} |x_j|, \qquad \|T\| = \max_{1 \le i \le n} \sum_{j=1}^{n} |T_{ij}|.$$

We then have $\|Tx\| \le \|T\| \, \|x\|$, and $\{x : \|x - a\| \le h\}$ is the cube of side length $2h$ centered at a.

Let Q be a cube in Ω, say $Q = \{x : \|x - a\| \le h\}$. By the mean value theorem, $g_j(x) - g_j(a) = \sum_j (x_j - a_j)(\partial g/\partial x_j)(y)$ for some y on the line segment joning x and a, so that for $x \in Q$, $\|G(x) - G(a)\| \le h(\sup_{y \in Q} \|D_y G\|)$. In other words, $G(Q)$ is contained in a cube of side length $\sup_{y \in Q} \|D_y G\|$ times that of Q, so that by Theorem 2.44, $m(G(Q)) \le (\sup_{y \in Q} \|D_y G\|)^n m(Q)$. If $T \in GL(n, \mathbb{R})$, we can apply this formula with G replaced by $T^{-1} \circ G$ together with Theorem 2.44 to obtain

$$\begin{aligned} m\big(G(Q)\big) &= |\det T| m\big(T^{-1}(G(Q))\big) \\ &\le |\det T| \Big(\sup_{y \in Q} \|T^{-1} D_y G\|\Big)^n m(Q). \end{aligned} \tag{2.48}$$

Since $D_y G$ is continuous in y, for any $\epsilon > 0$ we can choose $\delta > 0$ so that $\|(D_z G)^{-1} D_y G\|^n \le 1 + \epsilon$ if $y, z \in Q$ and $\|y - z\| \le \delta$. Let us now subdivide Q into subcubes $Q_1, \dots Q_N$ whose interiors are disjoint, whose side lengths are at most δ, and whose centers are $x_1, \dots x_N$. Applying (2.48) with Q replaced by Q_j and with $T = D_{x_j} G$, we obtain

$$\begin{aligned} m\big(G(Q)\big) &\le \sum_1^N m\big(G(Q_j)\big) \\ &\le \sum_1^N |\det D_{x_j} G| \Big(\sup_{y \in Q_j} \|(D_{x_j} G)^{-1} D_y G\|\Big)^n m(Q_j) \\ &\le (1 + \epsilon) \sum_1^N |\det D_{x_j} G| m(Q_j). \end{aligned}$$

This last sum is the integral of $\sum_1^N |\det D_{x_j} G| \chi_{Q_j}(x)$, which tends uniformly on Q to $|\det D_x G|$ as $\delta \to 0$ since $D_x G$ is continuous. Thus, letting $\delta \to 0$ and $\epsilon \to 0$, we find that

$$m\big(G(Q)\big) \le \int_Q |\det D_x G| \, dx.$$

We claim that this estimate holds with Q replaced by any Borel set in Ω. Indeed, if $U \subset \Omega$ is open, by Lemma 2.43 we can write $U = \bigcup_1^\infty Q_j$ where the Q_j's are cubes

with disjoint interiors. Since the boundaries of the cubes have Lebesgue measure zero, we have

$$m(G(U)) \le \sum_1^\infty m(G(Q_j)) \le \sum_1^\infty \int_{Q_j} |\det D_x G|\, dx = \int_U |\det D_x G|\, dx.$$

Moreover, if $E \subset \Omega$ is any Borel set of finite measure, by Theorem 2.40 there is a decreasing sequence of open sets $U_j \subset \Omega$ of finite measure such that $E \subset \bigcap_1^\infty U_j$ and $m(\bigcap_1^\infty U_j \setminus E) = 0$. Hence by the dominated convergence theorem,

$$\begin{aligned} m(G(E)) \le m\Big(G\Big(\bigcap_1^\infty U_j\Big)\Big) &= \lim m(G(U_j)) \\ &\le \lim \int_{U_j} |\det D_x G|\, dx = \int_E |\det D_x G|\, dx. \end{aligned}$$

Finally, since m is σ-finite, it follows from this that $m(G(E)) \le \int_E |\det D_x G|\, dx$ for any Borel set $E \subset \Omega$.

If $f = \sum a_j \chi_{A_j}$ is a nonnegative simple function on $G(\Omega)$, we therefore have

$$\begin{aligned} \int_{G(\Omega)} f(x)\, dx = \sum a_j m(A_j) &\le \sum a_j \int_{G^{-1}(A_j)} |\det D_x G|\, dx \\ &= \int_\Omega f \circ G(x) |\det D_x G|\, dx. \end{aligned}$$

Theorem 2.10 and the monotone convergence theorem then imply that

$$\int_{G(\Omega)} f(x)\, dx \le \int_\Omega f \circ G(x) |\det D_x G|\, dx$$

for any nonnegative measurable f. But the same reasoning applies with G replaced by G^{-1} and f replaced by $f \circ G$, so that

$$\begin{aligned} &\int_\Omega f \circ G(x) |\det D_x G|\, dx \\ &\quad \le \int_{G(\Omega))} f \circ G \circ G^{-1}(x) |\det D_{G^{-1}(x)} G|\, |\det D_x G^{-1}|\, dx = \int_{G(\Omega)} f(x)\, dx. \end{aligned}$$

This establishes (a) for $f \ge 0$, and the case $f \in L^1$ follows immediately. Since (b) is just the special case of (a) where $f = \chi_{G(E)}$, the proof is complete. ∎

Exercises

53. Fill in the details of the proof of Theorem 2.41.

54. How much of Theorem 2.44 remains valid if T is not invertible?

55. Let $E = [0,1] \times [0,1]$. Investigate the existence and equality of $\int_E f\,dm^2$, $\int_0^1 \int_0^1 f(x,y)\,dx\,dy$, and $\int_0^1 \int_0^1 f(x,y)\,dy\,dx$ for the following f.

a. $f(x,y) = (x^2 - y^2)(x^2 + y^2)^{-2}$.

b. $f(x,y) = (1 - xy)^{-a}$ $(a > 0)$.

c. $f(x,y) = (x - \frac{1}{2})^{-3}$ if $0 < y < |x - \frac{1}{2}|$, $f(x,y) = 0$ otherwise.

56. If f is Lebesgue integrable on $(0,a)$ and $g(x) = \int_x^a t^{-1} f(t)\,dt$, then g is integrable on $(0,a)$ and $\int_0^a g(x)\,dx = \int_0^a f(x)\,dx$.

57. Show that $\int_0^\infty e^{-sx} x^{-1} \sin x\,dx = \arctan(s^{-1})$ for $s > 0$ by integrating $e^{-sxy} \sin x$ with respect to x and y. (It may be useful to recall that $\tan(\frac{\pi}{2} - \theta) = (\tan\theta)^{-1}$. Cf. Exercise 31d.)

58. Show that $\int e^{-sx} x^{-1} \sin^2 x\,dx = \frac{1}{4}\log(1 + 4s^{-2})$ for $s > 0$ by integrating $e^{-sx} \sin 2xy$ with respect to x and y.

59. Let $f(x) = x^{-1} \sin x$.

a. Show that $\int_0^\infty |f(x)|\,dx = \infty$.

b. Show that $\lim_{b\to\infty} \int_0^b f(x)\,dx = \frac{1}{2}\pi$ by integrating $e^{-xy} \sin x$ with respect to x and y. (In view of part (a), some care is needed in passing to the limit as $b \to \infty$.)

60. $\Gamma(x)\Gamma(y)/\Gamma(x+y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$ for $x, y > 0$. (Recall that Γ was defined in §2.3. Write $\Gamma(x)\Gamma(y)$ as a double integral and use the argument of the exponential as a new variable of integration.)

61. If f is continuous on $[0,\infty)$, for $\alpha > 0$ and $x \geq 0$ let

$$I_\alpha f(x) = \frac{1}{\Gamma(\alpha)} \int_0^x (x-t)^{\alpha-1} f(t)\,dt.$$

$I_\alpha f$ is called the αth **fractional integral** of f.

a. $I_{\alpha+\beta} f = I_\alpha(I_\beta f)$ for all $\alpha, \beta > 0$. (Use Exercise 60.)

b. If $n \in \mathbb{N}$, $I_n f$ is an nth-order antiderivative of f.

2.7 INTEGRATION IN POLAR COORDINATES

The most important nonlinear coordinate systems in $\mathbb{R}^2$ and $\mathbb{R}^3$ are polar coordinates ($x = r\cos\theta$, $y = r\sin\theta$) and spherical coordinates ($x = r\sin\phi\cos\theta$, $y = r\sin\phi\sin\theta$, $z = r\cos\phi$). Theorem 2.47, applied to these coordinates, yields the familiar formulas (loosely stated) $dx\,dy = r\,dr\,d\theta$ and $dx\,dy\,dz = r^2 \sin\phi\,dr\,d\theta\,d\phi$. Similar coordinate systems exist in higher dimensions, but they become increasingly complicated as the dimension increases. (See Exercise 65.) For most purposes, however, it is sufficient to know that Lebesgue measure is effectively the product of the measure $r^{n-1}\,dr$ on $(0,\infty)$ and a certain "surface measure" on the unit sphere ($d\theta$ for $n = 2$, $\sin\phi\,d\theta\,d\phi$ for $n = 3$).

Our construction of this surface measure is motivated by a familiar fact from plane geometry. Namely, if S_θ is a sector of a disc of radius r with central angle θ (i.e., the region in the disc contained between the two sides of the angle), the area $m(S_\theta)$ is proportional to θ; in fact, $m(S_\theta) = \frac{1}{2}r^2\theta$. This equation can be solved for θ and hence used to *define* the angular measure θ in terms of the area $m(S_\theta)$. The same idea works in higher dimensions: We shall define the surface measure of a subset of the unit sphere in terms of the Lebesgue measure of the corresponding sector of the unit ball.

We shall denote the unit sphere $\{x \in \mathbb{R}^n : |x| = 1\}$ by S^{n-1}. If $x \in \mathbb{R}^n \setminus \{0\}$, the **polar coordinates** of x are

$$r = |x| \in (0, \infty), \qquad x' = \frac{x}{|x|} \in S^{n-1}.$$

The map $\Phi(x) = (r, x')$ is a continuous bijection from $\mathbb{R}^n \setminus \{0\}$ to $(0, \infty) \times S^{n-1}$ whose (continuous) inverse is $\Phi^{-1}(r, x') = rx'$. We denote by m_* the Borel measure on $(0, \infty) \times S^{n-1}$ induced by Φ from Lebesgue measure on $\mathbb{R}^n$, that is, $m_*(E) = m(\Phi^{-1}(E))$. Moreover, we define the measure $\rho = \rho_n$ on $(0, \infty)$ by $\rho(E) = \int_E r^{n-1}\, dr$.

2.49 Theorem. *There is a unique Borel measure $\sigma = \sigma_{n-1}$ on S^{n-1} such that $m_* = \rho \times \sigma$. If f is Borel measurable on $\mathbb{R}^n$ and $f \geq 0$ or $f \in L^1(m)$, then*

$$\int_{\mathbb{R}^n} f(x)\, dx = \int_0^\infty \int_{S^{n-1}} f(rx') r^{n-1}\, d\sigma(x')\, dr. \tag{2.50}$$

Proof. Equation (2.50), when f is a characteristic function of a set, is merely a restatement of the equation $m_* = \rho \times \sigma$, and it follows for general f by the usual linearity and approximation arguments. Hence we need only to construct σ.

If E is a Borel set in S^{n-1}, for $a > 0$ let

$$E_a = \Phi^{-1}\big((0, a] \times E\big) = \big\{rx' : 0 < r \leq a,\ x' \in E\big\}.$$

If (2.50) is to hold when $f = \chi_{E_1}$, we must have

$$m(E_1) = \int_0^1 \int_E r^{n-1}\, d\sigma(x')\, dr = \sigma(E) \int_0^1 r^{n-1}\, dr = \frac{\sigma(E)}{n}.$$

We therefore *define* $\sigma(E)$ to be $n \cdot m(E_1)$. Since the map $E \mapsto E_1$ takes Borel sets to Borel sets and commutes with unions, intersections, and complements, it is clear that σ is a Borel measure on S^{n-1}. Also, since E_a is the image of E_1 under the map $x \mapsto ax$, it follows from Theorem 2.44 that $m(E_a) = a^n m(E_1)$, and hence, if $0 < a < b$,

$$\begin{aligned} m_*\big((a, b] \times E\big) &= m(E_b \setminus E_a) = \frac{b^n - a^n}{n}\sigma(E) = \sigma(E) \int_a^b r^{n-1}\, dr \\ &= \rho \times \sigma\big((a, b] \times E\big). \end{aligned}$$

Fix $E \in \mathcal{B}_{S^{n-1}}$ and let $\mathcal{A}_E$ be the collection of finite disjoint unions of sets of the form $(a, b] \times E$. By Proposition 1.7, $\mathcal{A}_E$ is an algebra on $(0, \infty) \times E$ that generates the σ-algebra $\mathcal{M}_E = \{A \times E : A \in \mathcal{B}_{(0,\infty)}\}$. By the preceding calculation we have $m_* = \rho \times \sigma$ on $\mathcal{A}_E$, and hence by the uniqueness assertion of Theorem 1.14, $m_* = \rho \times \sigma$ on $\mathcal{M}_E$. But $\bigcup\{\mathcal{M}_E : E \in \mathcal{B}_{S^{n-1}}\}$ is precisely the set of Borel rectangles in $(0, \infty) \times S^{n-1}$, so another application of the uniqueness theorem shows that $m_* = \rho \times \sigma$ on all Borel sets. ∎

Of course, (2.50) can be extended to Lebesgue measurable functions by considering the completion of the measure σ. Details are left to the reader.

2.51 Corollary. *If f is a measurable function on $\mathbb{R}^n$, nonnegative or integrable, such that $f(x) = g(|x|)$ for some function g on $(0, \infty)$, then*

$$\int f(x)\, dx = \sigma(S^{n-1}) \int_0^\infty g(r) r^{n-1}\, dr.$$

2.52 Corollary. *Let c and C denote positive constants, and let $B = \{x \in \mathbb{R}^n : |x| < c\}$. Suppose that f is a measurable function on $\mathbb{R}^n$.*

a. If $|f(x)| \le C|x|^{-a}$ on B for some $a < n$, then $f \in L^1(B)$. However, if $|f(x)| \ge C|x|^{-n}$ on B, then $f \notin L^1(B)$.

b. If $|f(x)| \le C|x|^{-a}$ on B^c for some $a > n$, then $f \in L^1(B^c)$. However, if $|f(x)| \ge C|x|^{-n}$ on B^c, then $f \notin L^1(B^c)$.

Proof. Apply Corollary 2.51 to $|x|^{-a}\chi_B$ and $|x|^{-a}\chi_{B^c}$. ∎

We shall compute $\sigma(S^{n-1})$ shortly. Of course, we know that $\sigma(S^1) = 2\pi$; this is just the definition of 2π as the ratio of the circumference of a circle to its radius. Armed with this fact, we can compute a very important integral.

2.53 Proposition. *If $a > 0$,*

$$\int_{\mathbb{R}^n} \exp(-a|x|^2)\, dx = \left(\frac{\pi}{a}\right)^{n/2}.$$

Proof. Denote the integral on the left by I_n. For $n = 2$, by Corollary 2.51 we have

$$I_2 = 2\pi \int_0^\infty r e^{-ar^2}\, dr = -\left(\frac{\pi}{a}\right) e^{-ar^2}\Big|_0^\infty = \frac{\pi}{a}.$$

Since $\exp(-a|x|^2) = \prod_1^n \exp(-ax_j^2)$, Tonelli's theorem implies that $I_n = (I_1)^n$. In particular, $I_1 = (I_2)^{1/2}$, so $I_n = (I_2)^{n/2} = (\pi/a)^{n/2}$. ∎

Once we know this result, the device used in its proof can be turned around to compute $\sigma(S^{n-1})$ for all n in terms of the gamma function introduced in §2.3.

2.54 Proposition. $\sigma(S^{n-1}) = \dfrac{2\pi^{n/2}}{\Gamma(n/2)}$.

Proof. By Corollary 2.51, Proposition 2.53, and the substitution $s = r^2$,

$$\pi^{n/2} = \int_{\mathbb{R}^n} e^{-|x|^2}\,dx = \sigma(S^{n-1}) \int_0^\infty r^{n-1} e^{-r^2}\,dr$$
$$= \frac{\sigma(S^{n-1})}{2} \int_0^\infty s^{(n/2)-1} e^{-s}\,ds = \frac{\sigma(S^{n-1})}{2}\Gamma\left(\frac{n}{2}\right).$$

∎

2.55 Corollary. *If* $B^n = \{x \in \mathbb{R}^n : |x| < 1\}$, *then* $m(B^n) = \dfrac{\pi^{n/2}}{\Gamma(\frac{1}{2}n + 1)}$.

Proof. $m(B^n) = n^{-1}\sigma(S^{n-1})$ by definition of σ, and $\frac{1}{2}n\Gamma(\frac{1}{2}n) = \Gamma(\frac{1}{2}n + 1)$ by the functional equation for the gamma function. ∎

We observed in §2.3 that $\Gamma(n) = (n-1)!$. Now we can also evaluate the gamma function at the half-integers:

2.56 Proposition. $\Gamma(n + \frac{1}{2}) = (n - \frac{1}{2})(n - \frac{3}{2}) \cdots (\frac{1}{2})\sqrt{\pi}$.

Proof. We have $\Gamma(n + \frac{1}{2}) = (n - \frac{1}{2})(n - \frac{3}{2}) \cdots (\frac{1}{2})\Gamma(\frac{1}{2})$ by the functional equation, and by Proposition 2.53 and the substitution $s = r^2$,

$$\Gamma(\tfrac{1}{2}) = \int_0^\infty s^{-1/2} e^{-s}\,ds = 2\int_0^\infty e^{-r^2}\,dr = \int_{-\infty}^\infty e^{-r^2}\,dr = \sqrt{\pi}.$$

∎

An amusing consequence of Proposition 2.56 and the formula $\Gamma(n) = (n-1)!$ is that the surface measure of the unit sphere and the Lebesgue measure of the unit ball in $\mathbb{R}^n$ are always rational multiples of integer powers of π, and the power of π increases by 1 when n increases by 2.

Exercises

62. The measure σ on S^{n-1} is invariant under rotations.

63. The technique used to prove Proposition 2.54 can also be used to integrate any polynomial over S^{n-1}. In fact, suppose $f(x) = \prod_1^n x_j^{\alpha_j}$ ($\alpha_j \in \mathbb{N} \cup \{0\}$) is a monomial. Then $\int f\,d\sigma = 0$ if any α_j is odd, and if all α_j's are even,

$$\int f\,d\sigma = \frac{2\Gamma(\beta_1)\cdots\Gamma(\beta_n)}{\Gamma(\beta_1 + \cdots + \beta_n)}, \text{ where } \beta_j = \frac{\alpha_j + 1}{2}.$$

64. For which real values of a and b is $|x|^a |\log|x||^b$ integrable over $\{x \in \mathbb{R}^n : |x| < \frac{1}{2}\}$? Over $\{x \in \mathbb{R}^n : |x| > 2\}$?

65. Define $G : \mathbb{R}^n \to \mathbb{R}^n$ by $G(r, \phi_1, \ldots, \phi_{n-2}, \theta) = (x_1, \ldots, x_n)$ where

$$x_1 = r\cos\phi_1, \quad x_2 = r\sin\phi_1\cos\phi_2, \quad x_3 = r\sin\phi_1\sin\phi_2\cos\phi_3, \ldots,$$
$$x_{n-1} = r\sin\phi_1\cdots\sin\phi_{n-2}\cos\theta, \quad x_n = r\sin\phi_1\cdots\sin\phi_{n-2}\sin\theta.$$

a. G maps $\mathbb{R}^n$ onto $\mathbb{R}^n$, and $|G(r, \phi_1, \ldots, \phi_{n-2}, \theta)| = |r|$.
b. $\det D_{(r,\phi_1,\ldots,\phi_{n-2},\theta)} G = r^{n-1} \sin^{n-2} \phi_1 \sin^{n-3} \phi_2 \cdots \sin \phi_{n-2}$.
c. Let $\Omega = (0, \infty) \times (0, \pi)^{n-2} \times (0, 2\pi)$. Then $G|\Omega$ is a diffeomorphism and $m(\mathbb{R}^n \setminus G(\Omega)) = 0$.
d. Let $F(\phi_1, \ldots, \phi_{n-2}, \theta) = G(1, \phi_1, \ldots, \phi_{n-2}, \theta)$ and $\Omega' = (0, \pi)^{n-2} \times (0, 2\pi)$. Then $(F|\Omega')^{-1}$ defines a coordinate system on S^{n-1} except on a σ-null set, and the measure σ is given in these coordinates by

$$d\sigma(\phi_1, \ldots \phi_{n-2}, \theta) = \sin^{n-2} \phi_1 \sin^{n-3} \phi_2 \cdots \sin \phi_{n-2} \, d\phi_1 \cdots d\phi_{n-2} \, d\theta.$$

2.8 NOTES AND REFERENCES

The history of modern measure and integration theory can fairly be said to have begun with the publication of Lebesgue's thesis [91] in 1902, although of course Lebesgue was building on earlier works of other mathematicians, and some of his results were obtained independently by Vitali and W. H. Young. The theory of the Lebesgue integral was extensively developed by a number of mathematicians in the ensuing decade, during which time most of the results in this chapter were first derived. In particular, Lebesgue himself proved the dominated convergence theorem and deduced the monotone convergence theorem from it in the case when the limit function f is integrable; when $\int f = \infty$ the latter theorem is due to B. Levi.

Lebesgue [92] studied more general measures on $\mathbb{R}^n$ (which he called "additive set functions") in connection with the problem of generalizing the notion of indefinite integrals to functions of several variables. Radon [111] then developed the theory of integration with respect to what we now call regular Borel measures on $\mathbb{R}^n$, which in particular yields the Lebesgue-Stieltjes integrals when $n = 1$. Finally, in 1915 Fréchet [53] pointed out that many of Radon's ideas would work in the general setting of sets equipped with σ-algebras. Thus was abstract measure and integration theory born. It continued to develop until, by about 1950, it had assumed more or less the form in which we know it today. The first systematic modern treatise on the subject is Halmos [62].

For accounts of the prehistory and early history of the Lebesgue integral, see Hawkins [70]. References concerning the later development of the subject can be found in Saks [128] and Hahn and Rosenthal [61].

We have adopted the point of view of beginning with measures and deriving integrals from them. However, it is also possible to go the other way, a procedure first developed by Daniell [29]. Roughly speaking, one starts with an "elementary integral": a linear functional I defined on a suitable space of functions that satisfies some mild continuity conditions and is positive in the sense that $I(f) \geq 0$ whenever $f \geq 0$ (for example, the Riemann integral on the space of continuous functions on $[a, b]$). The Daniell theory provides an extension of I to a functional $\overline{I}$ defined on a larger class of functions. Under appropriate hypotheses, the collection $\mathcal{M}$ of sets E such that χ_E is in the domain of $\overline{I}$ is then a σ-algebra, the function $\mu(E) = \overline{I}(\chi_E)$ is a

measure on $\mathcal{M}$, and $\overline{I}$ is integration with respect to μ. See Royden [121] for a concise account of the Daniell theory and Pfeffer [108] for a comprehensive treatment, as well as König [86] for a somewhat different approach.

The Lebesgue theory is not the last word regarding integration on $\mathbb{R}$. Motivated partly by the problem of establishing the fundamental theorem of calculus in the greatest possible generality (about which we shall say more in §3.6), a number of theories of integration have been developed that include not only the Lebesgue integral but also certain "conditionally convergent" integrals. That is, they assign a meaning to $\int f(x)\,dx$ for certain measurable functions $f : \mathbb{R} \to \mathbb{R}$ such that $\int f^+ = \int f^- = \infty$, but for which the cancellation of positive and negative values in some way yields a reasonable definition of $\int f(x)\,dx$. (A standard example is $f(x) = x^{-1}\sin x$; see Exercise 59.) The first procedures for defining such integrals, due to Denjoy and Perron, were quite complicated. However, in the late 1950s, Henstock and Kurzweil independently discovered a modification of the classical Riemann integral that yields the same results.

The Henstock-Kurzweil integral on a bounded interval $[a, b]$ is defined as follows. A **tagged partition** of $[a, b]$ is a finite sequence $\{x_j\}_0^N$ such that $a = x_0 < \cdots < x_N = b$ (i.e., a partition in the sense of §2.3) together with another finite sequence $\{t_j\}_1^N$ such that $t_j \in [x_{j-1}, x_j]$. A **gauge** on $[a, b]$ is an (arbitrary!) function $\delta : [a, b] \to (0, \infty)$. If P is a tagged partition and δ is a gauge, P is called **δ-fine** if $x_j - x_{j-1} < \delta(t_j)$ for all j. The compactness of $[a, b]$ easily implies that for any gauge δ there is a δ-fine tagged partition of $[a, b]$.

Now suppose f is a real-valued function on $[a, b]$. If P is a tagged partition of $[a, b]$, the corresponding Riemann sum for f is $\Sigma_P f = \sum_1^N f(t_j)(x_j - x_{j-1})$. The function f is called **Henstock-Kurzweil integrable** on $[a, b]$ if there exists $c \in \mathbb{R}$ with the following property: For any $\epsilon > 0$ there is a gauge δ_ϵ such that if P is any δ_ϵ-fine tagged partition of $[a, b]$, then $|\Sigma_P f - c| < \epsilon$. In this case the number c is unique, and it is called the **Henstock-Kurzweil integral** of f. The ordinary Riemann integral of f, in contrast, can be defined in exactly the same way except that one allows only *constant* gauges.

It turns out that the Henstock-Kurzweil integral coincides with the integrals of Denjoy and Perron. In particular, it coincides with the Lebesgue integral for nonnegative functions, but its domain includes many functions that have both positive and negative values and are not in $L^1([a, b])$. The definition of the Henstock-Kurzweil integral is easily extended to unbounded intervals. It also admits an n-dimensional version: One simply defines an n-interval to be a product of n one-dimensional intervals and a tagged partition of an n-interval I to be a finite collection $\{I_j\}$ of n-intervals with disjoint interiors whose union is I together with a choice of $t_j \in I_j$ for each j; the definition of the integral then proceeds as above.

A good case can be made that the Henstock-Kurzweil integral ought to be *the* theory of integration on $\mathbb{R}^n$ that is generally taught to students, not just because of its added generality but (more cogently) because its definition is relatively simple and requires no measure theory to get started. On the other hand, it does not generalize as readily to spaces other than $\mathbb{R}^n$, and although it can be developed in a rather abstract setting, it loses much of its appealing simplicity there. Moreover, although conditionally

convergent integrals that cannot be obtained by a simple limiting procedure from absolutely convergent ones do turn up now and then in certain problems, their utility is not sufficently broad to make a compelling case for their study by nonspecialists.

In any case, in this book we shall content ourselves with the Lebesgue integral and the general theory of measure and integration of which it is a part. Readers who wish to learn more about the Henstock-Kurzweil integral can find a brief introduction in Bartle [13] and detailed treatments in McLeod [99] and Pfeffer [109]. See also Gordon [57] for a comprehensive account of the Denjoy, Perron, and Henstock-Kurzweil integrals on $[a, b]$, and Henstock [72] for a development of the theory in a more abstract setting.

§2.1: A **Borel isomorphism** between two measurable spaces $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ is a bijection $f : X \to Y$ such that f^{-1} is a bijection from $\mathcal{N}$ to $\mathcal{M}$. Unlike the related notion of homeomorphism for topological spaces (see Chapter 4) and notions of isomorphism in various other categories, the notion of Borel isomorphism is of limited utility, because it is too easy for two spaces to be Borel isomorphic. That this is so is clearly indicated by the single major theorem in the subject, due to Kuratowski:

> Suppose that $(X, \mathcal{M})$ is Borel isomorphic to a Borel subset E of a complete separable metric space Y (equipped with the σ-algebra $\{F \in \mathcal{B}_Y : F \subset E\}$). Then either X is countable and $\mathcal{M} = \mathcal{P}(X)$, or X is Borel isomorphic to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

A proof of this theorem, as well as much additional information about Borel sets, can be found in Srivastava [139].

There is a hierarchy of Borel measurable functions on a metric space that corresponds roughly to the hierarchy of Borel sets (open and closed, F_σ and G_δ, etc.). Namely, let B_0 be the space of all continuous functions, and for each countable ordinal α define B_α recursively as follows. If α has an immediate predecessor β, B_α is the set of all limits of pointwise convergent sequences in B_β; otherwise, $B_\alpha = \bigcup_{\beta<\alpha} B_\beta$. Functions in B_α are said to be of **Baire class** α. For example, if f is everywhere differentiable on $\mathbb{R}$, f' is of Baire class 1.

Exercise 11 is a result from Lebesgue's first published paper. See Rudin [123] for a discussion of it.

§2.3: The blurring of the distinction between individual measurable functions and equivalence classes of functions defined by almost-everywhere equality is often convenient and rarely disastrous. The most common situations where some care is needed involve the interplay of measurable and continuous functions (on $\mathbb{R}^n$, say), for a function that is equal a.e. to a continuous function will not be continuous in general. See Zaanen [165] for a careful discussion of this point.

§2.4: An interesting discussion of Egoroff's theorem, including some necessary and sufficient conditions for almost uniform convergence, can be found in Bartle [12]. For a simple proof of Lusin's theorem (Exercise 44) that does not depend on Egoroff's theorem, see Feldman [43]. We shall prove a more general form of this theorem in §7.2.

§2.5: The original theorems of Fubini and Tonelli pertained to Lebesgue measure in the plane. The theory of abstract product measures was developed independently by several people in the 1930s; the construction of $\mu \times \nu$ presented here is that of Hahn [60]. It is also possible to define a product measure on the product of an infinite family $\{(X_\alpha, \mathcal{M}_\alpha, \mu_\alpha)\}_{\alpha \in A}$ of measure spaces provided that $\mu_\alpha(X_\alpha) = 1$ for all but finitely many α; see Saeki [127], Halmos [62, §38], or Hewitt and Stromberg [76, §22]. We shall present a version of this result in §7.4 (Theorem 7.28).

Using the axiom of choice but not the continuum hypothesis, Sierpiński [134] has proved the existence of a Lebesgue nonmeasurable subset of $\mathbb{R}^2$ whose intersection with any straight line contains at most two points. This should be compared with Exercise 47 (which is also due to Sierpiński).

The following generalization of the notion of product measures is useful in a number of situations: One is given a measurable space $(X, \mathcal{M})$, a σ-finite measure space $(Y, \mathcal{N}, \nu)$, and a family $\{\mu_y : y \in Y\}$ of finite measures on X such that the function $y \mapsto \mu_y(E)$ is measurable on Y for each $E \in \mathcal{M}$. One can then define a measure λ on $X \times Y$ such that $\int f \, d\lambda = \iint f(x, y) \, d\mu_y(x) \, d\nu(y)$ for $f \in L^+(X \times Y)$. See Johnson [79].

§2.6: Our proof of Theorem 2.47 follows J. Schwartz [131]. This theorem can also be proved under slightly weaker hypotheses on the transformation G; see Rudin [125, Theorem 7.26].

3

Signed Measures and Differentiation

The principal theme of this chapter is the concept of differentiating a measure ν with respect to another measure μ on the same σ-algebra. We do this first on the abstract level, then obtain a more refined result when μ is Lebesgue measure on $\mathbb{R}^n$. When the latter is specialized to the case $n = 1$, it joins with classical real-variable theory to produce a version of the fundamental theorem of calculus for Lebesgue integrals.

In developing this program it is useful to generalize the notion of measure so as to allow measures to assume negative or even complex values. There are three reasons for this. First, in applications such "signed measures" can represent things such as electric charge that can be either positive or negative. Second, the differentiation theory proceeds more naturally in the more general seting. Finally, complex measures have a functional-analytic significance that will be explained in Chapter 7.

3.1 SIGNED MEASURES

Let $(X, \mathcal{M})$ be a measurable space. A **signed measure** on $(X, \mathcal{M})$ is a function $\nu : \mathcal{M} \to [-\infty, \infty]$ such that

- $\nu(\varnothing) = 0$;
- ν assumes at most one of the values $\pm\infty$;
- if $\{E_j\}$ is a sequence of disjoint sets in $\mathcal{M}$, then $\nu(\bigcup_1^\infty E_j) = \sum_1^\infty \nu(E_j)$, where the latter sum converges absolutely if $\nu(\bigcup_1^\infty E_j)$ is finite.

Thus every measure is a signed measure; for emphasis we shall sometimes refer to measures as **positive measures**.

Two examples of signed measures come readily to mind. First, if μ_1, μ_2 are measures on $\mathcal{M}$ and at least one of them is finite, then $\nu = \mu_1 - \mu_2$ is a signed measure. Second, if μ is a measure on $\mathcal{M}$ and $f : X \to [-\infty, \infty]$ is a measurable function such that at least one of $\int f^+ \, d\mu$ and $\int f^- \, d\mu$ is finite (in which case we shall call f an **extended μ-integrable** function), then the set function ν defined by $\nu(E) = \int_E f \, d\mu$ is a signed measure. In fact, we shall see shortly that these are really the *only* examples: Every signed measure can be represented in either of these two forms.

3.1 Proposition. *Let ν be a signed measure on $(X, \mathcal{M})$. If $\{E_j\}$ is an increasing sequence in $\mathcal{M}$, then $\nu(\bigcup_1^\infty E_j) = \lim_{j\to\infty} \nu(E_j)$. If $\{E_j\}$ is a decreasing sequence in $\mathcal{M}$ and $\nu(E_1)$ is finite, then $\nu(\bigcap_1^\infty E_j) = \lim_{j\to\infty} \nu(E_j)$.*

The proof is essentially the same as for positive measures (Theorem 1.8) and is left to the reader (Exercise 1).

If ν is a signed measure on $(X, \mathcal{M})$, a set $E \in \mathcal{M}$ is called **positive** (resp. **negative, null**) for ν if $\nu(F) \geq 0$ (resp. $\nu(F) \leq 0$, $\nu(F) = 0$) for all $F \in \mathcal{M}$ such that $F \subset E$. (Thus, in the example $\nu(E) = \int_E f \, d\mu$ described above, E is positive, negative, or null precisely when $f \geq 0$, $f \leq 0$, or $f = 0$ μ-a.e. on E.)

3.2 Lemma. *Any measurable subset of a positive set is positive, and the union of any countable family of positive sets is positive.*

Proof. The first assertion is obvious from the definition of positivity. If $P_1, P_2, \ldots$ are positive sets, let $Q_n = P_n \setminus \bigcup_1^{n-1} P_j$. Then $Q_n \subset P_n$, so Q_n is positive. Hence if $E \subset \bigcup_1^\infty P_j$, then $\nu(E) = \sum_1^\infty \nu(E \cap Q_j) \geq 0$, as desired. ∎

3.3 The Hahn Decomposition Theorem. *If ν is a signed measure on $(X, \mathcal{M})$, there exist a positive set P and a negative set N for ν such that $P \cup N = X$ and $P \cap N = \varnothing$. If P', N' is another such pair, then $P \triangle P'$ $(= N \triangle N')$ is null for ν.*

Proof. Without loss of generality, we assume that ν does not assume the value $-\infty$. (Otherwise, consider $-\nu$.) Let m be the supremum of $\nu(E)$ as E ranges over all positive sets; thus there is a sequence $\{P_j\}$ of positive sets such that $\nu(P_j) \to m$. Let $P = \bigcup_1^\infty P_j$. By Lemma 3.2 and Proposition 3.1, P is positive and $\nu(P) = m$; in particular, $m < \infty$. We claim that $N = X \setminus P$ is negative. To this end, we assume that N is not negative and derive a contradiction.

First, notice that N cannot contain any nonnull positive sets. Indeed, if $E \subset N$ is positive and $\nu(E) > 0$, then $E \cup P$ is positive and $\nu(E \cup P) = \nu(E) + \nu(P) > m$, which is impossible.

Second, if $A \subset N$ and $\nu(A) > 0$, there exists $B \subset A$ with $\nu(B) > \nu(A)$. Indeed, since A cannot be positive, there exists $C \subset A$ with $\nu(C) < 0$; thus if $B = A \setminus C$ we have $\nu(B) = \nu(A) - \nu(C) > \nu(A)$.

If N is not negative, then, we can specify a sequence of subsets $\{A_j\}$ of N and a sequence $\{n_j\}$ of positive integers as follows: n_1 is the smallest integer for which there exists a set $B \subset N$ with $\nu(B) > n_1^{-1}$, and A_1 is such a set. Proceeding

inductively, n_j is the smallest integer for which there exists a set $B \subset A_{j-1}$ with $\nu(B) > \nu(A_{j-1}) + n_j^{-1}$, and A_j is such a set.

Let $A = \bigcap_1^\infty A_j$. Then $\infty > \nu(A) = \lim \nu(A_j) > \sum_1^\infty n_j^{-1}$, so $n_j \to \infty$ as $j \to \infty$. But once again, there exists $B \subset A$ with $\nu(B) > \nu(A) + n^{-1}$ for some integer n. For j sufficiently large we have $n < n_j$, and $B \subset A_{j-1}$, which contradicts the construction of n_j and A_j. Thus the assumption that N is not negative is untenable.

Finally, if P', N' is another pair of sets as in the statement of the theorem, we have $P \setminus P' \subset P$ and $P \setminus P' \subset N'$, so that $P \setminus P'$ is both positive and negative, hence null; likewise for $P' \setminus P$. ∎

The decomposition $X = P \cup N$ if X as the disjoint union of a positive set and a negative set is called a **Hahn decomposition** for ν. It is usually not unique (ν-null sets can be transferred from P to N or from N to P), but it leads to a canonical representation of ν as the difference of two positive measures.

To state this result we need a new concept: We say that two signed measures μ and ν on $(X, \mathcal{M})$ are **mutually singular**, or that ν is **singular with respect to** μ, or vice versa, if there exist $E, F \in \mathcal{M}$ such that $E \cap F = \varnothing$, $E \cup F = X$, E is null for μ, and F is null for ν. Informally speaking, mutual singularity means that μ and ν "live on disjoint sets." We express this relationship symbolically with the perpendicularity sign:

$$\mu \perp \nu.$$

3.4 The Jordan Decomposition Theorem. *If ν is a signed measure, there exist unique positive measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$.*

Proof. Let $X = P \cup N$ be a Hahn decomposition for ν, and define $\nu^+(E) = \nu(E \cap P)$ and $\nu^-(E) = -\nu(E \cap N)$. Then clearly $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$. If also $\nu = \mu^+ - \mu^-$ and $\mu^+ \perp \mu^-$, let $E, F \in \mathcal{M}$ be such that $E \cap F = \varnothing$, $E \cup F = X$, and $\mu^+(F) = \mu^-(E) = 0$. Then $X = E \cup F$ is another Hahn decomposition for ν, so $P \triangle E$ is ν-null. Therefore, for any $A \in \mathcal{M}$, $\mu^+(A) = \mu^+(A \cap E) = \nu(A \cap E) = \nu(A \cap P) = \nu^+(A)$, and likewise $\nu^- = \mu^-$. ∎

The measures ν^+ and ν^- are called the **positive** and **negative variations** of ν, and $\nu = \nu^+ - \nu^-$ is called the **Jordan decomposition** of ν, by analogy with the representation of a function of bounded variation on $\mathbb{R}$ as the difference of two increasing functions (see §3.5). Furthermore, we define the **total variation** of ν to be the measure $|\nu|$ defined by

$$|\nu| = \nu^+ + \nu^-.$$

It is easily verified that $E \in \mathcal{M}$ is ν-null iff $|\nu|(E) = 0$, and $\nu \perp \mu$ iff $|\nu| \perp \mu$ iff $\nu^+ \perp \mu$ and $\nu^- \perp \mu$ (Exercise 2.)

We observe that if ν omits the value ∞ then $\nu^+(X) = \nu(P) < \infty$, so that ν^+ is a finite measure and ν is bounded above by $\nu^+(X)$; similarly if ν omits the value $-\infty$. In particular, if the range of ν is contained in $\mathbb{R}$, then ν is bounded. We observe also

that ν is of the form $\nu(E) = \int_E f\,d\mu$, where $\mu = |\nu|$ and $f = \chi_P - \chi_N$, $X = P \cup N$ being a Hahn decomposition for ν.

Integration with respect to a signed measure ν is defined in the obvious way: We set

$$L^1(\nu) = L^1(\nu^+) \cap L^1(\nu^-),$$
$$\int f\,d\nu = \int f\,d\nu^+ - \int f\,d\nu^- \qquad (f \in L^1(\nu)).$$

One more piece of terminology: a signed measure ν is called **finite** (resp. **σ-finite**) if $|\nu|$ is finite (resp. σ-finite).

Exercises

1. Prove Proposition 3.1.

2. If ν is a signed measure, E is ν-null iff $|\nu|(E) = 0$. Also, if ν and μ are signed measures, $\nu \perp \mu$ iff $|\nu| \perp \mu$ iff $\nu^+ \perp \mu$ and $\nu^- \perp \mu$.

3. Let ν be a signed measure on $(X, \mathcal{M})$.
a. $L^1(\nu) = L^1(|\nu|)$.
b. If $f \in L^1(\nu)$, $|\int f\,d\nu| \le \int |f|\,d|\nu|$.
c. If $E \in \mathcal{M}$, $|\nu|(E) = \sup\{|\int_E f\,d\nu| : |f| \le 1\}$.

4. If ν is a signed measure and λ, μ are positive measures such that $\nu = \lambda - \mu$, then $\lambda \ge \nu^+$ and $\mu \ge \nu^-$.

5. If ν_1, ν_2 are signed measures that both omit the value $+\infty$ or $-\infty$, then $|\nu_1 + \nu_2| \le |\nu_1| + |\nu_2|$. (Use Exercise 4.)

6. Suppose $\nu(E) = \int f\,d\mu$ where μ is a positive measure and f is an extended μ-integrable function. Describe the Hahn decompositions of ν and the positive, negative, and total variations of ν in terms of f and μ.

7. Suppose that ν is a signed measure on $(X, \mathcal{M})$ and $E \in \mathcal{M}$.
a. $\nu^+(E) = \sup\{\nu(F) : E \in \mathcal{M},\ F \subset E\}$ and $\nu^-(E) = -\inf\{\nu(F) : F \in \mathcal{M},\ F \subset E\}$.
b. $|\nu|(E) = \sup\{\sum_1^n |\nu(E_j)| : n \in \mathbb{N},\ E_1, \ldots, E_n$ are disjoint, and $\bigcup_1^n E_j = E\}$.

3.2 THE LEBESGUE-RADON-NIKODYM THEOREM

Suppose that ν is a signed measure and μ is a positive measure on $(X, \mathcal{M})$. We say that ν is **absolutely continuous** with respect to μ and write

$$\nu \ll \mu$$

if $\nu(E) = 0$ for every $E \in \mathcal{M}$ for which $\mu(E) = 0$. It is easily verified that $\nu \ll \mu$ iff $|\nu| \ll \mu$ iff $\nu^+ \ll \mu$ and $\nu^- \ll \mu$ (Exercise 8). Absolute continuity is in a sense the

antithesis of mutual singularity. More precisely, if $\nu \perp \mu$ and $\nu \ll \mu$, then $\nu = 0$, for if E and F are disjoint sets such that $E \cup F = X$ and $\mu(E) = |\nu|(F) = 0$, then the fact that $\nu \ll \mu$ implies that $|\nu|(E) = 0$, whence $|\nu| = 0$ and $\nu = 0$. One can extend the notion of absolute continuity to the case where μ is a signed measure (namely, $\nu \ll \mu$ iff $\nu \ll |\mu|$), but we shall have no need of this more general definition.

The term "absolute continuity" is derived from real-variable theory; see §3.5. For finite signed measures it is equivalent to another condition that is obviously a form of continuity.

3.5 Theorem. *Let ν be a finite signed measure and μ a positive measure on $(X, \mathcal{M})$. Then $\nu \ll \mu$ iff for every $\epsilon > 0$ there exists $\delta > 0$ such that $|\nu(E)| < \epsilon$ whenever $\mu(E) < \delta$.*

Proof. Since $\nu \ll \mu$ iff $|\nu| \ll \mu$ and $|\nu(E)| \leq |\nu|(E)$, it suffices to assume that $\nu = |\nu|$ is positive. Clearly the ϵ-δ condition implies that $\nu \ll \mu$. On the other hand, if the ϵ-δ condition is not satisfied, there exists $\epsilon > 0$ such that for all $n \in \mathbb{N}$ we can find $E_n \in \mathcal{M}$ with $\mu(E_n) < 2^{-n}$ and $\nu(E_n) \geq \epsilon$. Let $F_k = \bigcup_k^\infty E_n$ and $F = \bigcap_1^\infty F_k$. Then $\mu(F_k) < \sum_k^\infty 2^{-n} = 2^{1-k}$, so $\mu(F) = 0$; but $\nu(F_k) \geq \epsilon$ for all k and hence, since ν is finite, $\nu(F) = \lim \nu(F_k) \geq \epsilon$. Thus it is false that $\nu \ll \mu$. ∎

If μ is a measure and f is an extended μ-integrable function, the signed measure ν defined by $\nu(E) = \int_E f\, d\mu$ is clearly absolutely continuous with respect to μ; it is finite iff $f \in L^1(\mu)$. For any complex-valued $f \in L^1(\mu)$, the preceding theorem can be applied to $\operatorname{Re} f$ and $\operatorname{Im} f$, and we obtain the following useful result:

3.6 Corollary. *If $f \in L^1(\mu)$, for every $\epsilon > 0$ there exists $\delta > 0$ such that $|\int_E f\, d\mu| < \epsilon$ whenever $\mu(E) < \delta$.*

We shall use the following notation to express the relationship $\nu(E) = \int_E f\, d\mu$:

$$d\nu = f\, d\mu.$$

Sometimes, by a slight abuse of language, we shall refer to "the signed measure $f\, d\mu$."

We now come to the main theorem of this section, which gives a complete picture of the structure of a signed measure relative to a given positive measure. First, a technical lemma.

3.7 Lemma. *Suppose that ν and μ are finite measures on $(X, \mathcal{M})$. Either $\nu \perp \mu$, or there exist $\epsilon > 0$ and $E \in \mathcal{M}$ such that $\mu(E) > 0$ and $\nu \geq \epsilon\mu$ on E (that is, E is a positive set for $\nu - \epsilon\mu$).*

Proof. Let $X = P_n \cup N_n$ be a Hahn decomposition for $\nu - n^{-1}\mu$, and let $P = \bigcup_1^\infty P_n$ and $N = \bigcap_1^\infty N_n = P^c$. Then N is a negative set for $\nu - n^{-1}\mu$ for all n, i.e., $0 \leq \nu(N) \leq n^{-1}\mu(N)$ for all n, so $\nu(N) = 0$. If $\mu(P) = 0$, then $\nu \perp \mu$. If $\mu(P) > 0$, then $\mu(P_n) > 0$ for some n, and P_n is a positive set for $\nu - n^{-1}\mu$. ∎

3.8 The Lebesgue-Radon-Nikodym Theorem. *Let ν be a σ-finite signed measure and μ a σ-finite positive measure on $(X, \mathcal{M})$. There exist unique σ-finite signed measures λ, ρ on $(X, \mathcal{M})$ such that*

$$\lambda \perp \mu, \quad \rho \ll \mu, \quad \text{and} \quad \nu = \lambda + \rho.$$

Moreover, there is an extended μ-integrable function $f : X \to \mathbb{R}$ such that $d\rho = f\, d\mu$, and any two such functions are equal μ-a.e.

Proof. Case I: Suppose that ν and μ are finite positive measures. Let

$$\mathcal{F} = \left\{ f : X \to [0, \infty] : \int_E f\, d\mu \le \nu(E) \text{ for all } E \in \mathcal{M} \right\}.$$

$\mathcal{F}$ is nonempty since $0 \in \mathcal{F}$. Also, if $f, g \in \mathcal{F}$, then $h = \max(f, g) \in \mathcal{F}$, for if $A = \{x : f(x) > g(x)\}$, for any $E \in \mathcal{M}$ we have

$$\int_E h\, d\mu = \int_{E \cap A} f\, d\mu + \int_{E \setminus A} g\, d\mu \le \nu(E \cap A) + \nu(E \setminus A) = \nu(E).$$

Let $a = \sup\{\int f\, d\mu : f \in \mathcal{F}\}$, noting that $a \le \nu(X) < \infty$, and choose a sequence $\{f_n\} \subset \mathcal{F}$ such that $\int f_n\, d\mu \to a$. Let $g_n = \max(f_1, \dots, f_n)$ and $f = \sup_n f_n$. Then $g_n \in \mathcal{F}$, g_n increases pointwise to f, and $\int g_n\, d\mu \ge \int f_n\, d\mu$. It follows that $\lim \int g_n\, d\mu = a$ and hence, by the monotone convergence theorem, that $f \in \mathcal{F}$ and $\int f\, d\mu = a$. (In particular, $f < \infty$ a.e., so we may take f to be real-valued everywhere.)

We claim that the measure $d\lambda = d\nu - f\, d\mu$ (which is positive since $f \in \mathcal{F}$) is singular with respect to μ. If not, by Lemma 3.7 there exist $E \in \mathcal{M}$ and $\epsilon > 0$ such that $\mu(E) > 0$ and $\lambda \ge \epsilon\mu$ on E. But then $\epsilon\chi_E\, d\mu \le d\lambda = d\nu - f\, d\mu$, that is, $(f + e\chi_E)\, d\mu \le d\nu$, so $f + \epsilon\chi_E \in \mathcal{F}$ and $\int (f + \epsilon\chi_E)\, d\mu = a + \epsilon\mu(E) > a$, contradicting the definition of a.

Thus the existence of λ, f, and $d\rho = f\, d\mu$ is proved. As for uniqueness, if also $d\nu = d\lambda' + f' d\mu$, we have $d\lambda - d\lambda' = (f' - f)\, d\mu$. But $\lambda - \lambda' \perp \mu$ (see Exercise 9), while $(f' - f)\, d\mu \ll d\mu$; hence $d\lambda - d\lambda' = (f' - f)\, d\mu = 0$, so that $\lambda = \lambda'$ and (by Proposition 2.23) $f = f'$ μ-a.e. Thus we are done in the case when μ and ν are finite measures.

Case II: Suppose that μ and ν are σ-finite measures. Then X is a countable disjoint union of μ-finite sets and a countable disjoint union of ν-finite sets; by taking intersections of these we obtain a disjoint sequence $\{A_j\} \subset \mathcal{M}$ such that $\mu(A_j)$ and $\nu(A_j)$ are finite for all j and $X = \bigcup_1^\infty A_j$. Define $\mu_j(E) = \mu(E \cap A_j)$ and $\nu_j(E) = \nu(E \cap A_j)$. By the reasoning above, for each j we have $d\nu_j = d\lambda_j + f_j\, d\mu_j$ where $\lambda_j \perp \mu_j$. Since $\mu_j(A_j^c) = \nu_j(A_j^c) = 0$, we have $\lambda_j(A_j^c) = \nu_j(A_j^c) - \int_{A_j^c} f\, d\mu_j = 0$, and we may assume that $f_j = 0$ on A_j^c. Let $\lambda = \sum_1^\infty \lambda_j$ and $f = \sum_1^\infty f_j$. Then $d\nu = d\lambda + f\, d\mu$, $\lambda \perp \mu$ (see Exercise 9), and $d\lambda$ and $f\, d\mu$ are σ-finite, as desired. Uniqueness follows as before.

The General Case: If ν is a signed measure, we apply the preceding argument to ν^+ and ν^- and subtract the results. ∎

The decomposition $\nu = \lambda + \rho$ where $\lambda \perp \mu$ and $\rho \ll \mu$ is called the **Lebesgue decomposition** of ν with respect to μ. In the case where $\nu \ll \mu$, Theorem 3.8 says that $d\nu = f\,d\mu$ for some f. This result is usually known as the **Radon-Nikodym theorem**, and f is called the **Radon-Nikodym derivative** of ν with respect to μ. We denote it by $d\nu/d\mu$:

$$d\nu = \frac{d\nu}{d\mu}\,d\mu.$$

(Strictly speaking, $d\nu/d\mu$ should be construed as the as the class of functions equal to f μ-a.e.) The formulas suggested by the differential notation $d\mu/d\nu$ are generally correct. For example, it is obvious that $d(\nu_1 + \nu_2)/d\mu = (d\nu_1/d\mu) + (d\nu_2/d\mu)$, and we have the chain rule:

3.9 Proposition. *Suppose that ν is a σ-finite signed measure and μ, λ are σ-finite measures on $(X, \mathcal{M})$ such that $\nu \ll \mu$ and $\mu \ll \lambda$.*

a. If $g \in L^1(\nu)$, then $g(d\nu/d\mu) \in L^1(\mu)$ and

$$\int g\,d\nu = \int g\frac{d\nu}{d\mu}\,d\mu.$$

b. We have $\nu \ll \lambda$, and

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda} \quad \lambda\text{-a.e.}$$

Proof. By considering ν^+ and ν^- separately, we may assume that $\nu \geq 0$. The equation $\int g\,d\nu = \int g(d\nu/d\mu)\,d\mu$ is true when $g = \chi_E$ by definition of $d\nu/d\mu$. It is therefore true for simple functions by linearity, then for nonnegative measurable functions by the monotone convergence theorem, and finally for functions in $L^1(\nu)$ by linearity again. Replacing ν, μ by μ, λ and setting $g = \chi_E(d\nu/d\mu)$, we obtain

$$\nu(E) = \int_E \frac{d\nu}{d\mu}\,d\mu = \int_E \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}\,d\lambda$$

for all $E \in \mathcal{M}$, whence $(d\nu/d\lambda) = (d\nu/d\mu)(d\mu/d\lambda)$ λ-a.e. by Proposition 2.23. ∎

3.10 Corollary. *If $\mu \ll \lambda$ and $\lambda \ll \mu$, then $(d\lambda/d\mu)(d\mu/d\lambda) = 1$ a.e. (with respect to either λ or μ).*

Nonexample: Let μ be Lebesgue measure and ν the point mass at 0 on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Clearly $\nu \perp \mu$. The nonexistent Radon-Nikodym derivative $d\nu/d\mu$ is popularly known as the Dirac δ-finction.

We conclude this section with a simple but important observation:

3.11 Proposition. *If $\mu_1, \ldots, \mu_n$ are measures on $(X, \mathcal{M})$, there is a measure μ such that $\mu_j \ll \mu$ for all j — namely, $\mu = \sum_1^n \mu_j$.*

The proof is trivial.

Exercises

8. $\nu \ll \mu$ iff $|\nu| \ll \mu$ iff $\nu^+ \ll \mu$ and $\nu^- \ll \mu$.

9. Suppose $\{\nu_j\}$ is a sequence of positive measures. If $\nu_j \perp \mu$ for all j, then $\sum_1^\infty \nu_j \perp \mu$; and if $\nu_j \ll \mu$ for all j, then $\sum_1^\infty \nu_j \ll \mu$.

10. Theorem 3.5 may fail when ν is not finite. (Consider $d\nu(x) = dx/x$ and $d\mu(x) = dx$ on $(0,1)$, or ν = counting measure and $\mu(E) = \sum_{n\in E} 2^{-n}$ on $\mathbb{N}$.)

11. Let μ be a positive measure. A collection of functions $\{f_\alpha\}_{\alpha\in A} \subset L^1(\mu)$ is called **uniformly integrable** if for every $\epsilon > 0$ there exists $\delta > 0$ such that $|\int_E f_\alpha \, d\mu| < \epsilon$ for all $\alpha \in A$ whenever $\mu(E) < \delta$.

a. Any finite subset of $L^1(\mu)$ is uniformly integrable.

b. If $\{f_n\}$ is a sequence in $L^1(\mu)$ that converges in the L^1 metric to $f \in L^1(\mu)$, then $\{f_n\}$ is uniformly integrable.

12. For $j = 1, 2$, let μ_j, ν_j be σ-finite measures on $(X_j, \mathcal{M}_j)$ such that $\nu_j \ll \mu_j$. Then $\nu_1 \times \nu_2 \ll \mu_1 \times \mu_2$ and

$$\frac{d(\nu_1 \times \nu_2)}{d(\mu_1 \times \mu_2)}(x_1, x_2) = \frac{d\nu_1}{d\mu_1}(x_1)\frac{d\nu_2}{d\mu_2}(x_2).$$

13. Let $X = [0,1]$, $\mathcal{M} = \mathcal{B}_{[0,1]}$, m = Lebesgue measure, and μ = counting measure on $\mathcal{M}$.

a. $m \ll \mu$ but $dm \neq f\, d\mu$ for any f.

b. μ has no Lebesgue decomposition with respect to m.

14. If ν is an arbitrary signed measure and μ is a σ-finite measure on $(X, \mathcal{M})$ such that $\nu \ll \mu$, there exists an extended μ-integrable function $f : X \to [-\infty, \infty]$ such that $d\nu = f\, d\mu$. Hints:

a. It suffices to assume that μ is finite and ν is positive.

b. With these assumptions, there exists $E \in \mathcal{M}$ that is σ-finite for ν such that $\mu(E) \geq \mu(F)$ for all sets F that are σ-finite for ν.

c. The Radon-Nikodym theorem applies on E. If $F \cap E = \varnothing$, then either $\nu(F) = \mu(F) = 0$ or $\mu(F) > 0$ and $|\nu(F)| = \infty$.

15. A measure μ on $(X, \mathcal{M})$ is called **decomposable** if there is a family $\mathcal{F} \subset \mathcal{M}$ with the following properties: (i) $\mu(F) < \infty$ for all $F \in \mathcal{F}$; (ii) the members of $\mathcal{F}$ are disjoint and their union is X; (iii) if $\mu(E) < \infty$ then $\mu(E) = \sum_{F\in\mathcal{F}} \mu(E \cap F)$; (iv) if $E \subset X$ and $E \cap F \in \mathcal{M}$ for all $F \in \mathcal{F}$ then $E \in \mathcal{M}$.

a. Every σ-finite measure is decomposable.

b. If μ is decomposable and ν is any signed measure on $(X, \mathcal{M})$ such that $\nu \ll \mu$, there exists a measurable $f : X \to [-\infty, \infty]$ such that $\nu(E) = \int_E f\, d\mu$ for any E that is σ-finite for μ, and $|f| < \infty$ on any $F \in \mathcal{F}$ that is σ-finite for ν. (Use Exercise 14 if ν is not σ-finite.)

16. Suppose that μ, ν are measures on $(X, \mathcal{M})$ with $\nu \ll \mu$, and let $\lambda = \mu + \nu$. If $f = d\nu/d\lambda$, then $0 \leq f < 1$ μ-a.e. and $d\nu/d\mu = f/(1-f)$.

17. Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space, $\mathcal{N}$ a sub-σ-algebra of $\mathcal{M}$, and $\nu = \mu|\mathcal{N}$. If $f \in L^1(\mu)$, there exists $g \in L^1(\nu)$ (thus g is $\mathcal{N}$-measurable) such that $\int_E f\,d\mu = \int_E g\,d\nu$ for all $E \in \mathcal{N}$; if g' is another such function then $g = g'$ ν-a.e. (In probability theory, g is called the **conditional expectation** of f on $\mathcal{N}$.)

3.3 COMPLEX MEASURES

A **complex measure** on a measurable space $(X, \mathcal{M})$ is a map $\nu : \mathcal{M} \to \mathbb{C}$ such that

- $\nu(\varnothing) = 0$;
- if $\{E_j\}$ is a sequence of disjoint sets in $\mathcal{M}$, then $\nu(\bigcup_1^\infty E_j) = \sum_1^\infty \nu(E_j)$, where the series converges absolutely.

In particular, infinite values are not allowed, so a positive measure is a complex measure only if it is finite. Example: If μ is a positive measure and $f \in L^1(\mu)$, then $f\,d\mu$ is a complex measure.

If ν is a complex measure, we shall write ν_r and ν_i for the real and imaginary parts of ν. Thus ν_r and ν_i are signed measures that do not assume the values $\pm\infty$; hence they are finite, and so the range of ν is a bounded subset of $\mathbb{C}$.

The notions we have developed for signed measures generalize easily to complex measures. For example, we define $L^1(\nu)$ to be $L^1(\nu_r) \cap L^1(\nu_i)$, and for $f \in L^1(\nu)$, we set $\int f\,d\nu = \int f\,d\nu_r + i\int f\,d\nu_i$. If ν and μ are complex measures, we say that $\nu \perp \mu$ if $\nu_a \perp \mu_b$ for $a, b = r, i$, and if λ is a positive measure, we say that $\nu \ll \lambda$ if $\nu_r \ll \lambda$ and $\nu_i \ll \lambda$. The theorems of §3.2 also generalize; one has merely to apply them to apply them to the real and imaginary parts separately. In particular:

3.12 The Lebesgue-Radon-Nikodym Theorem. *If ν is a complex measure and μ is a σ-finite positive measure on $(X, \mathcal{M})$, there exist a complex measure λ and an $f \in L^1(\mu)$ such that $\lambda \perp \mu$ and $d\nu = d\lambda + f\,d\mu$. If also $\lambda' \perp \mu$ and $d\nu = d\lambda' + f'\,d\mu$, then $\lambda = \lambda'$ and $f = f'$ μ-a.e.*

As before, if $\nu \ll \mu$, we denote the f in Theorem 3.12 by $d\nu/d\mu$.

The **total variation** of a complex measure ν is the positive measure $|\nu|$ determined by the property that if $d\nu = f\,d\mu$ where μ is a positive measure, then $d|\nu| = |f|\,d\mu$. To see that this is well defined, we observe first that every ν is of the form $f\,d\mu$ for some finite measure μ and some $f \in L^1(\mu)$; indeed, we can take $\mu = |\nu_r| + |\nu_i|$ and use Theorem 3.12 to obtain f. Second, if $d\nu = f_1\,d\mu_1 = f_2\,d\mu_2$, let $\rho = \mu_1 + \mu_2$. Then by Proposition 3.9,

$$f_1 \frac{d\mu_1}{d\rho} d\rho = d\nu = f_2 \frac{d\mu_2}{d\rho} d\rho,$$

so that $f_1(d\mu_1/d\rho) = f_2(d\mu_2/d\rho)$ ρ-a.e. Since $d\mu_j/d\rho$ is nonnegative, we therefore have

$$|f_1| \frac{d\mu_1}{d\rho} = \left| f_1 \frac{d\mu_1}{d\rho} \right| = \left| f_2 \frac{d\mu_2}{d\rho} \right| = |f_2| \frac{d\mu_2}{d\rho} \quad \rho\text{-a.e.},$$

and thus

$$|f_1|\,d\mu_1 = |f_1|\frac{d\mu_1}{d\rho}\,d\rho = |f_2|\frac{d\mu_2}{d\rho}\,d\rho = |f_2|\,d\mu_2.$$

Hence the definition of $|\nu|$ is independent of the choice of μ and f. This definition agrees with the previous definition of ν when ν is a signed measure, for in that case $d\nu = (\chi_P - \chi_N)d|\nu|$ where $X = P \cup N$ is a Hahn decomposition, and $|\chi_P - \chi_N| = 1$.

3.13 Proposition. *Let ν be a complex measure on $(X, \mathcal{M})$.*

a. $|\nu(E)| \le |\nu|(E)$ *for all* $E \in \mathcal{M}$.

b. $\nu \ll |\nu|$, *and* $d\nu/d|\nu|$ *has absolute value* 1 $|\nu|$*-a.e.*

c. $L^1(\nu) = L^1(|\nu|)$, *and if* $f \in L^1(\nu)$, *then* $|\int f\,d\nu| \le \int |f|\,d|\nu|$.

Proof. Suppose $d\nu = f\,d\mu$ as in the definition of $|\nu|$. Then

$$|\nu(E)| = |\int_E f\,d\mu| \le \int_E |f|\,d\mu = |\nu|(E).$$

This proves (a) and shows that $\nu \ll |\nu|$. If $g = d\nu/d|\nu|$, then, we have $f\,d\mu = d\nu = g\,d|\nu| = g|f|\,d\mu$, so $g|f| = f$ μ-a.e. and hence $|\nu|$-a.e. But clearly $|f| > 0$ $|\nu|$-a.e., whence $|g| = 1$ $|\nu|$-a.e. Part (c) is left to the reader (Exercise 18). ∎

3.14 Proposition. *If ν_1, ν_2 are complex measures on $(X, \mathcal{M})$, then* $|\nu_1 + \nu_2| \le |\nu_1| + |\nu_2|$.

Proof. By Proposition 3.11 we can write $\nu_j = f_j\,d\mu$, with the same μ, for $j = 1, 2$. But then $d|\nu_1 + \nu_2| = |f_1 + f_2|\,d\mu \le |f_1|\,d\mu + |f_2|\,d\mu = d|\nu_1| + d|\nu_2|$. ∎

Exercises

18. Prove Proposition 3.13c.

19. If ν, μ are complex measures and λ is a positive measure, then $\nu \perp \mu$ iff $|\nu| \perp |\mu|$, and $\nu \ll \lambda$ iff $|\nu| \ll \lambda$.

20. If ν is a complex measure on $(X, \mathcal{M})$ and $\nu(X) = |\nu|(X)$, then $\nu = |\nu|$.

21. Let ν be a complex measure on $(X, \mathcal{M})$. If $E \in \mathcal{M}$, define

$$\mu_1(E) = \sup\Big\{\sum_1^n |\nu(E_j)| : n \in \mathbb{N},\ E_1, \dots, E_n \text{ disjoint}, E = \bigcup_1^n E_j\Big\},$$

$$\mu_2(E) = \sup\Big\{\sum_1^\infty |\nu(E_j)| : E_1, E_2, \dots \text{ disjoint}, E = \bigcup_1^\infty E_j\Big\},$$

$$\mu_3(E) = \sup\Big\{\Big|\int_E f\,d\mu\Big| : |f| \le 1\Big\}.$$

Then $\mu_1 = \mu_2 = \mu_3 = |\nu|$. (First show that $\mu_1 \le \mu_2 \le \mu_3$. To see that $\mu_3 = |\nu|$, let $f = d\nu/d|\nu|$ and apply Proposition 3.13. To see that $\mu_3 \le \mu_1$, approximate f by simple functions.)

3.4 DIFFERENTIATION ON EUCLIDEAN SPACE

The Radon-Nikodym theorem provides an abstract notion of the "derivative" of a signed or complex measure ν with respect to a measure μ. In this section we analyze more deeply the special case where $(X, \mathcal{M}) = (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ and $\mu = m$ is Lebesgue measure. Here one can define a *pointwise* derivative of ν with respect to m in the following way. Let $B(r, x)$ be the open ball of radius r about x in $\mathbb{R}^n$; then one can consider the limit

$$F(x) = \lim_{r \to 0} \frac{\nu(B(r,x))}{m(B(r,x))}$$

when it exists. (One can also replace the balls $B(r, x)$ by other sets which, in a suitable sense, shrink to x in a regular way; we shall examine this point later.) If $\nu \ll m$, so that $d\nu = f\, dm$, then $\nu(B(r,x))/m(B(r,x))$ is simply the average value of f on $B(r, x)$, so one would hope that $F = f$ m-a.e. This turns out to be the case provided that $\nu(B(r,x))$ is finite for all r, x. From the point of view of the function f, this may be regarded as a generalization of the fundamental theorem of calculus: The derivative of the indefinite integral of f (namely, ν) is f.

For the remainder of this section, terms such as "integrable" and "almost everywhere" refer to Lebesgue measure unless otherwise specified. We begin our analysis with a technical lemma that is of interest in its own right.

3.15 Lemma. *Let $\mathcal{C}$ be a collection of open balls in $\mathbb{R}^n$, and let $U = \bigcup_{B \in \mathcal{C}} B$. If $c < m(U)$, there exist disjoint $B_1, \ldots, B_k \in \mathcal{C}$ such that $\sum_1^k m(B_j) > 3^{-n}c$.*

Proof. If $c < m(U)$, by Theorem 2.40 there is a compact $K \subset U$ with $m(K) > c$, and finitely many of the balls in $\mathcal{C}$ — say, $A_1, \ldots, A_m$ — cover K. Let B_1 be the largest of the A_j's (that is, choose B_1 to have maximal radius), let B_2 be the largest of the A_j's that are disjoint from B_1, B_3 the largest of the A_j's that are disjoint from B_1 and B_2, and so on until the list of A_j's is exhausted. According to this construction, if A_i is not one of the B_j's, there is a j such that $A_i \cap B_j \neq \varnothing$, and if j is the smallest integer with this property, the radius of A_i is at most that of B_j. Hence $A_i \subset B_j^*$, where B_j^* is the ball concentric with B_j whose radius is three times that of B_j. But then $K \subset \bigcup_1^k B_j^*$, so

$$c < m(K) \le \sum_1^k m(B_j^*) = 3^n \sum_1^k m(B_j).$$

∎

A measurable function $f : \mathbb{R}^n \to \mathbb{C}$ is called **locally integrable** (with respect to Lebesgue measure) if $\int_K |f(x)|\, dx < \infty$ for every bounded measurable set $K \subset \mathbb{R}^n$.

We denote the space of locally integrable functions by L^1_{loc}. If $f \in L^1_{\text{loc}}$, $x \in \mathbb{R}^n$, and $r > 0$, we define $A_r f(x)$ to be the average value of f on $B(r,x)$:

$$A_r f(x) = \frac{1}{m(B(r,x))} \int_{B(r,x)} f(y)\, dy.$$

3.16 Lemma. *If $f \in L^1_{\text{loc}}$, $A_r f(x)$ is jointly continuous in r and x ($r > 0$, $x \in \mathbb{R}^n$).*

Proof. From the results in §2.7 we know that $m(B(r,x)) = cr^n$ where $c = m(B(1,0))$, and $m(S(r,x)) = 0$ where $S(r,x) = \{y : |y - x| = r\}$. Moreover, as $r \to r_0$ and $x \to x_0$, $\chi_{B(r,x)} \to \chi_{B(r_0,x_0)}$ pointwise on $\mathbb{R}^n \setminus S(r_0,x_0)$. Hence $\chi_{B(r,x)} \to \chi_{B(r_0,x_0)}$ a.e., and $|\chi_{B(r,x)}| \le \chi_{B(r_0+1,\, x_0)}$ if $r < r_0 + \frac{1}{2}$ and $|x - x_0| < \frac{1}{2}$. By the dominated convergence theorem, it follows that $\int_{B(r,x)} f(y)\, dy$ is continuous in r and x, and hence so is $A_r f(x) = c^{-1} r^{-n} \int_{B(r,x)} f(y)\, dy$. ∎

Next, if $f \in L^1_{\text{loc}}$, we define its **Hardy-Littlewood maximal function** Hf by

$$Hf(x) = \sup_{r>0} A_r|f|(x) = \sup_{r>0} \frac{1}{m(B(r,x))} \int_{B(r,x)} |f(y)|\, dy.$$

Hf is measurable, for $(Hf)^{-1}((a,\infty)) = \bigcup_{r>0} (A_r|f|)^{-1}((a,\infty))$ is open for any $a \in \mathbb{R}$, by Lemma 3.16.

3.17 The Maximal Theorem. *There is a constant $C > 0$ such that for all $f \in L^1$ and all $\alpha > 0$,*

$$m(\{x : Hf(x) > \alpha\}) \le \frac{C}{\alpha} \int |f(x)|\, dx.$$

Proof. Let $E_\alpha = \{x : Hf(x) > \alpha\}$. For each $x \in E_\alpha$ we can choose $r_x > 0$ such that $A_{r_x}|f|(x) > \alpha$. The balls $B(r_x, x)$ cover E_α, so by Lemma 3.15, if $c < m(E_\alpha)$ there exist $x_1, \dots, x_k \in E_\alpha$ such that the balls $B_j = B(r_{x_j}, x_j)$ are disjoint and $\sum_1^k m(B_j) > 3^{-n} c$. But then

$$c < 3^n \sum_1^k m(B_j) \le \frac{3^n}{\alpha} \sum_1^k \int_{B_j} |f(y)|\, dy \le \frac{3^n}{\alpha} \int_{\mathbb{R}^n} |f(y)|\, dy.$$

Letting $c \to m(E_\alpha)$, we obtain the desired result. ∎

With this tool in hand, we now present three successively sharper versions of the fundamental differentiation theorem. In the proofs we shall use the notion of limit superior for real-valued functions of a real variable,

$$\limsup_{r \to R} \phi(r) = \lim_{\epsilon \to 0} \sup_{0 < |r-R| < \epsilon} \phi(r) = \inf_{\epsilon > 0} \sup_{0 < |r-R| < \epsilon} \phi(r),$$

and the easily verified fact that

$$\lim_{r \to R} \phi(r) = c \quad \text{iff} \quad \limsup_{r \to R} |\phi(r) - c| = 0.$$

3.18 Theorem. *If $f \in L^1_{\text{loc}}$, then $\lim_{r\to 0} A_r f(x) = f(x)$ for a.e. $x \in \mathbb{R}^n$.*

Proof. It suffices to show that for $N \in \mathbb{N}$, $A_r f(x) \to f(x)$ for a.e. x with $|x| \le N$. But for $|x| \le N$ and $r \le 1$ the values $A_r f(x)$ depend only on the values $f(y)$ for $|y| \le N+1$, so by replacing f with $f\chi_{B(N+1,0)}$ we may assume that $f \in L^1$.

Given $\epsilon > 0$, by Theorem 2.41 we can find a continuous integrable function g such that $\int |g(y) - f(y)|\,dy < \epsilon$. Continuity of g implies that for every $x \in \mathbb{R}^n$ and $\delta > 0$ there exists $r > 0$ such that $|g(y) - g(x)| < \delta$ whenever $|y - x| < r$, and hence

$$\left|A_r g(x) - g(x)\right| = \frac{1}{m(B(r,x))}\left|\int_{B(r,x)} [g(y) - g(x)]\,dy\right| < \delta.$$

Therefore $A_r g(x) \to g(x)$ as $r \to 0$ for every x, so

$$\begin{aligned}\limsup_{r\to 0} &\left|A_r f(x) - f(x)\right| \\ &= \limsup_{r\to 0}\left|A_r(f-g)(x) + (A_r g - g)(x) + (g-f)(x)\right| \\ &\le H(f-g)(x) + 0 + |f-g|(x).\end{aligned}$$

Hence, if

$$E_\alpha = \{x : \limsup_{r\to 0} |A_r f(x) - f(x)| > \alpha\}, \qquad F_\alpha = \{x : |f-g|(x) > \alpha\},$$

we have

$$E_\alpha \subset F_{\alpha/2} \cup \{x : H(f-g)(x) > \alpha/2\}.$$

But $(\alpha/2)m(F_{\alpha/2}) \le \int_{F_{\alpha/2}} |f(x) - g(x)|\,dx < \epsilon$, so by the maximal theorem,

$$m(E_\alpha) \le \frac{2\epsilon}{\alpha} + \frac{2C\epsilon}{\alpha}.$$

Since ϵ is arbitrary, $m(E_\alpha) = 0$ for all $\alpha > 0$. But $\lim_{r\to 0} A_r f(x) = f(x)$ for all $x \notin \bigcup_1^\infty E_{1/n}$, so we are done. ∎

This result can be rephrased as follows: If $f \in L^1_{\text{loc}}$,

$$\lim_{r\to 0} \frac{1}{m(B(r,x))}\int_{B(r,x)} [f(y) - f(x)]\,dy = 0 \text{ for a.e. } x. \tag{3.19}$$

Actually, something stronger is true: (3.19) remains valid if one replaces the integrand by its absolute value. That is, let us define the **Lebesgue set** L_f of f to be

$$L_f = \left\{x : \lim_{r\to 0} \frac{1}{m(B(r,x))}\int_{B(r,x)} |f(y) - f(x)|\,dy = 0\right\}.$$

3.20 Theorem. *If $f \in L^1_{\text{loc}}$, then $m((L_f)^c) = 0$.*

Proof. For each $c \in \mathbb{C}$ we can apply Theorem 3.18 to $g_c(x) = |f(x) - c|$ to conclude that, except on a Lebesgue null set E_c, we have

$$\lim_{r\to 0} \frac{1}{m(B(r,x))} \int_{B(r,x)} |f(y) - c|\, dy = |f(x) - c|.$$

Let D be a countable dense subset of $\mathbb{C}$, and let $E = \bigcup_{c\in D} E_c$. Then $m(E) = 0$, and if $x \notin E$, for any $\epsilon > 0$ we can choose $c \in D$ with $|f(x) - c| < \epsilon$, so that $|f(y) - f(x)| < |f(y) - c| + \epsilon$, and it follows that

$$\limsup_{r\to 0} \frac{1}{m(B(r,x))} \int_{B(r,x)} |f(y) - f(x)|\, dy \le |f(x) - c| + \epsilon < 2\epsilon.$$

Since ϵ is arbitrary, the desired result follows. ∎

Finally, we consider families of sets more general than balls. A family $\{E_r\}_{r>0}$ of Borel subsets of $\mathbb{R}^n$ is said to **shrink nicely** to $x \in \mathbb{R}^n$ if

- $E_r \subset B(r,x)$ for each r;
- there is a constant $\alpha > 0$, independent of r, such that $m(E_r) > \alpha m(B(r,x))$.

The sets E_r need not contain x itself. For example, if U is any Borel subset of $B(1,0)$ such that $m(U) > 0$, and $E_r = \{x + ry : y \in U\}$, then $\{E_r\}$ shrinks nicely to x. Here, then, is the final version of the differentiation theorem.

3.21 The Lebesgue Differentiation Theorem. *Suppose $f \in L^1_{\text{loc}}$. For every x in the Lebesgue set of f — in particular, for almost every x — we have*

$$\lim_{r\to 0} \frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)|\, dy = 0 \text{ and } \lim_{r\to 0} \frac{1}{m(E_r)} \int_{E_r} f(y)\, dy = f(x)$$

for every family $\{E_r\}_{r>0}$ that shrinks nicely to x.

Proof. For some $\alpha > 0$ we have

$$\begin{aligned}\frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)|\, dy &\le \frac{1}{m(E_r)} \int_{B(r,x)} |f(y) - f(x)|\, dy \\ &\le \frac{1}{\alpha m(B(r,x))} \int_{B(r,x)} |f(y) - f(x)|\, dy.\end{aligned}$$

The first equality therefore follows from Theorem 3.20, and one sees immediately that it implies the second one by writing the latter in the form (3.19). ∎

We now return to the study of measures. A Borel measure ν on $\mathbb{R}^n$ will be called **regular** if

- $\nu(K) < \infty$ for every compact K;
- $\nu(E) = \inf\{\nu(U) : U \text{ open}, E \subset U\}$ for every $E \in \mathcal{B}_{\mathbb{R}^n}$.

(Condition (ii) is actually implied by condition (i). For $n = 1$ this follows from Theorems 1.16 and 1.18, and we shall prove it for arbitrary n in §7.2. For the time being, we assume (ii) explicitly.) We observe that by (i), every regular measure is σ-finite. A signed or complex Borel measure ν will be called **regular** if $|\nu|$ is regular.

For example, if $f \in L^+(\mathbb{R}^n)$, the measure $f\,dm$ is regular iff $f \in L^1_{\text{loc}}$. Indeed, the condition $f \in L^1_{\text{loc}}$ is clearly equivalent to (i). If this holds, (ii) may be verified directly as follows. Suppose that E is a bounded Borel set. Given $\delta > 0$, by Theorem 2.40 there is a bounded open $U \supset E$ such that $m(U) < m(E) + \delta$ and hence $m(U \setminus E) < \delta$. But then, given $\epsilon > 0$, by Corollary 3.6 there is an open $U \supset E$ such that $\int_{U\setminus E} f\,dm < \epsilon$ and hence $\int_U f\,dm < \int_E f\,dm + \epsilon$. The case of unbounded E follows easily by writing $E = \bigcup_1^\infty E_j$ where E_j is bounded and finding an open $U_j \supset E_j$ such that $\int_{U_j\setminus E_j} f\,dm < \epsilon 2^{-j}$.

3.22 Theorem. *Let ν be a regular signed or complex Borel measure on $\mathbb{R}^n$, and let $d\nu = d\lambda + f\,dm$ be its Lebesgue-Radon-Nikodym representation. Then for m-almost every $x \in \mathbb{R}^n$,*

$$\lim_{r\to 0} \frac{\nu(E_r)}{m(E_r)} = f(x)$$

for every family $\{E_r\}_{r>0}$ that shrinks nicely to x.

Proof. It is easily verified that $d|\nu| = d|\lambda| + |f|\,d\mu$, so the regularity of ν implies the regularity of both λ and $f\,dm$ (Exercise 26). In particular, $f \in L^1_{\text{loc}}$, so in view of Theorem 3.21, it suffices to show that if λ is regular and $\lambda \perp m$, then for m-almost every x, $\lambda(E_r)/m(E_r) \to 0$ as $r \to 0$ when E_r shrinks nicely to x. It also suffices to take $E_r = B(r, x)$ and to assume that λ is positive, since for some $\alpha > 0$ we have

$$\left|\frac{\lambda(E_r)}{m(E_r)}\right| \le \frac{|\lambda|(E_r)}{m(E_r)} \le \frac{|\lambda|(B(r,x))}{m(E_r)} \le \frac{|\lambda|(B(r,x))}{\alpha m(B(r,x))}.$$

Assuming $\lambda \ge 0$, then, let A be a Borel set such that $\lambda(A) = m(A^c) = 0$, and let

$$F_k = \left\{x \in A : \limsup_{r\to 0} \frac{\lambda(B(r,x))}{m(B(r,x))} > \frac{1}{k}\right\}.$$

We shall show that $m(F_k) = 0$ for all k, and this will complete the proof.

The argument is similar to the proof of the maximal theorem. By regularity of λ, given $\epsilon > 0$ there is an open $U_\epsilon \supset A$ such that $\lambda(U_\epsilon) < \epsilon$. Each $x \in F_k$ is the center of a ball $B_x \subset U_\epsilon$ such that $\lambda(B_x) > k^{-1}m(B_x)$. By Lemma 3.15, if

$V_\epsilon = \bigcup_{x \in F_k} B_x$ and $c < m(V_\epsilon)$ there exist $x_1, \ldots, x_J$ such that $B_{x_1}, \ldots, B_{x_J}$ are disjoint and

$$c < 3^n \sum_1^J m(B_{x_j}) \leq 3^n k \sum_1^J \lambda(B_{x_j}) \leq 3^n k \lambda(V_\epsilon) \leq 3^n k \lambda(U_\epsilon) \leq 3^n k \epsilon.$$

We conclude that $m(V_\epsilon) \leq 3^n k \epsilon$, and since $F_k \subset V_\epsilon$ and ϵ is arbitrary, $m(F_k) = 0$. ∎

Exercises

22. If $f \in L^1(\mathbb{R}^n)$, $f \neq 0$, there exist $C, R > 0$ such that $Hf(x) \geq C|x|^{-n}$ for $|x| > R$. Hence $m(\{x : Hf(x) > \alpha\}) \geq C'/\alpha$ when α is small, so the estimate in the maximal theorem is essentially sharp.

23. A useful variant of the Hardy-Littlewood maximal function is

$$H^* f(x) = \sup \left\{ \frac{1}{m(B)} \int_B |f(y)|\, dy : B \text{ is a ball and } x \in B \right\}.$$

Show that $Hf \leq H^* f \leq 2^n Hf$.

24. If $f \in L^1_{\text{loc}}$ and f is continuous at x, then x is in the Lebesgue set of f.

25. If E is a Borel set in $\mathbb{R}^n$, the **density** $D_E(x)$ of E at x is defined as

$$D_E(x) = \lim_{r \to 0} \frac{m(E \cap B(r,x))}{m(B(r,x))},$$

whenever the limit exists.

a. Show that $D_E(x) = 1$ for a.e. $x \in E$ and $D_E(x) = 0$ for a.e. $x \in E^c$.

b. Find examples of E and x such that $D_E(x)$ is a given number $\alpha \in (0,1)$, or such that $D_E(x)$ does not exist.

26. If λ and μ are positive, mutually singular Borel measures on $\mathbb{R}^n$ and $\lambda + \mu$ is regular, then so are λ and μ.

3.5 FUNCTIONS OF BOUNDED VARIATION

The theorems of the preceding section apply in particular on the real line, where, because of the correspondence between regular Borel measures and increasing functions that we established in §1.5, they yield results about differentiation and integration of functions. As in §1.5, we adopt the notation that if F is an increasing, right continuous function on $\mathbb{R}$, μ_F is the Borel measure determined by the relation $\mu_F((a,b]) = F(b) - F(a)$. Also, throughout this section the term "almost everywhere" will always refer to Lebesgue measure.

Our first result uses the Lebesgue differentiation theorem to prove the a.e. differentiability of increasing functions.

3.23 Theorem. *Let $F : \mathbb{R} \to \mathbb{R}$ be increasing, and let $G(x) = F(x+)$.*

a. The set of points at which F is discontinuous is countable.

b. F and G are differentiable a.e., and $F' = G'$ a.e.

Proof. Since F is increasing, the intervals $(F(x-), F(x+))$ $(x \in \mathbb{R})$ are disjoint, and for $|x| < N$ they lie in the interval $(F(-N), F(N))$. Hence

$$\sum_{|x|<N} [F(x+) - F(x-)] \leq F(N) - F(-N) < \infty,$$

which implies that $\{x \in (-N, N) : F(x+) \neq F(x-)\}$ is countable. As this is true for all N, (a) is proved.

Next, we observe that G is increasing and right continuous, and $G = F$ except perhaps where F is discontinuous. Moreover,

$$G(x+h) - G(x) = \begin{cases} \mu_G((x,\, x+h]) & \text{if } h > 0, \\ -\mu_G((x+h,\, x]) & \text{if } h < 0, \end{cases}$$

and the families $\{(x-r,\, x]\}$ and $\{(x,\, x+r]\}$ shrink nicely to x as $r = |h| \to 0$. Thus, an application of Theorem 3.22 to the measure μ_G (which is regular by Theorem 1.18) shows that $G'(x)$ exists for a.e. x. To complete the proof, it remains to show that if $H = G - F$, then H' exists and equals zero a.e.

Let $\{x_j\}$ be an enumeration of the points at which $H \neq 0$. Then $H(x_j) > 0$, and as above we have $\sum_{\{j:|x_j|<N\}} H(x_j) < \infty$ for any N. Let δ_j be the point mass at x_j and $\mu = \sum_j H(x_j)\delta_j$. Then μ finite on compact sets by the preceding sentence, and hence μ is regular by Theorems 1.16 and 1.18; also, $\mu \perp m$ since $m(E) = \mu(E^c) = 0$ where $E = \{x_j\}_1^\infty$. But then

$$\left| \frac{H(x+h) - H(x)}{h} \right| \leq \frac{H(x+h) + H(x)}{|h|} \leq 4\frac{\mu\big((x - 2|h|,\, x + 2|h|)\big)}{4|h|},$$

which tends to zero as $h \to 0$ for a.e. x, by Theorem 3.22. Thus $H' = 0$ a.e., and we are done. ∎

As positive measures on $\mathbb{R}$ are related to increasing functions, complex measures on $\mathbb{R}$ are related to so-called functions of bounded variation. The definition of the latter concept is a bit technical, so some motivation may be appropriate. Intuitively, if $F(t)$ represents the position of a particle moving along the real line at time t, the "total variation" of F over the interval $[a, b]$ is the total distance traveled from time a to time b, as shown on an odometer. If F has a continuous derivative, this is just the integral of the "speed," $\int_a^b |F'(t)|\, dt$. To define the total variation without any smoothness hypotheses on F requires a different approach; namely, one partitions $[a, b]$ into subintervals $[t_{j-1}, t_j]$ and approximates F on each subinterval by the linear function whose graph joins $(t_{j-1}, F(t_{j-1}))$ to $(t_j, F(t_j))$, and then passes to a limit.

In making this precise, we begin with a slightly different point of view, taking $a = -\infty$ and considering the total variation as a function of b. To wit, if $F : \mathbb{R} \to \mathbb{C}$

and $x \in \mathbb{R}$, we define

$$T_F(x) = \sup\Big\{\sum_1^n |F(x_j) - F(x_{j-1})| : n \in \mathbb{N},\ -\infty < x_0 < \cdots < x_n = x\Big\}.$$

T_F is called the **total variation function** of F. We observe that the sums in the definition of T_F are made bigger if the additional subdivision points x_j are added. Hence, if $a < b$, the definition of $T_F(b)$ is unaffected if we assume that a is always one of the subdivision points. It follows that

$$\begin{aligned} &T_F(b) - T_F(a) \\ &= \sup\Big\{\sum_1^n |F(x_j) - F(x_{j-1})| : n \in \mathbb{N},\ a = x_0 < \cdots < x_n = b\Big\}. \end{aligned} \tag{3.24}$$

Thus T_F is an increasing function with values in $[0, \infty]$. If $T_F(\infty) = \lim_{x\to\infty} T_F(x)$ is finite, we say that F is of **bounded variation** on $\mathbb{R}$, and we denote the space of all such F by BV.

More generally, the supremum on the right of (3.24) is called the **total variation** of F on $[a, b]$. It depends only on the values of F on $[a, b]$, so we may define $BV([a, b])$ to be the set of all functions on $[a, b]$ whose total variation on $[a, b]$ is finite. If $F \in BV$, the restriction of F to $[a, b]$ is in $BV([a, b])$ for all a, b; indeed, its total variation on $[a, b]$ is nothing but $T_F(b) - T_F(a)$. Conversely, if $F \in BV([a, b])$ and we set $F(x) = F(a)$ for $x < a$ and $F(x) = F(b)$ for $x > b$, then $F \in BV$. By this device the results that we shall prove for BV can also be applied to $BV([a, b])$.

3.25 Examples.

a. If $F : \mathbb{R} \to \mathbb{R}$ is bounded and increasing, then $F \in BV$ (in fact, $T_F(x) = F(x) - F(-\infty)$).
b. If $F, G \in BV$ and $a, b \in \mathbb{C}$, then $aF + bG \in BV$.
c. If F is differentiable on $\mathbb{R}$ and F' is bounded, then $F \in BV([a, b])$ for $-\infty < a < b < \infty$ (by the mean value theorem).
d. If $F(x) = \sin x$, then $F \in BV([a, b])$ for $-\infty < a < b < \infty$, but $F \notin BV$.
e. If $F(x) = x\sin(x^{-1})$ for $x \neq 0$ and $F(0) = 0$, then $F \notin BV([a, b])$ for $a \leq 0 < b$ or $a < 0 \leq b$.

The verification of these examples is left to the reader (Exercise 27).

3.26 Lemma. *If $F \in BV$ is real-valued, then $T_F + F$ and $T_F - F$ are increasing.*

Proof. If $x < y$ and $\epsilon > 0$, choose $x_0 < \cdots < x_n = x$ such that

$$\sum_1^n |F(x_j) - F(x_{j-1})| \geq T_F(x) - \epsilon.$$

Then $\sum_1^n |F(x_j) - F(x_{j-1})| + |F(y) - F(x)|$ is an approximating sum for $T_F(y)$, and $F(y) = [F(y) - F(x)] + F(x)$, so

$$\begin{aligned} T_F(y) \pm F(y) &\geq \sum_1^n |F(x_j) + F(x_{j-1})| \\ &\quad + |F(y) - F(x)| \pm [F(y) - F(x)] \pm F(x) \\ &\geq T_F(x) - \epsilon \pm F(x). \end{aligned}$$

Since ϵ is arbitrary, $T_F(y) \pm F(y) \geq T_F(x) \pm F(x)$, as desired. ∎

3.27 Theorem.

a. $F \in BV$ *iff* $\operatorname{Re} F \in BV$ *and* $\operatorname{Im} F \in BV$.

b. If $F : \mathbb{R} \to \mathbb{R}$, *then* $F \in BV$ *iff* F *is the difference of two bounded increasing functions; for* $F \in BV$ *these functions may be taken to be* $\frac{1}{2}(T_F + F)$ *and* $\frac{1}{2}(T_F - F)$.

c. If $F \in BV$, *then* $F(x+) = \lim_{y \searrow x} F(y)$ *and* $F(x-) = \lim_{y \nearrow x} F(y)$ *exist for all* $x \in \mathbb{R}$, *as do* $F(\pm\infty) = \lim_{y \to \pm\infty} F(y)$.

d. If $F \in BV$, *the set of points at which* F *is discontinuous is countable.*

e. If $F \in BV$ *and* $G(x) = F(x+)$, *then* F' *and* G' *exist and are equal a.e.*

Proof. (a) is obvious. For (b), the "if" implication is easy (see Examples 3.25a,b). To prove "only if," observe that by Lemma 3.26, the equation $F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F)$ expresses F as the difference of two increasing functions. Also, the inequalities

$$T_F(y) \pm F(y) \geq T_F(x) \pm F(x) \qquad (y > x)$$

imply that

$$|F(y) - F(x)| \leq T_F(y) - T_F(x) \leq T_F(\infty) - T_F(-\infty) < \infty,$$

so that F, and hence $T_F \pm F$, is bounded. Finally, (c), (d), and (e) follow from (a), (b), and Theorem 3.23. ∎

The representation $F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F)$ of a real-valued $F \in BV$ is called the **Jordan decomposition** of F, and $\frac{1}{2}(T_F + F)$ and $\frac{1}{2}(T_F - F)$ are called the **positive** and **negative variations** of F. Since $x^+ = \max(x, 0) = \frac{1}{2}(|x| + x)$ and $x^- = \max(-x, 0) = \frac{1}{2}(|x| - x)$ for $x \in \mathbb{R}$, we have

$$\begin{aligned} &\tfrac{1}{2}(T_F \pm F)(x) \\ &\qquad = \sup\Big\{\sum_1^n [F(x_j) - F(x_{j-1})]^{\pm} : x_0 < \cdots < x_n = x\Big\} \pm \tfrac{1}{2}F(-\infty), \end{aligned}$$

Theorem 3.27(a,b) leads to the connection between BV and the space of complex Borel measures on $\mathbb{R}$. To make this precise, we introduce the space NBV (N for "normalized") defined by

$$NBV = \big\{F \in BV : F \text{ is right continuous and } F(-\infty) = 0\big\}.$$

We observe that if $F \in BV$, then the function G defined by $G(x) = F(x+) - F(-\infty)$ is in NBV and $G' = F'$ a.e. (That $G \in BV$ follows easily from Theorem 3.27(a,b): if F is real and $F = F_1 - F_2$ where F_1, F_2 are increasing, then $G(x) = F_1(x+) - [F_2(x+) + F(-\infty)]$, which is again the difference of two increasing functions.)

3.28 Lemma. *If $F \in BV$, then $T_F(-\infty) = 0$. If F is also right continuous, then so is T_F.*

Proof. If $\epsilon > 0$ and $x \in \mathbb{R}$, choose $x_0 < \cdots < x_n = x$ so that

$$\sum_1^n |F(x_j) - F(x_{j-1})| \geq T_F(x) - \epsilon.$$

From (3.24) we see that $T_F(x) - T_F(x_0) \geq T_F(x) - \epsilon$, and hence $T_F(y) \leq \epsilon$ for $y \leq x_0$. Thus $T_F(-\infty) = 0$.

Now suppose that F is right continuous. Given $x \in \mathbb{R}$ and $\epsilon > 0$, let $\alpha = T_F(x+) - T_F(x)$, and choose $\delta > 0$ so that $|F(x+h) - F(x)| < \epsilon$ and $T_F(x+h) - T_F(x+) < \epsilon$ whenever $0 < h < \delta$. For any such h, by (3.24) there exist $x_0 < \cdots < x_n = x + h$ such that

$$\sum_1^n |F(x_j) - F(x_{j-1})| \geq \tfrac{3}{4}[T_F(x+h) - T_F(x)] \geq \tfrac{3}{4}\alpha,$$

and hence

$$\sum_2^n |F(x_j) - F(x_{j-1})| \geq \tfrac{3}{4}\alpha - |F(x_1) - F(x_0)| \geq \tfrac{3}{4}\alpha - \epsilon.$$

Likewise, there exist $x = t_0 < \cdots < t_m = x_1$ such that $\sum_1^n |F(t_j) - F(t_{j-1})| \geq \frac{3}{4}\alpha$, and hence

$$\begin{aligned} \alpha + \epsilon &> T_F(x+h) - T_F(x) \\ &\geq \sum_1^m |F(t_j) - F(t_{j-1})| + \sum_2^n |F(x_j) - F(x_{j-1})| \\ &\geq \tfrac{3}{2}\alpha - \epsilon. \end{aligned}$$

Thus $\alpha < 4\epsilon$, and since ϵ is arbitrary, $\alpha = 0$. ∎

3.29 Theorem. *If μ is a complex Borel measure on $\mathbb{R}$ and $F(x) = \mu((-\infty, x])$, then $F \in NBV$. Conversely, if $F \in NBV$, there is a unique complex Borel measure μ_F such that $F(x) = \mu_F((-\infty, x])$; moreover, $|\mu_F| = \mu_{T_F}$.*

Proof. If μ is a complex measure, we have $\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)$ where the $\mu_j^\pm$ are finite measures. If $F_j^\pm(x) = \mu_j^\pm((-\infty, x])$, then $F_j^\pm$ is increasing and right continuous, $F_j^\pm(-\infty) = 0$, and $F_j^\pm(\infty) = \mu_j^\pm(\mathbb{R}) < \infty$. By Theorem 3.27(a,b),

the function $F = F_1^+ - F_1^- + i(F_2^+ - F_2^-)$ is in NBV. Conversely, by Theorem 3.27 and Lemma 3.28, any $F \in NBV$ can be written in this form with the $F_j^\pm$ increasing and in NBV. Each $F_j^\pm$ gives rise to a measure $\mu_j^\pm$ according to Theorem 1.16, so $F(x) = \mu_F((-\infty, x])$ where $\mu_F = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)$. The proof that $|\mu_F| = \mu_{T_F}$ is outlined in Exercise 28. ∎

The next obvious question is: Which functions in NBV correspond to measures μ such that $\mu \perp m$ or $\mu \ll m$? One answer is the following:

3.30 Proposition. *If $F \in NBV$, then $F' \in L^1(m)$. Moreover, $\mu_F \perp m$ iff $F' = 0$ a.e., and $\mu_F \ll m$ iff $F(x) = \int_{-\infty}^x F'(t)\,dt$.*

Proof. We have merely to observe that $F'(x) = \lim_{r\to 0} \mu_F(E_r)/m(E_r)$ where $E_r = (x, x+r]$ or $(x-r, x]$ and apply Theorem 3.22. (The measure μ_F is automatically regular by Theorem 1.18.) ∎

The condition $\mu_F \ll m$ can also be expressed directly in terms of F, as follows. A function $F : \mathbb{R} \to \mathbb{C}$ is called **absolutely continuous** if for every $\epsilon > 0$ there exists $\delta > 0$ such that for any finite set of disjoint intervals $(a_1, b_1), \dots, (a_N, b_N)$,

$$\sum_1^N (b_j - a_j) < \delta \quad \Longrightarrow \quad \sum_1^N |F(b_j) - F(a_j)| < \epsilon. \tag{3.31}$$

More generally, F is said to be **absolutely continuous** on $[a, b]$ if this condition is satisfied whenever the intervals (a_j, b_j) all lie in $[a, b]$. Clearly, if F is absolutely continuous, then F is uniformly continuous (take $N = 1$ in (3.31)). On the other hand, if F is everywhere differentiable and F' is bounded, then F is absolutely continuous, for $|F(b_j) - F(a_j)| \le (\max |F'|)(b_j - a_j)$ by the mean value theorem.

3.32 Proposition. *If $F \in NBV$, then F is absolutely continuous iff $\mu_F \ll m$.*

Proof. If $\mu_F \ll m$, the absolute continuity of F follows by applying Theorem 3.5 to the sets $E = \bigcup_1^N (a_j, b_j]$. To prove the converse, suppose that E is a Borel set such that $m(E) = 0$. If ϵ and δ are as in the definition of absolute continuity of F, by Theorem 1.18 we can find open sets $U_1 \supset U_2 \supset \cdots \supset E$ such that $m(U_1) < \delta$ (and thus $\mu(U_j) < \delta$ for all j) and $\mu_F(U_j) \to \mu_F(E)$. Each U_j is a disjoint union of open intervals (a_j^k, b_j^k), and

$$\sum_{k=1}^N \left|\mu_F\big((a_j^k, b_j^k)\big)\right| \le \sum_{k=1}^N |F(b_j^k) - F(a_j^k)| < \epsilon$$

for all N. Letting $N \to \infty$, we obtain $|\mu_F(U_j)| < \epsilon$ and hence $|\mu_F(E)| \le \epsilon$. Since ϵ is arbitrary, $\mu_F(E) = 0$, which shows that $\mu_F \ll m$. ∎

3.33 Corollary. *If $f \in L^1(m)$, then the function $F(x) = \int_{-\infty}^x f(t)\,dt$ is in NBV and is absolutely continuous, and $f = F'$ a.e. Conversely, if $F \in NBV$ is absolutely continuous, then $F' \in L^1(m)$ and $F(x) = \int_{-\infty}^x F'(t)\,dt$.*

Proof. This follows immediately from Propositions 3.30 and 3.32. ∎

If we consider functions on bounded intervals, this result can be refined a bit.

3.34 Lemma. *If F is absolutely continuous on $[a, b]$, then $F \in BV([a, b])$.*

Proof. Let δ be as in the definition of absolute continuity, corresponding to $\epsilon = 1$, and let N be the greatest integer less than $\delta^{-1}(b - a) + 1$. If $a = x_0 < \cdots < x_n = b$, by inserting more subdivision points if necessary, we can collect the intervals (x_{j-1}, x_j) into at most N groups of consecutive intervals such that the sum of the lengths in each group is less than δ. The sum $\sum |F(x_j) - F(x_{j-1})|$ over each group is at most 1, and hence the total variation of F on $[a, b]$ is at most N. ∎

3.35 The Fundamental Theorem of Calculus for Lebesgue Integrals. *If $-\infty < a < b < \infty$ and $F : [a, b] \to \mathbb{C}$, the following are equivalent:*

a. F is absolutely continuous on $[a, b]$.

b. $F(x) - F(a) = \int_a^x f(t)\,dt$ for some $f \in L^1([a, b], m)$.

c. F is differentiable a.e. on $[a, b]$, $F' \in L^1([a, b], m)$, and $F(x) - F(a) = \int_a^x F'(t)dt$.

Proof. To prove that (a) implies (c), we may assume by subtracting a constant from F that $F(a) = 0$. If we set $F(x) = 0$ for $x < a$ and $F(x) = F(b)$ for $x > b$, then $F \in NBV$ by Lemma 3.34, so (c) follows from Corollary 3.33. That (c) implies (b) is trivial. Finally, (b) implies (a) by setting $f(t) = 0$ for $t \notin [a, b]$ and applying Corollary 3.33. ∎

The following decomposition of Borel measures on $\mathbb{R}^n$ is sometimes important. A complex Borel measure μ on $\mathbb{R}^n$ is called **discrete** if there is a countable set $\{x_j\} \subset \mathbb{R}^n$ and complex numbers c_j such that $\sum |c_j| < \infty$ and $\mu = \sum c_j \delta_{x_j}$, where δ_x is the point mass at x. On the other hand, μ is called **continuous** if $\mu(\{x\}) = 0$ for all $x \in \mathbb{R}^n$. Any complex measure μ can be written uniquely as $\mu = \mu_d + \mu_c$ where μ_d is discrete and μ_c is continuous. Indeed, let $E = \{x : \mu(\{x\}) \neq 0\}$. For any countable subset F of E the series $\sum_{x \in F} \mu(\{x\})$ converges absolutely (to $\mu(F)$), so $\{x \in E : |\mu(\{x\})| > k^{-1}\}$ is finite for all k, and it follows that E itself is countable. Hence $\mu_d(A) = \mu(A \cap E)$ is discrete and $\mu_c(A) = \mu(A \setminus E)$ is continuous.

Obviously, if μ is discrete, then $\mu \perp m$; and if $\mu \ll m$, then μ is continuous. Thus, by Theorem 3.22, any (regular) complex Borel measure on $\mathbb{R}^n$ can be written uniquely as

$$\mu = \mu_d + \mu_{ac} + \mu_{sc}$$

where μ_d is discrete, μ_{ac} is absolutely continuous with respect to m, and μ_{sc} is a "singular continuous" measure, that is, μ_{sc} is continuous but $\mu_{sc} \perp m$.

The existence of nonzero singular continuous measures in $\mathbb{R}^n$ is evident enough when $n > 1$; the surface measure on the unit sphere discussed in §2.7 is one example. Their existence when $n = 1$ is not quite so obvious; they correspond via Theorem 3.29 to nonconstant functions $F \in NBV$ such that F is continuous but $F' = 0$ a.e. One such function is the Cantor function constructed in §1.5 (extended to $\mathbb{R}$ by setting $F(x) = 0$ for $x < 0$ and $F(x) = 1$ for $x > 1$). More surprisingly, there exist strictly increasing continuous functions F such that $F' = 0$ a.e.; see Exercise 40.

If $F \in NBV$, it is customary to denote the integral of a function g with respect to the measure μ_F by $\int g\,dF$ or $\int g(x)\,dF(x)$; such integrals are called **Lebesgue-Stieltjes integrals**. We conclude by presenting an integration-by-parts formula for Lebesgue-Stieltjes integrals; other variants of this result can be found in Exercises 34 and 35.

3.36 Theorem. *If F and G are in NBV and at least one of them is continuous, then for $-\infty < a < b < \infty$,*

$$\int_{(a,b]} F\,dG + \int_{(a,b]} G\,dF = F(b)G(b) - F(a)G(a).$$

Proof. F and G are linear combinations of increasing functions in NBV by Theorem 3.27(a,b), so a simple calculation shows that it suffices to assume F and G increasing. Suppose for the sake of definiteness that G is continuous, and let $\Omega = \{(x,y) : a < x \le y \le b\}$. We use Fubini's theorem to compute $\mu_F \times \mu_G(\Omega)$ in two ways:

$$\begin{aligned}\mu_F \times \mu_G(\Omega) &= \int_{(a,b]} \int_{(a,y]} dF(x)\,dG(x) = \int_{(a,b]} [F(y) - F(a)]\,dG(y)\\ &= \int_{(a,b]} F\,dG - F(a)[G(b) - G(a)],\end{aligned}$$

and since $G(x) = G(x-)$,

$$\begin{aligned}\mu_F \times \mu_G(\Omega) &= \int_{(a,b]} \int_{[x,b]} dG(y)\,dF(x) = \int_{(a,b]} [G(b) - G(x)]\,dF(x)\\ &= G(b)[F(b) - F(a)] - \int_{(a,b]} G\,dF.\end{aligned}$$

Subtracting these two equations, we obtain the desired result. ∎

Exercises

27. Verify the assertions in Examples 3.25.

28. If $F \in NBV$, let $G(x) = |\mu_F|((-\infty, x])$. Prove that $|\mu_F| = \mu_{T_F}$ by showing that $G = T_F$ via the following steps.

a. From the definition of T_F, $T_F \le G$.

b. $|\mu_F(E)| \le \mu_{T_F(E)}$ when E is an interval, and hence when E is a Borel set.

c. $|\mu_F| \le \mu_{T_F}$, and hence $G \le T_F$. (Use Exercise 21.)

29. If $F \in NBV$ is real-valued, then $\mu_F^+ = \mu_P$ and $\mu_F^- = \mu_N$ where P and N are the positive and negative variations of F. (Use Exercise 28.)

30. Construct an increasing function on $\mathbb{R}$ whose set of discontinuities is $\mathbb{Q}$.

31. Let $F(x) = x^2 \sin(x^{-1})$ and $G(x) = x^2 \sin(x^{-2})$ for $x \neq 0$, and $F(0) = G(0) = 0$.

a. F and G are differentiable everywhere (including $x = 0$).

b. $F \in BV([-1,1])$, but $G \notin BV([-1,1])$.

32. If $F_1, F_2, \ldots, F \in NBV$ and $F_j \to F$ pointwise, then $T_F \leq \liminf T_{F_j}$.

33. If F is increasing on $\mathbb{R}$, then $F(b) - F(a) \geq \int_a^b F'(t)\,dt$.

34. Suppose $F, G \in NBV$ and $-\infty < a < b < \infty$.

a. By adapting the proof of Theorem 3.36, show that

$$\int_{[a,b]} \frac{F(x) + F(x-)}{2}\,dG(x) + \int_{[a,b]} \frac{G(x) + G(x-)}{2}\,dF(x) = F(b)G(b) - F(a-)G(a-).$$

b. If there are no points in $[a,b]$ where F and G are both discontinuous, then

$$\int_{[a,b]} F\,dG + \int_{[a,b]} G\,dF = F(b)G(b) - F(a-)G(a-).$$

35. If F and G are absolutely continuous on $[a,b]$, then so is FG, and

$$\int_a^b (FG' + GF')(x)\,dx = F(b)G(b) - F(a)G(a).$$

36. Let G be a continuous increasing function on $[a,b]$ and let $G(a) = c$, $G(b) = d$.

a. If $E \subset [c,d]$ is a Borel set, then $m(E) = \mu_G(G^{-1}(E))$. (First consider the case where E is an interval.)

b. If f is a Borel measurable and integrable function on $[c,d]$, then $\int_c^d f(y)\,dy = \int_a^b f(G(x))\,dG(x)$. In particular, $\int_c^d f(y)\,dy = \int_a^b f(G(x))G'(x)\,dx$ if G is absolutely continuous.

c. The validity of (b) may fail if G is merely right continuous rather than continuous.

37. Suppose $F : \mathbb{R} \to \mathbb{C}$. There is a constant M such that $|F(x) - F(y)| \leq M|x-y|$ for all $x, y \in \mathbb{R}$ (that is, F is **Lipschitz continuous**) iff F is absolutely continuous and $|F'| \leq M$ a.e.

38. If $f : [a,b] \to \mathbb{R}$, consider the graph of f as a subset of $\mathbb{C}$, namely, $\{t + if(t) : t \in [a,b]\}$. The length L of this graph is by definition the supremum of the lengths of all inscribed polygons. (An "inscribed polygon" is the union of the line segments joining $t_{j-1} + if(t_{j-1})$ to $t_j + if(t_j)$, $1 \leq j \leq n$, where $a = t_0 < \cdots < t_n = b$.)

a. Let $F(t) = t + if(t)$; then L is the total variation of F on $[a,b]$.

b. If f is absolutely continuous, $L = \int_a^b [1 + f'(t)^2]^{1/2}\,dt$.

39. If $\{F_j\}$ is a sequence of nonnegative increasing functions on $[a,b]$ such that $F(x) = \sum_1^\infty F_j(x) < \infty$ for all $x \in [a,b]$, then $F'(x) = \sum_1^\infty F_j'(x)$ for a.e. $x \in [a,b]$. (It suffices to assume $F_j \in NBV$. Consider the measures μ_{F_j}.)

40. Let F denote the Cantor function on $[0,1]$ (see §1.5), and set $F(x) = 0$ for $x < 0$ and $F(x) = 1$ for $x > 1$. Let $\{[a_n, b_n]\}$ be an enumeration of the closed subintervals of $[0,1]$ with rational endpoints, and let $F_n(x) = F((x - a_n)/(b_n - a_n))$. Then $G = \sum_1^\infty 2^{-n} F_n$ is continuous and strictly increasing on $[0,1]$, and $G' = 0$ a.e. (Use Exercise 39.)

41. Let $A \subset [0,1]$ be a Borel set such that $0 < m(A \cap I) < m(I)$ for every subinterval I of $[0,1]$ (Exercise 33, Chapter 1).

a. Let $F(x) = m([0,x] \cap A)$. Then F is absolutely continuous and strictly increasing on $[0,1]$, but $F' = 0$ on a set of positive measure.

b. Let $G(x) = m([0,x] \cap A) - m([0,x] \setminus A)$. Then G is absolutely continuous on $[0,1]$, but G is not monotone on any subinterval of $[0,1]$.

42. A function $F : (a,b) \to \mathbb{R}$ $(-\infty \le a < b \le \infty)$ is called **convex** if

$$F(\lambda s + (1-\lambda)t) \le \lambda F(s) + (1-\lambda)F(t)$$

for all $s, t \in (a,b)$ and $\lambda \in (0,1)$. (Geometrically, this says that the graph of F over the interval from s to t lies underneath the line segment joining $(s, F(s))$ to $(t, F(t))$.)

a. F is convex iff for all $s, t, s', t' \in (a,b)$ such that $s \le s' < t'$ and $s < t \le t'$,

$$\frac{F(t) - F(s)}{t - s} \le \frac{F(t') - F(s')}{t' - s'}.$$

b. F is convex iff F is absolutely continuous on every compact subinterval of (a,b) and F' is increasing (on the set where it is defined).

c. If F is convex and $t_0 \in (a,b)$, there exists $\beta \in \mathbb{R}$ such that $F(t) - F(t_0) \ge \beta(t - t_0)$ for all $t \in (a,b)$.

d. **(Jensen's Inequality)** If $(X, \mathcal{M}, \mu)$ is a measure space with $\mu(X) = 1$, $g : X \to (a,b)$ is in $L^1(\mu)$, and F is convex on (a,b), then

$$F\left(\int g\, d\mu\right) \le \int F \circ g\, d\mu.$$

(Let $t_0 = \int g\, d\mu$ and $t = g(x)$ in (c), and integrate.)

3.6 NOTES AND REFERENCES

§3.2: The Lebesgue-Radon-Nikodym theorem was proved by Lebesgue [92] in the case where μ is Lebesgue measure on $\mathbb{R}^n$. Under the hypothesis $\nu \ll \mu$, it was generalized by Radon [111] to arbitrary regular Borel measures on $\mathbb{R}^n$ and by Nikodym [107] to measures on abstract spaces. The Lebesgue decomposition in the abstract setting appears in Saks [128]. The proof of the Lebesgue-Radon-Nikodym theorem in the text is similar to, but more efficient than, the one in Halmos [62]; I learned it from L. Loomis.

§3.3: The characterization

$$|\nu|(E) = \sup\Big\{\sum_1^n |\nu(E_j)| : n \in \mathbb{N},\ E_1, \dots, E_n \text{ disjoint}, E = \bigcup_1^n E_j\Big\}$$

of the total variation of a complex measure ν (see Exercise 21) is usually taken as the definition of $|\nu|$. Our definition seems more generally useful, and it is certainly easier to compute with.

§3.4: Theorems 3.21 and 3.22 are due to Lebesgue [92], but the line of argument we have presented is essentially that of Wiener [161], and the maximal function Hf, in dimension one, was first studied in Hardy and Littlewood [65]. Our proof of Theorem 3.18 is illustrative of a general technique that has been much exploited in recent years, namely, controlling the limiting behavior of a family of operators by means of estimates on an appropriate maximal function.

Lemma 3.15, a simplified version of Wiener's covering lemma, is taken from Rudin [125]. There is also an older and more delicate covering theorem, due to Vitali, which is used for similar purposes:

> If $E \subset \mathbb{R}^n$ and $\mathcal{Q}$ is a family of cubes such that each $x \in E$ is contained in members of $\mathcal{Q}$ of arbitrarily small diameter, then there is a (finite or infinite) disjoint sequence $\{Q_j\} \subset \mathcal{Q}$ such that $m(E \setminus \bigcup_j Q_j) = 0$.

Proofs can be found in many books, for example, Cohn [27, §6.2], Falconer [39, §1.3], and Hewitt and Stromberg [76, §17].

§3.5: The main results of this section are due to Lebesgue and Vitali; see Hawkins [70] for detailed references. Exercise 36 gives one form of the change-of-variable formula for Lebesgue integrals; others can be found in Serrin and Varberg [133]. Exercise 39 is a theorem of Fubini, and the example in Exercise 40 is due to Brown [21].

The Stieltjes integral $\int_a^b g\,dF$ was originally defined, under the hypothesis that F is an increasing function on $[a, b]$, as a limit of Riemann sums $\sum g(t_j)[F(t_j) - F(t_{j-1})]$. The theory of such "Riemann-Stieltjes" integrals is much like that of the ordinary Riemann integral, but some care is needed to handle cases where g and F are both allowed to be discontinuous. See ter Horst [148], which contains the analogue of Theorem 2.28 for Stieltjes integrals.

The example of the Cantor function shows that a continuous, a.e.-differentiable function need not be the integral of its derivative. It is a highly nontrivial theorem that if F is continuous on $[a, b]$, $F'(x)$ exists for every $x \in [a, b] \setminus A$ where A is *countable,* and $F' \in L^1$, then F is absolutely continuous and hence can be recovered from F' by integration. A proof can be found in Cohn [27, §6.3]; see also Rudin [125, Theorem 7.26] for the somewhat easier case when $A = \varnothing$.

However, this is not the end of the story, for there exist everywhere differentiable functions F such that $F' \notin L^1$. Perhaps the simplest example is $F(x) = x^2 \sin(x^{-2})$ (see Exercise 31). Here the only trouble is at $x = 0$, so for $a \le 0 \le b$ one could consider $\int_a^b F'(t)\,dt$ as an improper integral, i.e., the limit of Lebesgue integrals over

$[a, b] \setminus [-\epsilon, \epsilon]$ as $\epsilon \to 0$. However, it is not hard to construct examples in which the singularities of F' are so complicated that F' is not Lebesgue integrable on any interval. In this situation the Lebesgue integral is simply insufficient. However, the Henstock-Kurzweil integral (or the Denjoy or Perron integral) that was discussed in §2.8 is powerful enough to integrate such F', and by using this integral one obtains the general fundamental theorem of calculus: If F is everywhere differentiable on $[a, b]$, then $F(b) - F(a) = \int_a^b F'(t)\, dt$.

4

Point Set Topology

The concepts of limit, convergence, and continuity are central to all of analysis, and it is useful to have a general framework for studying them that includes the classical manifestations as special cases. One such framework, which has the advantage of not requiring many ideas beyond those occurring in analysis on Euclidean space, is that of metric spaces. However, metric spaces are not sufficiently general to describe even some very classical modes of convergence, for example, pointwise convergence of functions on $\mathbb{R}$. A more flexible theory can be built by taking the open sets, rather than a metric, as the primitive data, and it is this theory that we shall explore in the present chapter.

4.1 TOPOLOGICAL SPACES

Let X be a nonempty set. A **topology** on X is a family $\mathcal{T}$ of subsets of X that contains $\varnothing$ and X and is closed under arbitrary unions and finite intersections (i.e., if $\{U_\alpha\}_{\alpha\in A} \subset \mathcal{T}$ then $\bigcup_{\alpha\in A} U_\alpha \in \mathcal{T}$, and if $U_1, \ldots, U_n \in \mathcal{T}$ then $\bigcap_1^n U_j \in \mathcal{T}$). The pair $(X, \mathcal{T})$ is called a **topological space**. If $\mathcal{T}$ is understood, we shall simply refer to the topological space X. Let us examine a few examples:

- If X is any nonempty set, $\mathcal{P}(X)$ and $\{\varnothing, X\}$ are topologies on X. They are called the **discrete topology** and the **trivial** (or **indiscrete**) **topology**, respectively.

- If X is an infinite set, $\{U \subset X : U = \varnothing \text{ or } U^c \text{ is finite}\}$ is a topology on X, called the **cofinite topology**.

- If X is a metric space, the collection of all open sets with respect to the metric is a topology on X.

- If $(X, \mathcal{T})$ is a topological space and $Y \subset X$, then $\mathcal{T}_Y = \{U \cap Y : U \in \mathcal{T}\}$ is a topology on Y, called the **relative topology** induced by $\mathcal{T}$.

We now present the basic terminology concerning topological spaces. Most of these concepts are already familiar in the context of metric spaces. Until further notice, $(X, \mathcal{T})$ will be a fixed topological space.

The members of $\mathcal{T}$ are called **open sets**, and their complements are called **closed sets**. If $Y \subset X$, the open (closed) subsets of Y in the relative topology are called **relatively** open (closed). We observe that, by deMorgan's laws, the family of closed sets is closed under arbitrary intersections and finite unions.

If $A \subset X$, the union of all open sets contained in A is called the **interior** of A, and the intersection of all closed sets containing A is called the **closure** of A. We denote the interior and closure of A by A^o and $\overline{A}$, respectively. Clearly A^o is the largest open set contained in A and $\overline{A}$ is the smallest closed set containing A, and we have $(A^o)^c = \overline{A^c}$ and $(\overline{A})^c = (A^c)^o$. The difference $\overline{A} \setminus A^o = \overline{A} \cap \overline{A^c}$ is called the **boundary** of A and is denoted by ∂A. If $\overline{A} = X$, A is called **dense** in X. On the other hand, if $(\overline{A})^o = \varnothing$, A is called **nowhere dense**.

If $x \in X$ (or $E \subset X$), a **neighborhood** of x (or E) is a set $A \subset X$ such that $x \in A^o$ (or $E \subset A^o$). Thus, a set A is open iff it is a neighborhood of itself. (Some authors require neighborhoods to be open sets; we do not.) A point x is called an **accumulation point** of A if $A \cap (U \setminus \{x\}) \neq \varnothing$ for every neighborhood U of x. (Other terms sometimes used for the same concept are "cluster point" and "limit point." We shall use "cluster point" to mean something a bit different below.)

4.1 Proposition. *If $A \subset X$, let* $\operatorname{acc}(A)$ *be the set of accumulation points of A. Then* $\overline{A} = A \cup \operatorname{acc}(A)$, *and A is closed iff* $\operatorname{acc}(A) \subset A$.

Proof. If $x \notin \overline{A}$, then A^c is a neighborhood of x that does not intersect A, so $x \notin \operatorname{acc}(A)$; thus $A \cup \operatorname{acc}(A) \subset \overline{A}$. If $x \notin A \cup \operatorname{acc}(A)$, there is an open U containing x such that $U \cap A = \varnothing$, so that $\overline{A} \subset U^c$ and $x \notin \overline{A}$. Thus $\overline{A} \subset A \cup \operatorname{acc}(A)$. Finally, A is closed iff $A = \overline{A}$, and this happens iff $\operatorname{acc}(A) \subset A$. ∎

If $\mathcal{T}_1$ and $\mathcal{T}_2$ are topologies on X such that $\mathcal{T}_1 \subset \mathcal{T}_2$, we say that $\mathcal{T}_1$ is **weaker** (or **coarser**) than $\mathcal{T}_2$, or that $\mathcal{T}_2$ is **stronger** (or **finer**) than $\mathcal{T}_1$. Clearly the trivial topology is the weakest topology on X, while the discrete topology is the strongest. If $\mathcal{E} \subset \mathcal{P}(X)$, there is a unique weakest topology $\mathcal{T}(\mathcal{E})$ on X that contains $\mathcal{E}$, namely the intersection of all topologies on X containing $\mathcal{E}$. It is called the topology **generated** by $\mathcal{E}$, and $\mathcal{E}$ is sometimes called a **subbase** for $\mathcal{T}(\mathcal{E})$.

If $\mathcal{T}$ is a topology on X, a **neighborhood base** for $\mathcal{T}$ at $x \in X$ is a family $\mathcal{N} \subset \mathcal{T}$ such that

- $x \in V$ for all $V \in \mathcal{N}$;

- if $U \in \mathcal{T}$ and $x \in U$, there exists $V \in \mathcal{N}$ such that $x \in V$ and $V \subset U$.

A **base** for $\mathcal{T}$ is a family $\mathcal{B} \subset \mathcal{T}$ that contains a neighborhood base for $\mathcal{T}$ at each $x \in X$. For example, if X is a metric space, the collection of open balls centered at x is a neighborhood base for the metric topology at x, and the collection of all open balls in X is a base.

4.2 Proposition. *If $\mathcal{T}$ is a topology on X and $\mathcal{E} \subset \mathcal{T}$, then $\mathcal{E}$ is a base for $\mathcal{T}$ iff every nonempty $U \in \mathcal{T}$ is a union of members of $\mathcal{E}$.*

Proof. If $\mathcal{E}$ is a base, $U \in \mathcal{T}$, and $x \in U$, there exists $V_x \in \mathcal{E}$ with $x \in V_x \subset U$, so $U = \bigcup_{x \in U} V_x$. Conversely, if every nonempty $U \in \mathcal{T}$ is a union of members of $\mathcal{E}$, then $\{V \in \mathcal{E} : x \in V\}$ is clearly a neighborhood base at x, so $\mathcal{E}$ is a base. ∎

4.3 Proposition. *If $\mathcal{E} \subset \mathcal{P}(X)$, in order for $\mathcal{E}$ to be a base for a topology on X it is necessary and sufficient that the following two conditions be satisfied:*

a. each $x \in X$ is contained in some $V \in \mathcal{E}$;

b. if $U, V \in \mathcal{E}$ and $x \in U \cap V$, there exists $W \in \mathcal{E}$ with $x \in W \subset (U \cap V)$.

Proof. The necessity is clear, since if U, V are open, then so is $U \cap V$. To prove the sufficiency, let

$$\mathcal{T} = \{U \subset X : \text{for every } x \in U \text{ there exists } V \in \mathcal{E} \text{ with } x \in V \subset U\}.$$

Then $X \in \mathcal{T}$ by condition (a) and $\varnothing \in \mathcal{T}$ trivially, and $\mathcal{T}$ is obviously closed under unions. If $U_1, U_2 \in \mathcal{T}$ and $x \in U_1 \cap U_2$, there exist $V_1, V_2 \in \mathcal{E}$ with $x \in V_1 \subset U_1$ and $x \in V_2 \subset U_2$, and by condition (b) there exists $W \in \mathcal{E}$ with $x \in W \subset (V_1 \cap V_2)$. Thus $U_1 \cap U_2 \in \mathcal{T}$, so by induction $\mathcal{T}$ is closed under finite intersections. Therefore $\mathcal{T}$ is a topology, and $\mathcal{E}$ is clearly a base for $\mathcal{T}$. ∎

4.4 Proposition. *If $\mathcal{E} \subset \mathcal{P}(X)$, the topology $\mathcal{T}(\mathcal{E})$ generated by $\mathcal{E}$ consists of $\varnothing$, X, and all unions of finite intersections of members of $\mathcal{E}$.*

Proof. The family of finite intersections of sets in $\mathcal{E}$, together with X, satisfies the conditions of Proposition 4.3, so by Propostiion 4.2 the family of all unions of such sets, together with $\varnothing$, is a topology. It is obviously contained in $\mathcal{T}(\mathcal{E})$, hence equal to $\mathcal{T}(\mathcal{E})$. ∎

Note how the simplicity of this proposition contrasts with the corresponding result for σ-algebras (Proposition 1.23). What makes life easier here is that only *finite* intersections are involved.

The concept of topological space is general enough to include a great profusion of interesting examples, but — by the same token — too general to yield many interesting theorems. To build a reasonable theory one must usually restrict the class of spaces under consideration. The remainder of this section is devoted to a discussion of two types of restrictions that are commonly made, the so-called countability and separation axioms.

A topological space $(X, \mathcal{T})$ satisfies the **first axiom of countability**, or is **first countable**, if there is a countable neighborhood base for $\mathcal{T}$ at every point of X. (It is useful to observe that if X is first countable, for every $x \in X$ there is a neighborhood base $\{U_j\}_1^\infty$ at x such that $U_j \supset U_{j+1}$ for all j. Indeed, if $\{V_j\}_1^\infty$ is any countable neighborhood base at x, we can take $U_j = \bigcap_1^j V_i$.) The space $(X, \mathcal{T})$ satisfies the **second axiom of countability**, or is **second countable**, if $\mathcal{T}$ has a countable base. Also, $(X, \mathcal{T})$ is **separable** if X has a countable dense subset. Every metric space is first countable (the balls of rational radius about x are a neighborhood base at x), and a metric space is second countable iff it is separable (Exercise 5). The latter fact can be partly generalized:

4.5 Proposition. *Every second countable space is separable.*

Proof. If X is second countable, let $\mathcal{E}$ be a countable base for the topology, and for each $U \in \mathcal{E}$ pick a point $x_U \in U$. Then the complement of the closure of $\{x_U : U \in \mathcal{E}\}$ is an open set that does not include any $U \in \mathcal{E}$; hence it is empty and $\{x_U : U \in \mathcal{E}\}$ is dense. ∎

A sequence $\{x_j\}$ in a topological space X **converges** to $x \in X$ (in symbols: $x_j \to x$) if for every neighborhood U of x there exists $J \in \mathbb{N}$ such that $x_j \in U$ for all $j > J$. First countable spaces have the pleasant property that such things as closure and continuity can be characterized in terms of sequential convergence — which is not the case in more general spaces, as we shall see. For example:

4.6 Proposition. *If X is first countable and $A \subset X$, then $x \in \overline{A}$ iff there is a sequence $\{x_j\}$ in A that converges to x.*

Proof. Let $\{U_j\}$ be a countable neighborhood base at x with $U_j \supset U_{j+1}$ for all j. If $x \in \overline{A}$, then $U_j \cap A \neq \varnothing$ for all j. Pick $x_j \in U_j \cap A$; since $U_k \subset U_j$ for $k > j$ and every neighborhood of x contains some U_j, it is clear that $x_j \to x$. On the other hand, if $x \notin \overline{A}$ and $\{x_j\}$ is any sequence in A, then $(\overline{A})^c$ is a neighborhood of x containing no x_j, so $x_j \not\to x$. ∎

Lastly, we discuss the separation axioms. These are properties of a topological space, labeled $T_0, \ldots, T_4$, that guarantee the existence of open sets that separate points or closed sets from each other. If X has the property T_j, we say that X is a T_j space or that the topology on X is T_j.

T_0: If $x \neq y$, there is an open set containing x but not y *or* an open set containing y but not x.

T_1: If $x \neq y$, there is an open set containing y but not x.

T_2: If $x \neq y$, there are disjoint open sets U, V with $x \in U$ and $y \in V$.

T_3: X is a T_1 space, and for any closed set $A \subset X$ and any $x \in A^c$ there are disjoint open sets U, V with $x \in U$ and $a \subset V$.

T_4: X is a T_1 space, and for any disjoint closed sets A, B in X there are disjoint open sets U, V with $A \subset U$ and $B \subset V$.

T_2, T_3, and T_4 also have other names: A T_2 space is a **Hausdorff** space, a T_3 space is a **regular** space, and a T_4 space is a **normal** space. (Some authors do not require regular and normal spaces to be T_1.) There is an additional useful separation condition, intermediate between T_3 and T_4, that we shall discuss in §4.2.

The following characterization of T_1 spaces is useful. It shows in particular that every normal space is regular and that every regular space is Hausdorff.

4.7 Proposition. *X is a T_1 space iff $\{x\}$ is closed for every $x \in X$.*

Proof. If X is T_1 and $x \in X$, for each $y \neq x$ there is an open U_y containing y but not x; thus $\{x\}^c = \bigcup_{y \neq x} U_y$ is open and $\{x\}$ is closed. Conversely, if $\{x\}$ is closed, then $\{x\}^c$ is an open set containing every $y \neq x$. ∎

The vast majority of topological space that arise in practice (and, in particular, in this book) are Hausdorff, or become Hausdorff after simple modifications. (This last phrase refers to spaces such as the space of integrable functions on a measure space, which becomes a Hausdorff space with the L^1 metric when we identify two functions that are equal a.e.) However, two classes of (usually) non-Hausdorff topologies are of sufficient importance to warrant special mention: the quotient topology on a space of equivalence classes, discussed in Exercises 28 and 29 (§4.2), and the Zariski topology on an algebraic variety. Without attempting to give the definition of an algebraic variety, we shall describe the Zariski topology on a vector space.

Let k be a field, and let $k(X_1, \ldots, X_n)$ be the ring of polynomials in n variables over x. Each $P \in k(X_1, \ldots, X_n)$ determines a polynomial map $p : k^n \to k$ by substituting elements of k for the formal indeterminates $X_1, \ldots, X_n$. The correspondence $P \to p$ is one-to-one precisely when k is infinite. The collection of all sets $p^{-1}(\{0\})$ in k^n, as p ranges over all polynomial maps, is closed under finite unions, since $p^{-1}(\{0\}) \cup q^{-1}(\{0\}) = (pq)^{-1}(\{0\})$, and it contains k^n itself (take $p = 0$). Hence, by Propositions 4.2 and 4.3, the collection of all sets of the form $\bigcap_{\alpha \in A} p_\alpha^{-1}(\{0\})$ (p_α being a polynomial map for each α) is the collection of closed sets for a topology on k^n, called the **Zariski topology**. The Zariski topology is T_1 by Proposition 4.7, for if $a = (a_1, \ldots, a_n) \in k^n$ then $\{a\} = \bigcap_1^n p_j^{-1}(\{0\})$ where $p_j(X_1, \ldots, X_n) = X_j - a_j$. If k is finite the Zariski topology is discrete, but if k is infinite the Zariski topology is not Hausdorff; in fact, any two nonempty open sets have nonempty intersection. This is just a restatement of the fact that $k(X_1, \ldots, X_n)$ is an integral domain, that is, if P and Q are nonzero polynomials, then PQ is nonzero. (For $n = 1$, the Zariski topology is the cofinite topology.)

Other examples illustrating the separation and countability axioms will be found in the exercises.

Exercises

1. If $\operatorname{card}(X) \geq 2$, there is a topology on X that is T_0 but not T_1.

2. If X is an infinite set, the cofinite topology on X is T_1 but not T_2, and is first countable iff X is countable.

3. Every metric space is normal. (If A, B are closed sets in the metric space (X, ρ), consider the sets of points x where $\rho(x, A) < \rho(x, B)$ or $\rho(x, A) > \rho(x, B)$.)

4. Let $X = \mathbb{R}$, and let $\mathcal{T}$ be the family of all subsets of $\mathbb{R}$ of the form $U \cup (V \cap \mathbb{Q})$ where U, V are open in the usual sense. Then $\mathcal{T}$ is a topology that is Hausdorff but not regular. (In view of Exercise 3, this shows that a topology stronger than a normal topology need not be normal or even regular.)

5. Every separable metric space is second countable.

6. Let $\mathcal{E} = \{(a, b] : -\infty < a < b < \infty\}$.

a. $\mathcal{E}$ is a base for a topology $\mathcal{T}$ on $\mathbb{R}$ in which the members of $\mathcal{E}$ are both open and closed.

b. $\mathcal{T}$ is first countable but not second countable. (If $x \in \mathbb{R}$, every neighborhood base at x contains a set whose supremum is x.)

c. $\mathbb{Q}$ is dense in $\mathbb{R}$ with respect to $\mathcal{T}$. (Thus the converse of Proposition 4.5 is false.)

7. If X is a topological space, a point $x \in X$ is called a **cluster point** of the sequence $\{x_j\}$ if for every neighborhood U of x, $x_j \in U$ for infinitely many j. If X is first countable, x is a cluster point of $\{x_j\}$ iff some subsequence of $\{x_j\}$ converges to x.

8. If X is an infinite set with the cofinite topology and $\{x_j\}$ is a sequence of distinct points in X, then $x_j \to x$ for every $x \in X$.

9. If X is a linearly ordered set, the topology $\mathcal{T}$ generated by the sets $\{x : x < a\}$ and $\{x : x > a\}$ $(a \in X)$ is called the **order topology**.

a. If $a, b \in X$ and $a < b$, there exist $U, V \in \mathcal{T}$ with $a \in U$, $b \in V$, and $x < y$ for all $x \in U$ and $y \in V$. The order topology is the weakest topology with this property.

b. If $Y \subset X$, the order topology on Y is never stronger than, but may be weaker than, the relative topology on Y induced by the order topology on X.

c. The order topology on $\mathbb{R}$ is the usual topology.

10. A topological space X is called **disconnected** if there exist nonempty open sets U, V such that $U \cap V = \varnothing$ and $U \cup V = X$; otherwise X is **connected**. When we speak of connected or disconnected subsets of X, we refer to the relative topology on them.

a. X is connected iff $\varnothing$ and X are the only subsets of X that are both open and closed.

b. If $\{E_\alpha\}_{\alpha \in A}$ is a collection of connected subsets of X such that $\bigcap_{\alpha \in A} E_\alpha \neq \varnothing$, then $\bigcup_{\alpha \in A} E_\alpha$ is connected.

c. If $A \subset X$ is connected, then $\overline{A}$ is connected.

d. Every point $x \in X$ is contained in a unique maximal connected subset of X, and this subset is closed. (It is called the **connected component** of x.)

11. If $E_1, \ldots, E_n$ are subsets of a topological space, the closure of $\bigcup_1^n E_j$ is $\bigcup_1^n \overline{E}_j$.

12. Let X be a set. A **Kuratowski closure operator** on X is a map $A \mapsto A^*$ from $\mathcal{P}(X)$ to itself satisfying (i) $\varnothing^* = \varnothing$, (ii) $A \subset A^*$ for all A, (iii) $(A^*)^* = A^*$ for all A, and (iv) $(A \cup B)^* = A^* \cup B^*$ for all A, B.

a. If X is a topological space, the map $A \mapsto \overline{A}$ is a Kuratowski closure operator. (Use Exercise 11.)

b. Conversely, given a Kuratowski closure operator, let $\mathcal{F} = \{A \subset X : A = A^*\}$ and $\mathcal{T} = \{U \subset X : U^c \in \mathcal{F}\}$. Then $\mathcal{T}$ is a topology, and for any set $A \subset X$, A^* is its closure with respect to $\mathcal{T}$.

13. If X is a topological space, U is open in X, and A is dense in X, then $\overline{U} = \overline{U \cap A}$.

4.2 CONTINUOUS MAPS

Topological spaces are the natural setting for the concept of continuity, which can be described in either global or local terms as follows. Let X and Y be topological spaces and f a map from X to Y. Then f is called **continuous** if $f^{-1}(V)$ is open in X for every open $V \subset Y$. (Since $f^{-1}(A^c) = [f^{-1}(A)]^c$, an equivalent condition is that $f^{-1}(A)$ is closed in X for every closed $A \subset Y$.) If $x \in X$, f is called **continuous at** x if for every neighborhood V of $f(x)$ there is a neighborhood U of x such that $f(U) \subset V$, or equivalently, if $f^{-1}(V)$ is a neighborhood of x for every neighborhood V of $f(x)$. Clearly, if $f : X \to Y$ and $g : Y \to Z$ are continuous (or f is continuous at x and g is continuous at $f(x)$), then $g \circ f$ is continuous (at x). We shall denote the set of continuous maps from X to Y by $C(X, Y)$.

4.8 Proposition. *The map $f : X \to Y$ is continuous iff f is continuous at every $x \in X$.*

Proof. If f is continuous and V is a neighborhood of $f(x)$, $f^{-1}(V^o)$ is an open set containing x, so f is continuous at x. Conversely, suppose that f is continuous at each $x \in X$. If $V \subset Y$ is open, V is a neighborhood of each of its points, so $f^{-1}(V)$ is a neighborhood of each of its points. Thus $f^{-1}(V)$ is open, so f is continuous. ∎

4.9 Proposition. *If the topology on Y is generated by a family of sets $\mathcal{E}$, then $f : X \to Y$ is continuous iff $f^{-1}(V)$ is open in X for every $V \in \mathcal{E}$.*

Proof. This is clear from Proposition 4.4 and the fact that the set mapping f^{-1} commutes with unions and intersections. ∎

If $f : X \to Y$ is bijective and f and f^{-1} are both continuous, f is called a **homeomorphism**, and X and Y are said to be **homeomorphic**. In this case the set mapping f^{-1} is a bijection from the open sets in Y to the open sets in X, so

X and Y may be considered identical as far as their topological properties go. If $f : X \to Y$ is injective but not surjective, and $f : X \to f(X)$ is a homeomorphism when $f(X) \subset Y$ is given the relative topology, f is called an **embedding**.

If X is any set and $\{f_\alpha : X \to Y_\alpha\}_{\alpha \in A}$ is a family of maps from X into some topological spaces Y_α, there is a unique weakest topology $\mathcal{T}$ on X that makes all the f_α continuous; it is called the **weak topology** generated by $\{f_\alpha\}_{\alpha \in A}$. Namely, $\mathcal{T}$ is the topology generated by sets of the form $f_\alpha^{-1}(U_\alpha)$ where $\alpha \in A$ and U_α is open in Y_α.

The most important example of this construction is the Cartesian product of topological spaces. If $\{X_\alpha\}_{\alpha \in A}$ is any family of topological spaces, the **product topology** on $X = \prod_{\alpha \in A} X_\alpha$ is the weak topology generated by the coordinate maps $\pi_\alpha : X \to X_\alpha$. When we consider a Cartesian product of topological spaces, we *always* endow it with the product topology unless we specify otherwise. By Proposition 4.4, a base for the product topology is given by the sets of the form $\bigcap_1^n \pi_{\alpha_j}^{-1}(U_{\alpha_j})$ where $n \in \mathbb{N}$ and U_{α_j} is open in X_{α_j} for $1 \le j \le n$. These sets can also be written as $\prod_{\alpha \in A} U_\alpha$ where $U_\alpha = X_\alpha$ if $\alpha \neq \alpha_1, \ldots, \alpha_n$. Notice, in particular, that if A is infinite, a product of nonempty open sets $\prod_{\alpha \in A} U_\alpha$ is open in $\prod_{\alpha \in A} X_\alpha$ iff $U_\alpha = X_\alpha$ for all but finitely many α.

4.10 Proposition. *If X_α is Hausdorff for each $\alpha \in A$, then $X = \prod_{\alpha \in A} X_\alpha$ is Hausdorff.*

Proof. If x and y are distinct points of X, we must have $\pi_\alpha(x) \neq \pi_\alpha(y)$ for some α. Let U and V be disjoint neighborhoods of $\pi_\alpha(x)$ and $\pi_\alpha(y)$ in X_α. Then $\pi_\alpha^{-1}(U)$ and $\pi_\alpha^{-1}(V)$ are disjoint neighborhoods of x and y in X. ∎

4.11 Proposition. *If X_α ($\alpha \in A$) and Y are topological spaces and $X = \prod_{\alpha \in A} X_\alpha$, then $f : Y \to X$ is continuous iff $\pi_\alpha \circ f$ is continuous for each α.*

Proof. If $\pi_\alpha \circ f$ is continuous for each α, then $f^{-1}(\pi_\alpha^{-1}(U_\alpha))$ is open in Y for each open U_α in X_α. By Proposition 4.9, f is continuous. The converse is obvious. ∎

If the spaces X_α are all equal to some fixed space X, the product $\prod_{\alpha \in A} X_\alpha$ is just the set X^A of mappings from A to X, and the product topology is just the topology of pointwise convergence. More precisely:

4.12 Proposition. *If X is a topological space, A is a nonempty set, and $\{f_n\}$ is a sequence in X^A, then $f_n \to f$ in the product topology iff $f_n \to f$ pointwise.*

Proof. The sets

$$N(U_1, \ldots, U_k) = \bigcap_1^k \pi_{\alpha_j}^{-1}(U_j) = \{g \in X^A : g(\alpha_j) \in U_j \text{ for } 1 \le j \le k\},$$

where $k \in \mathbb{N}$ and U_j is a neighborhood of $f(\alpha_j)$ in X for each j, form a neighborhood base for the product topology at f. If $f_n \to f$ pointwise, then $f_n(\alpha_j) \in U_j$ for

$n \geq N_j$ and hence $f_n \in N(U_1, \ldots, U_k)$ for $n \geq \max(N_1, \ldots, N_k)$; therefore $f_n \to f$ in the product topology. Conversely, if $f_n \to f$ in the product topology, $\alpha \in A$, and U is a neighborhood of $f(\alpha)$, then $f_n \in N(U) = \pi_\alpha^{-1}(U)$ for large n; hence $f_n(\alpha) \in U$ for large n, and so $f_n(\alpha) \to f(\alpha)$. ∎

We shall be particularly interested in real- and complex-valued functions on topological spaces. If X is any set, we denote by $B(X, \mathbb{R})$ (resp. $B(X, \mathbb{C})$) the space of all bounded real- (resp. complex-) valued functions on X. If X is a topological space, we also have the spaces $C(X, \mathbb{R})$ and $C(X, \mathbb{C})$ of countinuous functions on X, and we define

$$BC(X, F) = B(X, F) \cap C(X, F) \qquad (F = \mathbb{R} \text{ or } \mathbb{C}).$$

In speaking of complex-valued functions we shall usually omit the $\mathbb{C}$ and simply write $B(X)$, $C(X)$, and $BC(X)$. Since addition and multiplication are continuous from $\mathbb{C} \times \mathbb{C}$ to $\mathbb{C}$, $C(X)$ and $BC(X)$ are complex vector spaces.

If $f \in B(X)$, we define the **uniform norm** of f to be

$$\|f\|_u = \sup\{|f(x)| : x \in X\}.$$

The function $\rho(f, g) = \|f - g\|_u$ is easily seen to be a metric on $B(X)$, and convergence with respect to this metric is simply uniform convergence on X. $B(X)$ is obviously complete in the uniform metric: If $\{f_n\}$ is uniformly Cauchy, then $\{f_n(x)\}$ is Cauchy for each x, and if we set $f(x) = \lim_n f_n(x)$, it is easily verified that $\|f_n - f\|_u \to 0$.

4.13 Proposition. *If X is a topological space, $BC(X)$ is a closed subspace of $B(X)$ in the uniform metric; in particular, $BC(X)$ is complete.*

Proof. Suppose $\{f_n\} \subset BC(X)$ and $\|f_n - f\|_u \to 0$. Given $\epsilon > 0$, choose N so large that $\|f_n - f\|_u < \epsilon/3$ for $n > N$. Given $n > N$ and $x \in X$, since f_n is continuous at x there is a neighborhood U of x such that $|f_n(y) - f_n(x)| < \epsilon/3$ for $y \in U$. But then

$$|f(y) - f(x)| \leq |f(y) - f_n(y)| + |f_n(y) - f_n(x)| + |f_n(x) - f(x)| < \epsilon,$$

so f is continuous at x. By Proposition 4.8, f is continuous. ∎

For a given topological space X it may happen that $C(X)$ consists only of constant functions. This is obviously the case, for example, if X has the trivial topology, but it can happen even when X is regular. Normal spaces, however, always have plenty of continuous functions, as the following fundamental theorems show.

4.14 Lemma. *Suppose that A and B are disjoint closed subsets of the normal space X, and let $\Delta = \{k2^{-n} : n \geq 1 \text{ and } 0 < k < 2^n\}$ be the set of dyadic rational numbers in $(0, 1)$. There is a family $\{U_r : r \in \Delta\}$ of open sets in X such that $A \subset U_r \subset B^c$ for all $r \in \Delta$ and $\overline{U}_r \subset U_s$ for $r < s$.*

Proof. By normality, there exist disjoint open sets V, W such that $A \subset V$, $B \subset W$. Let $U_{1/2} = V$. Then since W^c is closed,

$$A \subset U_{1/2} \subset \overline{U}_{1/2} \subset W^c \subset B^c.$$

We now select U_r for $r = k2^{-n}$ by induction on n. Suppose that we have chosen U_r for $r = k2^{-n}$ when $0 < k < 2^n$ and $n \leq N - 1$. To find U_r for $r = (2j+1)2^{-N}$ ($0 \leq j < 2^{N-1}$), observe that $\overline{U}_{j2^{1-N}}$ and $(U_{(j+1)2^{1-N}})^c$ are disjoint closed sets (where we set $\overline{U}_0 = A$ and $U_1^c = B$), so as above we can choose an open U_r with

$$A \subset \overline{U}_{j2^{1-N}} \subset U_r \subset \overline{U}_r \subset U_{(j+1)2^{1-N}} \subset B^c.$$

These U_r's clearly have the desired properties. ∎

4.15 Urysohn's Lemma. *Let X be a normal space. If A and B are disjoint closed sets in X, there exists $f \in C(X, [0,1])$ such that $f = 0$ on A and $f = 1$ on B.*

Proof. Let U_r be as in Lemma 4.14 for $r \in \Delta$, and set $U_1 = X$. For $x \in X$, define $f(x) = \inf\{r : x \in U_r\}$. Since $A \subset U_r \subset B^c$ for $0 < r < 1$, we clearly have $f(x) = 0$ for $x \in A$ and $f(x) = 1$ for $x \in B$, and $0 \leq f(x) \leq 1$ for all $x \in X$. It remains to show that f is continuous. To this end, observe that $f(x) < \alpha$ iff $x \in U_r$ for some $r < \alpha$ iff $x \in \bigcup_{r<\alpha} U_r$, so $f^{-1}((-\infty, \alpha)) = \bigcup_{r<\alpha} U_r$ is open. Also $f(x) > \alpha$ iff $x \notin U_r$ for some $r > \alpha$ iff $x \notin \overline{U}_s$ for some $s > \alpha$ (since $\overline{U}_s \subset U_r$ for $s < r$) iff $x \in \bigcup_{s>\alpha} (\overline{U}_s)^c$, so $f^{-1}((a, \infty)) = \bigcup_{s>\alpha} (\overline{U}_s)^c$ is open. Since the open half-lines generate the topology on $\mathbb{R}$, f is continuous by Proposition 4.9. ∎

The proof of Urysohn's lemma may seem somewhat opaque at first, but there is a simple geometric intuition behind it. If one pictures X as the plane $\mathbb{R}^2$ and the sets U_r as regions bounded by curves, the curves ∂U_r form a "topographic map" of the function f.

4.16 The Tietze Extension Theorem. *Let X be a normal space. If A is a closed subset of X and $f \in C(A, [a,b])$, there exists $F \in C(X, [a,b])$ such that $F|A = f$.*

Proof. Replacing f by $(f-a)/(b-a)$, we may assume that $[a,b] = [0,1]$. We claim that there is a sequence $\{g_n\}$ of continuous functions on X such that $0 \leq g_n \leq 2^{n-1}/3^n$ on X and $0 \leq f - \sum_1^n g_j \leq (2/3)^n$ on A. To begin with, let $B = f^{-1}([0, 1/3])$ and $C = f^{-1}([2/3, 1])$. These are closed subsets of A, and since A itself is closed, they are closed in X. By Urysohn's lemma there is a continuous $g_1 : X \to [0, 1/3]$ with $g_1 = 0$ on B and $g_1 = 1/3$ on C; it follows that $0 \leq f - g_1 \leq 2/3$ on A. Having found $g_1, \ldots, g_{n-1}$, by the same reasoning we can find $g_n : X \to [0, 2^{n-1}/3^n]$ such that $g_n = 0$ on the set where $f - \sum_1^{n-1} g_j \leq 2^{n-1}/3^n$ and $g_n = 2^{n-1}/3^n$ on the set where $f - \sum_1^{n-1} g_j \geq (2/3)^n$. Let $F = \sum_1^\infty g_n$. Since $\|g_n\|_u \leq 2^{n-1}/3^n$, the partial sums of this series converge uniformly, so F is continuous by Proposition 4.13. Moreover, on A we have $0 \leq f - F \leq (2/3)^n$ for all n, whence $F = f$ on A. ∎

4.17 Corollary. *If X is normal, $A \subset X$ is closed, and $f \in C(A)$, there exists $F \in C(X)$ such that $F|A = f$.*

Proof. By considering real and imaginary parts separately, it suffices to assume that f is real-valued. Let $g = f/(1 + |f|)$. Then $g \in C(A, (-1, 1))$, so there exists $G \in C(X, [-1, 1])$ with $G|A = g$. Let $B = G^{-1}(\{-1, 1\})$. By Urysohn's lemma there exists $h \in C(X, [0, 1])$ with $h = 1$ on A, $h = 0$ on B. Then $hG = G$ on A and $|hG| < 1$ everywhere, so $F = hG/(1 - |hG|)$ does the job. ∎

A topological space X is called **completely regular** if X is T_1 and for each closed $A \subset X$ and each $x \notin A$ there exists $f \in C(X, [0, 1])$ such that $f(x) = 1$ and $f = 0$ on A. Completely regular spaces are also called **Tychonoff** spaces or $T_{3\frac{1}{2}}$ spaces. The latter terminology is justified, for every completely regular space is T_3 (if A, x, f are as above, then $f^{-1}((\frac{1}{2}, \infty))$ and $f^{-1}((-\infty, \frac{1}{2}))$ are disjoint neighborhoods of x and A), and Urysohn's lemma shows that every T_4 space is completely regular.

Exercises

14. If X and Y are topological spaces, $f : X \to Y$ is continuous iff $f(\overline{A}) \subset \overline{f(A)}$ for all $A \subset X$ iff $\overline{f^{-1}(B)} \subset f^{-1}(\overline{B})$ for all $B \subset Y$.

15. If X is a topological space, $A \subset X$ is closed, and $g \in C(A)$ satisfies $g = 0$ on ∂A, then the extension of g to X defined by $g(x) = 0$ for $x \in A^c$ is continuous.

16. Let X be a topological space, Y a Hausdorff space, and f, g continuous maps from X to Y.

a. $\{x : f(x) = g(x)\}$ is closed.

b. If $f = g$ on a dense subset of X, then $f = g$ on all of X.

17. If X is a set, $\mathcal{F}$ a collection of real-valued functions on X, and $\mathcal{T}$ the weak topology generated by $\mathcal{F}$, then $\mathcal{T}$ is Hausdorff iff for every $x, y \in X$ with $x \neq y$ there exists $f \in \mathcal{F}$ with $f(x) \neq f(y)$.

18. If X and Y are topological spaces and $y_0 \in Y$, then X is homeomorphic to $X \times \{y_0\}$ where the latter has the relative topology as a subset of $X \times Y$.

19. If $\{X_\alpha\}$ is a family of topological spaces, $X = \prod_\alpha X_\alpha$ (with the product topology) is uniquely determined up to homeomorphism by the following property: There exist continuous maps $\pi_\alpha : X \to X_\alpha$ such that if Y is any topological space and $f_\alpha \in C(Y, X_\alpha)$ for each α, there is a unique $F \in C(Y, X)$ such that $f_\alpha = \pi_\alpha \circ F$. (Thus X is the category-theoretic product of the X_α's in the category of topological spaces.)

20. If A is a countable set and X_α is a first (resp. second) countable space for each $\alpha \in A$, then $\prod_{\alpha \in A} X_\alpha$ is first (resp. second) countable.

21. If X is an infinite set with the cofinite topology, then every $f \in C(X)$ is constant.

22. Let X be a topological space, (Y, ρ) a complete metric space, and $\{f_n\}$ a sequence in Y^X such that $\sup_{x \in X} \rho(f_n(x), f_m(x)) \to 0$ as $m, n \to \infty$. There is

a unique $f \in Y^X$ such that $\sup_{x \in X} \rho(f_n(x), f(x)) \to 0$ as $n \to \infty$. If each f_n is continuous, so is f.

23. Give an elementary proof of the Tietze extension theorem for the case $X = \mathbb{R}$.

24. A Hausdorff space X is normal iff X satisfies the conclusion of Urysohn's lemma iff X satisfies the conclusion of the Tietze extension theorem.

25. If $(X, \mathcal{T})$ is completely regular, then $\mathcal{T}$ is the weak topology generated by $C(X)$.

26. Let X and Y be topological spaces.

a. If X is connected (see Exercise 10) and $f \in C(X, Y)$, then $f(X)$ is connected.

b. X is called **arcwise connected** if for all $x_0, x_1 \in X$ there exists $f \in C([0,1], X)$ with $f(0) = x_0$ and $f(1) = x_1$. Every arcwise connected space is connected.

c. Let $X = \{(s,t) \in \mathbb{R}^2 : t = \sin(s^{-1})\} \cup \{(0,0)\}$, with the relative topology induced from $\mathbb{R}^2$. Then X is connected but not arcwise connected.

27. If X_α is connected for each $\alpha \in A$ (see Exercise 10), then $X = \prod_{\alpha \in A} X_\alpha$ is connected. (Fix $x \in X$ and let Y be the connected component of x in X. Show that Y includes $\{y \in X : \pi_\alpha(y) = \pi_\alpha(x)$ for all but finitely many $\alpha\}$ and that the latter set is dense in X. Use Exercises 10 and 18.)

28. Let X be a topological space equipped with an equivalence relation, $\widetilde{X}$ the set of equivalence classes, $\pi : X \to \widetilde{X}$ the map taking each $x \in X$ to its equivalence class, and $\mathcal{T} = \{U \subset \widetilde{X} : \pi^{-1}(U)$ is open in $X\}$.

a. $\mathcal{T}$ is a topology on $\widetilde{X}$. (It is called the **quotient topology**.)

b. If Y is a topological space, $f : \widetilde{X} \to Y$ is continuous iff $f \circ \pi$ is continuous.

c. $\widetilde{X}$ is T_1 iff every equivalence class is closed.

29. If X is a topological space and G is a group of homeomorphisms from X to itself, G induces an equivalence relation on X, namely, $x \sim y$ iff $y = g(x)$ for some $g \in G$. Let $X = \mathbb{R}^2$; describe the quotient space $\widetilde{X}$ and the quotient topology on it (as in Exercise 28) for each of the following groups of invertible linear maps. In particular, show that in (a) the quotient space is homeomorphic to $[0, \infty)$; in (b) it is T_1 but not Hausdorff; in (c) it is T_0 but not T_1, and in (d) it is not T_0. (In fact, in (d) $\widetilde{X}$ is uncountable, but there are only six open sets and there are points $p \in \widetilde{X}$ such that $\overline{\{p\}} = \widetilde{X}$.)

a. $\left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbb{R} \right\}$

b. $\left\{ \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} : a \in \mathbb{R} \right\}$

c. $\left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} : a > 0,\ b \in \mathbb{R} \right\}$

d. $\left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} : a, b \in \mathbb{Q} \setminus \{0\} \right\}$

4.3 NETS

As we have hinted above, sequential convergence does not play the same central role in general topological spaces as it does in metric spaces. The reasons for this may be illustrated by the following example. Consider the space $\mathbb{C}^{\mathbb{R}}$ of all complex-valued functions on $\mathbb{R}$, with the product topology (i.e., the topology of pointwise convergence), and its subspace $C(\mathbb{R})$. On the one hand, by Corollary 2.9, if $\{f_n\} \subset C(\mathbb{R})$ and $f_n \to f$ pointwise, then f is Borel measurable, so the set of limits of convergent sequences in $C(\mathbb{R})$ is a proper subset of $\mathbb{C}^{\mathbb{R}}$. Nonetheless, $C(\mathbb{R})$ is dense in $\mathbb{C}^{\mathbb{R}}$. Indeed, if $f \in \mathbb{C}^{\mathbb{R}}$, the sets

$$\{g \in \mathbb{C}^{\mathbb{R}} : |g(x_j) - f(x_j)| < \epsilon \text{ for } j = 1, \ldots, n\}$$
$$(n \in \mathbb{N},\ x_1, \ldots, x_n \in \mathbb{R},\ \epsilon > 0)$$

form a neighborhood base at f, and each of these sets clearly contains continuous functions.

There is, however, a generalization of the notion of sequence that works well in arbitrary topological spaces; the key idea is to use index sets more general than $\mathbb{N}$. The precise definitions are as follows.

A **directed set** is a set A equipped with a binary relation $\lesssim$ such that

- $\alpha \lesssim \alpha$ for all $\alpha \in A$;
- if $\alpha \lesssim \beta$ and $\beta \lesssim \gamma$ then $\alpha \lesssim \gamma$;
- for any $\alpha, \beta \in A$ there exists $\gamma \in A$ such that $\alpha \lesssim \gamma$ and $\beta \lesssim \gamma$.

If $\alpha \lesssim \beta$, we shall also write $\beta \gtrsim \alpha$. A **net** in a set X is a mapping $\alpha \mapsto x_\alpha$ from a directed set A into X. We shall usually denote such a mapping by $\langle x_\alpha \rangle_{\alpha \in A}$, or just by $\langle x_\alpha \rangle$ if A is understood, and we say that $\langle x_\alpha \rangle$ is indexed by A.

Here are some examples of directed sets:

i. The set of positive integers $\mathbb{N}$, with $j \lesssim k$ iff $j \leq k$.

ii. The set $\mathbb{R} \setminus \{a\}$ $(a \in \mathbb{R})$, with $x \lesssim y$ iff $|x - a| \geq |y - a|$.

iii. The set of all partitions $\{x_j\}_0^n$ of the interval $[a, b]$ (i.e., $a = x_0 < \cdots < x_n = b$), with $\{x_j\}_0^n \lesssim \{y_k\}_0^m$ iff $\max(x_j - x_{j-1}) \geq \max(y_k - y_{k-1})$.

iv. The set $\mathcal{N}$ of all neighborhoods of a point x in a topological space X, with $U \lesssim V$ iff $U \supset V$. (We say that $\mathcal{N}$ is directed by **reverse inclusion**.)

v. The Cartesian product $A \times B$ of two directed sets, with $(\alpha, \beta) \lesssim (\alpha', \beta')$ iff $\alpha \lesssim \alpha'$ and $\beta \lesssim \beta'$. (This is *always* the way we make $A \times B$ into a directed set.)

Examples (i)–(iii) occur in elementary analysis: A net indexed by $\mathbb{N}$ is just a sequence, and the nets indexed by the sets in (ii) and (iii) occur in defining limits of

real variables and Riemann integrals. Example (iv) is of fundamental importance in topology, and we shall see several uses of the construction in (v).

Let X be a topological space and E a subset of X. A net $\langle x_\alpha \rangle_{\alpha \in A}$ is **eventually** in E if there exists $\alpha_0 \in A$ such that $x_\alpha \in E$ for $\alpha \gtrsim \alpha_0$, and $\langle x_\alpha \rangle$ is **frequently** in E if for every $\alpha \in A$ there exists $\beta \gtrsim \alpha$ such that $x_\beta \in E$. A point $x \in X$ is a **limit** of $\langle x_\alpha \rangle$ (or $\langle x_\alpha \rangle$ **converges** to x, or $x_\alpha \to x$) if for every neighborhood U of x, $\langle x_\alpha \rangle$ is eventually in U, and x is a **cluster point** of $\langle x_\alpha \rangle$ if for every neighborhod U of x, $\langle x_\alpha \rangle$ is frequently in U.

The next three propositions show that nets are a good substitute for sequences.

4.18 Proposition. *If X is a topological space, $E \subset X$, and $x \in X$, then x is an accumulation point of E iff there is a net in $E \setminus \{x\}$ that converges to x, and $x \in \overline{E}$ iff there is a net in E that converges to x.*

Proof. If x is an accumulation point of E, let $\mathcal{N}$ be the set of neighborhoods of x, directed by reverse inclusion. For each $U \in \mathcal{N}$, pick $x_U \in (U \setminus \{x\}) \cap E$. Then $x_U \to x$. Conversely, if $x_\alpha \in E \setminus \{x\}$ and $x_\alpha \to x$, then every punctured neighborhood of x contains some x_α, so x is an accumulation point of E. Likewise, if $x_\alpha \to x$ where $x_\alpha \in E$, then $x \in \overline{E}$, and the converse follows from Proposition 4.1. ∎

4.19 Proposition. *If X and Y are topological spaces and $f : X \to Y$, then f is continuous at $x \in X$ iff for every net $\langle x_\alpha \rangle$ converging to x, $\langle f(x_\alpha) \rangle$ converges to $f(x)$.*

Proof. If f is continuous at x and V is a neighborhood of $f(x)$, then $f^{-1}(V)$ is a neighborhood of x. Hence, if $x_\alpha \to x$ then $\langle x_\alpha \rangle$ is eventually in $f^{-1}(V)$, so $\langle f(x_\alpha) \rangle$ is eventually in V, and thus $f(x_\alpha) \to f(x)$. On the other hand, if f is not continuous at x, there is a neighborhood V of $f(x)$ such that $f^{-1}(V)$ is not a neighborhood of x, that is, $x \notin (f^{-1}(V))^o$, or equivalently, $x \in \overline{f^{-1}(V^c)}$. By Proposition 4.18, there is a net $\langle x_\alpha \rangle$ in $f^{-1}(V^c)$ that converges to x. But then $f(x_\alpha) \notin V$, so $f(x_\alpha) \not\to f(x)$. ∎

A **subnet** of a net $\langle x_\alpha \rangle_{\alpha \in A}$ is a net $\langle y_\beta \rangle_{\beta \in B}$ together with a map $\beta \mapsto \alpha_\beta$ from B to A such that:

- for every $\alpha_0 \in A$ there exists $\beta_0 \in B$ such that $\alpha_\beta \gtrsim \alpha_0$ whenever $\beta \gtrsim \beta_0$;
- $y_\beta = x_{\alpha_\beta}$.

Clearly if $\langle x_\alpha \rangle$ converges to a point x, then so does any subnet $\langle x_{\alpha_\beta} \rangle$.

Warning: The name "subnet" is used because subnets perform much the same functions as subsequences, but it should not be taken too literally, as the mapping $\beta \mapsto \alpha_\beta$ need not be injective. In particular, the index set B may well have larger cardinality than the index set A, and a subnet of a sequence need not be a subsequence.

4.20 Proposition. *If $\langle x_\alpha \rangle_{\alpha \in A}$ is a net in a topological space X, then $x \in X$ is a cluster point of $\langle x_\alpha \rangle$ iff $\langle x_\alpha \rangle$ has a subnet that converges to x.*

Proof. If $\langle y_\beta \rangle = \langle x_{\alpha_\beta} \rangle$ is a subnet converging to x and U is a neighborhood of x, choose $\beta_1 \in B$ such that $y_\beta \in U$ for $\beta \gtrsim \beta_1$. Also, given $\alpha \in A$, choose $\beta_2 \in B$ such that $\alpha_\beta \gtrsim \alpha$ for $\beta \gtrsim \beta_2$. Then there exists $\beta \in B$ with $\beta \gtrsim \beta_1$ and $\beta \gtrsim \beta_2$, and we have $\alpha_\beta \gtrsim \alpha$ and $x_{\alpha_\beta} = y_\beta \in U$. Thus $\langle x_\alpha \rangle$ is frequently in U, so x is a cluster point of $\langle x_\alpha \rangle$. Conversely, if x is a cluster point of $\langle x_\alpha \rangle$, let $\mathcal{N}$ be the set of neighborhoods of x and make $\mathcal{N} \times A$ into a directed set by declaring that $(U, \alpha) \lesssim (U', \alpha')$ iff $U \supset U'$ and $\alpha \lesssim \alpha'$. For each $(U, \gamma) \in \mathcal{N} \times A$ we can choose $\alpha_{(U,\gamma)} \in A$ such that $\alpha_{(U,\gamma)} \gtrsim \gamma$ and $x_{\alpha_{(U,\gamma)}} \in U$. Then if $(U', \gamma') \gtrsim (U, \gamma)$ we have $\alpha_{(U',\gamma')} \gtrsim \gamma' \gtrsim \gamma$ and $x_{\alpha_{(U',\gamma')}} \in U' \subset U$, whence it follows that $\langle x_{\alpha_{(U,\gamma)}} \rangle$ is a subnet of $\langle x_\alpha \rangle$ that converges to x. ∎

Exercises

30. If A is a directed set, a subset B of A is called **cofinal** in A if for each $\alpha \in A$ there exists $\beta \in B$ such that $\beta \gtrsim \alpha$.

a. If B is cofinal in A and $\langle x_\alpha \rangle_{\alpha \in A}$ is a net, the inclusion map $B \to A$ makes $\langle x_\beta \rangle_{\beta \in B}$ a subnet of $\langle x_\alpha \rangle_{\alpha \in A}$.

b. If $\langle x_\alpha \rangle_{\alpha \in A}$ is a net in a topological space, then $\langle x_\alpha \rangle$ converges to x iff for every cofinal $B \subset A$ there is a cofinal $C \subset B$ such that $\langle x_\gamma \rangle_{\gamma \in C}$ converges to x.

31. Let $\langle x_n \rangle_{n \in \mathbb{N}}$ be a sequence.

a. If $k \to n_k$ is a map from $\mathbb{N}$ to itself, then $\langle x_{n_k} \rangle_{k \in \mathbb{N}}$ is a subnet of $\langle x_n \rangle$ iff $n_k \to \infty$ as $k \to \infty$, and it is a subsequence (as defined in §0.1) iff n_k is strictly increasing in k.

b. There is a natural one-to-one correspondence between the subsequences of $\langle x_n \rangle$ and the subnets of $\langle x_n \rangle$ defined by cofinal sets as in Exercise 30.

32. A topological space X is Hausdorff iff every net in X converges to at most one point. (If X is not Hausdorff, let x and y be distinct points with no disjoint neighborhoods, and consider the directed set $\mathcal{N}_x \times \mathcal{N}_y$ where $\mathcal{N}_x, \mathcal{N}_y$ are the families of neighborhoods of x, y.)

33. Let $\langle x_\alpha \rangle_{\alpha \in A}$ be a net in a topological space, and for each $\alpha \in A$ let $E_\alpha = \{x_\beta : \beta \gtrsim \alpha\}$. Then x is a cluster point of $\langle x_\alpha \rangle$ iff $x \in \bigcap_{\alpha \in A} \overline{E}_\alpha$.

34. If X has the weak topology generated by a family $\mathcal{F}$ of functions, then $\langle x_\alpha \rangle$ converges to $x \in X$ iff $\langle f(x_\alpha) \rangle$ converges to $f(x)$ for all $f \in \mathcal{F}$. (In particular, if $X = \prod_{i \in I} X_i$, then $x_\alpha \to x$ iff $\pi_i(x_\alpha) \to \pi_i(x)$ for all $i \in I$.)

35. Let X be a set and $\mathcal{A}$ the collection of all finite subsets of X, directed by inclusion. Let $f : X \to \mathbb{R}$ be an arbitrary function, and for $A \in \mathcal{A}$, let $z_A = \sum_{x \in A} f(x)$. Then the net $\langle z_A \rangle$ converges in $\mathbb{R}$ iff $\{x : f(x) \neq 0\}$ is a countable set $\{x_n\}_{n \in \mathbb{N}}$ and $\sum_1^\infty |f(x_n)| < \infty$, in which case $z_A \to \sum_1^\infty f(x_n)$. (Cf. Proposition 0.20.)

36. Let X be the set of Lebesgue measurable complex-valued functions on $[0, 1]$. There is no topology $\mathcal{T}$ on X such that a sequence $\langle f_n \rangle$ converges to f with respect to $\mathcal{T}$ iff $f_n \to f$ a.e. (Use Corollary 2.32 and Exercises 30b and 31b.)

4.4 COMPACT SPACES

In §0.6 we gave three equivalent characterizations of compactness for metric spaces: the Heine-Borel property, the Bolzano-Weierstrass property, and completeness plus total boundedness. Only the first two of these make sense for general topological spaces, and it is the first one that turns out to be the most useful. Accordingly, we define a topological space X to be **compact** if whenever $\{U_\alpha\}_{\alpha\in A}$ is an open cover of X — that is, a collection of open sets such that $X = \bigcup_{\alpha\in A} U_\alpha$ — there is a finite subset B of A such that $X = \bigcup_{\alpha\in B} U_\alpha$. To be brief (although somewhat sylleptic, since the adjectives "open" and "finite" refer to different things), we say: X is compact if every open cover of X has a finite subcover.

A subset Y of a topological space X is called **compact** if it is compact in the relative topology; thus $Y \subset X$ is compact iff whenever $\{U_\alpha\}_{\alpha\in A}$ is a collection of open subsets of X with $Y \subset \bigcup_{\alpha\in A} U_\alpha$, there is a finite $B \subset A$ with $Y \subset \bigcup_{\alpha\in B} U_\alpha$. Furthermore, Y is called **precompact** if its closure is compact.

DeMorgan's laws lead to the following characterization of compactness in terms of closed sets. A family $\{F_\alpha\}_{\alpha\in A}$ of subsets of X is said to have the **finite intersection property** if $\bigcap_{\alpha\in B} F_\alpha \neq \varnothing$ for all finite $B \subset A$.

4.21 Proposition. *A topological space X is compact iff for every family $\{F_\alpha\}_{\alpha\in A}$ of closed sets with the finite intersection property, $\bigcap_{\alpha\in A} F_\alpha \neq \varnothing$.*

Proof. Let $U_\alpha = (F_\alpha)^c$. Then U_α is open, $\bigcap_{\alpha\in A} F_\alpha \neq \varnothing$ iff $\bigcup_{\alpha\in A} U_\alpha \neq X$, and $\{F_\alpha\}$ has the finite intersection property iff no finite subfamily of $\{U_\alpha\}$ covers X. The result follows. ∎

We now list several basic facts about compact spaces.

4.22 Proposition. *A closed subset of a compact space is compact.*

Proof. If X is compact, $F \subset X$ is closed, and $\{U_\alpha\}_{\alpha\in A}$ is a family of open sets in X with $F \subset \bigcup_{\alpha\in A} U_\alpha$, then $\{U_\alpha\}_{\alpha\in A} \cup \{F^c\}$ is an open cover of X. It has a finite subcover, so by discarding F^c from the latter if necessary, we obtain a finite subcollection of $\{U_\alpha\}_{\alpha\in A}$ that covers F. ∎

4.23 Proposition. *If F is a compact subset of a Hausdorff space X and $x \notin F$, there are disjoint open sets U, V such that $x \in U$ and $F \subset V$.*

Proof. For each $y \in F$, choose disjoint open U_y and V_y with $x \in U_y$ and $y \in V_y$. $\{V_y\}_{y\in F}$ is an open cover of F, so it has a finite subcover $\{V_{y_j}\}_1^n$. Then $U = \bigcap_1^n U_{y_j}$ and $V = \bigcup_1^n V_{y_j}$ have the desired properties. ∎

4.24 Proposition. *Every compact subset of a Hausdorff space is closed.*

Proof. According to Proposition 4.23, if F is compact then F^c is a neighborhood of each of its points, hence is open. ∎

We remark that in a non-Hausdorff space, compact sets need not be closed (for example, every subset of a space with the trivial topology is compact), and the intersection of compact sets need not be compact; see Exercise 37. Of course, in a Hausdorff space the intersection of any family of compact sets is compact by Propositions 4.22 and 4.24. Moreover, in an arbitrary topological space a finite union of compact sets is always compact. (If $K_1, \dots K_n$ are compact and $\{U_\alpha\}$ is an open cover of $\bigcup_1^n K_j$, choose a finite subcover of each K_j and combine them.)

4.25 Proposition. *Every compact Hausdorff space is normal.*

Proof. Suppose that X is compact Hausdorff and E, F are disjoint closed subsets of X. By Proposition 4.23, for each $x \in E$ there exist disjoint open sets U_x, V_x with $x \in U_x$, $F \subset V_x$. By Proposition 4.22, E is compact, and $\{U_x\}_{x \in E}$ is an open cover of E, so there is a finite subcover $\{U_{x_j}\}_1^n$. Let $U = \bigcup_1^n U_{x_j}$ and $V = \bigcap_1^n V_{x_j}$. Then U and V are disjoint open sets with $E \subset U$ and $F \subset V$. ∎

4.26 Proposition. *If X is compact and $f : X \to Y$ is continuous, then $f(X)$ is compact.*

Proof. Let $\{V_\alpha\}$ be an open cover of $f(X)$ in Y. Then $\{f^{-1}(V_\alpha)\}$ is an open cover of X, so it has a finite subcover $\{f^{-1}(V_{\alpha_j})\}$, and $\{V_{\alpha_j}\}$ is then a finite subcover of $f(X)$. ∎

4.27 Corollary. *If X is compact, then $C(X) = BC(X)$.*

4.28 Proposition. *If X is compact and Y is Hausdorff, then any continuous bijection $f : X \to Y$ is a homeomorphism.*

Proof. If $E \subset X$ is closed, then E is compact, hence $f(E)$ is compact, hence $f(E)$ is closed, by Propositions 4.22, 4.26, and 4.24. This means that f^{-1} is continuous, so f is a homeomorphism. ∎

We now show that a version of the Bolzano-Weierstrass property holds for compact topological spaces. As one might suspect, it is merely necessary to replace sequences by nets.

4.29 Theorem. *If X is a topological space, the following are equivalent:*

a. X is compact.

b. Every net in X has a cluster point.

c. Every net in X has a convergent subnet.

Proof. The equivalence of (b) and (c) follows from Proposition 4.20. If X is compact and $\langle x_\alpha \rangle$ is a net in X, let $E_\alpha = \{x_\beta : \beta \gtrsim \alpha\}$. Since for any $\alpha, \beta \in A$ there exists $\gamma \in A$ with $\gamma \gtrsim \alpha$ and $\gamma \gtrsim \beta$, the family $\{E_\alpha\}_{\alpha \in A}$ has the finite intersection property, so by Proposition 4.21, $\bigcap_{\alpha \in A} \overline{E}_\alpha \neq \varnothing$. If $x \in \bigcap_{\alpha \in A} \overline{E}_\alpha$ and U is a neighborhood of x, then U intersects each E_α, which means that $\langle x_\alpha \rangle$

is frequently in U, so x is a cluster point of $\langle x_\alpha \rangle$. On the other hand, if X is not compact, let $\{U_\beta\}_{\beta \in B}$ be an open cover of X with no finite subcover. Let $\mathcal{A}$ be the collection of finite subsets of B, directed by inclusion, and for each $A \in \mathcal{A}$ let x_A be a point in $(\bigcup_{\beta \in A} U_\beta)^c$. Then $\langle x_A \rangle_{A \in \mathcal{A}}$ is a net with no cluster point. Indeed, if $x \in X$, choose $\beta \in B$ with $x \in U_\beta$. If $A \in \mathcal{A}$ and $A \gtrsim \{\beta\}$ then $x_A \notin U_\beta$, so x is not a cluster point of $\langle x_A \rangle$. ∎

We conclude by mentioning two other useful concepts related to compactness. A topological space X is called **countably compact** if every countable open cover of X has a finite subcover, and **sequentially compact** if every sequence in X has a convergent subsequence. Of course, every compact space is countably compact, and for metric spaces compactness and sequential compactness are equivalent. However, in general there is no relation between compactness and sequential compactness. See Exercises 39–43 for further results and examples.

Exercises

37. Let $0'$ denote a point that is is not an element of $(-1,1)$, and let $X = (-1,1) \cup \{0'\}$. Let $\mathcal{T}$ be the topology on X generated by the sets $(-1,a)$, $(a,1)$, $[(-1,b) \setminus \{0\}] \cup \{0'\}$, and $[(c,1) \setminus \{0\}] \cup \{0'\}$ where $-1 < a < 1, 0 < b < 1$, and $-1 < c < 0$. (One should picture X as $(-1,1)$ with the point 0 split in two.)

a. Define $f, g : (-1,1) \to X$ by $f(x) = x$ for all x, $g(x) = x$ for $x \neq 0$, and $g(0) = 0'$. Then f and g are homeomorphisms onto their ranges.

b. X is T_1 but not Hausdorff, although each point of X has a neighborhood that is homeomorphic to $(-1,1)$ (and hence is Hausdorff).

c. The sets $[-\frac{1}{2}, \frac{1}{2}]$ and $([-\frac{1}{2}, \frac{1}{2}] \setminus \{0\}) \cup \{0'\}$ are compact but not closed in X, and their intersection is not compact.

38. Suppose that $(X, \mathcal{T})$ is a compact Hausdorff space and $\mathcal{T}'$ is another topology on X. If $\mathcal{T}'$ is strictly stronger than $\mathcal{T}$, then $(X, \mathcal{T}')$ is Hausdorff but not compact. If $\mathcal{T}'$ is strictly weaker than $\mathcal{T}$, then $(X, \mathcal{T}')$ is compact but not Hausdorff.

39. Every sequentially compact space is countably compact.

40. If X is countably compact, then every sequence in X has a cluster point. If X is also first countable, then X is sequentially compact.

41. A T_1 space X is countably compact iff every infinite subset of X has an accumulation point.

42. The set of countable ordinals (§0.4) with the order topology (Exercise 9) is sequentially compact and first countable but not compact. (To prove sequential compactness, use Proposition 0.19.)

43. For $x \in [0,1)$, let $\sum_1^\infty a_n(x) 2^{-n}$ ($a_n(x) = 0$ or 1) be the base-2 decimal expansion of x. (If x is a dyadic rational, choose the expansion such that $a_n(x) = 0$ for n large.) Then the sequence $\langle a_n \rangle$ in $\{0,1\}^{[0,1)}$ has no pointwise convergent subsequence. (Hence $\{0,1\}^{[0,1)}$, with the product topology arising from the discrete

topology on $\{0, 1\}$, is not sequentially compact. It is, however, compact, as we shall show in §4.6.)

44. If X is countably compact and $f : X \to Y$ is continuous, then $f(X)$ is countably compact.

45. If X is normal, then X is countably compact iff $C(X) = BC(X)$. (Use Exercises 40 and 44. If $\langle x_n \rangle$ is a sequence in X with no cluster point, then $\{x_n : n \in \mathbb{N}\}$ is closed, and Corollary 4.17 applies.)

4.5 LOCALLY COMPACT HAUSDORFF SPACES

A topological space is called **locally compact** if every point has a compact neighborhood. We shall be mainly concerned with locally compact Hausdorff spaces, which we call **LCH** spaces for short.

4.30 Proposition. *If X is an LCH space, $U \subset X$ is open, and $x \in U$, there is a compact neighborhood N of x such that $N \subset U$.*

Proof. We may assume $\overline{U}$ is compact; otherwise, replace U by $U \cap F^o$ where F is a compact neighborhood of x. By Proposition 4.23, there are disjoint relatively open sets V, W in $\overline{U}$ with $x \in V$ and $\partial U \subset W$. Then V is open in X since $V \subset U$, and $\overline{V}$ is a closed and hence compact subset of $U \setminus W$. Thus we may take $N = \overline{V}$. ∎

4.31 Proposition. *If X is an LCH space and $K \subset U \subset X$ where K is compact and U is open, there exists a precompact open V such that $K \subset V \subset \overline{V} \subset U$.*

Proof. By Proposition 4.30, for each $x \in K$ we can choose a compact neighborhood N_x of x with $N_x \subset U$. Then $\{N_x^o\}_{x \in K}$ is an open cover of K, so there is a finite subcover $\{N_{x_j}^o\}_1^n$. Let $V = \bigcup_1^n N_{x_j}^o$; then $K \subset V$ and $\overline{V} = \bigcup_1^n N_{x_j}$ is compact and contained in U. ∎

4.32 Urysohn's Lemma, Locally Compact Version. *If X is an LCH space and $K \subset U \subset X$ where K is compact and U is open, there exists $f \in C(X, [0, 1])$ such that $f = 1$ on K and $f = 0$ outside a compact subset of U.*

Proof. Let V be as in Proposition 4.31. Then $\overline{V}$ is normal by Proposition 4.25, so by Urysohn's lemma 4.15 there exists $f \in C(\overline{V}, [0, 1])$ such that $f = 1$ on K and $f = 0$ on ∂V. We extend f to X by setting $f = 0$ on $\overline{V}^c$. Suppose that $E \subset [0, 1]$ is closed. If $0 \notin E$ we have $f^{-1}(E) = (f|\overline{V})^{-1}(E)$, and if $0 \in E$ we have $f^{-1}(E) = (f|\overline{V})^{-1}(E) \cup \overline{V}^c = (f|\overline{V})^{-1}(E) \cup V^c$ since $(f|\overline{V})^{-1}(E) \supset \partial V$. In either case, $f^{-1}(E)$ is closed, so f is continuous. ∎

4.33 Corollary. *Every LCH space is completely regular.*

4.34 Tietze Extension Theorem, Locally Compact Version. *Suppose that X is an LCH space and $K \subset X$ is compact. If $f \in C(K)$, there exists $F \in C(X)$ such that $F|K = f$. Moreover, F may be taken to vanish outside a compact set.*

The proof is similar to that of Theorem 4.32; details are left to the reader (Exercise 46).

The preceding results show that LCH spaces have a rich supply of continuous functions that vanish outside compact sets. Let us introduce some terminology: If X is a topological space and $f \in C(X)$, the **support** of f, denoted by $\operatorname{supp}(f)$, is the smallest closed set outside of which f vanishes, that is, the closure of $\{x : f(x) \neq 0\}$. If $\operatorname{supp}(f)$ is compact, we say that f is **compactly supported**, and we define

$$C_c(X) = \{f \in C(X) : \operatorname{supp}(f) \text{ is compact}\}.$$

Moreover, if $f \in C(X)$, we say that f **vanishes at infinity** if for every $\epsilon > 0$ the set $\{x : |f(x)| \geq \epsilon\}$ is compact, and we define

$$C_0(X) = \{f \in C(X) : f \text{ vanishes at infinity}\}.$$

Clearly $C_c(X) \subset C_0(X)$. Moreover, $C_0(X) \subset BC(X)$, because for $f \in C_0(X)$ the image of the set $\{x : |f(x)| \geq \epsilon\}$ is compact, and $|f| < \epsilon$ on its complement.

4.35 Proposition. *If X is an LCH space, $C_0(X)$ is the closure of $C_c(X)$ in the uniform metric.*

Proof. If $\{f_n\}$ is a sequence in $C_c(X)$ that converges uniformly to $f \in C(X)$, for each $\epsilon > 0$ there exists $n \in \mathbb{N}$ such that $\|f_n - f\|_u < \epsilon$. Then $|f(x)| < \epsilon$ if $x \notin \operatorname{supp}(f_n)$, so $f \in C_0(X)$. Conversely, if $f \in C_0(X)$, for $n \in \mathbb{N}$ let $K_n = \{x : |f(x)| \geq n^{-1}\}$. Then K_n is compact, so by Theorem 4.32 there exists $g_n \in C_c(X)$ with $0 \leq g_n \leq 1$ and $g_n = 1$ on K_n. Let $f_n = g_n f$. Then $f_n \in C_c(X)$ and $\|f_n - f\|_u \leq n^{-1}$, so $f_n \to f$ uniformly. ∎

If X is a noncompact LCH space, it is possible to make X into a compact space by adding a single point "at infinity" in such a way that the functions in $C_0(X)$ are precisely those continuous functions f such that $f(x) \to 0$ as x approaches the point at infinity. More precisely, let ∞ denote a point that is not an element of X, let $X^* = X \cup \{\infty\}$, and let $\mathcal{T}$ be the collection of all subsets of X^* such that either (i) U is an open subset of X, or (ii) $\infty \in U$ and U^c is a compact subset of X.

4.36 Proposition. *If X, X^*, and $\mathcal{T}$ are as above, then $(X^*, \mathcal{T})$ is a compact Hausdorff space, and the inclusion map $i : X \to X^*$ is an embedding. Moreover, if $f \in C(X)$, then f extends continuously to X^* iff $f = g + c$ where $g \in C_0(X)$ and c is a constant, in which case the continuous extension is given by $f(\infty) = c$.*

The proof is straightforward and is left to the reader (Exercise 47). The space X^* is called the **one-point compactification** or **Alexandroff compactification** of X.

If X is a topological space, the space $\mathbb{C}^X$ of all complex-valued functions on X can be topologized in various ways. One way, of course, is the product topology,

that is, the topology of pointwise convergence. Another is the **topology of uniform convergence**, which is generated by the sets

$$\left\{g \in \mathbb{C}^X : \sup_{x \in X} |g(x) - f(x)| < n^{-1}\right\} \qquad (n \in \mathbb{N},\ f \in \mathbb{C}^X).$$

The proof of Proposition 4.13 shows that $C(X)$ is a closed subspace of $\mathbb{C}^X$ in the topology of uniform convergence. Intermediate between these two topologies is the **topology of uniform convergence on compact sets**, which is generated by the sets

$$\left\{g \in \mathbb{C}^X : \sup_{x \in K} |g(x) - f(x)| < n^{-1}\right\} \qquad (n \in \mathbb{N},\ f \in \mathbb{C}^X,\ K \subset X \text{ compact}).$$

We shall now examine this topology in the case where X is an LCH space.

4.37 Lemma. *If X is an LCH space and $E \subset X$, then E is closed iff $E \cap K$ is closed for every compact $K \subset X$.*

Proof. If E is closed, then $E \cap K$ is closed by Propositions 4.22 and 4.24. If E is not closed, pick $x \in \overline{E} \setminus E$ and let K be a compact neighborhood of x. Then x is an accumulation point of $E \cap K$ but is not in $E \cap K$, so by Proposition 4.1 $E \cap K$ is not closed. ∎

4.38 Proposition. *If X is an LCH space, $C(X)$ is a closed subspace of $\mathbb{C}^X$ in the topology of uniform convergence on compact sets.*

Proof. If f is in the closure of $C(X)$, then f is a uniform limit of continuous functions on each compact $K \subset X$, so $f|K$ is continuous. If $E \subset \mathbb{C}$ is closed, $f^{-1}(E) \cap K = (f|K)^{-1}(E)$ is thus closed for each compact K, so by Lemma 4.37 $f^{-1}(E)$ is closed, whence f is continuous. ∎

A topological space X is called **σ-compact** if it is a countable union of compact sets. To appreciate the significance of the next two propositions, see Exercise 54.

4.39 Proposition. *If X is a σ-compact LCH space, there is a sequence $\{U_n\}$ of precompact open sets such that $\overline{U}_n \subset U_{n+1}$ for all n and $X = \bigcup_1^\infty U_n$.*

Proof. Suppose $X = \bigcup_1^\infty K_n$ where each K_n is compact. Every compact subset of X has a precompact open neighborhood by Proposition 4.31. Thus we may take U_1 to be a precompact open neighborhood of K_1, and then, proceeding inductively, take U_n to be a precompact open neighborhood of $\overline{U}_{n-1} \cup K_n$. ∎

4.40 Proposition. *If X is a σ-compact LCH space and $\{U_n\}$ is as in Proposition 4.39, then for each $f \in \mathbb{C}^X$ the sets*

$$\left\{g \in \mathbb{C}^X : \sup_{x \in \overline{U}_n} |g(x) - f(x)| < m^{-1}\right\} \qquad (m, n \in \mathbb{N})$$

form a neighborhood base for f in the topology of uniform convergence on compact sets. Hence this topology is first countable, and $f_j \to f$ uniformly on compact sets iff $f_j \to f$ uniformly on each $\overline{U}_n$.

Proof. These assertions follow easily from the observation that if $K \subset X$ is compact, then $\{U_n\}_1^\infty$ is an open cover of K and hence $K \subset \overline{U}_n$ for some n. Details are left to the reader (Exercise 48). ∎

We close this section with a construction that is useful in a number of situations. If X is a topological space and $E \subset X$, a **partition of unity** on E is a collection $\{h_\alpha\}_{\alpha \in A}$ of functions in $C(X, [0,1])$ such that

- each $x \in X$ has a neighborhood on which only finitely many h_α's are nonzero;
- $\sum_{\alpha \in A} h_\alpha(x) = 1$ for $x \in E$.

A partition of unity $\{h_\alpha\}$ is **subordinate** to an open cover $\mathcal{U}$ of E if for each α there exists $U \in \mathcal{U}$ with $\operatorname{supp}(h_\alpha) \subset U$.

4.41 Proposition. *Let X be an LCH space, K a compact subset of X, and $\{U_j\}_1^n$ an open cover of K. There is a partition of unity on K subordinate to $\{U_j\}_1^n$ consisting of compactly supported functions.*

Proof. By Proposition 4.30, each $x \in K$ has a compact neighborhood N_x such that $N_x \subset U_j$ for some j. Since $\{N_x^o\}$ is an open cover of K, there exist $x_1, \ldots, x_m$ such that $K \subset \bigcup_1^m N_{x_k}$. Let F_j be the union of those N_{x_k}'s that are subsets of U_j. Then F_j is a compact subset of U_j, so by Urysohn's lemma there exist $g_1, \ldots, g_n \in C_c(X, [0,1])$ with $g_j = 1$ on F_j and $\operatorname{supp}(g_j) \subset U_j$. Since the F_j's cover K we have $\sum_1^n g_k \geq 1$ on K, so by Urysohn again there exists $f \in C_c(X, [0,1])$ with $f = 1$ on K and $\operatorname{supp}(f) \subset \{x : \sum_1^n g_k(x) > 0\}$. Let $g_{n+1} = 1 - f$, so that $\sum_1^{n+1} g_k > 0$ everywhere, and for $j = 1, \ldots, n$ let $h_j = g_j / \sum_1^{n+1} g_k$. Then $\operatorname{supp}(h_j) = \operatorname{supp}(g_j) \subset U_j$ and $\sum_1^n h_j = 1$ on K. ∎

A generalization of this result may be found in Exercise 57.

Exercises

46. Prove Theorem 4.34.

47. Prove Proposition 4.36. Also, show that if X is Hausdorff but not locally compact, Proposition 4.36 remains valid except that X^* is not Hausdorff.

48. Complete the proof of Proposition 4.40.

49. Let X be a compact Hausdorff space and $E \subset X$.

a. If E is open, then E is locally compact in the relative topology.

b. If E is dense in X and locally compact in the relative topology, then E is open. (Use Exercise 13.)

c. E is locally compact in the relative topology iff E is relatively open in $\overline{E}$.

50. Let U be an open subset of a compact Hausdorff space X and U^* its one-point compactification (see Exercise 49a). If $\phi : X \to U^*$ is defined by $\phi(x) = x$ if $x \in U$ and $\phi(x) = \infty$ if $x \in U^c$, then ϕ is continuous.

51. If X and Y are topological spaces, $\phi \in C(X, Y)$ is called **proper** if $\phi^{-1}(K)$ is compact in X for every compact $K \subset Y$. Suppose that X and Y are LCH spaces and X^* and Y^* are their one-point compactifications. If $\phi \in C(X, Y)$, then ϕ is proper iff ϕ extends continuously to a map from X^* to Y^* by setting $\phi(\infty_X) = \infty_Y$.

52. The one-point compactification of $\mathbb{R}^n$ is homeomorphic to the n-sphere $\{x \in \mathbb{R}^{n+1} : |x| = 1\}$.

53. Lemma 4.37 remains true if the assumption that X is locally compact is replaced by the assumption that X is first countable.

54. Let $\mathbb{Q}$ have the relative topology induced from $\mathbb{R}$.

a. $\mathbb{Q}$ is not locally compact.

b. $\mathbb{Q}$ is σ-compact (it is a countable union of singleton sets), but uniform convergence on singletons (i.e., pointwise convergence) does not imply uniform convergence on compact subsets of $\mathbb{Q}$.

55. Every open set in a second countable LCH space is σ-compact.

56. Define $\Phi : [0, \infty] \to [0, 1]$ by $\Phi(t) = t/(t + 1)$ for $t \in [0, \infty)$ and $\Phi(\infty) = 1$.

a. Φ is strictly increasing and $\Phi(t + s) \leq \Phi(t) + \Phi(s)$.

b. If (Y, ρ) is a metric space, then $\Phi \circ \rho$ is a bounded metric on Y that defines the same topology as ρ.

c. If X is a topological space, the function $\rho(f, g) = \Phi(\sup_{x \in X} |f(x) - g(x)|)$ is a metric on $\mathbb{C}^X$ whose associated topology is the topology of uniform convergence.

d. If X is a σ-compact LCH space and $\{U_n\}_1^\infty$ is as in Proposition 4.39, the function

$$\rho(f, g) = \sum_1^\infty 2^{-n} \Phi\left(\sup_{x \in \overline{U}_n} |f(x) - g(x)| \right)$$

is a metric on $\mathbb{C}^X$ whose associated topology is the topology of uniform convergence on compact sets.

57. An open cover $\mathcal{U}$ of a topological space X is called **locally finite** if each $x \in X$ has a neighborhood that intersects only finitely many members of $\mathcal{U}$. If $\mathcal{U}, \mathcal{V}$ are open covers of X, $\mathcal{V}$ is a **refinement** of $\mathcal{U}$ if for each $V \in \mathcal{V}$ there exists $U \in \mathcal{U}$ with $V \subset U$. X is called **paracompact** if every open cover of X has a locally finite refinement.

a. If X is a σ-compact LCH space, then X is paracompact. In fact, every open cover $\mathcal{U}$ has locally finite refinements $\{V_\alpha\}$, $\{W_\alpha\}$ such that $\overline{V}_\alpha$ is compact and $\overline{W}_\alpha \subset V_\alpha$ for all α. (Let $\{U_n\}_1^\infty$ be as in Proposition 4.39. For each n, $\{E \cap (U_{n+2} \setminus \overline{U}_{n-1}) : E \subset \mathcal{U}\}$ is an open cover of $\overline{U}_{n+1} \setminus U_n$. Choose a finite

subcover to obtain $\{V_\alpha\}$ and mimic the beginning of the proof of Proposition 4.41 to obtain $\{W_\alpha\}$.)

b. If X is a σ-compact LCH space, for any open cover $\mathcal{U}$ of X there is a partition of unity on X subordinate to $\mathcal{U}$ and consisting of compactly supported functions.

4.6 TWO COMPACTNESS THEOREMS

The geometric objects on which one does analysis (Euclidean spaces, manifolds, etc.) tend to be compact or locally compact. However, in infinite-dimensional spaces such as spaces of functions, compactness is a rather rare phenomenon and is to be greatly prized when it is available. Almost all compactness results in such situations are obtained via two basic theorems, Tychonoff's theorem and the Arzelà-Ascoli theorem, which we present in this section.

Tychonoff's theorem has to do with compactness of Cartesian products. To prepare for it, we introduce some notation. Recall that an element x of $X = \prod_{\alpha\in A} X_\alpha$ is, strictly speaking, a mapping from A into $\bigcup_{\alpha\in A} X_\alpha$; namely, $x(\alpha) \in X_\alpha$ is the αth coordinate of x, which we generally denote by $\pi_\alpha(x)$. If $B \subset A$, there is a natural map $\pi_B : X \to \prod_{\alpha\in B} X_\alpha$; namely, $\pi_B(x)$ is the restriction of the map x to B. (In particular, $\pi_{\{\alpha\}}$ is essentially identical to π_α, and we shall not distinguish between them.) If $p \in \prod_{\alpha\in B} X_\alpha$ and $q \in \prod_{\alpha\in C} X_\alpha$, we shall say that q is an extension of p if q extends p as a mapping, that is, if $B \subset C$ and $p(\alpha) = q(\alpha)$ for $\alpha \in B$.

4.42 Tychonoff's Theorem. *If $\{X_\alpha\}_{\alpha\in A}$ is any family of compact topological spaces, then $X = \prod_{\alpha\in A} X_\alpha$ (with the product topology) is compact.*

Proof. By Theorem 4.29, it is enough to show that any net $\langle x_i\rangle_{i\in I}$ in X has a cluster point. We shall do this by examining cluster points of the nets $\langle \pi_B(x_i)\rangle$ in the subproducts of X. To wit, let

$$\mathcal{P} = \bigcup_{B\subset A} \Big\{ p \in \prod_{\alpha\in B} X_\alpha : p \text{ is a cluster point of } \langle \pi_B(x_i)\rangle \Big\}.$$

$\mathcal{P}$ is nonempty, because each X_α is compact and so $\langle \pi_B(x_i)\rangle$ has cluster points when $B = \{\alpha\}$. Moreover, $\mathcal{P}$ is partially ordered by extension; that is, $p \le q$ if q is an extension of p as defined above.

Suppose that $\{p_l : l \in L\}$ is a linearly ordered subset of $\mathcal{P}$, where $p_l \in \prod_{\alpha\in B_l} X_\alpha$. Let $B^* = \bigcup_{l\in L} B_l$, and let p^* be the unique element of $\prod_{\alpha\in B^*} X_\alpha$ that extends every p_l. We claim that $p^* \in \mathcal{P}$. Indeed, from the definition of the product topology, any neighborhood of p^* contains a set of the form $\prod_{\alpha\in B^*} U_\alpha$ where each U_α is open in X_α and $U_\alpha = X_\alpha$ for all but finitely many α, say $\alpha_1, \dots, \alpha_n$. Each of these α_j's belongs to some B_l, so by linearity of the ordering they all belong to a single B_l. But then $\prod_{\alpha\in B_l} U_\alpha$ is a neighborhood of p_l, so $\langle \pi_{B_l}(x_i)\rangle$ is frequently in $\prod_{\alpha\in B_l} U_\alpha$; hence $\langle \pi_{B^*}(x_i)\rangle$ is frequently in $\prod_{\alpha\in B^*} U_\alpha$, so p^* is a cluster point of $\langle \pi_{B^*}(x_i)\rangle$. Therefore p^* is an upper bound for $\{p_l\}$ in $\mathcal{P}$.

By Zorn's lemma, then, $\mathcal{P}$ has a maximal element $\overline{p} \in \prod_{\alpha \in \overline{B}} X_\alpha$. We claim that $\overline{B} = A$. If not, pick $\gamma \in A \setminus \overline{B}$. By Proposition 4.20 there is a subnet $\{\pi_{\overline{B}}(x_{i(j)})\rangle_{j \in J}$ of $\langle \pi_{\overline{B}}(x_i)\rangle$ that converges to $\overline{p}$, and since X_γ is compact, there is a subnet $\langle \pi_\gamma(x_{i(j(k))})\rangle_{k \in K}$ of $\langle \pi_\gamma(x_{i(j)})\rangle$ that converges to some $p_\gamma \in X_\gamma$. Let q be the unique element of $\prod_{\alpha \in \overline{B} \cup \{\gamma\}} X_\alpha$ that extends both $\overline{p}$ and p_γ; then the net $\langle \pi_{\overline{B} \cup \{\gamma\}}(x_{i(j(k))})\rangle_{k \in K}$ converges to q and hence q is a cluster point of $\langle \pi_{\overline{B} \cup \{\gamma\}}(x_i)\rangle$, contradicting the maximality of $\overline{p}$. Therefore $\overline{p}$ is a cluster point of $\langle x_i \rangle$, and we are done. ∎

We now turn to the Arzelà-Ascoli theorem, which has to do with compactness in spaces of continuous mappings. There are several variants of this result; the theorems below are two of the most useful ones. See also Exercise 61.

If X is a topological space and $\mathcal{F} \subset C(X)$, $\mathcal{F}$ is called **equicontinuous at** $x \in X$ if for every $\epsilon > 0$ there is a neighborhood U of x such that $|f(y) - f(x)| < \epsilon$ for all $y \in U$ and all $f \in \mathcal{F}$, and $\mathcal{F}$ is called **equicontinuous** if it is equicontinuous at each $x \in X$. Also, $\mathcal{F}$ is said to be **pointwise bounded** if $\{f(x) : f \in \mathcal{F}\}$ is a bounded subset of $\mathbb{C}$ for each $x \in X$.

4.43 Arzelà-Ascoli Theorem I. *Let X be a compact Hausdorff space. If $\mathcal{F}$ is an equicontinuous, pointwise bounded subset of $C(X)$, then $\mathcal{F}$ is totally bounded in the uniform metric, and the closure of $\mathcal{F}$ in $C(X)$ is compact.*

Proof. Suppose $\epsilon > 0$. Since $\mathcal{F}$ is equicontinuous, for each $x \in X$ there is an open neighborhood U_x of x such that $|f(y) - f(x)| < \frac{1}{4}\epsilon$ for all $y \in U_x$ and all $f \in \mathcal{F}$. Since X is compact, we can choose $x_1, \ldots, x_n \in X$ such that $\bigcup_1^n U_{x_j} = X$. Then by pointwise boundedness, $\{f(x_j) : f \in \mathcal{F},\ 1 \le j \le n\}$ is a bounded subset of $\mathbb{C}$, so there is a finite set $\{z_1, \ldots, z_m\} \subset \mathbb{C}$ that is $\frac{1}{4}\epsilon$-dense in it — that is, each $f(x_j)$ is at a distance less than $\frac{1}{4}\epsilon$ from some z_k. Let $A = \{x_1, \ldots, x_n\}$ and $B = \{z_1, \ldots, z_m\}$; then the set B^A of functions from A to B is finite. For each $\phi \in B^A$, let

$$\mathcal{F}_\phi = \left\{f \in \mathcal{F} : |f(x_j) - \phi(x_j)| < \tfrac{1}{4}\epsilon \text{ for } 1 \le j \le n\right\}.$$

Then clearly $\bigcup_{\phi \in B^A} \mathcal{F}_\phi = \mathcal{F}$, and we claim that each $\mathcal{F}_\phi$ has diameter at most ϵ, so we obtain a finite ϵ-dense subset of $\mathcal{F}$ by picking one f from each nonempty $\mathcal{F}_\phi$. To prove the claim, suppose $f, g \in \mathcal{F}_\phi$. Since $|f - \phi| < \frac{1}{4}\epsilon$ and $|g - \phi| < \frac{1}{4}\epsilon$ on A, we have $|f - g| < \frac{1}{2}\epsilon$ on A. If $x \in X$, we have $x \in U_{x_j}$ for some j, and then

$$|f(x) - g(x)| \le |f(x) - f(x_j)| + |f(x_j) - g(x_j)| + |g(x_j) - g(x)| < \epsilon.$$

This shows that $\mathcal{F}$ is totally bounded. Since the closure of a totally bounded set is totally bounded and $C(X)$ is complete, the theorem is proved. ∎

4.44 Arzelà-Ascoli Theorem II. *Let X be a σ-compact LCH space. If $\{f_n\}$ is an equicontinuous, pointwise bounded sequence in $C(X)$, there exist $f \in C(X)$ and a subsequence of $\{f_n\}$ that converges to f uniformly on compact sets.*

Proof. By Proposition 4.39 there is a sequence $\{U_k\}$ of precompact open sets such that $\overline{U}_k \subset U_{k+1}$ and $X = \bigcup_1^\infty U_k$. By Theorem 4.43 there is a subsequence $\{f_{n_j}\}_{j=1}^\infty$ of $\{f_n\}$ that is uniformly Cauchy on $\overline{U}_1$; we denote it by $\{f_j^1\}_{j=1}^\infty$. Proceeding inductively, for $k \in \mathbb{N}$ we obtain a subsequence $\{f_j^k\}_{j=1}^\infty$ of $\{f_j^{k-1}\}_{j=1}^\infty$ that is uniformly Cauchy on $\overline{U}_k$. Let $g_k = f_k^k$; then $\{g_k\}$ is a subsequence of $\{f_n\}$ which is (except for the first $k-1$ terms) a subsequence of $\{f_j^k\}$ and hence is uniformly Cauchy on each $\overline{U}_k$. Let $f = \lim g_k$. Then $f \in C(X)$ and $g_k \to f$ uniformly on compact sets by Propositions 4.38 and 4.40. ∎

Exercises

58. If $\{X_\alpha\}_\alpha \in A$ is a family of topological spaces of which infinitely many are noncompact, then every closed compact subset of $\prod_{\alpha \in A} X_\alpha$ is nowhere dense.

59. The product of finitely many locally compact spaces is locally compact.

60. The product of countably many sequentially compact spaces is sequentially compact. (Use the "diagonal trick" as in the proof of Theorem 4.44.)

61. Theorem 4.43 remains valid for maps from a compact Hausdorff space X into a complete metric space Y provided the hypothesis of pointwise boundedness is replaced by pointwise total boundedness. (Make this statement precise and then prove it.)

62. Rephrase Theorem 4.44 in a form similar to Theorem 4.43 by using the metric in Exercise 56d.

63. Let $K \in C([0,1] \times [0,1])$. For $f \in C([0,1])$, let $Tf(x) = \int_0^1 K(x,y)f(y)\,dy$. Then $Tf \in C([0,1])$, and $\{Tf : \|f\|_u \le 1\}$ is precompact in $C([0,1])$.

64. Let (X, ρ) be a metric space. A function $f \in C(X)$ is called **Hölder continuous of exponent α** ($\alpha > 0$) if the quantity

$$N_\alpha(f) = \sup_{x \neq y} \frac{|f(x) - f(y)|}{\rho(x,y)^\alpha}$$

is finite. If X is compact, $\{f \in C(X) : \|f\|_u \le 1 \text{ and } N_\alpha(f) \le 1\}$ is compact in $C(X)$.

65. Let U be an open subset of $\mathbb{C}$, and let $\{f_n\}$ be a sequence of holomorphic functions on U. If $\{f_n\}$ is uniformly bounded on compact subsets of U, there is a subsequence that converges uniformly to a holomorphic function on compact subsets of U. (Use the Cauchy integral formula to obtain equicontinuity.)

4.7 THE STONE-WEIERSTRASS THEOREM

In this section we prove a far-reaching generalization of the well-known theorem of Weierstrass to the effect that any continuous function on a compact interval $[a, b]$ is

the uniform limit of polynomials on $[a, b]$. Throughout this section, X will denote a compact Hausdorff space, and we equip the space $C(X)$ with the uniform metric.

A subset $\mathcal{A}$ of $C(X, \mathbb{R})$ or $C(X)$ is said to **separate points** if for every $x, y \in X$ with $x \neq y$ there exists $f \in \mathcal{A}$ such that $f(x) \neq f(y)$. $\mathcal{A}$ is called an **algebra** if it is a real (resp. complex) vector subspace of $C(X, \mathbb{R})$ (resp. $C(X)$) such that $fg \in \mathcal{A}$ whenever $f, g \in \mathcal{A}$. If $\mathcal{A} \subset C(X, \mathbb{R})$, $\mathcal{A}$ is called a **lattice** if $\max(f, g)$ and $\min(f, g)$ are in $\mathcal{A}$ whenever $f, g \in \mathcal{A}$. Since the algebra and lattice operations are continuous, one easily sees that if $\mathcal{A}$ is an algebra or a lattice, so is its closure $\overline{\mathcal{A}}$ in the uniform metric.

4.45 The Stone-Weierstrass Theorem. *Let X be a compact Hausdorff space. If $\mathcal{A}$ is a closed subalgebra of $C(X, \mathbb{R})$ that separates points, then either $\mathcal{A} = C(X, \mathbb{R})$ or $\mathcal{A} = \{f \in C(X, \mathbb{R}) : f(x_0) = 0\}$ for some $x_0 \in X$. The first alternative holds iff $\mathcal{A}$ contains the constant functions.*

The proof will require several lemmas. The first one, in effect, proves the theorem when X consists of two points, and the second one is a special case of the classical Weierstrass theorem for $X = [-1, 1]$. After these two we return to the general case.

4.46 Lemma. *Consider $\mathbb{R}^2$ as an algebra under coordinatewise addition and multiplication. Then the only subalgebras of $\mathbb{R}^2$ are $\mathbb{R}^2$, $\{(0, 0)\}$, and the linear spans of $(1, 0)$, $(0, 1)$, and $(1, 1)$.*

Proof. The subspaces of $\mathbb{R}^2$ listed above are evidently subalgebras. If $\mathcal{A} \subset \mathbb{R}^2$ is a nonzero algebra and $(0, 0) \neq (a, b) \in \mathcal{A}$, then $(a^2, b^2) \in \mathcal{A}$. If $a \neq 0$, $b \neq 0$, and $a \neq b$, then (a, b) and (a^2, b^2) are linearly independent, so $\mathcal{A} = \mathbb{R}^2$. The other possibilities — $a \neq 0 = b$, $a = 0 \neq b$, and $a = b \neq 0$ for all nonzero $(a, b) \in \mathcal{A}$ — give the other three subalgebras. ∎

4.47 Lemma. *For any $\epsilon > 0$ there is a polynomial P on $\mathbb{R}$ such that $P(0) = 0$ and $\big| |x| - P(x) \big| < \epsilon$ for $x \in [-1, 1]$.*

Proof. Consider the Maclaurin series for $(1 - t)^{1/2}$:

$$\begin{aligned}(1-t)^{1/2} &= 1 + \sum_1^\infty \left(-\frac{1}{2}\right)\left(\frac{1}{2}\right)\cdots\left(\frac{2n-3}{2}\right)\frac{t^n}{n!} \\ &= 1 - \sum_1^\infty c_n t^n \qquad (c_n > 0).\end{aligned}$$

By the ratio test, this series converges for $|t| < 1$; a proof that its sum is actually $(1 - t)^{1/2}$ is outlined in Exercise 66. Moreover, by the monotone convergence theorem (applied to counting measure on $\mathbb{N}$),

$$\sum_1^\infty c_n = \lim_{t \nearrow 1} \sum_1^\infty c_n t^n = 1 - \lim_{t \nearrow 1} (1 - t)^{1/2} = 1.$$

It follows from the finiteness of $\sum_1^\infty c_n$ that the series $1 - \sum_1^\infty c_n t^n$ converges absolutely and uniformly on $[-1,1]$, and its sum is $(1-t)^{1/2}$ there. Therefore, given $\epsilon > 0$, by taking a suitable partial sum of this series we obtain a polynomial Q such that $|(1-t)^{1/2} - Q(t)| < \frac{1}{2}\epsilon$ for $t \in [-1,1]$. Setting $t = 1 - x^2$ and $R(x) = Q(1-x^2)$, we obtain a polynomial R such that $\big|\,|x| - R(x)\big| < \frac{1}{2}\epsilon$ for $x \in [-1,1]$. In particular, $|R(0)| < \frac{1}{2}\epsilon$, so if we set $P(x) = R(x) - R(0)$, P is a polynomial such that $P(0) = 0$ and $\big|\,|x| - P(x)\big| < \epsilon$ for $x \in [-1,1]$. ∎

4.48 Lemma. *If $\mathcal{A}$ is a closed subalgebra of $C(X,\mathbb{R})$, then $|f| \in \mathcal{A}$ whenever $f \in \mathcal{A}$, and $\mathcal{A}$ is a lattice.*

Proof. If $f \in \mathcal{A}$ and $f \neq 0$, let $h = f/\|f\|_u$. Then h maps X into $[-1,1]$, so if $\epsilon > 0$ and P is as in Lemma 4.47, we have $\|\,|h| - P \circ h\|_u < \epsilon$. Since $P(0) = 0$, P has no constant term, so $P \circ h \in \mathcal{A}$ since $\mathcal{A}$ is an algebra. Since $\mathcal{A}$ is closed and ϵ is arbitrary, we have $|h| \in \mathcal{A}$ and hence $|f| = \|f\|_u |h| \in \mathcal{A}$. This proves the first assertion, and the second one follows because

$$\max(f,g) = \tfrac{1}{2}(f + g + |f-g|), \qquad \min(f,g) = \tfrac{1}{2}(f + g - |f-g|).$$

∎

4.49 Lemma. *Suppose $\mathcal{A}$ is a closed lattice in $C(X,\mathbb{R})$ and $f \in C(X,\mathbb{R})$. If for every $x, y \in X$ there exists $g_{xy} \in \mathcal{A}$ such that $g_{xy}(x) = f(x)$ and $g_{xy}(y) = f(y)$, then $f \in \mathcal{A}$.*

Proof. Given $\epsilon > 0$, for each $x, y \in X$ let $U_{xy} = \{z \in X : f(z) < g_{xy}(z) + \epsilon\}$ and $V_{xy} = \{z \in X : f(z) > g_{xy}(z) - \epsilon\}$. These sets are open and contain x and y. Fix y; then $\{U_{xy} : x \in X\}$ covers X, so there is a finite subcover $\{U_{x_j y}\}_1^n$. Let $g_y = \max(g_{x_1 y}, \ldots, g_{x_n y})$; then $f < g_y + \epsilon$ on X and $f > g_y - \epsilon$ on $V_y = \bigcap_1^n V_{x_j y}$, which is open and contains y. Thus $\{V_y\}_{y \in X}$ is another open cover of X, so there is a finite subcover $\{V_{y_j}\}_1^m$. Let $g = \min(g_{y_1}, \ldots, g_{y_m})$; then $\|f - g\|_u < \epsilon$. Since $\mathcal{A}$ is a lattice, $g \in \mathcal{A}$, and since $\mathcal{A}$ is closed and ϵ is arbitrary, $f \in \mathcal{A}$. ∎

Proof of Theorem 4.45. Given $x \neq y \in X$, let $\mathcal{A}_{xy} = \{(f(x), f(y)) : f \in \mathcal{A}\}$. Then $\mathcal{A}_{xy}$ is a subalgebra of $\mathbb{R}^2$ as in Lemma 4.46 because $f \mapsto (f(x), f(y))$ is an algebra homomorphism. If $\mathcal{A}_{xy} = \mathbb{R}^2$ for all x, y, then Lemmas 4.48 and 4.49 imply that $\mathcal{A} = C(X,\mathbb{R})$. Otherwise, there exist x, y for which $\mathcal{A}_{xy}$ is a proper subalgebra of $\mathbb{R}^2$. It cannot be $\{(0,0)\}$ or the linear span of $(1,1)$ because $\mathcal{A}$ separates points, so by Lemma 4.46 $\mathcal{A}_{xy}$ is the linear span of $(1,0)$ or $(0,1)$. In either case there exists $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathcal{A}$. There is only one such x_0 since $\mathcal{A}$ separates points, so if neither x nor y is x_0, we have $\mathcal{A}_{xy} = \mathbb{R}^2$. Lemmas 4.48 and 4.49 now imply that $\mathcal{A} = \{f \in C(X,\mathbb{R}) : f(x_0) = 0\}$. Finally, if $\mathcal{A}$ contains constant functions, there is no x_0 such that $f(x_0) = 0$ for all $f \in \mathcal{A}$, so $\mathcal{A}$ must equal $C(X,\mathbb{R})$. ∎

We have stated the Stone-Weierstrass theorem in the form that is most natural for the proof. However, in applications one is typically dealing with a subalgebra $\mathcal{B}$ of $C(X, \mathbb{R})$ that is not closed, and one applies the theorem to $\mathcal{A} = \overline{\mathcal{B}}$. The resulting restatement of the theorem is as follows:

4.50 Corollary. *Suppose $\mathcal{B}$ is a subalgebra of $C(X, \mathbb{R})$ that separates points. If there exists $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathcal{B}$, then $\mathcal{B}$ is dense in $\{f \in C(X, \mathbb{R}) : f(x_0) = 0\}$. Otherwise, $\mathcal{B}$ is dense in $C(X, \mathbb{R})$.*

The classical Weierstrass approximation theorem is the special case of this corollary where X is a compact subset of $\mathbb{R}^n$ and $\mathcal{B}$ is the algebra of polynomials on $\mathbb{R}^n$ (restricted to X); here $\mathcal{B}$ contains the constant functions, so the conclusion is that it is dense in $C(X, \mathbb{R})$.

The Stone-Weierstrass theorem, as it stands, is false for complex-valued functions. For example, the algebra of polynomials in one complex variable is not dense in $C(K)$ for most compact subsets K of $\mathbb{C}$. (In particular, if $K^o \neq \varnothing$, any uniform limit of polynomials on K must be holomorphic on K^o.) Here we shall give a simple proof that the function $f(z) = \overline{z}$ cannot be approximated uniformly by polynomials on the unit circle $\{e^{it} : t \in [0, 2\pi]\}$. If $P(z) = \sum_0^n a_j z^j$, then

$$\int_0^{2\pi} \overline{f}(e^{it}) P(e^{it})\, dt = \sum_0^n a_j \int_0^{2\pi} e^{i(j+1)t}\, dt = 0.$$

Thus, abbreviating $f(e^{it})$ and $P(e^{it})$ by f and P, since $|f| = 1$ on the unit circle we have

$$\begin{aligned} 2\pi = \left| \int_0^{2\pi} f\overline{f}\, dt \right| &\leq \left| \int_0^{2\pi} (f - P)\overline{f}\, dt \right| + \left| \int_0^{2\pi} \overline{f} P\, dt \right| \\ &= \left| \int_0^{2\pi} (f - P)\overline{f}\, dt \right| \leq \int_0^{2\pi} |f - P|\, dt \leq 2\pi \|f - P\|_u. \end{aligned}$$

Therefore, $\|f - P\|_u \geq 1$ for any polynomial P.

There is, however, a complex version of the Stone-Weierstrass theorem.

4.51 The Complex Stone-Weierstrass Theorem. *Let X be a compact Hausdorff space. If $\mathcal{A}$ is a closed complex subalgebra of $C(X)$ that separates points and is closed under complex conjugation, then either $\mathcal{A} = C(X)$ or $\mathcal{A} = \{f \in C(X) : f(x_0) = 0\}$ for some $x_0 \in X$.*

Proof. Since $\operatorname{Re} f = (f + \overline{f})/2$ and $\operatorname{Im} f = (f - \overline{f})/2i$, the set $\mathcal{A}_{\mathbb{R}}$ of real and imaginary parts of functions in $\mathcal{A}$ is a subalgebra of $C(X, \mathbb{R})$ to which the Stone-Weierstrass theorem applies. Since $\mathcal{A} = \{f + ig : f, g \in \mathcal{A}_{\mathbb{R}}\}$, the desired result follows. ∎

There is also a version of the Stone-Weierstrass theorem for noncompact LCH spaces. We state this result for real functions; the corresponding analogue of Theorem 4.51 for complex functions is an immediate consequence.

4.52 Theorem. *Let X be a noncompact LCH space. If $\mathcal{A}$ is a closed subalgebra of $C_0(X,\mathbb{R})$ $(= C_0(X) \cap C(X,\mathbb{R}))$ that separates points, then either $\mathcal{A} = C_0(X,\mathbb{R})$ or $\mathcal{A} = \{f \in C_0(X,\mathbb{R}) : f(x_0) = 0\}$ for some $x_0 \in X$.*

The proof is outlined in Exercise 67.

Exercises

66. Let $1 - \sum_1^\infty c_n t^n$ be the Maclaurin series for $(1-t)^{1/2}$.

a. The series converges absolutely and uniformly on compact subsets of $(-1,1)$, as does the termwise differentiated series $-\sum_1^\infty nc_n t^{n-1}$. Thus, if $f(t) = 1 - \sum_1^\infty c_n t^n$, then $f'(t) = -\sum_1^\infty nc_n t^{n-1}$.

b. By explicit calculation, $f(t) = -2(1-t)f'(t)$, from which it follows that $(1-t)^{-1/2}f(t)$ is constant. Since $f(0) = 1$, $f(t) = (1-t)^{1/2}$.

67. Prove Theorem 4.52. (If there exists $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathcal{A}$, let Y be the one-point compactification of $X \setminus \{x_0\}$; otherwise let Y be the one-point compactification of X. Apply Proposition 4.36 and the Stone-Weierstrass theorem on Y.)

68. Let X and Y be compact Hausdorff spaces. The algebra generated by functions of the form $f(x,y) = g(x)h(y)$, where $g \in C(X)$ and $h \in C(Y)$, is dense in $C(X \times Y)$.

69. Let A be a nonempty set, and let $X = [0,1]^A$. The algebra generated by the coordinate maps $\pi_\alpha : X \to [0,1]$ $(\alpha \in A)$ and the constant function 1 is dense in $C(X)$.

70. Let X be a compact Hausdorff space. An **ideal** in $C(X,\mathbb{R})$ is a subalgebra $\mathcal{I}$ of $C(X,\mathbb{R})$ such that if $f \in \mathcal{I}$ and $g \in C(X,\mathbb{R})$ then $fg \in \mathcal{I}$.

a. If $\mathcal{I}$ is an ideal in $C(X,\mathbb{R})$, let $h(\mathcal{I}) = \{x \in X : f(x) = 0 \text{ for all } f \in \mathcal{I}\}$. Then $h(\mathcal{I})$ is a closed subset of X, called the **hull** of $\mathcal{I}$.

b. If $E \subset X$, let $k(E) = \{f \in C(X,\mathbb{R}) : f(x) = 0 \text{ for all } x \in E\}$. Then $k(E)$ is a closed ideal in $C(X,\mathbb{R})$, called the **kernel** of E.

c. If $E \subset X$, then $h(k(E)) = \overline{E}$.

d. If $\mathcal{I}$ is an ideal in $C(X,\mathbb{R})$, then $k(h(\mathcal{I})) = \overline{\mathcal{I}}$. ((Hint: $k(h(\mathcal{I}))$ may be identified with a subalgebra of $C_0(U,\mathbb{R})$ where $U = X \setminus h(\mathcal{I})$.)

e. The closed subsets of X are in one-to-one correspondence with the closed ideals of $C(X,\mathbb{R})$.

71. (This is a variation on the theme of Exercise 70; it does not use the Stone-Weierstrass theorem.) Let X be a compact Hausdorff space, and let M be the set of all nonzero algebra homomorphisms from $C(X,\mathbb{R})$ to $\mathbb{R}$. Each $x \in X$ defines an element $\widehat{x}$ of M by $\widehat{x}(f) = f(x)$.

a. If $\phi \in M$, then $\{f \in C(X,\mathbb{R}) : \phi(f) = 0\}$ is a maximal proper ideal in $C(X,\mathbb{R})$.

b. If $\mathcal{I}$ is a proper ideal in $C(X, \mathbb{R})$, there exists $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in \mathcal{I}$. (Suppose not; construct an $f \in \mathcal{I}$ with $f > 0$ everywhere and conclude that $1 \in \mathcal{I}$. This requires no deep theorems.)

c. The map $x \to \widehat{x}$ is a bijection from X to M.

d. If M is equipped with the topology of pointwise convergence, then the map $x \to \widehat{x}$ is a homeomorphism from X to M. (Since M is defined purely algebraically, it follows that the topological structure of X is completely determined by the algebraic structure of $C(X, \mathbb{R})$.)

4.8 EMBEDDINGS IN CUBES

We now present a technique for embedding topological spaces in products of intervals and discuss some of its applications. (These results will not be used elsewhere in this book.) Throughout this section we shall denote the unit interval $[0, 1]$ by I, and if A is any nonempty set, we shall call the product space I^A a **cube**.

If X is a topological space and $\mathcal{F} \subset C(X, I)$, we say that $\mathcal{F}$ **separates points and closed sets** if for every closed $E \subset X$ and every $x \in E^c$ there exists $f \in \mathcal{F}$ such that $f(x) \notin \overline{f(E)}$. If $\mathcal{F}$ separates points and closed sets, there is another family $\mathcal{G} \subset C(X, I)$ with the following slightly stronger property: For every closed set $E \subset X$ and every $x \in E^c$ there exists $g \in \mathcal{G}$ with $g(x) = 1$ and $g = 0$ on E. (Indeed, if $f \in \mathcal{F}$ satisfies $f(x) \notin \overline{f(E)}$, take $g = \phi \circ f$ where $\phi \in C(I, I)$, $\phi(f(x)) = 1$, and $\phi = 0$ on $\overline{f(E)}$.) It follows that a T_1 space X admits a family $\mathcal{F}$ that separates points and closed sets iff X is completely regular.

Each nonempty $\mathcal{F} \subset C(X, I)$ canonically induces a map $e : X \to I^{\mathcal{F}}$ by the formula $\pi_f(e(x)) = f(x)$, where $\pi_f : I^{\mathcal{F}} \to I$ is the coordinate map. We call e the map from X into the cube $I^{\mathcal{F}}$ **associated to** $\mathcal{F}$. (Evidently this construction can be generalized to target spaces other than I; see Exercise 20.)

4.53 Proposition. *Let X be a topological space, $\mathcal{F} \subset C(X, I)$, and $e : X \to I^{\mathcal{F}}$ be the map associated to $\mathcal{F}$. Then*

a. e is continuous.

b. If $\mathcal{F}$ separates points, then e is injective.

c. If X is T_1 and $\mathcal{F}$ separates points and closed sets, then e is an embedding.

Proof. (a) follows from Proposition 4.11, and (b) is obvious. Next, observe that if $\mathcal{F}$ separates points and closed sets and X is T_1, then e is injective by (b) and Proposition 4.7. To prove the continuity of the inverse, suppose that U is open in X. If $x \in U$, choose $f \in \mathcal{F}$ with $f(x) \notin \overline{f(U^c)}$ and let

$$V = \pi_f^{-1}\left[\overline{f(U^c)}\right]^c = \{p \in I^{\mathcal{F}} : \pi_f(p) \notin \overline{f(U^c)}\}.$$

Then V is open in $I^{\mathcal{F}}$ and $e(x) \in V \cap e(X) \subset e(U)$. Thus $e(U)$ is a neighborhood of $e(x)$ in $e(X)$ at every $x \in U$, so $e(U)$ is open in $e(X)$. It follows that e^{-1} is continuous. ∎

4.54 Corollary. *Every compact Hausdorff space is homeomorphic to a closed subset of a cube.*

Proof. By Proposition 4.25 and Urysohn's lemma, we can take $\mathcal{F} = C(X, I)$. ∎

4.55 Corollary. *A topological space is completely regular iff it is homeomorphic to a subset of a compact Hausdorff space.*

Proof. Proposition 4.53, with $\mathcal{F} = C(X, I)$, gives the "only if" implication; the converse is left to the reader (Exercise 72). ∎

A **compactification** of a topological space X is a pair (Y, ϕ) where Y is a compact Hausdorff space and ϕ is a homeomorphism from X onto a dense subset of Y. (Frequently one identifies X with its image $\phi(X) \subset Y$ and then speaks simply of "the compactification Y of X.") For example, $([-1, 1], \tanh)$ is a compactification of $\mathbb{R}$, and the one-point compactification (X^*, i) of an LCH space X is a compactification in the present sense, where $i : X \to X^*$ is the inclusion map.

Suppose X is completely regular. According to Proposition 4.53, if $\mathcal{F} \subset C(X, I)$ separates points and closed sets, $e : X \to I^{\mathcal{F}}$ is the associated embedding, and Y is the closure of $e(X)$ in $I^{\mathcal{F}}$, then (Y, e) is a compactification of X. It has the property that if we identify X with its image $e(X)$, every $f \in \mathcal{F}$ has a continuous extension to Y, which is unique since X is dense in Y. Indeed, the identification of X with $e(X)$ turns f into the coordinate map $\pi_f|e(X)$, which extends to $\pi_f|Y$. Moreover, if f and g are bounded continuous functions on X that extend continuously to Y, obviously so are $f + g$ and fg, and if $\{f_n\}$ is a uniformly convergent sequence of functions on X that extend continuously to Y, their extensions converge uniformly on Y since X is dense in Y, so $f = \lim f_n$ also extends continuously. We have proved:

4.56 Proposition. *Suppose that $\mathcal{F} \subset C(X, I)$ separates points and closed sets. Let (Y, e) be the compactification of X associated to $\mathcal{F}$, and let $\mathcal{A}$ be the smallest closed subalgebra of $BC(X)$ that contains $\mathcal{F}$. Then every $f \in \mathcal{A}$ has a continuous extension to Y.*

This result has a converse: see Exercise 73.

If X is a completely regular space, the compactification of X associated to $\mathcal{F} = C(X, I)$ is called the **Stone-Čech compactification** of X and is denoted by $(\beta X, e)$, or simply by βX if we identify X with $e(X)$. Every $f \in BC(X)$ extends continuously to βX; in fact, a much more general result holds:

4.57 Theorem. *If X is a completely regular space, Y is a compact Hausdorff space, and $\phi \in C(X, Y)$, then ϕ has a unique continuous extension $\widetilde{\phi}$ to βX — that is, there is a unique $\widetilde{\phi} \in C(\beta X, Y)$ such that $\widetilde{\phi} \circ e = \phi$. If (Y, ϕ) is a compactification of X, then $\widetilde{\phi}$ is surjective; if also every $f \in BC(X)$ extends continuously to Y (i.e., $f = g \circ \phi$ for some $g \in C(Y)$), then $\widetilde{\phi}$ is a homeomorphism.*

Proof. Let $\mathcal{F} = C(X, I)$ and $\mathcal{G} = C(Y, I)$, and let $(\beta Y, i)$ be the Stone-Čech compactification of Y. (That is, $i : Y \to I^{\mathcal{G}}$ is the embedding associated to $\mathcal{G}$,

and $\beta Y = i(Y)$; βY is homeomorphic to Y since Y is compact.) Given $\phi \in C(X, Y)$, define $\Phi : I^{\mathcal{F}} \to I^{\mathcal{G}}$ by $\pi_g(\Phi(p)) = \pi_{g\circ\phi}(p)$. The map Φ is continuous by Proposition 4.11, and

$$\pi_g(\Phi(e(x))) = \pi_{g\circ\phi}(e(x)) = g(\phi(x)) = \pi_g(i(\phi(x))),$$

that is, $\Phi \circ e = i \circ \phi$. It follows that $\Phi(e(X)) = i(\phi(X)) \subset \beta Y$ and hence that $\Phi(\beta X) \subset \overline{\beta Y} = \beta Y$. The situation is summarized in the following commutative diagram:

$$\begin{array}{ccccc} X & \xrightarrow{e} & \beta X & \hookrightarrow & I^{\mathcal{F}} \\ {\scriptstyle\phi}\downarrow & & {\scriptstyle\Phi|\beta X}\downarrow & & {\scriptstyle\Phi}\downarrow \\ Y & \xrightarrow{i} & \beta Y & \hookrightarrow & I^{\mathcal{G}} \end{array}$$

Let $\widetilde{\phi} = i^{-1} \circ (\Phi|\beta X)$. Then $\widetilde{\phi} \circ e = i^{-1} \circ \Phi \circ e = \phi$, and uniqueness of $\widetilde{\phi}$ is clear since $e(X)$ is dense in βX; thus the first assertion is proved. If (Y, ϕ) is a compactification of X, then $\phi(X)$ is dense in Y; but then $\widetilde{\phi}(\beta X)$ is dense in Y and also compact, so that $\widetilde{\phi}(\beta X) = Y$. Finally, if every $f \in BC(X)$ is of the form $g \circ \phi$ for some $g \in C(Y)$, then Φ is injective; hence $\widetilde{\phi}$ is bijective and therefore, by Proposition 4.28, a homeomorphism. ∎

This theorem shows that βX is the "largest" compactification of a completely regular space X, in the sense that every other compactification is a continuous image of it. At the other end of the scale, if X is locally compact, then $\mathcal{F} = C_c(X) \cap C(X, I)$ separates points and closed sets by Urysohn's lemma. A glance at the construction of the compactification (Y, e) associated to this $\mathcal{F}$ shows that Y consists of $e(X)$ together with the single point of $I^{\mathcal{F}}$ all of whose coordinates are zero. It is then easy to verify that Y is homeomorphic to the one-point compactification of X constructed in §4.5.

As a final application of the embedding $e : X \to I^{\mathcal{F}}$, we give a partial answer to the question: When is a topological space metrizable, that is, when is its topology defined by a metric? A necessary condition for X to be metrizable is that X be normal (Exercise 3). On the other hand:

4.58 The Urysohn Metrization Theorem. *Every second countable normal space is metrizable.*

Since every subset of a metrizable space is metrizable (with the same metric), this theorem is an immediate consequence of Proposition 4.53 and the following two facts, whose proofs are outlined in Exercises 76 and 77:

- If X is normal and second countable, there is a countable family $\mathcal{F} \subset C(X, I)$ that separates points and closed sets.
- If $\mathcal{F}$ is countable, $I^{\mathcal{F}}$ is metrizable.

Exercises

72. Every subset of a completely regular space is completely regular in the relative topology.

73. If X is a completely regular space, a subalgebra $\mathcal{A}$ of $BC(X)$ is called **completely regular** if (i) it is closed and contains the constant functions, and (ii) $\mathcal{A} \cap C(X, I)$ separates points and closed sets.

a. If (Y, e) is a Hausdorff compactification of X, $\mathcal{A}_Y = \{f \circ e : f \in C(Y)\}$ is a completely regular subalgebra of $BC(X)$.

b. If (Y, e) and (Y', e') are Hausdorff compactifications of X such that $\mathcal{A}_Y = \mathcal{A}_{Y'}$, there is a homeomorphism $\phi : Y \to Y'$ such that $\phi \circ e = e'$. (Adapt the proof of Theorem 4.57, which deals with the case $Y = \beta X$.)

c. If (Y, e) is the compactification of X associated to $\mathcal{F} \subset C(X, I)$, then $\mathcal{A}_Y$ is the smallest closed subalgebra of $BC(X)$ that contains $\mathcal{F}$. (Use Exercise 69.)

d. The Hausdorff compactifications of X are in one-to-one correspondence with the completely regular subalgebras of $BC(X)$.

74. Consider $\mathbb{N}$ (with the discrete topology) as a subset of its Stone-Čech compactification $\beta\mathbb{N}$.

a. If A and B are disjoint subsets of $\mathbb{N}$, their closures in $\beta\mathbb{N}$ are disjoint. (Hint: $\chi_A \in C(\mathbb{N}, I)$.)

b. No sequence in $\mathbb{N}$ converges in $\beta\mathbb{N}$ unless it is eventually constant (so $\beta\mathbb{N}$ is emphatically *not* sequentially compact).

75. Suppose X is a completely regular space. The set M of nonzero algebra homomorphisms from $BC(X, \mathbb{R})$ to $\mathbb{R}$, equipped with the topology of pointwise convergence, is homeomorphic to βX. (See Exercise 71. This realization of βX is the natural one from the point of view of Banach algebra theory.)

76. If X is normal and second countable, there is a countable family $\mathcal{F} \subset C(X, I)$ that separates points and closed sets. (Let $\mathcal{B}$ be a countable base for the topology. Consider the set of pairs $(U, V) \in \mathcal{B} \times \mathcal{B}$ such that $\overline{U} \subset V$, and use Urysohn's lemma.)

77. Let $\{(X_n, \rho_n)\}_1^\infty$ be a countable family of metric spaces whose metrics take values in $[0, 1]$. (The latter restriction can always be satisfied; see Exercise 56b.) Let $X = \prod_1^\infty X_n$. If $x, y \in X$, say $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2, \ldots)$, define $\rho(x, y) = \sum_1^\infty 2^{-n} \rho_n(x_n, y_n)$. Then ρ is a metric that defines the product topology on X.

4.9 NOTES AND REFERENCES

The germ of the concept of topological space is clearly present in Riemann's lecture [113] on the foundations of geometry, delivered in 1854, but another half century passed before the mathematical world was ready to consider abstract spaces in a

systematic way. The first attempt to construct an abstract framework for the study of limits and continuity was made in 1906 by Fréchet [52], who introduced metric spaces as well as a more general class of quasi-topological spaces whose properties were defined in terms of sequential convergence. A few years later Hausdorff [68] devised axioms for neighborhoods of points that amount to the definition of a Hausdorff space, and he deduced from them many of the basic results of general topology. The usefulness of his point of view was quickly recognized, and it became the foundation for the further development of the subject.

There are several good books to which the reader may refer for a more comprehensive treatment of point set topology, including Bourbaki [20], Dugundji [34], Engelking [38], Kelley [83], and Nagata [102]. Engelking [38] contains extensive references and historical notes.

§4.2: Urysohn's lemma and the Tietze extension theorem were both first proved in Urysohn [152]. Special cases of the latter had previously been obtained by several authors, including Tietze (see [152] for references). Examples of completely regular spaces that are not normal and regular spaces that are not completely regular, which are all rather complicated, were first constructed by Tychonoff [151]. Particularly noteworthy is the existence of a regular space that admits no nonconstant continuous functions, a result due to Hewitt [73]. Examples may also be found in the books cited above.

§4.3: The theory of nets is sometimes called the Moore-Smith theory of convergence, after its originators [101]. Another general theory of convergence, invented by H. Cartan and publicized by Bourbaki, is based on the notion of filters. A **filter** in a set X is a family $\mathcal{F} \subset \mathcal{P}(X)$ with the following properties:

- If $F \in \mathcal{F}$ and $E \supset F$, then $E \in \mathcal{F}$.
- If $E \in \mathcal{F}$ and $F \in \mathcal{F}$, then $E \cap F \in \mathcal{F}$.
- $\varnothing \notin \mathcal{F}$.

If X is a topological space, a filter $\mathcal{F}$ in X **converges** to $x \in X$ if every neighborhood of x belongs to $\mathcal{F}$. Filters and nets are related as follows. If $\langle x_\alpha \rangle_{\alpha \in A}$ is a net in X, its **derived filter** is the collection of all $E \subset X$ such that $\langle x_\alpha \rangle$ is eventually in E. On the other hand, if $\mathcal{F}$ is a filter, then $\mathcal{F}$ is a directed set under reverse inclusion, and a net $\langle x_F \rangle_{F \in \mathcal{F}}$ indexed by $\mathcal{F}$ is said to be **associated** to $\mathcal{F}$ if $x_F \in F$ for all $F \in \mathcal{F}$. It is then easy to verify that a net $\langle x_\alpha \rangle$ converges to x iff its derived filter converges to x, and a filter $\mathcal{F}$ converges to x iff all of its associated nets converge to x. See Bourbaki [20] or Dugundji [34] for more information.

§4.4: The usage of the term "compact" is not completely standardized. In many older works the terms "compact" and "bicompact" were used to mean countably compact and compact, respectively, and some authors use "compact" and "quasi-compact" to mean compact Hausdorff and compact, respectively. Synonyms for "precompact" that are frequently found in the literature are "conditionally compact" and "relatively compact"; the latter one is infelicitous because it suggests compactness in the relative topology, which is quite different.

§4.6: Tychonoff [151] proved that $[0,1]^A$ is compact for any set A; together with Corollary 4.54, which is in the same paper, this easily implies that any product of compact Hausdorff spaces is compact. The Tychonoff theorem in full generality is due to Čech [23]. The proof we have presented, which is simpler and more elegant than the older ones, is due to Chernoff [24].

The axiom of choice, usually in the form of Zorn's lemma, is an essential ingredient in all the proofs of Tychonoff's theorem. It is an intriguing fact, discovered by Kelley [82], that Tychonoff's theorem in turn implies the axiom of choice. Here is the proof:

Suppose that $\{X_\alpha\}_{\alpha\in A}$ is a nonempty collection of nonempty sets. Pick a point ω that is not an element of any X_α, set $X_\alpha^* = X_\alpha \cup \{\omega\}$, and define a topology on X_α^* by declaring the open sets to be $\varnothing$, X_α, $\{\omega\}$, and X_α^*. Evidently X_α^* is compact, so Tychonoff's theorem implies that $X^* = \prod_{\alpha\in A} X_\alpha^*$ is compact. Let $F_\alpha = \pi_\alpha^{-1}(X_\alpha)$. The sets F_α are closed, and by the axiom of choice for finite collections of sets — which is provable from the other standard axioms of set theory — they have the finite intersection property. Indeed, given a finite set $B \subset A$, pick $x_\beta \in X_\beta$ for $\beta \in B$; then $\bigcap_{\beta\in B} F_\beta$ contains the point $x \in X$ such that $\pi_\beta(x) = x_\beta$ for $\beta \in B$ and $\pi_\alpha(x) = \omega$ for $\alpha \notin B$. By Proposition 4.21, $\bigcap_{\alpha\in A} F_\alpha$, which is precisely $\prod_{\alpha\in A} X_\alpha$, is nonempty.

By elaboration of this argument, one can deduce the axiom of choice from the special case of Tychonoff's theorem that X^A is compact for any A if X is compact; see Ward [156].

The original results of Arzelà and Ascoli had to do with functions on $\mathbb{R}$; see Arzelà [6]. Other versions of the Arzelà-Ascoli theorem, pertaining to the compactness of subsets of $C(X,Y)$ under various hypotheses on X and Y, can be found in the books cited above and in Royden [121].

§4.7. The Stone-Weierstrass theorem first appeared in the middle of a lengthy and difficult paper of Stone [144]. Later Stone [145] wrote a much-simplified exposition of the theorem and some of its applications, which still makes good reading.

§4.8. The history of this material begins with Urysohn [153], where the metrization theorem is proved, essentially by the method we have outlined. The technique of embedding spaces in cubes is implicit in this paper, but it was first developed explicitly in Tychonoff [151]. The Stone-Čech compactification, in turn, is implicit in the latter paper, but it was first described explicitly and investigated by Stone [144] and Čech [23].

It is not hard to show that every second countable regular space is normal (see Kelley [82, Lemma 4.1]; consequently, the hypothesis of normality in the Urysohn metrization theorem can be replaced by regularity. Necessary and sufficient conditions are known for an arbitrary topological space to be metrizable, but they are not as readily verifiable as the conditions in Urysohn's theorem. See the books cited above.

Occasionally the term "compactification" is used to mean a continuous injection $\phi : X \to Y$ from a topological space X onto a dense subset of a compact space Y without the requirement that it be an embedding. Such "compactifications" arise from subalgebras of $C(X)$ that separate points but are not completely regular in the sense of Exercise 73. An example is provided by the algebra of "uniformly

almost periodic" functions on $\mathbb{R}$, which is the algebra generated by the functions $f_\lambda(x) = e^{i\lambda x}$, $\lambda \in \mathbb{R}$; the associated "compactification" of $\mathbb{R}$ is known as the Bohr compactification. See Folland [47, §4.7].

5

Elements of Functional Analysis

"Functional analysis" is the traditional name for the study of infinite-dimensional vector spaces over $\mathbb{R}$ or $\mathbb{C}$ and the linear maps between them. What distinguishes this from mere linear algebra is the importance of topological considerations. On finite-dimensional vector spaces there is only one reasonable topology, and linear maps are automatically continuous, but in infinite dimensions things are not so simple. (As we have already observed, if $\{f_n\}$ is a sequence of functions on $\mathbb{R}$, there are many things one can mean by the statement "$f_n \to f$.") As our aim in this chapter is only to give a brief introduction to the subject, we shall restrict attention — except in §5.4 — to topologies defined by norms on vector spaces.

5.1 NORMED VECTOR SPACES

Let K denote either $\mathbb{R}$ or $\mathbb{C}$, and let $\mathcal{X}$ be a vector space over K. We denote the zero element of $\mathcal{X}$ simply by 0, relying on context to distinguish it from the scalar $0 \in K$. By a **subspace** we shall always mean a vector subspace. If $x \in \mathcal{X}$, we denote by Kx the one-dimensional subspace spanned by x. Also, if $\mathcal{M}$ and $\mathcal{N}$ are subspaces of $\mathcal{X}$, $\mathcal{M} + \mathcal{N}$ denotes the subspace $\{x + y : x \in \mathcal{M},\ y \in \mathcal{N}\}$.

A **seminorm** on $\mathcal{X}$ is a function $x \mapsto \|x\|$ from $\mathcal{X}$ to $[0, \infty)$ such that

- $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathcal{X}$ (the **triangle inequality**),
- $\|\lambda x\| = |\lambda|\, \|x\|$ for all $x \in \mathcal{X}$ and $\lambda \in K$.

The second property clearly implies that $\|0\| = 0$. A seminorm such that $\|x\| = 0$ only when $x = 0$ is called a **norm**, and a vector space equipped with a norm is called a **normed vector space** (or **normed linear space**).

If $\mathcal{X}$ is a normed vector space, the function $\rho(x, y) = \|x - y\|$ is a metric on $\mathcal{X}$, since

$$\|x - z\| \leq \|x - y\| + \|y - z\|, \qquad \|x - y\| = \|(-1)(x - y)\| = \|y - x\|.$$

The topology it defines is called the **norm topology** on $\mathcal{X}$. Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on $\mathcal{X}$ are called **equivalent** if there exist $C_1, C_2 > 0$ such that

$$C_1\|x\|_1 \leq \|x\|_2 \leq C_2\|x\|_1 \qquad (x \in \mathcal{X}).$$

Equivalent norms define equivalent metrics and hence the same topology and the same Cauchy sequences.

A normed vector space that is complete with respect to the norm metric is called a **Banach space**. (Every normed vector space can be embedded in a Banach space as a dense subspace. One way to do this is to mimic the construction of $\mathbb{R}$ from $\mathbb{Q}$ via Cauchy sequences; we shall present a simpler way in §5.2.) The following is a useful criterion for completeness of a normed vector space. If $\{x_n\}$ is a sequence in $\mathcal{X}$, the series $\sum_1^\infty x_n$ is said to **converge** to x if $\sum_1^N x_n \to x$ as $N \to \infty$, and it is called **absolutely convergent** if $\sum_1^\infty \|x_n\| < \infty$.

5.1 Theorem. *A normed vector space $\mathcal{X}$ is complete iff every absolutely convergent series in $\mathcal{X}$ converges.*

Proof. If $\mathcal{X}$ is complete and $\sum_1^\infty \|x_n\| < \infty$, let $S_N = \sum_1^N x_n$. Then for $N > M$ we have

$$\|S_N - S_M\| \leq \sum_{M+1}^{N} \|x_n\| \to 0 \text{ as } M, N \to \infty,$$

so the sequence $\{S_N\}$ is Cauchy and hence convergent. Conversely, suppose that every absolutely convergent series converges, and let $\{x_n\}$ be a Cauchy sequence. We can choose $n_1 < n_2 < \cdots$ such that $\|x_n - x_m\| < 2^{-j}$ for $m, n \geq n_j$. Let $y_1 = x_{n_1}$ and $y_j = x_{n_j} - x_{n_{j-1}}$ for $j > 1$. Then $\sum_1^k y_j = x_{n_k}$, and

$$\sum_1^\infty \|y_j\| \leq \|y_1\| + \sum_1^\infty 2^{-j} = \|y_1\| + 1 < \infty,$$

so $\lim x_{n_k} = \sum_1^\infty y_j$ exists. But since $\{x_n\}$ is Cauchy, it is easily verified that $\{x_n\}$ converges to the same limit as $\{x_{n_k}\}$. ∎

We have already seen some examples of Banach spaces. First, if X is a topological space, $B(X)$ and $BC(X)$ are Banach spaces with the uniform norm $\|f\|_u = \sup_{x \in X} |f(x)|$. Second, if $(X, \mathcal{M}, \mu)$ is a measure space, $L^1(\mu)$ is a Banach space

with the L^1 norm $\|f\|_1 = \int |f|\, d\mu$. (Observe that $\|\cdot\|_1$ is only a seminorm if we think of $L^1(\mu)$ as consisting of individual functions, but it becomes a norm if we identify functions that are equal a.e.) That $L^1(\mu)$ is complete follows from Theorems 2.25 and 5.1. Indeed, if $\sum_1^\infty \|f_n\|_1 < \infty$, Theorem 2.25 shows that $f = \sum_1^\infty f_n$ exists a.e., and

$$\int \Big| f - \sum_1^N f_n \Big|\, d\mu \le \sum_{N+1}^\infty \int |f_n|\, d\mu \to 0 \text{ as } N \to \infty.$$

More examples will be found in Exercises 8–11 and in subsequent sections.

If $\mathcal{X}$ and $\mathcal{Y}$ are normed vector spaces, $\mathcal{X} \times \mathcal{Y}$ becomes a normed vector space when equipped with the **product norm**

$$\|(x, y)\| = \max(\|x\|, \|y\|).$$

(Here, of course, $\|x\|$ refers to the norm on $\mathcal{X}$ while $\|y\|$ refers to the norm on $\mathcal{Y}$.) Sometimes other norms equivalent to this one, such as $\|(x, y)\| = \|x\| + \|y\|$ or $\|(x, y)\| = (\|x\|^2 + \|y\|^2)^{1/2}$, are used instead.

A related construction is that of quotient spaces. If $\mathcal{M}$ is a vector subspace of the vector space $\mathcal{X}$, it defines an equivalence relation on $\mathcal{X}$ as follows: $x \sim y$ iff $x - y \in \mathcal{M}$. The equivalence class of $x \in \mathcal{X}$ is denoted by $x + \mathcal{M}$, and the set of equivalence classes, or **quotient space**, is denoted by $\mathcal{X}/\mathcal{M}$. $\mathcal{X}/\mathcal{M}$ is a vector space with vector operations $(x+\mathcal{M})+(y+\mathcal{M}) = (x+y)+\mathcal{M}$ and $\lambda(x+\mathcal{M}) = (\lambda x)+\mathcal{M}$. If $\mathcal{X}$ is a normed vector space and $\mathcal{M}$ is closed, $\mathcal{X}/\mathcal{M}$ inherits a norm from $\mathcal{X}$ called the **quotient norm**, namely

$$\|x + \mathcal{M}\| = \inf_{y \in \mathcal{M}} \|x + y\|.$$

See Exercise 12 for a more detailed discussion.

A linear map $T : \mathcal{X} \to \mathcal{Y}$ between two normed vector spaces is called **bounded** if there exists $C \ge 0$ such that

$$\|Tx\| \le C\|x\| \text{ for all } x \in \mathcal{X}.$$

This is different from the notion of boundedness for functions on a set, according to which T would be bounded if $\|Tx\| \le C$ for all x. Clearly no nonzero linear map can satisfy the latter condition, since $T(\lambda x) = \lambda Tx$ for all scalars λ. The present definition means that T is bounded on bounded subsets of $\mathcal{X}$.

5.2 Proposition. *If $\mathcal{X}$ and $\mathcal{Y}$ are normed vector spaces and $T : \mathcal{X} \to \mathcal{Y}$ is a linear map, the following are equivalent:*

a. T is continuous.

b. T is continuous at 0.

c. T is bounded.

Proof. That (a) implies (b) is trivial. If T is continuous at $0 \in \mathcal{X}$, there is a neighborhood U of 0 such that $T(U) \subset \{y \in \mathcal{Y} : \|y\| \le 1\}$, and U must contain

a ball $B = \{x \in \mathcal{X} : \|x\| \le \delta\}$ about 0; thus $\|Tx\| \le 1$ when $\|x\| \le \delta$. Since T commutes with scalar multiplication, it follows that $\|Tx\| \le a\delta^{-1}$ whenever $\|x\| \le a$, that is, $\|Tx\| \le \delta^{-1}\|x\|$. This shows that (b) implies (c). Finally, if $\|Tx\| \le C\|x\|$ for all x, then $\|Tx_1 - Tx_2\| = \|T(x_1 - x_2)\| \le \epsilon$ whenever $\|x_1 - x_2\| \le C^{-1}\epsilon$, so that T is continuous. ∎

If $\mathcal{X}$ and $\mathcal{Y}$ are normed vector spaces, we denote the space of all bounded linear maps from $\mathcal{X}$ to $\mathcal{Y}$ by $L(\mathcal{X}, \mathcal{Y})$. It is easily verified that $L(\mathcal{X}, \mathcal{Y})$ is a vector space and that the function $T \mapsto \|T\|$ defined by

$$\begin{aligned} \|T\| &= \sup\{\|Tx\| : \|x\| = 1\} \\ &= \sup\left\{\frac{\|Tx\|}{\|x\|} : x \neq 0\right\} \\ &= \inf\{C : \|Tx\| \le C\|x\| \text{ for all } x\} \end{aligned} \tag{5.3}$$

is a norm on $L(\mathcal{X}, \mathcal{Y})$, called the **operator norm** (Exercise 2). We always assume $L(\mathcal{X}, \mathcal{Y})$ to be equipped with this norm unless we specify otherwise.

5.4 Proposition. *If $\mathcal{Y}$ is complete, so is $L(\mathcal{X}, \mathcal{Y})$.*

Proof. Let $\{T_n\}$ be a Cauchy sequence in $L(\mathcal{X}, \mathcal{Y})$. If $x \in \mathcal{X}$, then $\{T_n x\}$ is Cauchy in $\mathcal{Y}$ because $\|T_n x - T_m x\| \le \|T_n - T_m\| \, \|x\|$. Define $T : \mathcal{X} \to \mathcal{Y}$ by $Tx = \lim T_n x$. We leave it to the reader (Exercise 3) to verify that $T \in L(\mathcal{X}, \mathcal{Y})$ (in fact, $\|T\| = \lim \|T_n\|$) and that $\|T_n - T\| \to 0$. ∎

Another useful property of the operator norm is the following. If $T \in L(\mathcal{X}, \mathcal{Y})$ and $S \in L(\mathcal{Y}, \mathcal{Z})$, then

$$\|STx\| \le \|S\| \, \|Tx\| \le \|S\| \, \|T\| \, \|x\|,$$

so that $ST \in L(\mathcal{X}, \mathcal{Z})$ and $\|ST\| \le \|S\| \, \|T\|$. In particular, $L(\mathcal{X}, \mathcal{X})$ is an algebra. If $\mathcal{X}$ is complete, $L(\mathcal{X}, \mathcal{X})$ is in fact a **Banach algebra**: a Banach space that is also an algebra, such that the norm of a product is at most the product of the norms. (Another example of a Banach algebra is $BC(X)$, where X is a topological space, with pointwise multiplication and the uniform norm.)

If $T \in L(\mathcal{X}, \mathcal{Y})$, T is said to be **invertible**, or an **isomorphism**, if T is bijective and T^{-1} is bounded (in other words, $\|Tx\| \ge C\|x\|$ for some $C > 0$). T is called an **isometry** if $\|Tx\| = \|x\|$ for all $x \in \mathcal{X}$. An isometry is injective but not necessarily surjective; it is, however, an isomorphism onto its range.

Exercises

1. If $\mathcal{X}$ is a normed vector space over K ($= \mathbb{R}$ or $\mathbb{C}$), then addition and scalar multiplication are continuous from $\mathcal{X} \times \mathcal{X}$ and $K \times \mathcal{X}$ to $\mathcal{X}$. Moreover, the norm is continuous from $\mathcal{X}$ to $[0, \infty)$; in fact, $\big| \, \|x\| - \|y\| \, \big| \le \|x - y\|$.

2. $L(\mathcal{X}, \mathcal{Y})$ is a vector space and the function $\|\cdot\|$ defined by (5.3) is a norm on it. In particular, the three expressions on the right of (5.3) are always equal.

3. Complete the proof of Proposition 5.4.

4. If $\mathcal{X}, \mathcal{Y}$ are normed vector spaces, the map $(T, x) \mapsto Tx$ is continuous from $L(\mathcal{X}, \mathcal{Y}) \times \mathcal{X}$ to $\mathcal{Y}$. (That is, if $T_n \to T$ and $x_n \to x$ then $T_n x_n \to Tx$.)

5. If $\mathcal{X}$ is a normed vector space, the closure of any subspace of $\mathcal{X}$ is a subspace.

6. Suppose that $\mathcal{X}$ is a finite-dimensional vector space. Let $e_1, \ldots, e_n$ be a basis for $\mathcal{X}$, and define $\| \sum_1^n a_j e_j \|_1 = \sum_1^n |a_j|$.

a. $\| \cdot \|_1$ is a norm on $\mathcal{X}$.

b. The map $(a_1, \ldots, a_n) \mapsto \sum_1^n a_j e_j$ is continuous from K^n with the usual Euclidean topology to $\mathcal{X}$ with the topology defined by $\| \cdot \|_1$.

c. $\{x \in \mathcal{X} : \|x\|_1 = 1\}$ is compact in the topology defined by $\| \cdot \|_1$.

d. All norms on $\mathcal{X}$ are equivalent. (Compare any norm to $\| \cdot \|_1$.)

7. Let $\mathcal{X}$ be a Banach space.

a. If $T \in L(\mathcal{X}, \mathcal{X})$ and $\|I - T\| < 1$ where I is the identity operator, then T is invertible; in fact, the series $\sum_0^\infty (I - T)^n$ converges in $L(\mathcal{X}, \mathcal{X})$ to T^{-1}.

b. If $T \in L(\mathcal{X}, \mathcal{X})$ is invertible and $\|S - T\| < \|T^{-1}\|^{-1}$, then S is invertible. Thus the set of invertible operators is open in $L(\mathcal{X}, \mathcal{X})$.

8. Let $(X, \mathcal{M})$ be a measurable space, and let $M(X)$ be the space of complex measures on $(X, \mathcal{M})$. Then $\|\mu\| = |\mu|(X)$ is a norm on $M(X)$ that makes $M(X)$ into a Banach space. (Use Theorem 5.1.)

9. Let $C^k([0,1])$ be the space of functions on $[0,1]$ possessing continuous derivatives up to order k on $[0,1]$, including one-sided derivatives at the endpoints.

a. If $f \in C([0,1])$, then $f \in C^k([0,1])$ iff f is k times continuously differentiable on $(0,1)$ and $\lim_{x \searrow 0} f^{(j)}(x)$ and $\lim_{x \nearrow 1} f^{(j)}(x)$ exist for $j \leq k$. (The mean value theorem is useful.)

b. $\|f\| = \sum_0^k \|f^{(j)}\|_u$ is a norm on $C^k([0,1])$ that makes $C^k([0,1])$ into a Banach space. (Use induction on k. The essential point is that if $\{f_n\} \subset C^1([0,1])$, $f_n \to f$ uniformly, and $f_n' \to g$ uniformly, then $f \in C^1([0,1])$ and $f' = g$. The easy way to prove this is to show that $f(x) - f(0) = \int_0^x g(t)\, dt$.)

10. Let $L_k^1([0,1])$ be the space of all $f \in C^{k-1}([0,1])$ such that $f^{(k-1)}$ is absolutely continuous on $[0,1]$ (and hence $f^{(k)}$ exists a.e. and is in $L^1([0,1])$). Then $\|f\| = \sum_0^k \int_0^1 |f^{(j)}(x)|\, dx$ is a norm on $L_k^1([0,1])$ that makes $L_k^1([0,1])$ into a Banach space. (See Exercise 9 and its hint.)

11. If $0 < \alpha \leq 1$, let $\Lambda_\alpha([0,1])$ be the space of Hölder continuous functions of exponent α on $[0,1]$. That is, $f \in \Lambda_\alpha([0,1])$ iff $\|f\|_{\Lambda_\alpha} < \infty$, where

$$\|f\|_{\Lambda_\alpha} = |f(0)| + \sup_{x,y \in [0,1],\ x \neq y} \frac{|f(x) - f(y)|}{|x - y|^\alpha}.$$

a. $\| \cdot \|_{\Lambda_\alpha}$ is a norm that makes $\Lambda_\alpha([0,1])$ into a Banach space.

b. Let $\lambda_\alpha([0,1])$ be the set of all $f \in \Lambda_\alpha([0,1])$ such that

$$\frac{|f(x)-f(y)|}{|x-y|^\alpha} \to 0 \text{ as } x \to y, \text{ for all } y \in [0,1].$$

If $\alpha < 1$, $\lambda_\alpha([0,1])$ is an infinite-dimensional closed subspace of $\Lambda_\alpha([0,1])$. If $\alpha = 1$, $\lambda_\alpha([0,1])$ contains only constant functions.

12. Let $\mathcal{X}$ be a normed vector space and $\mathcal{M}$ a proper closed subspace of $\mathcal{X}$.

a. $\|x+\mathcal{M}\| = \inf\{\|x+y\| : y \in \mathcal{M}\}$ is a norm on $\mathcal{X}/\mathcal{M}$.

b. For any $\epsilon > 0$ there exists $x \in \mathcal{X}$ such that $\|x\| = 1$ and $\|x+\mathcal{M}\| \geq 1-\epsilon$.

c. The projection map $\pi(x) = x + \mathcal{M}$ from $\mathcal{X}$ to $\mathcal{X}/\mathcal{M}$ has norm 1.

d. If $\mathcal{X}$ is complete, so is $\mathcal{X}/\mathcal{M}$. (Use Theorem 5.1.)

e. The topology defined by the quotient norm is the quotient topology as defined in Exercise 28 in §4.2.

13. If $\|\cdot\|$ is a seminorm on the vector space $\mathcal{X}$, let $\mathcal{M} = \{x \in \mathcal{X} : \|x\| = 0\}$. Then $\mathcal{M}$ is a subspace, and the map $x + \mathcal{M} \mapsto \|x\|$ is a norm on $\mathcal{X}/\mathcal{M}$.

14. If $\mathcal{X}$ is a normed vector space and $\mathcal{M}$ is a nonclosed subspace, then $\|x+\mathcal{M}\|$, as defined in Exercise 12, is a seminorm on $\mathcal{X}/\mathcal{M}$. If one divides by its nullspace as in Exercise 13, the resulting quotient space is isometrically isomorphic to $\mathcal{X}/\overline{\mathcal{M}}$. (Cf. Exercise 5.)

15. Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are normed vector spaces and $T \in L(\mathcal{X},\mathcal{Y})$. Let $\mathcal{N}(T) = \{x \in \mathcal{X} : Tx = 0\}$.

a. $\mathcal{N}(T)$ is a closed subspace of $\mathcal{X}$.

b. There is a unique $S \in L(\mathcal{X}/\mathcal{N}(T), \mathcal{Y})$ such that $T = S \circ \pi$ where $\pi : \mathcal{X} \to \mathcal{X}/\mathcal{M}$ is the projection (see Exercise 12). Moreover, $\|S\| = \|T\|$.

16. The purpose of this exercise is to develop a theory of integration for functions with values in a separable Banach space. Let $(X, \mathcal{M}, \mu)$ be a measure space, $\mathcal{Y}$ a separable Banach space, and $L_\mathcal{Y}$ the space of all $(\mathcal{M}, \mathcal{B}_\mathcal{Y})$-measurable maps from X to $\mathcal{Y}$, and $F_\mathcal{Y}$ the set of maps $f : X \to \mathcal{Y}$ of the form $f(x) = \sum_1^n \chi_{E_j}(x) y_j$ where $n \in \mathbb{N}$, $y_j \in \mathcal{Y}$, $E_j \in \mathcal{M}$, and $\mu(E_j) < \infty$. If $f \in L_\mathcal{Y}$, since $y \mapsto \|y\|$ is continuous (Exercise 1), $x \mapsto \|f(x)\|$ is $(\mathcal{M}, \mathcal{B}_\mathbb{R})$-measurable, and we define $\|f\|_1 = \int \|f(x)\|\, d\mu(x)$. Finally, let $L^1_\mathcal{Y} = \{f \in L_\mathcal{Y} : \|f\|_1 < \infty\}$.

a. $L_\mathcal{Y}$ is a vector space, $F_\mathcal{Y}$ and $L^1_\mathcal{Y}$ are subspaces of it, $F_\mathcal{Y} \subset L^1_\mathcal{Y}$, and $\|\cdot\|_1$ is a seminorm on $L^1_\mathcal{Y}$ that becomes a norm if we identify two functions that are equal a.e.

b. Let $\{y_n\}_1^\infty$ be a countable dense set in $\mathcal{Y}$. Given $\epsilon > 0$, let $B_n^\epsilon = \{y \in \mathcal{Y} : \|y - y_n\| < \epsilon\|y_n\|\}$. Then $\bigcup_1^\infty B_n^\epsilon \supset \mathcal{Y} \setminus \{0\}$.

c. If $f \in L^1_\mathcal{Y}$, there is a sequence $\{h_n\} \subset F_\mathcal{Y}$ with $h_n \to f$ a.e. and $\|h_n - f\|_1 \to 0$. (With notation as in (b), let $A_{nj} = B_n^{1/j} \setminus \bigcup_{m=1}^{n-1} B_m^{1/j}$ and $E_{nj} = f^{-1}(A_{nj})$, and consider $g_j = \sum_{n=1}^\infty y_n \chi_{E_{nj}}$.)

d. There is a unique linear map $\int : L^1_\mathcal{Y} \to \mathcal{Y}$ such that $\int y\chi_E = \mu(E)y$ for $y \in \mathcal{Y}$ and $E \in \mathcal{M}$ ($\mu(E) < \infty$), and $\|\int f\| \leq \|f\|_1$.

e. The dominated convergence theorem: If $\{f_n\}$ is a sequence in $L^1_{\mathcal{Y}}$ such that $f_n \to f$ a.e., and there exists $g \in L^1$ such that $\|f_n(x)\| \le g(x)$ for all n and a.e. x, then $\int f_n \to \int f$.
f. If $\mathcal{Z}$ is a separable Banach space, $T \in L(\mathcal{Y}, \mathcal{Z})$, and $f \in L^1_{\mathcal{Y}}$, then $T \circ f \in L^1_{\mathcal{Z}}$ and $\int T \circ f = T(\int f)$.

5.2 LINEAR FUNCTIONALS

Let $\mathcal{X}$ be a vector space over K, where $K = \mathbb{R}$ or $\mathbb{C}$. A linear map from $\mathcal{X}$ to K is called a **linear functional** on $\mathcal{X}$. If $\mathcal{X}$ is a normed vector space, the space $L(\mathcal{X}, K)$ of bounded linear functionals on $\mathcal{X}$ is called the **dual space** of $\mathcal{X}$ and is denoted by $\mathcal{X}^*$. According to Proposition 5.4, $\mathcal{X}^*$ is a Banach space with the operator norm.

If $\mathcal{X}$ is a vector space over $\mathbb{C}$, it is also a vector space over $\mathbb{R}$, and we can consider both real and complex linear functionals on $\mathcal{X}$, that is, maps $f : \mathcal{X} \to \mathbb{R}$ that are linear over $\mathbb{R}$ and maps $f : \mathcal{X} \to \mathbb{C}$ that are linear over $\mathbb{C}$. The relationship between the two is as follows:

5.5 Proposition. *Let $\mathcal{X}$ be a vector space over $\mathbb{C}$. If f is a complex linear functional on $\mathcal{X}$ and $u = \operatorname{Re} f$, then u is a real linear functional, and $f(x) = u(x) - iu(ix)$ for all $x \in \mathcal{X}$. Conversely, if u is a real linear functional on $\mathcal{X}$ and $f : \mathcal{X} \to \mathbb{C}$ is defined by $f(x) = u(x) - iu(ix)$, then f is complex linear. In this case, if $\mathcal{X}$ is normed, we have $\|u\| = \|f\|$.*

Proof. If f is complex linear and $u = \operatorname{Re} f$, u is clearly real linear and $\operatorname{Im} f(x) = -\operatorname{Re}[if(x)] = -u(ix)$, so $f(x) = u(x) - iu(ix)$. On the other hand, if u is real linear and $f(x) = u(x) - iu(ix)$, then f is clearly linear over $\mathbb{R}$, and $f(ix) = u(ix) - iu(-x) = u(ix) + iu(x) = if(x)$, so f is also linear over $\mathbb{C}$. Finally, if $\mathcal{X}$ is normed, since $|u(x)| = |\operatorname{Re} f(x)| \le |f(x)|$ we have $\|u\| \le \|f\|$. On the other hand, if $f(x) \ne 0$, let $\alpha = \overline{\operatorname{sgn} f(x)}$. Then $|f(x)| = \alpha f(x) = f(\alpha x) = u(\alpha x)$ (since $f(\alpha x)$ is real), so $|f(x)| \le \|u\|\,\|\alpha x\| = \|u\|\,\|x\|$, whence $\|f\| \le \|u\|$. ∎

It is not obvious that there are any nonzero bounded linear functionals on an arbitrary normed vector space. The fact that such functionals exist in great abundance is one of the fundamental theorems of functional analysis. We shall now present this result in a more general form that has other important applications.

If $\mathcal{X}$ is a real vector space, a **sublinear functional** on $\mathcal{X}$ is a map $p : \mathcal{X} \to \mathbb{R}$ such that

$$p(x + y) \le p(x) + p(y) \text{ and } p(\lambda x) = \lambda p(x) \text{ for all } x, y \in \mathcal{X} \text{ and } \lambda \ge 0.$$

For example, every seminorm is a sublinear functional.

5.6 The Hahn-Banach Theorem. *Let $\mathcal{X}$ be a real vector space, p a sublinear functional on $\mathcal{X}$, $\mathcal{M}$ a subspace of $\mathcal{X}$, and f a linear functional on $\mathcal{M}$ such that $f(x) \le p(x)$ for all $x \in \mathcal{M}$. Then there exists a linear functional F on $\mathcal{X}$ such that $F(x) \le p(x)$ for all $x \in \mathcal{X}$ and $F|\mathcal{M} = f$.*

Proof. We begin by showing that if $x \in \mathcal{X} \setminus \mathcal{M}$, f can be extended to a linear functional g on $\mathcal{M} + \mathbb{R}x$ satisfying $g(y) \le p(y)$ there. If $y_1, y_2 \in \mathcal{M}$, we have

$$f(y_1) + f(y_2) = f(y_1 + y_2) \le p(y_1 + y_2) \le p(y_1 - x) + p(x + y_2),$$

or

$$f(y_1) - p(y_1 - x) \le p(x + y_2) - f(y_2).$$

Hence

$$\sup\{f(y) - p(y - x) : y \in \mathcal{M}\} \le \inf\{p(x + y) - f(y) : y \in \mathcal{M}\}.$$

Let α be any number satisfying

$$\sup\{f(y) - p(y - x) : y \in \mathcal{M}\} \le \alpha \le \inf\{p(x + y) - f(y) : y \in \mathcal{M}\}$$

and define $g : \mathcal{M} + \mathbb{R}x \to \mathbb{R}$ by $g(y + \lambda x) = f(y) + \lambda\alpha$. Then g is clearly linear, and $g|\mathcal{M} = f$, so that $g(y) \le p(y)$ for $y \in \mathcal{M}$. Moreover, if $\lambda > 0$ and $y \in \mathcal{M}$,

$$g(y + \lambda x) = \lambda[f(y/\lambda) + \alpha] \le \lambda[f(y/\lambda) + p(x + (y/\lambda)) - f(y/\lambda)] = p(y + \lambda x),$$

whereas if $\lambda = -\mu < 0$,

$$g(y + \lambda x) = \mu[f(y/\mu) - \alpha] \le \mu[f(y/\mu) - f(y/\mu) + p((y/\mu) - x)] = p(y + \lambda x).$$

Thus $g(z) \le p(z)$ for all $z \in \mathcal{M} + \mathbb{R}x$.

Evidently the same reasoning can be applied to any linear extension F of f satisfying $F \le p$ on its domain, and it shows that the domain of a maximal linear extension satisfying $F \le p$ must be the whole space $\mathcal{X}$. But the family $\mathcal{F}$ of all linear extensions F of f satisfying $F \le p$ is partially ordered by inclusion (maps from subspaces of $\mathcal{X}$ to $\mathbb{R}$ being regarded as subsets of $\mathcal{X} \times \mathbb{R}$). Since the union of any increasing family of subspaces of $\mathcal{X}$ is again a subspace, one easily sees that the union of a linearly ordered subfamily of $\mathcal{F}$ lies in $\mathcal{F}$. The proof is therefore completed by invoking Zorn's lemma. ∎

If p is a seminorm and $f : \mathcal{X} \to \mathbb{R}$ is linear, the inequality $f \le p$ is equivalent to the inequality $|f| \le p$, because $|f(x)| = \pm f(x) = f(\pm x)$ and $p(-x) = p(x)$. In this situation the Hahn-Banach theorem also applies to complex linear functionals:

5.7 The Complex Hahn-Banach Theorem. *Let $\mathcal{X}$ be a complex vector space, p a seminorm on $\mathcal{X}$, $\mathcal{M}$ a subspace of $\mathcal{X}$, and f a complex linear functional on $\mathcal{M}$ such that $|f(x)| \le p(x)$ for $x \in \mathcal{M}$. Then there exists a complex linear functional F on $\mathcal{X}$ such that $|F(x)| \le p(x)$ for all $x \in \mathcal{X}$ and $F|\mathcal{M} = f$.*

Proof. Let $u = \operatorname{Re} f$. By Theorem 5.6 there is a real linear extension U of u to $\mathcal{X}$ such that $|U(x)| \le p(x)$ for all $x \in \mathcal{X}$. Let $F(x) = U(x) - iU(ix)$ as in Proposition 5.5. Then F is a complex linear extension of f, and as in the proof of Proposition 5.5, if $\alpha = \overline{\operatorname{sgn} F(x)}$ we have $|F(x)| = \alpha F(x) = F(\alpha x) = U(\alpha x) \le p(\alpha x) = p(x)$. ∎

From now on until §5.5, all of our results apply equally to real or complex vector spaces, but for the sake of definiteness we shall assume that the scalar field is $\mathbb{C}$. The principal applications of the Hahn-Banach theorem to normed vector spaces are summarized in the following theorem.

5.8 Theorem. *Let $\mathcal{X}$ be a normed vector space.*

a. If $\mathcal{M}$ is a closed subspace of $\mathcal{X}$ and $x \in \mathcal{X} \setminus \mathcal{M}$, there exists $f \in \mathcal{X}^*$ such that $f(x) \neq 0$ and $f|\mathcal{M} = 0$. In fact, if $\delta = \inf_{y \in \mathcal{M}} \|x - y\|$, f can be taken to satisfy $\|f\| = 1$ and $f(x) = \delta$.

b. If $x \neq 0 \in \mathcal{X}$, there exists $f \in \mathcal{X}^*$ such that $\|f\| = 1$ and $f(x) = \|x\|$.

c. The bounded linear functionals on $\mathcal{X}$ separate points.

d. If $x \in \mathcal{X}$, define $\widehat{x} : \mathcal{X}^* \to \mathbb{C}$ by $\widehat{x}(f) = f(x)$. Then the map $x \mapsto \widehat{x}$ is a linear isometry from $\mathcal{X}$ into $\mathcal{X}^{**}$ (the dual of $\mathcal{X}^*$).

Proof. To prove (a), define f on $\mathcal{M} + \mathbb{C}x$ by $f(y + \lambda x) = \lambda\delta$ ($y \in \mathcal{M}$, $\lambda \in \mathbb{C}$). Then $f(x) = \delta$, $f|\mathcal{M} = 0$, and for $\lambda \neq 0$, $|f(y + \lambda x)| = |\lambda|\delta \leq |\lambda| \, \|\lambda^{-1}y + x\| = \|y + \lambda x\|$. Thus the Hahn-Banach theorem can be applied, with $p(x) = \|x\|$ and $\mathcal{M}$ replaced by $\mathcal{M} + \mathbb{C}x$. (b) is the special case of (a) with $\mathcal{M} = \{0\}$, and (c) follows immediately: if $x \neq y$, there exists $f \in \mathcal{X}^*$ with $f(x - y) \neq 0$, i.e., $f(x) \neq f(y)$. As for (d), obviously $\widehat{x}$ is a linear functional on $\mathcal{X}^*$ and the map $x \mapsto \widehat{x}$ is linear. Moroever, $|\widehat{x}(f)| = |f(x)| \leq \|f\| \, \|x\|$, so $\|\widehat{x}\| \leq \|x\|$. On the other hand, (b) implies that $\|\widehat{x}\| \geq \|x\|$. ∎

With notation as in Theorem 5.8d, let $\widehat{\mathcal{X}} = \{\widehat{x} : x \in \mathcal{X}\}$. Since $\mathcal{X}^{**}$ is always complete, the closure $\overline{\widehat{\mathcal{X}}}$ of $\widehat{\mathcal{X}}$ in $\mathcal{X}^{**}$ is a Banach space, and the map $x \mapsto \widehat{x}$ embeds $\mathcal{X}$ into $\overline{\widehat{\mathcal{X}}}$ as a dense subspace. $\overline{\widehat{\mathcal{X}}}$ is called the **completion** of $\mathcal{X}$. In particular, if $\mathcal{X}$ is itself a Banach space then $\overline{\widehat{\mathcal{X}}} = \widehat{\mathcal{X}}$.

If $\mathcal{X}$ is finite-dimensional, then of course $\widehat{\mathcal{X}} = \mathcal{X}^{**}$, since these spaces have the same dimension. For infinite-dimensional Banach spaces it may or may not happen that $\widehat{\mathcal{X}} = \mathcal{X}^{**}$; if it does, $\mathcal{X}$ is called **reflexive**. The examples of Banach spaces we have examined so far are not reflexive except in trivial cases where they turn out to be finite-dimensional. We shall prove some cases of this assertion and present examples of reflexive Banach spaces in later sections.

Usually we shall identify $\widehat{x}$ with x and thus regard $\mathcal{X}^{**}$ as a superspace of $\mathcal{X}$; reflexivity then means that $\mathcal{X}^{**} = \mathcal{X}$.

Exercises

17. A linear functional f on a normed vector space $\mathcal{X}$ is bounded iff $f^{-1}(\{0\})$ is closed. (Use Exercise 12b.)

18. Let $\mathcal{X}$ be a normed vector space.

a. If $\mathcal{M}$ is a closed subspace and $x \in \mathcal{X} \setminus \mathcal{M}$ then $\mathcal{M} + \mathbb{C}x$ is closed. (Use Theorem 5.8a.)

b. Every finite-dimensional subspace of $\mathcal{X}$ is closed.

19. Let $\mathcal{X}$ be an infinite-dimensional normed vector space.

a. There is a sequence $\{x_j\}$ in $\mathcal{X}$ such that $\|x_j\| = 1$ for all j and $\|x_j - x_k\| \geq \frac{1}{2}$ for $j \neq k$. (Construct x_j inductively, using Exercises 12b and 18.)

b. $\mathcal{X}$ is not locally compact.

20. If $\mathcal{M}$ is a finite-dimensional subspace of a normed vector space $\mathcal{X}$, there is a closed subspace $\mathcal{N}$ such that $\mathcal{M} \cap \mathcal{N} = \{0\}$ and $\mathcal{M} + \mathcal{N} = \mathcal{X}$.

21. If $\mathcal{X}$ and $\mathcal{Y}$ are normed vector spaces, define $\alpha : \mathcal{X}^* \times \mathcal{Y}^* \to (\mathcal{X} \times \mathcal{Y})^*$ by $\alpha(f,g)(x,y) = f(x) + g(y)$. Then α is an isomorphism which is isometric if we use the norm $\|(x,y)\| = \max(\|x\|, \|y\|)$ on $\mathcal{X} \times \mathcal{Y}$, the corresponding operator norm on $(\mathcal{X} \times \mathcal{Y})^*$, and the norm $\|(f,g)\| = \|f\| + \|g\|$ on $\mathcal{X}^* \times \mathcal{Y}^*$.

22. Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are normed vector spaces and $T \in L(\mathcal{X}, \mathcal{Y})$.

a. Define $T^\dagger : \mathcal{Y}^* \to \mathcal{X}^*$ by $T^\dagger f = f \circ T$. Then $T^\dagger \in L(\mathcal{Y}^*, \mathcal{X}^*)$ and $\|T^\dagger\| = \|T\|$. $T^\dagger$ is called the **adjoint** or **transpose** of T.

b. Applying the construction in (a) twice, one obtains $T^{\dagger\dagger} \in L(\mathcal{X}^{**}, \mathcal{Y}^{**})$. If $\mathcal{X}$ and $\mathcal{Y}$ are identified with their natural images $\widehat{\mathcal{X}}$ and $\widehat{\mathcal{Y}}$ in $\mathcal{X}^{**}$ and $\mathcal{Y}^{**}$, then $T^{\dagger\dagger}|\mathcal{X} = T$.

c. $T^\dagger$ is injective iff the range of T is dense in $\mathcal{Y}$.

d. If the range of $T^\dagger$ is dense in $\mathcal{X}^*$, then T is injective; the converse is true if $\mathcal{X}$ is reflexive.

23. Suppose that $\mathcal{X}$ is a Banach space. If $\mathcal{M}$ is a closed subspace of $\mathcal{X}$ and $\mathcal{N}$ is a closed subspace of $\mathcal{X}^*$, let $\mathcal{M}^0 = \{f \in \mathcal{X}^* : f|\mathcal{M} = 0\}$ and $\mathcal{N}^\perp = \{x \in \mathcal{X} : f(x) = 0$ for all $f \in \mathcal{N}\}$. (Thus, if we identify $\mathcal{X}$ with its image in $\mathcal{X}^{**}$, $\mathcal{N}^\perp = \mathcal{N}^0 \cap \mathcal{X}$.)

a. $\mathcal{M}^0$ and $\mathcal{N}^\perp$ are closed subspaces of $\mathcal{X}^*$ and $\mathcal{X}$, respectively.

b. $(\mathcal{M}^0)^\perp = \mathcal{M}$ and $(\mathcal{N}^\perp)^0 \supset \mathcal{N}$. If $\mathcal{X}$ is reflexive, $(\mathcal{N}^\perp)^0 = \mathcal{N}$.

c. Let $\pi : \mathcal{X} \to \mathcal{X}/\mathcal{M}$ be the natural projection, and define $\alpha : (\mathcal{X}/\mathcal{M})^* \to \mathcal{X}^*$ by $\alpha(f) = f \circ \pi$. Then α is an isometric isomorphism from $(\mathcal{X}/\mathcal{M})^*$ onto $\mathcal{M}^0$, where $\mathcal{X}/\mathcal{M}$ has the quotient norm.

d. Define $\beta : \mathcal{X}^* \to \mathcal{M}^*$ by $\beta(f) = f|\mathcal{M}$; then β induces a map $\overline{\beta} : \mathcal{X}^*/\mathcal{M}^0 \to \mathcal{M}^*$ as in Exercise 15, and $\overline{\beta}$ is an isometric isomorphism.

24. Suppose that $\mathcal{X}$ is a Banach space.

a. Let $\widehat{\mathcal{X}}$, $(\mathcal{X}^*)\widehat{\ }$ be the natural images of $\mathcal{X}$, $\mathcal{X}^*$ in $\mathcal{X}^{**}$, $\mathcal{X}^{***}$, and let $\widehat{\mathcal{X}}^0 = \{F \in \mathcal{X}^{***} : F|\widehat{\mathcal{X}} = 0\}$. Then $(\mathcal{X}^*)\widehat{\ } \cap \widehat{\mathcal{X}}^0 = \{0\}$ and $(\mathcal{X}^*)\widehat{\ } + \widehat{\mathcal{X}}^0 = \mathcal{X}^{***}$.

b. $\mathcal{X}$ is reflexive iff $\mathcal{X}^*$ is reflexive.

25. If $\mathcal{X}$ is a Banach space and $\mathcal{X}^*$ is separable, then $\mathcal{X}$ is separable. (Let $\{f_n\}_1^\infty$ be a countable dense subset of $\mathcal{X}^*$. For each n choose $x_n \in \mathcal{X}$ with $\|x_n\| = 1$ and $|f_n(x_n)| \geq \frac{1}{2}\|f_n\|$. Then the linear combinations of $\{x_n\}_1^\infty$ are dense in $\mathcal{X}$.) *Note:* Separability of $\mathcal{X}$ does not imply separability of $\mathcal{X}^*$.

26. Let $\mathcal{X}$ be a real vector space and let P be a subset of $\mathcal{X}$ such that (i) if $x, y \in P$, then $x + y \in P$, (ii) if $x \in P$ and $\lambda \geq 0$, then $\lambda x \in P$, (iii) if $x \in P$ and $-x \in P$,

then $x = 0$. (Example: If $\mathcal{X}$ is a space of real-valued functions, P can be the set of nonnegative functions in $\mathcal{X}$.)

a. The relation $\leq$ defined by $x \leq y$ iff $y - x \in P$ is a partial ordering on $\mathcal{X}$.

b. (**The Krein Extension Theorem**) Suppose that $\mathcal{M}$ is a subspace of $\mathcal{X}$ such that for each $x \in \mathcal{X}$ there exists $y \in \mathcal{M}$ with $x \leq y$. If f is a linear functional on $\mathcal{M}$ such that $f(x) \geq 0$ for $x \in \mathcal{M} \cap P$, there is a linear functional F on $\mathcal{X}$ such that $F(x) \geq 0$ for $x \in P$ and $F|\mathcal{M} = f$. (Consider $p(x) = \inf\{f(y) : y \in \mathcal{M} \text{ and } x \leq y\}$.)

5.3 THE BAIRE CATEGORY THEOREM AND ITS CONSEQUENCES

In this section we present an important theorem about complete metric spaces and use it to obtain some fundamental results concerning linear maps between Banach spaces.

5.9 The Baire Category Theorem. *Let X be a complete metric space.*

a. If $\{U_n\}_1^\infty$ is a sequence of open dense subsets of X, then $\bigcap_1^\infty U_n$ is dense in X.

b. X is not a countable union of nowhere dense sets.

Proof. For part (a), we must show that if W is a nonempty open set in X, then W intersects $\bigcap_1^\infty U_n$. Since $U_1 \cap W$ is open and nonempty, it contains a ball $B(r_0, x_0)$, and we can assume that $0 < r_0 < 1$. For $n > 0$, we choose $x_n \in X$ and $r_n \in (0, \infty)$ inductively as follows: Having chosen x_j and r_j for $j < n$, we observe that $U_n \cap B(r_{n-1}, x_{n-1})$ is open and nonempty, so we can choose x_n, r_n so that $0 < r_n < 2^{-n}$ and $\overline{B(r_n, x_n)} \subset U_n \cap B(r_{n-1}, x_{n-1})$. Then if $n, m \geq N$, we see that $x_n, x_m \in B(r_N, x_N)$, and since $r_n \to 0$, the sequence $\{x_n\}$ is Cauchy. As X is complete, $x = \lim x_n$ exists. Since $x_n \in B(r_N, x_N)$ for $n \geq N$ we have

$$x \in \overline{B(r_N, x_N)} \subset U_N \cap B(r_1, x_1) \subset U_N \cap W$$

for all N, and the proof is complete.

As for (b), if $\{E_n\}$ is a sequence of nowhere dense sets in X, then $\{(\overline{E}_n)^c\}$ is a sequence of open dense sets. Since $\bigcap(\overline{E}_n)^c \neq \varnothing$, we have $\bigcup E_n \subset \bigcup \overline{E}_n \neq X$. ∎

We remark that since the conclusions of the Baire category theorem are purely topological, it suffices for X to be homeomorphic to a complete metric space. For example, the theorem applies to $X = (0, 1)$, which is not complete with the usual metric but is homeomorphic to $\mathbb{R}$.

The name of this theorem comes from Baire's terminology for sets: If X is a topological space, a set $E \subset X$ is **of the first category**, according to Baire, if E is a countable union of nowhere dense sets; otherwise E is **of the second category**. Thus Baire's theorem asserts that every complete metric space is of the second category in itself. A more modern and more descriptive synonym for "of the first category" is **meager**. The complement of a meager set is called **residual**.

The Baire category theorem is often used to prove existence results: One shows that objects having a certain property exist by showing that the set of objects *not* having the property (within a suitable complete metric space) is meager. For example, one can prove the existence of nowhere differentiable continuous functions in this way; see Exercise 42.

We turn to the applications of the Baire category theorem in the theory of linear maps. Some terminology: If X and Y are topological spaces, a map $f : X \to Y$ is called **open** if $f(U)$ is open in Y whenever U is open in X. If X and Y are metric spaces, this amounts to requiring that if B is a ball centered at $x \in X$, then $f(B)$ contains a ball centered at $f(x)$. Specializing still further, if X and Y are normed linear spaces and f is linear, then f commutes with translations and dilations; it follows that f is open iff $f(B)$ contains a ball centered at 0 in Y when B is the ball of radius 1 about 0 in X.

5.10 The Open Mapping Theorem. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces. If $T \in L(\mathcal{X}, \mathcal{Y})$ is surjective, then T is open.*

Proof. Let B_r denote the (open) ball of radius r about 0 in $\mathcal{X}$. By the preceding remarks, it will suffice to show that $T(B_1)$ contains a ball about 0 in $\mathcal{Y}$. Since $\mathcal{X} = \bigcup_1^\infty B_n$ and T is surjective, we have $\mathcal{Y} = \bigcup_1^\infty T(B_n)$. But $\mathcal{Y}$ is complete and the map $y \mapsto ny$ is a homeomorphism of $\mathcal{Y}$ that maps $T(B_1)$ to $T(B_n)$, so Baire's theorem implies that $T(B_1)$ cannot be nowhere dense. That is, there exist $y_0 \in \mathcal{Y}$ and $r > 0$ such that the ball $B(4r, y_0)$ is contained in $\overline{T(B_1)}$. Pick $y_1 = Tx_1 \in T(B_1)$ such that $\|y_1 - y_0\| < 2r$; then $B(2r, y_1) \subset B(4r, y_0) \subset \overline{T(B_1)}$, so if $\|y\| < 2r$,

$$y = Tx_1 + (y - y_1) \in \overline{T(x_1 + B_1)} \subset \overline{T(B_2)}.$$

Dividing both sides by 2, we conclude that there exists $r > 0$ such that if $\|y\| < r$ then $y \in \overline{T(B_1)}$. If we could replace $\overline{T(B_1)}$ by $T(B_1)$, perhaps shrinking r at the same time, the proof would be complete; we now proceed to accomplish this.

Since T commutes with dilations, it follows that if $\|y\| < r2^{-n}$, then $y \in \overline{T(B_{2^{-n}})}$. Suppose $\|y\| < r/2$; we can find $x_1 \in B_{1/2}$ such that $\|y - Tx_1\| < r/4$, and proceeding inductively, we can find $x_n \in B_{2^{-n}}$ such that $\|y - \sum_1^n Tx_j\| < r2^{-n-1}$. Since $\mathcal{X}$ is complete, by Theorem 5.1 the series $\sum_1^\infty x_n$ converges, say to x. But then $\|x\| < \sum_1^\infty 2^{-n} = 1$ and $y = Tx$. In other words, $T(B_1)$ contains all y with $\|y\| < r/2$, so we are done. ∎

5.11 Corollary. *If $\mathcal{X}$ and $\mathcal{Y}$ are Banach spaces and $T \in L(\mathcal{X}, \mathcal{Y})$ is bijective, then T is an isomorphism; that is, $T^{-1} \in L(\mathcal{Y}, \mathcal{X})$.*

Proof. If T is bijective, continuity of T^{-1} is equivalent to the openness of T. ∎

For the next results we need some more terminology. If $\mathcal{X}$ and $\mathcal{Y}$ are normed vector spaces and T is a linear map from $\mathcal{X}$ to $\mathcal{Y}$, we define the **graph** of T to be

$$\Gamma(T) = \{(x, y) \in \mathcal{X} \times \mathcal{Y} : y = Tx\},$$

which is a subspace of $\mathcal{X} \times \mathcal{Y}$. (From a strict set-theoretic point of view, of course, T and $\Gamma(T)$ are identical; the distinction is a psychological one.) We say that T is **closed** if $\Gamma(T)$ is a closed subspace of $\mathcal{X} \times \mathcal{Y}$. Clearly, if T is continuous, then T is closed, and if $\mathcal{X}$ and $\mathcal{Y}$ are complete the converse is also true:

5.12 The Closed Graph Theorem. *If $\mathcal{X}$ and $\mathcal{Y}$ are Banach spaces and $T : \mathcal{X} \to \mathcal{Y}$ is a closed linear map, then T is bounded.*

Proof. Let π_1 and π_2 be the projections of $\Gamma(T)$ onto $\mathcal{X}$ and $\mathcal{Y}$, that is, $\pi_1(x, Tx) = x$ and $\pi_2(x, Tx) = Tx$. Obviously $\pi_1 \in L(\Gamma(T), \mathcal{X})$ and $\pi_2 \in L(\Gamma(T), \mathcal{Y})$. Since $\mathcal{X}$ and $\mathcal{Y}$ are complete, so is $\mathcal{X} \times \mathcal{Y}$, and hence so is $\Gamma(T)$ since T is closed. The map π_1 is a bijection from $\Gamma(T)$ to $\mathcal{X}$, so by Corollary 5.11, π_1^{-1} is bounded. But then $T = \pi_2 \circ \pi_1^{-1}$ is bounded. ∎

Continuity of a linear map $T : \mathcal{X} \to \mathcal{Y}$ means that if $x_n \to x$ then $Tx_n \to Tx$, whereas closedness means that if $x_n \to x$ and $Tx_n \to y$ then $y = Tx$. Thus the significance of the closed graph theorem is that in verifying that $Tx_n \to Tx$ when $x_n \to x$, we may assume that Tx_n converges to *something,* and we need only to show that the limit is the right thing. This frequently saves a lot of trouble.

The completeness of $\mathcal{X}$ and $\mathcal{Y}$ was used in a crucial way in proving the open mapping theorem and hence also in proving the closed graph theorem. In fact, the conclusions of both of these theorems may fail if either $\mathcal{X}$ or $\mathcal{Y}$ is incomplete; see Exercises 29–31.

Our final result in this section is a theorem of almost magical power that allows one to deduce uniform estimates from pointwise estimates in certain situations.

5.13 The Uniform Boundedness Principle. *Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are normed vector spaces and $\mathcal{A}$ is a subset of $L(\mathcal{X}, \mathcal{Y})$.*

a. If $\sup_{T \in \mathcal{A}} \|Tx\| < \infty$ for all x in some nonmeager subset of $\mathcal{X}$, then $\sup_{T \in \mathcal{A}} \|T\| < \infty$.

b. If $\mathcal{X}$ is a Banach space and $\sup_{T \in \mathcal{A}} \|Tx\| < \infty$ for all $x \in \mathcal{X}$, then $\sup_{T \in \mathcal{A}} \|T\| < \infty$.

Proof. Let

$$E_n = \{x \in \mathcal{X} : \sup_{T \in \mathcal{A}} \|Tx\| \le n\} = \bigcap_{T \in \mathcal{A}} \{x \in \mathcal{X} : \|Tx\| \le n\}.$$

Then the E_n's are closed, so under the hypothesis of (a) some E_n must contain a nontrivial closed ball $\overline{B(r, x_0)}$. But then $E_{2n} \supset \overline{B(r, 0)}$, for if $\|x\| \le r$, then $x - x_0 \in E_n$ and hence

$$\|Tx\| \le \|T(x - x_0)\| + \|Tx_0\| \le 2n.$$

In other words, $\|Tx\| \le 2n$ whenever $T \in \mathcal{A}$ and $\|x\| \le r$, so $\sup_{T \in \mathcal{A}} \|T\| \le 2n/r$. This proves (a), and (b) follows by the Baire category theorem. ∎

Exercises

27. There exist meager subsets of $\mathbb{R}$ whose complements have Lebesgue measure zero.

28. The Baire category theorem remains true if X is assumed to be an LCH space rather than a complete metric space. (The proof is similar; the substitute for completeness is Proposition 4.21.)

29. Let $\mathcal{Y} = L^1(\mu)$ where μ is counting measure on $\mathbb{N}$, and let $\mathcal{X} = \{f \in \mathcal{Y} : \sum_1^\infty n|f(n)| < \infty\}$, equipped with the L^1 norm.

a. $\mathcal{X}$ is a proper dense subspace of $\mathcal{Y}$; hence $\mathcal{X}$ is not complete.

b. Define $T : \mathcal{X} \to \mathcal{Y}$ by $Tf(n) = nf(n)$. Then T is closed but not bounded.

c. Let $S = T^{-1}$. Then $S : \mathcal{Y} \to \mathcal{X}$ is bounded and surjective but not open.

30. Let $\mathcal{Y} = C([0,1])$ and $\mathcal{X} = C^1([0,1])$, both equipped with the uniform norm.

a. $\mathcal{X}$ is not complete.

b. The map $(d/dx) : \mathcal{X} \to \mathcal{Y}$ is closed (see Exercise 9) but not bounded.

31. Let $\mathcal{X}, \mathcal{Y}$ be Banach spaces and let $S : \mathcal{X} \to \mathcal{Y}$ be an unbounded linear map (for the existence of which, see §5.6). Let $\Gamma(S)$ be the graph of S, a subspace of $\mathcal{X} \times \mathcal{Y}$.

a. $\Gamma(S)$ is not complete.

b. Define $T : \mathcal{X} \to \Gamma(S)$ by $Tx = (x, Sx)$. Then T is closed but not bounded.

c. $T^{-1} : \Gamma(S) \to \mathcal{X}$ is bounded and surjective but not open.

32. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be norms on the vector space $\mathcal{X}$ such that $\|\cdot\|_1 \le \|\cdot\|_2$. If $\mathcal{X}$ is complete with respect to both norms, then the norms are equivalent.

33. There is no slowest rate of decay of the terms of an absolutely convergent series; that is, there is no sequence $\{a_n\}$ of positive numbers such that $\sum a_n|c_n| < \infty$ iff $\{c_n\}$ is bounded. (The set of bounded sequences is the space $B(\mathbb{N})$ of bounded functions on $\mathbb{N}$, and the set of absolutely summable sequences is $L^1(\mu)$ where μ is counting measure on $\mathbb{N}$. If such an $\{a_n\}$ exists, consider $T : B(\mathbb{N}) \to L^1(\mu)$ defined by $Tf(n) = a_n f(n)$. The set of f such that $f(n) = 0$ for all but finitely many n is dense in $L^1(\mu)$ but not in $B(\mathbb{N})$.)

34. With reference to Exercises 9 and 10, show that the inclusion map of $L^1_k([0,1])$ into $C^{k-1}([0,1])$ is continuous (a) by using the closed graph theorem, and (b) by direct calculation. (This is to illustrate the use of the closed graph theorem as a labor-saving device.)

35. Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces, $T \in L(\mathcal{X}, \mathcal{Y})$, $\mathcal{N}(T) = \{x : Tx = 0\}$, and $\mathcal{M} = \text{range}(T)$. Then $\mathcal{X}/\mathcal{N}(T)$ is isomorphic to $\mathcal{M}$ iff $\mathcal{M}$ is closed. (See Exercise 15.)

36. Let $\mathcal{X}$ be a separable Banach space and let μ be counting measure on $\mathbb{N}$. Suppose that $\{x_n\}_1^\infty$ is a countable dense subset of the unit ball of $\mathcal{X}$, and define $T : L^1(\mu) \to \mathcal{X}$ by $Tf = \sum_1^\infty f(n)x_n$.

a. T is bounded.

b. T is surjective.

c. $\mathcal{X}$ is isomorphic to a quotient space of $L^1(\mu)$. (Use Exercise 35.)

37. Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces. If $T : \mathcal{X} \to \mathcal{Y}$ is a linear map such that $f \circ T \in \mathcal{X}^*$ for every $f \in \mathcal{Y}^*$, then T is bounded.

38. Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces, and let $\{T_n\}$ be a sequence in $L(\mathcal{X}, \mathcal{Y})$ such that $\lim T_n x$ exists for every $x \in X$. Let $Tx = \lim T_n x$; then $T \in L(\mathcal{X}, \mathcal{Y})$.

39. Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be Banach spaces and let $B : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a separately continuous bilinear map; that is, $B(x, \cdot) \in L(\mathcal{Y}, \mathcal{Z})$ for each $x \in \mathcal{X}$ and $B(\cdot, y) \in L(\mathcal{X}, \mathcal{Z})$ for each $y \in \mathcal{Y}$. Then B is jointly continuous, that is, continuous from $\mathcal{X} \times \mathcal{Y}$ to $\mathcal{Z}$. (Reduce the problem to proving that $\|B(x, y)\| \leq C\|x\| \, \|y\|$ for some $C > 0$.)

40. (**The Principle of Condensation of Singularities**) Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and $\{T_{jk} : j, k \in \mathbb{N}\} \subset L(\mathcal{X}, \mathcal{Y})$. Suppose that for each k there exists $x \in \mathcal{X}$ such that $\sup\{\|T_{jk}x\| : j \in \mathbb{N}\} = \infty$. Then there is an x (indeed, a residual set of x's) such that $\sup\{\|T_{jk}x\| : j \in \mathbb{N}\} = \infty$ for all k.

41. Let $\mathcal{X}$ be a vector space of countably infinite dimension (that is, every element is a *finite* linear combination of members of a countably infinite linearly independent set). There is no norm on $\mathcal{X}$ with respect to which $\mathcal{X}$ is complete. (Given a norm on $\mathcal{X}$, apply Exercise 18b and the Baire category theorem.)

42. Let E_n be the set of all $f \in C([0, 1])$ for which there exists $x_0 \in [0, 1]$ (depending on f) such that $|f(x) - f(x_0)| \leq n|x - x_0|$ for all $x \in [0, 1]$.

a. E_n is nowhere dense in $C([0, 1])$. (Any real $f \in C([0, 1])$ can be uniformly approximated by a piecewise linear function g whose linear pieces, finite in number, have slope $\pm 2n$. If $\|h - g\|_u$ is sufficiently small, then $h \notin E_n$.)

b. The set of nowhere differentiable functions is residual in $C([0, 1])$.

5.4 TOPOLOGICAL VECTOR SPACES

It is frequently useful to consider topologies on vector spaces other than those defined by norms, the only crucial requirement being that the topology should be well behaved with respect to the vector operations. Precisely, a **topological vector space** is a vector space $\mathcal{X}$ over the field K ($= \mathbb{R}$ or $\mathbb{C}$) which is endowed with a topology such that the maps $(x, y) \to x + y$ and $(\lambda, x) \to \lambda x$ are continuous from $\mathcal{X} \times \mathcal{X}$ and $K \times \mathcal{X}$ to $\mathcal{X}$. A topological vector space is called **locally convex** if there is a base for the topology consisting of convex sets (that is, sets A such that if $x, y \in A$ then $tx + (1 - t)y \in A$ for $0 < t < 1$). Most topological vector spaces that arise in practice are locally convex and Hausdorff.

The most common way of defining locally convex topologies on vector spaces is in terms of seminorms. Namely, if we are given a family of seminorms on $\mathcal{X}$, the "balls" that they define can be used to generate a topology in the same way that the

balls defined by a norm generate the topology on a normed vector space. The precise result is as follows:

5.14 Theorem. *Let $\{p_\alpha\}_{\alpha\in A}$ be a family of seminorms on the vector space $\mathcal{X}$. If $x \in \mathcal{X}$, $\alpha \in A$, and $\epsilon > 0$, let*

$$U_{x\alpha\epsilon} = \{y \in \mathcal{X} : p_\alpha(y - x) < \epsilon\},$$

and let $\mathcal{T}$ be the topology generated by the sets $U_{x\alpha\epsilon}$.

a. For each $x \in \mathcal{X}$, the finite intersections of the sets $U_{x\alpha\epsilon}$ ($\alpha \in A$, $\epsilon > 0$) form a neighborhood base at x.

b. If $\langle x_i\rangle_{i\in I}$ is a net in $\mathcal{X}$, then $x_i \to x$ iff $p_\alpha(x_i - x) \to 0$ for all $\alpha \in A$.

c. $(\mathcal{X}, \mathcal{T})$ is a locally convex topological vector space.

Proof. (a) If $x \in \bigcap_1^k U_{x_j\alpha_j\epsilon_j}$, let $\delta_j = \epsilon_j - p_\alpha(x - x_j)$. By the triangle inequality, we have $x \in \bigcap_1^k U_{x\alpha_j\delta_j} \subset \bigcap_1^k U_{x_j\alpha_j\epsilon_j}$. Thus the assertion follows from Proposition 4.4.

(b) In view of (a), it suffices to observe that $p_\alpha(x_i - x) \to 0$ iff $\langle x_i\rangle$ is eventually in $U_{x\alpha\epsilon}$ for every $\epsilon > 0$.

(c) The continuity of the vector operations follows easily from Proposition 4.19 and part (b). Indeed, if $x_i \to x$ and $y_i \to y$, then

$$p_\alpha\big((x_i + y_i) - (x + y)\big) \le p_\alpha(x_i - x) + p_\alpha(y_i - y) \to 0,$$

so $x_i + y_i \to x + y$. If also $\lambda_i \to \lambda$, then eventually $|\lambda_i| \le C = |\lambda| + 1$, so

$$p_\alpha(\lambda_i x_i - \lambda x) \le p_\alpha\big(\lambda_i(x_i - x)\big) + p_\alpha\big((\lambda_i - \lambda)x\big) \le Cp_\alpha(x_i - x) + |\lambda_i - \lambda| p_\alpha(x),$$

and it follows that $\lambda_i x_i \to \lambda x$. Moreover, the sets $U_{x\alpha\epsilon}$ are convex, for if $y, z \in U_{x\alpha\epsilon}$, then

$$p_\alpha\big(x - [ty + (1-t)z]\big) \le p_\alpha(tx - ty) + p_\alpha\big((1-t)x + (1-t)z\big) < t\epsilon + (1-t)\epsilon = \epsilon.$$

The local convexity of the topology therefore follows from (a). ∎

In this context there is an analogue of Proposition 5.2:

5.15 Proposition. *Suppose $\mathcal{X}$ and $\mathcal{Y}$ are vector spaces with topologies defined, respectively, by the families $\{p_\alpha\}_{\alpha\in A}$ and $\{q_\beta\}_{\beta\in B}$ of seminorms, and $T : \mathcal{X} \to \mathcal{Y}$ is a linear map. Then T is continuous iff for each $\beta \in B$ there exist $\alpha_1, \dots, \alpha_k \in A$ and $C > 0$ such that $q_\beta(Tx) \le C \sum_1^k p_{\alpha_j}(x)$.*

Proof. If the latter condition holds and $\langle x_i\rangle$ is a net converging to $x \in \mathcal{X}$, by Theorem 5.14b we have $p_\alpha(x_i - x) \to 0$ for all α, hence $q_\beta(Tx_i - Tx) \to 0$ for all β, hence $Tx_i \to Tx$. By Proposition 4.19, T is continuous. Conversely, if T is continuous, for every $\beta \in B$ there is a neighborhood U of 0 in $\mathcal{X}$ such that $q_\beta(Tx) < 1$ for $x \in U$. By Theorem 5.14a we may assume that $U = \bigcap_1^k U_{x\alpha_j\epsilon_j}$. Let

$\epsilon = \min(\epsilon_1, \ldots, \epsilon_k)$; then $q_\beta(Tx) < 1$ whenever $p_{\alpha_j}(x) < \epsilon$ for all j. Now, given $x \in \mathcal{X}$, there are two possibilities. If $p_{\alpha_j}(x) > 0$ for some j, let $y = \epsilon x / \sum_1^k p_{\alpha_j}(x)$. Then $p_{\alpha_j}(y) < \epsilon$ for all j, so

$$q_\beta(Tx) = \sum_1^k \epsilon^{-1} p_{\alpha_j}(x) q_\beta(Ty) \leq \epsilon^{-1} \sum_1^k p_{\alpha_j}(x).$$

On the other hand, if $p_{\alpha_j}(x) = 0$ for all j, then $p_{\alpha_j}(rx) = 0$ for all j and all $r > 0$, hence $rq_\beta(Tx) = q_\beta(T(rx)) < 1$ for all $r > 0$, hence $q_\beta(Tx) = 0$. Thus $q_\beta(Tx) \leq \epsilon^{-1} \sum_1^k p_{\alpha_j}(x)$ in this case too, and we are done. ∎

The proof of the following proposition is left to the reader (Exercise 43).

5.16 Proposition. *Let $\mathcal{X}$ be a vector space equipped with the topology defined by a family $\{p_\alpha\}_{\alpha \in A}$ of seminorms.*

a. $\mathcal{X}$ is Hausdorff iff for each $x \neq 0$ there exists $\alpha \in A$ such that $p_\alpha(x) \neq 0$.

b. If $\mathcal{X}$ is Hausdorff and A is countable, then $\mathcal{X}$ is metrizable with a translation-invariant metric (i.e., $\rho(x, y) = \rho(x + z, y + z)$ for all $x, y, z \in \mathcal{X}$).

If $\mathcal{X}$ has the topology defined by the seminorms $\{p_\alpha\}_{\alpha \in A}$, by Proposition 5.15 a linear functional f on $\mathcal{X}$ is continuous iff $|f(x)| \leq C \sum_1^k p_{\alpha_j}(x)$ for some $C > 0$ and $\alpha_1, \ldots, \alpha_k \in A$. Since a finite sum of seminorms is again a seminorm, the Hahn-Banach theorem guarantees the existence of lots of continuous linear functionals on $\mathcal{X}$ — enough to separate points, if $\mathcal{X}$ is Hausdorff. The set of all such functionals is denoted, as before, by $\mathcal{X}^*$. There are various ways of making $\mathcal{X}^*$ into a topological vector space, but we shall not consider this question systematically. The simplest way is to impose the weakest topology that makes all the evaluation maps $f \mapsto f(x)$ ($x \in \mathcal{X}$) continuous, an idea that we shall discuss further below.

In a topological vector space $\mathcal{X}$ the notion of Cauchy sequence or Cauchy net makes sense. Namely, a net $\langle x_i \rangle_{i \in I}$ in $\mathcal{X}$ is called **Cauchy** if the net $\langle x_i - x_j \rangle_{(i,j) \in I \times I}$ converges to zero. (Here $I \times I$ is directed in the usual way: $(i, j) \lesssim (i', j')$ iff $i \lesssim i'$ and $j \lesssim j'$.) Naturally, $\mathcal{X}$ is called **complete** if every Cauchy net converges. Completeness is of most interest when $\mathcal{X}$ is first countable, in which case it is equivalent to the condition that every Cauchy *sequence* converges (Exercise 44). More particularly, if $\mathcal{X}$ is Hausdorff and its topology is defined by a countable family of seminorms, then this topology is first countable by Theorem 5.14a; indeed, it is given by a translation-invariant metric ρ by Proposition 5.16b, and a sequence is Cauchy according to the definition just given iff it is Cauchy with respect to ρ. A complete Hausdorff topological vector space whose topology is defined by a countable family of seminorms is called a **Fréchet space**.

Let us now consider some interesting examples of topological vector spaces whose topologies are defined by families of seminorms rather than by single norms. We have already met a couple of them in previous chapters:

- Let X be an LCH space. On $\mathbb{C}^X$, the topology of uniform convergence on compact sets is defined by the seminorms $p_K(f) = \sup_{x \in K} |f(x)|$ as

K ranges over compact subsets of X. If X is σ-compact and $\{U_n\}$ are as in Propositions 4.39 and 4.40, this topology is defined by the seminorms $p_n(f) = \sup_{x \in \overline{U}_n} |f(x)|$. In this case, $\mathbb{C}^X$ is easily seen to be complete, so it is a Fréchet space; by Proposition 4.38, so is $C(X)$.

- The space $L^1_{\text{loc}}(\mathbb{R}^n)$, defined in §3.4, is a Fréchet space with the topology defined by the seminorms $p_k(f) = \int_{|x| \le k} |f(x)|\, dx$. (Completeness follows easily from the completeness of L^1.) An obvious generalization of this construction yields a locally convex topological vector space $L^1_{\text{loc}}(X, \mu)$ where X is any LCH space and μ is a Borel measure on X that is finite on compact sets.

Another class of topological vector spaces arises naturally in connection with the theory of differential equations. One often wishes to study the operator d/dx, or more complicated operators constructed from it, acting on various spaces of functions. Unfortunately, it is virtually impossible to define norms on most infinite-dimensional functions spaces so that d/dx becomes a bounded operator. Here is one precise result along these lines: There is no norm on the space $C^\infty([0,1])$ of infinitely differentiable functions on $[0,1]$ with respect to which d/dx is bounded. Indeed, if $f_\lambda(x) = e^{\lambda x}$, then $(d/dx)f_\lambda = \lambda f_\lambda$, so $\|d/dx\| \ge |\lambda|$ for all λ no matter what norm is used on $C^\infty([0,1])$.

In view of this difficulty, three courses of action are available. First, one can consider differentiation as an unbounded operator from $\mathcal{X}$ to $\mathcal{Y}$ where $\mathcal{Y}$ is a suitable Banach space and $\mathcal{X}$ is a dense subspace of $\mathcal{Y}$, as in Exercise 30. Second, one can consider differentiation as a bounded linear map from one Banach space $\mathcal{X}$ to a different one $\mathcal{Y}$, such as $\mathcal{X} = C^k([0,1])$ and $\mathcal{Y} = C^{k-1}([0,1])$ in Exercise 9. Finally, one can consider differentiation as a continuous operator on a locally convex space $\mathcal{X}$ whose topology is not given by a norm. All of these points of view have their uses, but it is the last one that concerns us here. It is easy to construct families of seminorms on spaces of smooth functions such that differentiation becomes continuous almost by definition. For example, the seminorms $p_k(f) = \sup_{0 \le x \le 1} |f^{(k)}(x)|$ ($k = 0, 1, 2, \ldots$) make $C^\infty([0,1])$ into a Fréchet space (the completeness is proved as in Exercise 9), and d/dx is continuous on this space by Proposition 5.15 since $p_k(f') = p_{k+1}(f)$. Other examples are considered in Exercise 45 and in Chapter 9.

One of the most useful procedures for constructing topologies on vector spaces is by requiring the continuity of certain linear maps. Namely, suppose that $\mathcal{X}$ is a vector space, $\mathcal{Y}$ is a normed linear space, and $\{T_\alpha\}_{\alpha \in A}$ is a collection of linear maps from $\mathcal{X}$ to $\mathcal{Y}$. Then the weak topology $\mathcal{T}$ generated by $\{T_\alpha\}$ makes $\mathcal{X}$ into a locally convex topological vector space. Indeed, $\mathcal{T}$ is just the topology $\mathcal{T}'$ defined by the seminorms $p_\alpha(x) = \|T_\alpha x\|$ according to Theorem 5.14. ($\mathcal{T}$ is generated by sets of the form $\{x : \|T_\alpha x - y_0\| < \epsilon\}$ with $y_0 \in \mathcal{Y}$, whereas $\mathcal{T}'$ is generated by sets of the form $\{x : \|T_\alpha x - T_\alpha x_0\| < \epsilon\}$ with $x_0 \in \mathcal{X}$. If the T_α's are surjective, these are obviously the same; the general case is left as Exercise 46.) The topology on $C^\infty([0,1])$ in the preceding paragraph is an example of this construction, with $\mathcal{Y} = C([0,1])$ and $T_k f = f^{(k)}$. We now present some more.

First, let $\mathcal{X}$ be a normed vector space. The weak topology generated by $\mathcal{X}^*$ is known simply as the **weak topology** on $\mathcal{X}$, and convergence with respect to this

topology is known as **weak convergence**. Thus, if $\langle x_\alpha \rangle$ is a net in $\mathcal{X}$, $x_\alpha \to x$ weakly iff $f(x_\alpha) \to f(x)$ for all $f \in \mathcal{X}^*$. When $\mathcal{X}$ is infinite-dimensional, the weak topology is always weaker than the norm topology; see Exercise 49.

Next, let $\mathcal{X}$ be a normed vector space, $\mathcal{X}^*$ its dual space. The weak topology on $\mathcal{X}^*$ as defined above is the topology generated by $\mathcal{X}^{**}$; of more interest is the topology generated by $\mathcal{X}$ (considered as a subspace of $\mathcal{X}^{**}$), which is called the **weak* topology** (read "weak star topology") on $\mathcal{X}^*$. $\mathcal{X}^*$ is a space of functions on $\mathcal{X}$, and the weak* topology is simply the topology of pointwise convergence: $f_\alpha \to f$ iff $f_\alpha(x) \to f(x)$ for all $x \in \mathcal{X}$. The weak* topology is even weaker than the weak topology on $\mathcal{X}^*$; the two coincide precisely when $\mathcal{X}$ is reflexive.

Finally, let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces. The topology on $L(\mathcal{X}, \mathcal{Y})$ generated by the evaluation maps $T \mapsto Tx$ ($x \in \mathcal{X}$) is called the **strong operator topology** on $L(\mathcal{X}, \mathcal{Y})$, and the topology generated by the linear functionals $T \mapsto f(Tx)$ ($x \in \mathcal{X}$, $f \in \mathcal{Y}^*$) is called the **weak operator topology** on $L(\mathcal{X}, \mathcal{Y})$. Again, these topologies are best understood in terms of convergence: $T_\alpha \to T$ strongly iff $T_\alpha x \to Tx$ in the norm topology of $\mathcal{Y}$ for each $x \in \mathcal{X}$, whereas $T_\alpha \to T$ weakly iff $T_\alpha x \to Tx$ in the weak topology of $\mathcal{Y}$ for each $x \in \mathcal{X}$. Thus the strong operator topology is stronger than the weak operator topology but weaker than the norm topology on $L(\mathcal{X}, \mathcal{Y})$.

The following result concerning strong convergence is almost trivial but extremely useful:

5.17 Proposition. *Suppose* $\{T_n\}_1^\infty \subset L(\mathcal{X}, \mathcal{Y})$, $\sup_n \|T_n\| < \infty$, *and* $T \in L(\mathcal{X}, \mathcal{Y})$. *If* $\|T_n x - Tx\| \to 0$ *for all* x *in a dense subset* D *of* $\mathcal{X}$, *then* $T_n \to T$ *strongly.*

Proof. Let $C = \sup\{\|T\|, \|T_1\|, \|T_2\|, \ldots\}$. Given $x \in \mathcal{X}$ and $\epsilon > 0$, choose $x' \in D$ such that $\|x - x'\| < \epsilon/3C$. If n is large enough so that $\|T_n x' - Tx'\| < \epsilon/3$, we have

$$\begin{aligned}\|T_n x - Tx\| &\le \|T_n x - T_n x'\| + \|T_n x' - Tx'\| + \|Tx' - Tx\| \\ &\le 2C\|x - x'\| + \tfrac{1}{3}\epsilon < \epsilon,\end{aligned}$$

so that $T_n x \to Tx$. ∎

Our final result in this section is a compactness theorem that is one of the main reasons for the usefulness of the weak* topology on a dual space. The idea of the proof is similar to the techniques discussed in §4.8.

5.18 Alaoglu's Theorem. *If* $\mathcal{X}$ *is a normed vector space, the closed unit ball* $B^* = \{f \in \mathcal{X}^* : \|f\| \le 1\}$ *in* $\mathcal{X}^*$ *is compact in the weak* topology.*

Proof. For each $x \in \mathcal{X}$ let $D_x = \{z \in \mathbb{C} : |z| \le \|x\|\}$, and let $D = \prod_{x \in \mathcal{X}} D_x$. Then D is compact by Tychonoff's theorem. The elements of D are precisely those complex-valued functions ϕ on $\mathcal{X}$ such that $|\phi(x)| \le \|x\|$ for all $x \in \mathcal{X}$, and B^* consists of those elements of D that are linear. Moreover, the relative topologies that B^* inherits from the product topology on D and the weak* topology on $\mathcal{X}^*$ both coincide with the topology of pointwise convergence, so it suffices to see that B^* is

closed in D. But this is easy: If $\langle f_\alpha \rangle$ is a net in B^* that converges to $f \in D$, for any $x, y \in \mathcal{X}$ and $a, b \in \mathbb{C}$ we have

$$f(ax + by) = \lim f_\alpha(ax + by) = \lim[af_\alpha(x) + bf_\alpha(y)] = af(x) + bf(y),$$

so that $f \in B^*$. ∎

Warning: Alaoglu's theorem does not imply that $\mathcal{X}^*$ is locally compact in the weak* topology; see Exercise 49b.

Exercises

43. Prove Proposition 5.16. (For part (b), proceed as in Exercise 56d in §4.5.)

44. If $\mathcal{X}$ is a first countable topological vector space and every Cauchy sequence in $\mathcal{X}$ converges, then every Cauchy net in $\mathcal{X}$ converges.

45. The space $C^\infty(\mathbb{R})$ of all infinitely differentiable functions on $\mathbb{R}$ has a Fréchet space topology with respect to which $f_n \to f$ iff $f_n^{(k)} \to f^{(k)}$ uniformly on compact sets for all $k \geq 0$.

46. If $\mathcal{X}$ is a vector space, $\mathcal{Y}$ a normed linear space, $\mathcal{T}$ the weak topology on $\mathcal{X}$ generated by a family of linear maps $\{T_\alpha : \mathcal{X} \to \mathcal{Y}\}$, and $\mathcal{T}'$ the topology defined by the seminorms $\{x \mapsto \|T_\alpha x\|\}$, then $\mathcal{T} = \mathcal{T}'$.

47. Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are Banach spaces.

a. If $\{T_n\}_1^\infty \subset L(\mathcal{X}, \mathcal{Y})$ and $T_n \to T$ weakly (or strongly), then $\sup_n \|T_n\| < \infty$.

b. Every weakly convergent sequence in $\mathcal{X}$, and every weak*-convergent sequence in $\mathcal{X}^*$, is bounded (with respect to the norm).

48. Suppose that $\mathcal{X}$ is a Banach space.

a. The norm-closed unit ball $B = \{x \in \mathcal{X} : \|x\| \leq 1\}$ is also weakly closed. (Use Theorem 5.8d.)

b. If $E \subset \mathcal{X}$ is bounded (with respect to the norm), so is its weak closure.

c. If $F \subset \mathcal{X}^*$ is bounded (with respect to the norm), so is its weak* closure.

d. Every weak*-Cauchy sequence in $\mathcal{X}^*$ converges. (Use Exercise 38.)

49. Suppose that $\mathcal{X}$ is an infinite-dimensional Banach space.

a. Every nonempty weakly open set in $\mathcal{X}$, and every nonempty weak*-open set in $\mathcal{X}^*$, is unbounded (with respect to the norm).

b. Every bounded subset of $\mathcal{X}$ is nowhere dense in the weak topology, and every bounded subset of $\mathcal{X}^*$ is nowhere dense in the weak* topology. (Use Exercise 48b,c.)

c. $\mathcal{X}$ is meager in itself with respect to the weak topology, and $\mathcal{X}^*$ is meager in itself with respect to the weak* topology.

d. The weak* topology on $\mathcal{X}^*$ is not defined by any translation-invariant metric. (Use Exercise 48d.)

50. If $\mathcal{X}$ is a separable normed linear space, the weak* topology on the closed unit ball in $\mathcal{X}^*$ is second countable and hence metrizable. (But cf. Exercise 49d.)

51. A vector subspace of a normed vector space $\mathcal{X}$ is norm-closed iff it is weakly closed. (However, a norm-closed subspace of $\mathcal{X}^*$ need not be weak*-closed unless $\mathcal{X}$ is reflexive; see Exercise 52d.)

52. Let $\mathcal{X}$ be a Banach space and let $f_1, \ldots, f_n$ be linearly independent elements of $\mathcal{X}^*$.

a. Define $T : \mathcal{X} \to \mathbb{C}^n$ by $Tx = (f_1(x), \ldots, f_n(x))$. If $\mathcal{N} = \{x : Tx = 0\}$ and $\mathcal{M}$ is the linear span of $f_1, \ldots, f_n$, then $\mathcal{M} = \mathcal{N}^0$ in the notation of Exercise 23 and hence $\mathcal{M}^*$ is isomorphic to $(\mathcal{X}/\mathcal{N})^*$.

b. If $F \in \mathcal{X}^{**}$, for any $\epsilon > 0$ there exists $x \in \mathcal{X}$ such that $F(f_j) = f_j(x)$ for $j = 1, \ldots, n$ and $\|x\| \leq (1 + \epsilon)\|F\|$. ($F|\mathcal{M}$ can be identified with an element of $(\mathcal{X}/\mathcal{N})^{**}$ and hence with an element of $\mathcal{X}/\mathcal{N}$ since the latter is finite-dimensional.)

c. If $\mathcal{X}$ is considered as a subspace of $\mathcal{X}^{**}$, the relative topology on $\mathcal{X}$ induced by the weak* topology on $\mathcal{X}^{**}$ is the weak topology on $\mathcal{X}$.

d. In the weak* topology on $\mathcal{X}^{**}$, $\mathcal{X}$ is dense in $\mathcal{X}^{**}$ and the closed unit ball in $\mathcal{X}$ is dense in the closed unit ball in $\mathcal{X}^{**}$.

e. $\mathcal{X}$ is reflexive iff its closed unit ball is weakly compact.

53. Suppose that $\mathcal{X}$ is a Banach space and $\{T_n\}$, $\{S_n\}$ are sequences in $L(\mathcal{X}, \mathcal{X})$ such that $T_n \to T$ strongly and $S_n \to S$ strongly.

a. If $\{x_n\} \subset \mathcal{X}$ and $\|x_n - x\| \to 0$, then $\|T_n x_n - Tx\| \to 0$. (Use Exercise 47a.)

b. $T_n S_n \to TS$ strongly.

5.5 HILBERT SPACES

The most important Banach spaces, and the ones on which the most refined analysis can be done, are the Hilbert spaces, which are a direct generalization of finite-dimensional Euclidean spaces. Before defining them, we need to introduce a few concepts.

Let $\mathcal{H}$ be a complex vector space. An **inner product** (or **scalar product**) on $\mathcal{H}$ is a map $(x, y) \mapsto \langle x, y\rangle$ from $\mathcal{X} \times \mathcal{X} \to \mathbb{C}$ such that:

i. $\langle ax + by,\ z\rangle = a\langle x, z\rangle + b\langle y, z\rangle$ for all $x, y, z \in \mathcal{H}$ and $a, b \in \mathbb{C}$.

ii. $\langle y, x\rangle = \overline{\langle x, y\rangle}$ for all $x, y \in \mathcal{H}$.

iii. $\langle x, x\rangle \in (0, \infty)$ for all nonzero $x \in \mathcal{X}$.

We observe that (i) and (ii) imply that

$$\langle x,\ ay + bz\rangle = \overline{a}\langle x, y\rangle + \overline{b}\langle x, z\rangle \text{ for all } x, y, z \in \mathcal{H} \text{ and } a, b \in \mathbb{C}.$$

(One can also define inner products on real vector spaces: $\langle x, y\rangle$ is then real, a and b are assumed real in (i), and (ii) becomes $\langle y, x\rangle = \langle x, y\rangle$.)

A complex vector space equipped with an inner product is called a **pre-Hilbert space**. If $\mathcal{H}$ is a pre-Hilbert space, for $x \in \mathcal{H}$ we define

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

5.19 The Schwarz Inequality. $|\langle x, y\rangle| \leq \|x\|\,\|y\|$ *for all* $x, y \in \mathcal{H}$, *with equality iff* x *and* y *are linearly dependent.*

Proof. If $\langle x, y\rangle = 0$, the result is obvious. If $\langle x, y\rangle \neq 0$ (and in particular $y \neq 0$), let $\alpha = \operatorname{sgn}\langle x, y\rangle$ and $z = \alpha y$, so that $\langle x, z\rangle = \langle z, x\rangle = |\langle x, y\rangle|$ and $\|z\| = \|y\|$. Then for $t \in \mathbb{R}$ we have

$$0 \leq \langle x - tz,\, x - tz\rangle = \|x\|^2 - 2t|\langle x, y\rangle| + t^2\|y\|^2.$$

The expression on the right is a quadratic function of t whose absolute minimum occurs at $t = \|y\|^{-2}|\langle x, y\rangle|$. Setting t equal to this value, we obtain

$$0 \leq \|x - tz\|^2 = \|x\|^2 - \|y\|^{-2}|\langle x, y\rangle|^2$$

with equality iff $x - tz = x - \alpha t y = 0$, from which the desired result is immediate. ∎

5.20 Proposition. *The function* $x \mapsto \|x\|$ *is a norm on* $\mathcal{H}$.

Proof. That $\|x\| = 0$ iff $x = 0$ and that $\|\lambda x\| = |\lambda|\,\|x\|$ are obvious from the definition. As for the triangle inequality, we have

$$\|x + y\|^2 = \langle x + y,\, x + y\rangle = \|x\|^2 + 2\operatorname{Re}\langle x, y\rangle + \|y\|^2,$$

so by the Schwarz inequality,

$$\|x + y\|^2 \leq \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,$$

as desired. ∎

A pre-Hilbert space that is complete with respect to the norm $\|x\| = \sqrt{\langle x, x\rangle}$ is called a **Hilbert space**. (One can also consider real Hilbert spaces with real inner products. However, Hilbert spaces are usually assumed to be complex unless otherwise specified.)

Example: Let $(X, \mathcal{M}, \mu)$ be a measure space, and let $L^2(\mu)$ be the set of all measurable functions $f : X \to \mathbb{C}$ such that $\int |f|^2\, d\mu < \infty$ (where, as usual, we identify two functions that are equal a.e.). From the inequality $ab \leq \frac{1}{2}(a^2 + b^2)$, valid for all $a, b \geq 0$, we see that if $f, g \in L^2(\mu)$ then $|f\overline{g}| \leq \frac{1}{2}(|f|^2 + |g|^2)$, so that $f\overline{g} \in L^1(\mu)$. It follows easily that the formula

$$\langle f, g\rangle = \int f\overline{g}\, d\mu$$

defines an inner product on $L^2(\mu)$. In fact, $L^2(\mu)$ is a Hilbert space for any measure μ. We shall prove completeness in Theorem 6.6; for the present we shall take this result for granted.

An important special case of this construction is obtained by taking μ to be counting measure on $(A, \mathcal{P}(A))$, where A is any nonempty set; in this situation $L^2(\mu)$ is usually denoted by $l^2(A)$. Thus, $l^2(A)$ is the set of functions $f : A \to \mathbb{C}$ such that the sum $\sum_{\alpha \in A} |f(\alpha)|^2$ (as defined in §0.5) is finite. The completeness of $l^2(A)$ is rather easy to prove directly (Exercise 54).

For the remainder of this section, $\mathcal{H}$ will denote a Hilbert space.

5.21 Proposition. *If $x_n \to x$ and $y_n \to y$, then $\langle x_n, y_n \rangle \to \langle x, y \rangle$.*

Proof. By the Schwarz inequality,

$$\begin{aligned} \big|\langle x_n, y_n \rangle - \langle x, y \rangle\big| &= \big|\langle x_n - x,\, y_n \rangle + \langle x,\, y_n - y \rangle\big| \\ &\le \|x_n - x\|\, \|y_n\| + \|x\|\, \|y_n - y\|, \end{aligned}$$

which tends to zero since $\|y_n\| \to \|y\|$. ∎

5.22 The Parallelogram Law. *For all $x, y \in \mathcal{H}$,*

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

("The sum of the squares of the diagonals of a parallelogram is the sum of the squares of the four sides.")

Proof. Add the two formulas $\|x \pm y\|^2 = \|x\|^2 \pm 2\,\mathrm{Re}\langle x, y \rangle + \|y\|^2$. ∎

If $x, y \in \mathcal{X}$, we say that x is **orthogonal** to y and write $x \perp y$ if $\langle x, y \rangle = 0$. If $E \subset \mathcal{H}$, we define

$$E^\perp = \big\{x \in \mathcal{H} : \langle x, y \rangle = 0 \text{ for all } y \in E\big\}.$$

It is immediate from Proposition 5.21 and the linearity of the inner product in its first argument that $E^\perp$ is a closed subspace of $\mathcal{H}$.

5.23 The Pythagorean Theorem. *If $x_1, \ldots, x_n \in \mathcal{H}$ and $x_j \perp x_k$ for $j \neq k$,*

$$\Big\|\sum_1^n x_j\Big\|^2 = \sum_1^n \|x_j\|^2.$$

Proof. $\|\sum x_j\|^2 = \langle \sum x_j, \sum x_j \rangle = \sum_{j,k} \langle x_j, x_k \rangle$. The terms with $k \neq j$ are all zero, leaving only $\sum \langle x_j, x_j \rangle = \sum \|x_j\|^2$. ∎

5.24 Theorem. *If $\mathcal{M}$ is a closed subspace of $\mathcal{H}$, then $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$; that is, each $x \in \mathcal{H}$ can be expressed uniquely as $x = y + z$ where $y \in \mathcal{M}$ and $z \in \mathcal{M}^\perp$. Moreover, y and z are the unique elements of $\mathcal{M}$ and $\mathcal{M}^\perp$ whose distance to x is minimal.*

Proof. Given $x \in \mathcal{H}$, let $\delta = \inf\{\|x - y\| : y \in \mathcal{M}\}$, and let $\{y_n\}$ be a sequence in $\mathcal{M}$ such that $\|x - y_n\| \to \delta$. By the paralellogram law,

$$2\big(\|y_n - x\|^2 + \|y_m - x\|^2\big) = \|y_n - y_m\|^2 + \|y_n + y_m - 2x\|^2,$$

so since $\frac{1}{2}(y_n + y_m) \in \mathcal{M}$,

$$\begin{aligned}\|y_n - y_m\|^2 &= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\|\tfrac{1}{2}(y_n + y_m) - x\|^2 \\ &\le 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\delta^2.\end{aligned}$$

As $m, n \to \infty$ this last quantity tends to zero, so $\{y_n\}$ is a Cauchy sequence. Let $y = \lim y_n$ and $z = x - y$. Then $y \in \mathcal{M}$ since $\mathcal{M}$ is closed, and $\|x - y\| = \delta$.

We claim that $z \in \mathcal{M}^\perp$. Indeed, if $u \in \mathcal{M}$, after multiplying u by a nonzero scalar we may assume that $\langle z, u\rangle$ is real. Then the function

$$f(t) = \|z + tu\|^2 = \|z\|^2 + 2t\langle z, u\rangle + t^2\|u\|^2$$

is real for $t \in \mathbb{R}$, and is has a minimum (namely, δ^2) at $t = 0$ because $z + tu = x - (y - tu)$ and $y - tu \in \mathcal{M}$. Thus $2\langle z, u\rangle = f'(0) = 0$, so $z \in \mathcal{M}^\perp$. Moreover, if z' is another element of $\mathcal{M}^\perp$, by the Pythagorean theorem (since $x - z = y \in \mathcal{M}$) we have

$$\|x - z'\|^2 = \|x - z\|^2 + \|z - z'\|^2 \ge \|x - z\|^2,$$

with equality iff $z = z'$. The same reasoning shows that y is the unique element of $\mathcal{M}$ closest to x.

Finally, if $x = y' + z'$ with $y' \in \mathcal{M}$ and $z' \in \mathcal{M}^\perp$, then $y - y' = z' - z \in \mathcal{M} \cap \mathcal{M}^\perp$, so $y - y'$ and $z' - z$ are orthogonal to themselves and hence are zero. ∎

If $y \in \mathcal{H}$, the Schwarz inequality shows that the formula $f_y(x) = \langle x, y\rangle$ defines a bounded linear functional on $\mathcal{H}$ such that $\|f_y\| = \|y\|$. Thus, the map $y \to f_y$ is a conjugate-linear isometry of $\mathcal{H}$ into $\mathcal{H}^*$. It is a fundamental fact that this map is surjective:

5.25 Theorem. *If $f \in \mathcal{H}^*$, there is a unique $y \in \mathcal{H}$ such that $f(x) = \langle x, y\rangle$ for all $x \in \mathcal{X}$.*

Proof. Uniqueness is easy: If $\langle x, y\rangle = \langle x, y'\rangle$ for all x, by taking $x = y - y'$ we conclude that $\|y - y'\|^2 = 0$ and hence $y = y'$. If f is the zero functional, then obviously $y = 0$. Otherwise, let $\mathcal{M} = \{x \in \mathcal{H} : f(x) = 0\}$. Then $\mathcal{M}$ is a proper closed subspace of $\mathcal{X}$, so $\mathcal{M}^\perp \neq \{0\}$ by Theorem 5.24. Pick $z \in \mathcal{M}^\perp$ with $\|z\| = 1$. If $u = f(x)z - f(z)x$ then $u \in \mathcal{M}$, so

$$0 = \langle u, z\rangle = f(x)\|z\|^2 - f(z)\langle x, z\rangle = f(x) - \langle x, \overline{f(z)}z\rangle.$$

Hence $f(x) = \langle x, y\rangle$ where $y = \overline{f(z)}z$. ∎

Thus, Hilbert spaces are reflexive in a very strong sense: Not only is $\mathcal{H}$ naturally isomorphic to $\mathcal{H}^{**}$, it is naturally isomorphic (via a conjugate-linear map) to $\mathcal{H}^*$.

A subset $\{u_\alpha\}_{\alpha\in A}$ of $\mathcal{H}$ is called **orthonormal** if $\|u_\alpha\| = 1$ for all α and $u_\alpha \perp u_\beta$ whenever $\alpha \neq \beta$. If $\{x_n\}_1^\infty$ is a linearly independent sequence in $\mathcal{H}$, there is a standard inductive procedure, called the **Gram-Schmidt process**, for converting $\{x_n\}$ into an orthonormal sequence $\{u_n\}$ such that the linear span of $\{x_n\}_1^N$ coincides with the linear span of $\{u_n\}_1^N$ for all N. Namely, the first step is to set $u_1 = x_1/\|x_1\|$. Having defined $u_1, \dots, u_{N-1}$, we set $v_N = x_N - \sum_1^{N-1}\langle x_N, u_n\rangle u_n$. Then v_N is nonzero because x_N is not in the linear span of $x_1, \dots, x_{N-1}$ and hence of $u_1, \dots, u_{N-1}$, and $\langle v_N, u_m\rangle = \langle x_N, u_m\rangle - \langle x_N, u_m\rangle = 0$ for all $m < N$. We can therefore take $u_N = v_N/\|v_N\|$.

5.26 Bessel's Inequality. *If $\{u_\alpha\}_{\alpha\in A}$ is an orthonormal set in $\mathcal{H}$, then for any $x \in \mathcal{H}$,*

$$\sum_{\alpha\in A} |\langle x, u_\alpha\rangle|^2 \leq \|x\|^2.$$

In particular, $\{\alpha : \langle x, u_\alpha\rangle \neq 0\}$ is countable.

Proof. It suffices to show that $\sum_{\alpha\in F} |\langle x, u_\alpha\rangle|^2 \leq \|x\|^2$ for any finite $F \subset A$. But

$$\begin{aligned}
0 &\leq \Big\|x - \sum_{\alpha\in F}\langle x, u_\alpha\rangle u_\alpha\Big\|^2 \\
&= \|x\|^2 - 2\operatorname{Re}\Big\langle x, \sum_{\alpha\in F}\langle x, u_\alpha\rangle u_\alpha\Big\rangle + \Big\|\sum_{\alpha\in F}\langle x, u_\alpha\rangle u_\alpha\Big\|^2 \\
&= \|x\|^2 - 2\sum_{\alpha\in F}|\langle x, u_\alpha\rangle|^2 + \sum_{\alpha\in F}|\langle x, u_\alpha\rangle|^2 \\
&= \|x\|^2 - \sum_{\alpha\in F}|\langle x, u_\alpha\rangle|^2,
\end{aligned}$$

where the Pythagorean theorem was used in the third line. ∎

5.27 Theorem. *If $\{u_\alpha\}_{\alpha\in A}$ is an orthonormal set in $\mathcal{H}$, the following are equivalent:*

a. **(Completeness)** *If $\langle x, u_\alpha\rangle = 0$ for all α, then $x = 0$.*

b. **(Parseval's Identity)** $\|x\|^2 = \sum_{\alpha\in A}|\langle x, u_\alpha\rangle|^2$ *for all $x \in \mathcal{H}$.*

c. For each $x \in \mathcal{H}$, $x = \sum_{\alpha\in A}\langle x, u_\alpha\rangle u_\alpha$, where the sum on the right has only countably many nonzero terms and converges in the norm topology no matter how these terms are ordered.

Proof. (a) implies (c): If $x \in \mathcal{H}$, let $\alpha_1, \alpha_2, \dots$ be any enumeration of the α's for which $\langle x, u_\alpha\rangle \neq 0$. By Bessel's inequality the series $\sum|\langle x, u_{\alpha_j}\rangle|^2$ converges, so by the Pythagorean theorem,

$$\Big\|\sum_n^m \langle x, u_{\alpha_j}\rangle u_{\alpha_j}\Big\|^2 = \sum_n^m |\langle x, u_{\alpha_j}\rangle|^2 \to 0 \text{ as } m, n \to \infty.$$

The series $\sum\langle x, u_{\alpha_j}\rangle u_{\alpha_j}$ therefore converges since $\mathcal{H}$ is complete. If $y = x - \sum\langle x, u_{\alpha_j}\rangle u_{\alpha_j}$, then clearly $\langle y, u_\alpha\rangle = 0$ for all α, so by (a), $y = 0$.

(c) implies (b): With notation as above, as in the proof of Bessel's inequality we have

$$\|x\|^2 - \sum_1^n |\langle x, u_{\alpha_j}\rangle|^2 = \left\|x - \sum_1^n \langle x, u_{\alpha_j}\rangle u_{\alpha_j}\right\|^2 \to 0 \text{ as } n \to \infty.$$

Finally, that (b) implies (a) is obvious. ∎

An orthonormal set having the properties (a–c) in Theorem 5.27 is called an **orthonormal basis** for $\mathcal{H}$. For example, let $\mathcal{H} = l^2(A)$. For each $\alpha \in A$, define $e_\alpha \in l^2(A)$ by $e_\alpha(\beta) = 1$ if $\beta = \alpha$, $e_\alpha(\beta) = 0$ otherwise. The set $\{e_\alpha\}_{\alpha\in A}$ is clearly orthonormal, and for any $f \in l^2(A)$ we have $\langle f, e_\alpha\rangle = f(\alpha)$, from which it follows that $\{e_\alpha\}$ is an orthonormal basis.

5.28 Proposition. *Every Hilbert space has an orthonormal basis.*

Proof. A routine application of Zorn's lemma shows that the collection of orthonormal sets, ordered by inclusion, has a maximal element; and maximality is equivalent to property (a) in Theorem 5.27. ∎

5.29 Proposition. *A Hilbert space $\mathcal{H}$ is separable iff it has a countable orthonormal basis, in which case every orthonormal basis for $\mathcal{H}$ is countable.*

Proof. If $\{x_n\}$ is a countable dense set in $\mathcal{H}$, by discarding recursively any x_n that is in the linear span of $x_1, \ldots, x_{n-1}$ we obtain a linearly independent sequence $\{y_n\}$ whose linear span is dense in $\mathcal{H}$. Application of the Gram-Schmidt process to $\{y_n\}$ yields an orthonormal sequence $\{u_n\}$ whose linear span is dense in $\mathcal{H}$ and which is therefore a basis. Conversely, if $\{u_n\}$ is a countable orthonormal basis, the finite linear combinations of the u_n's with coefficients in a countable dense subset of $\mathbb{C}$ form a countable dense set in $\mathcal{H}$. Moreover, if $\{v_\alpha\}_{\alpha\in A}$ is another orthonormal basis, for each n the set $A_n = \{\alpha \in A : \langle u_n, v_\alpha\rangle \neq 0\}$ is countable. By completeness of $\{u_n\}$, $A = \bigcup_1^\infty A_n$, so A is countable. ∎

Most Hilbert spaces that arise in practice are separable. We discuss some examples in Exercises 60–62.

If $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces with inner products $\langle\cdot,\cdot\rangle_1$ and $\langle\cdot,\cdot\rangle_2$, a **unitary map** from $\mathcal{H}_1$ to $\mathcal{H}_2$ is an invertible linear map $U : \mathcal{H}_1 \to \mathcal{H}_2$ that preserves inner products:

$$\langle Ux, Uy\rangle_2 = \langle x, y\rangle_1 \text{ for all } x, y \in \mathcal{H}_1.$$

By taking $y = x$, we see that every unitary map is an isometry: $\|Ux\|_2 = \|x\|_1$. Conversely, every surjective isometry is unitary (Exercise 55). Unitary maps are the true "isomorphisms" in the category of Hilbert spaces; they preserve not only the linear structure and the topology but also the norm and the inner product. From the point of view of this abstract structure, every Hilbert space looks like an l^2 space:

5.30 Proposition. *Let $\{e_\alpha\}_{\alpha\in A}$ be an orthonormal basis for $\mathcal{X}$. Then the correspondence $x \mapsto \widehat{x}$ defined by $\widehat{x}(\alpha) = \langle x, u_\alpha\rangle$ is a unitary map from $\mathcal{H}$ to $l^2(A)$.*

Proof. The map $x \mapsto \widehat{x}$ is clearly linear, and it is an isometry from $\mathcal{H}$ to $l^2(A)$ by the Parseval identity $\|x\|^2 = \sum |\widehat{x}(\alpha)|^2$. If $f \in l^2(A)$ then $\sum |f(\alpha)|^2 < \infty$, so the Pythagorean theorem shows that the partial sums of the series $\sum f(\alpha)u_\alpha$ (of which only countably many terms are nonzero) are Cauchy; hence $x = \sum f(\alpha)u_\alpha$ exists in $\mathcal{H}$ and $\widehat{x} = f$. By Exercise 55b, $x \mapsto \widehat{x}$ is unitary. ∎

Exercises

54. For any nonempty set A, $l^2(A)$ is complete.

55. Let $\mathcal{H}$ be a Hilbert space.

a. (The polarization identity) For any $x, y \in \mathcal{H}$,

$$\langle x, y\rangle = \tfrac{1}{4}\big(\|x+y\|^2 + \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2\big).$$

(Completeness is not needed here.)

b. If $\mathcal{H}'$ is another Hilbert space, a linear map from $\mathcal{H}$ to $\mathcal{H}'$ is unitary iff it is isometric and surjective.

56. If E is a subset of a Hilbert space $\mathcal{H}$, $(E^\perp)^\perp$ is the smallest closed subspace of $\mathcal{H}$ containing E.

57. Suppose that $\mathcal{H}$ is a Hilbert space and $T \in L(\mathcal{H}, \mathcal{H})$.

a. There is a unique $T^* \in L(\mathcal{H}, \mathcal{H})$, called the **adjoint** of T, such that $\langle Tx, y\rangle = \langle x, T^*y\rangle$ for all $x, y \in \mathcal{H}$. (Cf. Exercise 22. We have $T^* = V^{-1}T^\dagger V$ where V is the conjugate-linear isomorphism from $\mathcal{H}$ to $\mathcal{H}^*$ in Theorem 5.25, $(Vy)(x) = \langle x, y\rangle$.)

b. $\|T^*\| = \|T\|$, $\|T^*T\| = \|T\|^2$, $(aS + bT)^* = \overline{a}S^* + \overline{b}T^*$, $(ST)^* = T^*S^*$, and $T^{**} = T$.

c. Let $\mathcal{R}$ and $\mathcal{N}$ denote range and nullspace; then $\mathcal{R}(T)^\perp = \mathcal{N}(T^*)$ and $\mathcal{N}(T)^\perp = \overline{\mathcal{R}(T^*)}$.

d. T is unitary iff T is invertible and $T^{-1} = T^*$.

58. Let $\mathcal{M}$ be a closed subspace of the Hilbert space $\mathcal{H}$, and for $x \in \mathcal{H}$ let Px be the element of $\mathcal{M}$ such that $x - Px \in \mathcal{M}^\perp$ as in Theorem 5.24.

a. $P \in L(\mathcal{H}, \mathcal{H})$, and in the notation of Exercise 57 we have $P^* = P$, $P^2 = P$, $\mathcal{R}(P) = \mathcal{M}$, and $\mathcal{N}(P) = \mathcal{M}^\perp$. P is called the **orthogonal projection** onto $\mathcal{M}$.

b. Conversely, suppose that $P \in L(\mathcal{H}, \mathcal{H})$ satisfies $P^2 = P^* = P$. Then $\mathcal{R}(P)$ is closed and P is the orthogonal projection onto $\mathcal{R}(P)$.

c. If $\{u_\alpha\}$ is an orthonormal basis for $\mathcal{M}$, then $Px = \sum \langle x, u_\alpha\rangle u_\alpha$.

59. Every closed convex set K in a Hilbert space has a unique element of minimal norm. (If $0 \in K$, the result is trivial; otherwise, adapt the proof of Theorem 5.24.)

60. Let $(X, \mathcal{M}, \mu)$ be a measure space. If $E \in \mathcal{M}$, we identify $L^2(E, \mu)$ with the subspace of $L^2(X, \mu)$ consisting of functions that vanish outside E. If $\{E_n\}$ is

a disjoint sequence in $\mathcal{M}$ with $X = \bigcup_1^\infty E_n$, then $\{L^2(E_n, \mu)\}$ is a sequence of mutually orthogonal subspaces of $L^2(X, \mu)$, and every $f \in L^2(X, \mu)$ can be written uniquely as $f = \sum_1^\infty f_n$ (the series converging in norm) where $f_n \in L^2(E_n, \mu)$. If $L^2(E_n, \mu)$ is separable for every n, so is $L^2(X, \mu)$.

61. Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces such that $L^2(\mu)$ and $L^2(\nu)$ are separable. If $\{f_m\}$ and $\{g_n\}$ are orthonormal bases for $L^2(\mu)$ and $L^2(\nu)$ and $h_{mn}(x, y) = f_m(x)g_n(y)$, then $\{h_{mn}\}$ is an orthonormal basis for $L^2(\mu \times \nu)$.

62. In this exercise the measure defining the L^2 spaces is Lebesgue measure.

a. $C([0, 1])$ is dense in $L^2([0, 1])$. (Adapt the proof of Theorem 2.26.)

b. The set of polynomials is dense in $L^2([0, 1])$.

c. $L^2([0, 1])$ is separable.

d. $L^2(\mathbb{R})$ is separable. (Use Exercise 60.)

e. $L^2(\mathbb{R}^n)$ is separable. (Use Exercise 61.)

63. Let $\mathcal{H}$ be an infinite-dimensional Hilbert space.

a. Every orthonormal sequence in $\mathcal{H}$ converges weakly to 0.

b. The unit sphere $S = \{x : \|x\| = 1\}$ is weakly dense in the unit ball $B = \{x : \|x\| \leq 1\}$. (In fact, every $x \in B$ is the weak limit of a sequence in S.)

64. Let $\mathcal{H}$ be a separable infinite-dimensional Hilbert space with orthonormal basis $\{u_n\}_1^\infty$.

a. For $k \in \mathbb{N}$, define $L_k \in L(\mathcal{H}, \mathcal{H})$ by $L_k(\sum_1^\infty a_n u_n) = \sum_k^\infty a_n u_{n-k}$. Then $L_k \to 0$ in the strong operator topology but not in the norm topology.

b. For $k \in \mathbb{N}$, define $R_k \in L(\mathcal{H}, \mathcal{H})$ by $R_k(\sum_1^\infty a_n u_n) = \sum_1^\infty a_n u_{n+k}$. Then $R_k \to 0$ in the weak operator topology but not in the strong operator topology.

c. $R_k L_k \to 0$ in the strong operator topology, but $L_k R_k = I$ for all k. (Use Exercise 53b.)

65. $l^2(A)$ is unitarily isomorphic to $l^2(B)$ iff $\operatorname{card}(A) = \operatorname{card}(B)$.

66. Let $\mathcal{M}$ be a closed subspace of $L^2([0, 1], m)$ that is contained in $C([0, 1])$.

a. There exists $C > 0$ such that $\|f\|_u \leq C\|f\|_{L^2}$ for all $f \in \mathcal{M}$. (Use the closed graph theorem.)

b. For each $x \in [0, 1]$ there exists $g_x \in \mathcal{M}$ such that $f(x) = \langle f, g_x \rangle$ for all $f \in \mathcal{M}$, and $\|g_x\|_{L^2} \leq C$.

c. The dimension of $\mathcal{M}$ is at most C^2. (Hint: If $\{f_j\}$ is an orthonormal sequence in $\mathcal{M}$, $\sum |f_j(x)|^2 \leq C^2$ for all $x \in [0, 1]$.)

67. (**The Mean Ergodic Theorem**). Let U be a unitary operator on the Hilbert space $\mathcal{H}$, $\mathcal{M} = \{x : Ux = x\}$, P the orthogonal projection onto $\mathcal{M}$ (Exercise 58), and $S_n = n^{-1} \sum_0^{n-1} U^j$. Then $S_n \to P$ in the strong operator topology. (If $x \in \mathcal{M}$, then $S_n x = x$; if $x = y - Uy$ for some y, then $S_n x \to 0$. By Exercise 57d, $\mathcal{M} = \{x : U^* x = x\}$. Apply Exercise 57c with $T = I - U$.)

5.6 NOTES AND REFERENCES

Functional analysis is a vast subject of which we have barely scratched the surface here. For the reader who wishes to learn more, Reed and Simon [112] and Rudin [126] are good places to start; one should also familiarize oneself with the treatises of Dunford and Schwartz [35] and Yosida [163].

Functional analysis has roots in a number of classical problems, particularly in the theory of differential and integral equations. The study of particular infinite-dimensional function spaces began in earnest around 1907 with work of F. Riesz, Fréchet, Schmidt, Helly, and others, and the notion of an abstract normed vector space appeared in papers by several authors about 1920. The research of the succeeding decade culminated in Banach's classic book [9], which marked the emergence of functional analysis as an established discipline. Detailed historical accounts can be found in Dieudonné [33] and in the notes in Dunford and Schwartz [35].

§5.1: The integral for vector-valued functions developed in Exercise 16 is called the **Bochner integral**. The hypothesis that $\mathcal{Y}$ is separable can be dropped, but the functions in $L^1_{\mathcal{Y}}$ must then be required to have separable range (after modification on a null set). A more detailed account can be found in Cohn [27] or Yosida [163].

Another approach to vector-valued integrals is as follows. Suppose that $(X, \mathcal{M}, \mu)$ is a measure space and $\mathcal{Y}$ is a topological vector space on which the continuous linear functionals separate points. A function $f : X \to \mathcal{Y}$ is called **weakly integrable** if (i) $\phi \circ f \in L^1(\mu)$ for all $\phi \in \mathcal{Y}^*$, and (ii) there exists $y \in \mathcal{Y}$ (necessarily unique) such that $\int \phi \circ f \, d\mu = \phi(y)$ for all $\phi \in \mathcal{Y}^*$. In this case we set $\int f \, d\mu = y$. If $\mathcal{Y}$ is a separable Banach space, this notion of integral coincides with the Bochner integral. See Yosida [163] and Rudin [126].

§5.3: The open mapping and closed graph theorems are due to Banach [9]. See Grabiner [58] for an interesting comment on the relation between the proofs of the open mapping theorem and the Tietze extension theorem.

The uniform boundedness principle, as we have stated it, is due to Banach and Steinhaus [10]; however, the second part of the theorem — that if $\mathcal{X}$ is a Banach space and $\sup_{T \in \mathcal{A}} \|Tx\| < \infty$ for all $x \in \mathcal{X}$, then $\sup_{T \in \mathcal{A}} \|T\| < \infty$ — had been proved previously by what Dieudonné [33] calls the "method of the gliding hump." This rather pretty (and elementary) argument has been largely neglected in recent years, but a modern exposition of it can be found in Hennefeld [71].

It is simple to construct examples of unbounded linear maps $T : \mathcal{X} \to \mathcal{Y}$ from one normed vector space to another when $\mathcal{X}$ is incomplete (see Exercises 29 and 30), but virtually impossible to do so when $\mathcal{X}$ is complete without using the axiom of choice. The standard method is as follows: Start with an unbounded $T : \mathcal{X}_0 \to \mathcal{Y}$ where $\mathcal{X}_0$ is incomplete, and let $\mathcal{X}$ be the completion of $\mathcal{X}_0$. Pick a basis $\{u_\alpha\}_{\alpha \in A}$ for $\mathcal{X}_0$ (meaning that every $x \in \mathcal{X}_0$ is a *finite* linear combination of the u_α's), and extend it to a basis $\{u_\alpha\}_{\alpha \in B}$ $(B \supset A)$ for $\mathcal{X}$. (This is where the axiom of choice comes in.) Let $\mathcal{M}$ be the linear span of $\{u_\alpha\}_{\alpha \in B \setminus A}$, so that each $x \in \mathcal{X}$ can be written uniquely as $x = x_0 + x_1$ where $x_1 \in \mathcal{X}_0$ and $x_1 \in \mathcal{M}$. Then T can be extended to $\mathcal{X}$ by setting $T(x_0 + x_1) = Tx_0$.

§5.4: Treves [150] contains a readable account of the general theory of topological vector spaces, with many concrete examples.

Alaoglu's theorem, which was first announced in Alaoglu [3] and proved in detail in Alaoglu [4], supersedes a number of earlier results dealing with special cases. It was discovered independently by Bourbaki [19].

§5.5: The space envisaged by Hilbert himself was $l^2(\mathbb{N})$; the notion of an abstract Hilbert space was introduced by von Neumann [154] in his work on the mathematics of quantum mechanics. Theorem 5.25 is originally due to F. Riesz [115] in the setting of L^2 spaces. It is one of several representation theorems for linear functionals on various spaces that bear his name, the others being Theorems 6.15, 7.2, and 7.17. To avoid confusion, we reserve the name "Riesz representation theorem" for the latter two, which are closely related.

In the literature of quantum physics, scalar products are customarily denoted by $\langle x|y\rangle$ and are taken to be linear in the second variable and conjugate-linear in the first.

6

L^p Spaces

L^p spaces are a class of Banach spaces of functions whose norms are defined in terms of integrals and which generalize the L^1 spaces discussed in Chapter 2. They furnish interesting examples of the general theory of Chapter 5 and play a central role in modern analysis.

6.1 BASIC THEORY OF L^P SPACES

In this chapter we shall be working on a fixed measure space $(X, \mathcal{M}, \mu)$. If f is a measurable function on X and $0 < p < \infty$, we define

$$\|f\|_p = \left[\int |f|^p \, d\mu\right]^{1/p}$$

(allowing the possibility that $\|f\|_p = \infty$), and we define

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

We abbreviate $L^p(X, \mathcal{M}, \mu)$ by $L^p(\mu)$, $L^p(X)$, or simply L^p when this will cause no confusion. As we have done with L^1, we consider two functions to define the same element of L^p when they are equal almost everywhere.

If A is any nonempty set, we define $l^p(A)$ to be $L^p(\mu)$ where μ is counting measure on $(A, \mathcal{P}(A))$, and we denote $l^p(\mathbb{N})$ simply by l^p.

L^p is a vector space, for if $f, g \in L^p$, then

$$|f + g|^p \leq \left[2\max(|f|, |g|)\right]^p \leq 2^p(|f|^p + |g|^p),$$

so that $f + g \in L^p$. Our notation suggests that $\|\cdot\|_p$ is a norm on L^p. Indeed, it is obvious that $\|f\|_p = 0$ iff $f = 0$ a.e. and $\|cf\|_p = |c|\,\|f\|_p$, so the only question is the triangle inequality. It turns out that the latter is valid precisely when $p \geq 1$, so our attention will be focused almost exclusively on this case.

Before proceeding further, however, let us see why the triangle inequality fails for $p < 1$. Suppose $a > 0$, $b > 0$, and $0 < p < 1$. For $t > 0$ we have $t^{p-1} > (a+t)^{p-1}$, and by integrating from 0 to b we obtain $a^p + b^p > (a+b)^p$. Thus, if E and F are disjoint sets of positive finite measure in X and we set $a = \mu(E)^{1/p}$ and $b = \mu(F)^{1/p}$, we see that

$$\|\chi_E + \chi_F\|_p = (a^p + b^p)^{1/p} > a + b = \|\chi_E\|_p + \|\chi_F\|_p.$$

The cornerstone of the theory of L^p spaces is Hölder's inequality, which we now derive.

6.1 Lemma. *If $a \geq 0$, $b \geq 0$, and $0 < \lambda < 1$, then*

$$a^\lambda b^{1-\lambda} \leq \lambda a + (1-\lambda)b,$$

with equality iff $a = b$.

Proof. The result is obvious if $b = 0$; otherwise, dividing both sides by b and setting $t = a/b$, we are reduced to showing that $t^\lambda \leq \lambda t + (1-\lambda)$ with equality iff $t = 1$. But by elementary calculus, $t^\lambda - \lambda t$ is strictly increasing for $t < 1$ and strictly decreasing for $t > 1$, so its maximum value, namely $1 - \lambda$, occurs at $t = 1$. ∎

6.2 Hölder's Inequality. *Suppose $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$ (that is, $q = p/(p-1)$). If f and g are measurable functions on X, then*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q. \tag{6.3}$$

In particular, if $f \in L^p$ and $g \in L^q$, then $fg \in L^1$, and in this case equality holds in (6.3) iff $\alpha|f|^p = \beta|g|^q$ a.e. for some constants α, β with $\alpha\beta \neq 0$.

Proof. The result is trivial if $\|f\|_p = 0$ or $\|g\|_q = 0$ (since then $f = 0$ or $g = 0$ a.e.), or if $\|f\|_p = \infty$ or $\|g\|_q = \infty$. Moreover, we observe that if (6.3) holds for a particular f and g, then it also holds for all scalar multiples of f and g, for if f and g are replaced by af and bg, both sides of (6.3) change by a factor of $|ab|$. It therefore suffices to prove that (6.3) holds when $\|f\|_p = \|g\|_q = 1$ with equality iff $|f|^p = |g|^q$ a.e. To this end, we apply Lemma 6.1 with $a = |f(x)|^p$, $b = |g(x)|^q$, and $\lambda = p^{-1}$ to obtain

$$|f(x)g(x)| \leq p^{-1}|f(x)|^p + q^{-1}|g(x)|^q. \tag{6.4}$$

Integration of both sides yields

$$\|fg\|_1 \leq p^{-1}\int |f|^p + q^{-1}\int |g|^q = p^{-1} + q^{-1} = 1 = \|f\|_p\|g\|_q.$$

Equality holds here iff it holds a.e. in (6.4), and by Lemma 6.1 this happens precisely when $|f|^p = |g|^q$ a.e. ∎

The condition $p^{-1}+q^{-1}=1$ occurring in Hölder's inequality turns up frequently in L^p theory. If $1<p<\infty$, the number $q=p/(p-1)$ such that $p^{-1}+q^{-1}=1$ is called the **conjugate exponent** to p.

6.5 Minkowski's Inequality. *If $1 \le p < \infty$ and $f, g \in L^p$, then*

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

Proof. The result is obvious if $p=1$ or if $f+g=0$ a.e. Otherwise, we observe that

$$|f+g|^p \le (|f|+|g|)|f+g|^{p-1}$$

and apply Hölder's inequality, noting that $(p-1)q=p$ when q is the conjugate exponent to p:

$$\begin{aligned}\int |f+g|^p &\le \|f\|_p \| \, |f+g|^{p-1}\|_q + \|g\|_p \| \, |f+g|^{p-1}\|_q \\ &= (\|f\|_p + \|g\|_p)\left(\int |f+g|^p\right)^{1/q}.\end{aligned}$$

Therefore,

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

∎

This result shows that, for $p \ge 1$, L^p is a normed vector space. More is true:

6.6 Theorem. *For $1 \le p < \infty$, L^p is a Banach space.*

Proof. We use Theorem 5.1. Suppose $\{f_k\} \subset L^p$ and $\sum_1^\infty \|f_k\|_p = B < \infty$. Let $G_n = \sum_1^n |f_k|$ and $G = \sum_1^\infty |f_k|$. Then $\|G_n\|_p \le \sum_1^n \|f_k\|_p \le B$ for all n, so by the monotone convergence theorem, $\int G^p = \lim \int G_n^p \le B^p$. Hence $G \in L^p$, and in particular $G(x) < \infty$ a.e., which implies that the series $\sum_1^\infty f_k$ converges a.e. Denoting its sum by F, we have $|F| \le G$ and hence $F \in L^p$; moreover, $|F - \sum_1^n f_k|^p \le (2G)^p \in L^1$, so by the dominated convergence theorem,

$$\left\|F - \sum_1^n f_k\right\|_p^p = \int \left|F - \sum_1^n f_k\right|^p \to 0.$$

Thus the series $\sum_1^\infty f_k$ converges in the L^p norm. ∎

6.7 Proposition. *For $1 \le p < \infty$, the set of simple functions $f = \sum_1^n a_j \chi_{E_j}$, where $\mu(E_j) < \infty$ for all j, is dense in L^p.*

Proof. Clearly such functions are in L^p. If $f \in L^p$, choose a sequence $\{f_n\}$ of simple functions such that $f_n \to f$ a.e. and $|f_n| \le |f|$, according to Theorem 2.10. Then $f_n \in L^p$ and $|f_n - f|^p \le 2^p|f|^p \in L^1$, so by the dominated convergence theorem, $\|f_n - f\|_p \to 0$. Moreover, if $f_n = \sum a_j \chi_{E_j}$ where the E_j are disjoint and the a_j are nonzero, we must have $\mu(E_j) < \infty$ since $\sum |a_j|^p \mu(E_j) = \int |f_n|^p < \infty$. ∎

To complete the picture of L^p spaces, we introduce a space corresponding to the limiting value $p = \infty$. If f is a measurable function on X, we define

$$\|f\|_\infty = \inf\Big\{a \geq 0 : \mu\big(\{x : |f(x)| > a\}\big) = 0\Big\},$$

with the convention that $\inf \varnothing = \infty$. We observe that the infimum is actually attained, for

$$\{x : |f(x)| > a\} = \bigcup_1^\infty \{x : |f(x)| > a + n^{-1}\},$$

and if the sets on the right are null, so is the one on the left. $\|f\|_\infty$ is called the **essential supremum** of $|f|$ and is sometimes written

$$\|f\|_\infty = \operatorname{ess\,sup}_{x \in X} |f(x)|.$$

We now define

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

with the usual convention that two functions that are equal a.e. define the same element of L^∞. Thus $f \in L^\infty$ iff there is a bounded measurable function g such that $f = g$ a.e.; we can take $g = f\chi_E$ where $E = \{x : |f(x)| \leq \|f\|_\infty\}$.

Two remarks: First, for fixed X and $\mathcal{M}$, $L^\infty(X, \mathcal{M}, \mu)$ depends on μ only insofar as μ determines which sets have measure zero; if μ and ν are mutually absolutely continuous, then $L^\infty(\mu) = L^\infty(\nu)$. Second, if μ is not semifinite, for some purposes it is appropriate to adopt a slightly different definition of L^∞. This point will be explored in Exercises 23–25.

The results we have proved for $1 \leq p < \infty$ extend easily to the case $p = \infty$, as follows:

6.8 Theorem.

a. If f and g are measurable functions on X, then $\|fg\|_1 \leq \|f\|_1 \|g\|_\infty$. If $f \in L^1$ and $g \in L^\infty$, $\|fg\|_1 = \|f\|_1 \|g\|_\infty$ iff $|g(x)| = \|g\|_\infty$ a.e. on the set where $f(x) \neq 0$.

b. $\|\cdot\|_\infty$ is a norm on L^∞.

c. $\|f_n - f\|_\infty \to 0$ iff there exists $E \in \mathcal{M}$ such that $\mu(E^c) = 0$ and $f_n \to f$ uniformly on E.

d. L^∞ is a Banach space.

e. The simple functions are dense in L^∞.

The proof is left to the reader (Exercise 2).

In view of Theorem 6.8a and the formal equality $1^{-1} + \infty^{-1} = 1$, it is natural to regard 1 and ∞ as conjugate exponents of each other, and we do so henceforth.

Theorem 6.8c shows that $\|\cdot\|_\infty$ is closely related to, but usually not identical with, the uniform norm $\|\cdot\|_u$. However, if we are dealing with Lebesgue measure, or more generally any Borel measure that assigns positive values to all open sets, then

$\|f\|_\infty = \|f\|_u$ whenever f is continuous, since $\{x : |f(x)| > a\}$ is open. In this situation we may use the notations $\|f\|_\infty$ and $\|f\|_u$ interchangeably, and we may regard the space of bounded continuous functions as a (closed!) subspace of L^∞.

In general we have $L^p \not\subset L^q$ for all $p \neq q$; to see what is at issue, it is instructive to consider the following simple examples on $(0, \infty)$ with Lebesgue measure. Let $f_a(x) = x^{-a}$, where $a > 0$. Elementary calculus shows that $f_a\chi_{(0,1)} \in L^p$ iff $p < a^{-1}$, and $f_a\chi_{(1,\infty)} \in L^p$ iff $p > a^{-1}$. Thus we see two reasons why a function f may fail to be in L^p: either $|f|^p$ blows up too rapidly near some point, or it fails to decay sufficiently rapidly at infinity. In the first situation the behavior of $|f|^p$ becomes worse as p increases, while in the second it becomes better. In other words, if $p < q$, functions in L^p can be locally more singular than functions in L^q, whereas functions in L^q can be globally more spread out than functions in L^p. These somewhat imprecisely expressed ideas are actually a rather accurate guide to the general situation, concerning which we now give four precise results. The last two show that inclusions $L^p \subset L^q$ can be obtained under conditions on the measure space that disallow one of the types of bad behavior described above; for a more general result, see Exercise 5.

6.9 Proposition. *If $0 < p < q < r \leq \infty$, then $L^q \subset L^p + L^r$; that is, each $f \in L^q$ is the sum of a function in L^p and a function in L^r.*

Proof. If $f \in L^q$, let $E = \{x : |f(x)| > 1\}$ and set $g = f\chi_E$ and $h = f\chi_{E^c}$. Then $|g|^p = |f|^p\chi_E \leq |f|^q\chi_E$, so $g \in L^p$, and $|h|^r = |f|^r\chi_{E^c} \leq |f|^q\chi_{E^c}$, so $h \in L^r$. (For $r = \infty$, obviously $\|h\|_\infty \leq 1$.) ∎

6.10 Proposition. *If $0 < p < q < r \leq \infty$, then $L^p \cap L^r \subset L^q$ and $\|f\|_q \leq \|f\|_p^\lambda \|f\|_r^{1-\lambda}$, where $\lambda \in (0,1)$ is defined by*

$$q^{-1} = \lambda p^{-1} + (1-\lambda)r^{-1}, \text{ that is, } \lambda = \frac{q^{-1} - r^{-1}}{p^{-1} - r^{-1}}.$$

Proof. If $r = \infty$, we have $|f|^q \leq \|f\|_\infty^{q-p}|f|^p$ and $\lambda = p/q$, so

$$\|f\|_q \leq \|f\|_p^{p/q}\|f\|_\infty^{1-(p/q)} = \|f\|_p^\lambda \|f\|_\infty^{1-\lambda}.$$

If $r < \infty$, we use Hölder's inequality, taking the pair of conjugate exponents to be $p/\lambda q$ and $r/(1-\lambda)q$:

$$\begin{aligned}\int |f|^q &= \int |f|^{\lambda q}|f|^{(1-\lambda)q} \leq \| \, |f|^{\lambda q}\|_{p/\lambda q} \| \, |f|^{(1-\lambda)q}\|_{r/(1-\lambda)q} \\ &= \left[\int |f|^p\right]^{\lambda q/p} \left[\int |f|^r\right]^{(1-\lambda)q/r} = \|f\|_p^{\lambda q}\|f\|_r^{(1-\lambda)q}.\end{aligned}$$

Taking qth roots, we are done. ∎

6.11 Proposition. *If A is any set and $0 < p < q \leq \infty$, then $l^p(A) \subset l^q(A)$ and $\|f\|_q \leq \|f\|_p$.*

Proof. Obviously $\|f\|_\infty^p = \sup_\alpha |f(\alpha)|^p \leq \sum_\alpha |f(\alpha)|^p$, so that $\|f\|_\infty \leq \|f\|_p$. The case $q < \infty$ then follows from Proposition 6.10: if $\lambda = p/q$,

$$\|f\|_q \leq \|f\|_p^\lambda \|f\|_\infty^{1-\lambda} \leq \|f\|_p.$$

∎

6.12 Proposition. *If $\mu(X) < \infty$ and $0 < p < q \leq \infty$, then $L^p(\mu) \supset L^q(\mu)$ and $\|f\|_p \leq \|f\|_q \mu(X)^{(1/p)-(1/q)}$.*

Proof. If $q = \infty$, this is obvious:

$$\|f\|_p^p = \int |f|^p \leq \|f\|_\infty^p \int 1 = \|f\|_\infty^p \mu(X).$$

If $q < \infty$, we use Hölder's inequality with the conjugate exponents q/p and $q/(q-p)$:

$$\|f\|_p^p = \int |f|^p \cdot 1 \leq \| \, |f|^p \|_{q/p} \|1\|_{q/(q-p)} = \|f\|_q^p \mu(X)^{(q-p)/q}.$$

∎

We conclude this section with a few remarks about the significance of the L^p spaces. The three most obviously important ones are L^1, L^2, and L^∞. With L^1 we are already familiar; L^2 is special because it is a Hilbert space; and the topology on L^∞ is closely related to the topology of uniform convergence. Unfortunately, L^1 and L^∞ are pathological in many respects, and it is more fruitful to deal with the intermediate L^p spaces. One manifestation of this is the duality theory in §6.2; another is the fact that many operators of interest in Fourier analysis and differential equations are bounded on L^p for $1 < p < \infty$ but not on L^1 or L^∞. (Some examples are mentioned in §9.4.)

Exercises

1. When does equality hold in Minkowski's inequality? (The answer is different for $p = 1$ and for $1 < p < \infty$. What about $p = \infty$?)

2. Prove Theorem 6.8.

3. If $1 \leq p < r \leq \infty$, $L^p \cap L^r$ is a Banach space with norm $\|f\| = \|f\|_p + \|f\|_r$, and if $p < q < r$, the inclusion map $L^p \cap L^r \to L^q$ is continuous.

4. If $1 \leq p < r \leq \infty$, $L^p + L^r$ is a Banach space with norm $\|f\| = \inf\{\|g\|_p + \|h\|_r : f = g + h\}$, and if $p < q < r$, the inclusion map $L^q \to L^p + L^r$ is continuous.

5. Suppose $0 < p < q < \infty$. Then $L^p \not\subset L^q$ iff X contains sets of arbitrarily small positive measure, and $L^q \not\subset L^p$ iff X contains sets of arbitrarily large finite measure.

(For the "if" implication: In the first case there is a disjoint sequence $\{E_n\}$ with $0 < \mu(E_n) < 2^{-n}$, and in the second case there is a disjoint sequence $\{E_n\}$ with $1 \leq \mu(E_n) < \infty$. Consider $f = \sum a_n \chi_{E_n}$ for suitable constants a_n.) What about the case $q = \infty$?

6. Suppose $0 < p_0 < p_1 \leq \infty$. Find examples of functions f on $(0, \infty)$ (with Lebesgue measure), such that $f \in L^p$ iff (a) $p_0 < p < p_1$, (b) $p_0 \leq p \leq p_1$, (c) $p = p_0$. (Consider functions of the form $f(x) = x^{-a}|\log x|^b$.)

7. If $f \in L^p \cap L^\infty$ for some $p < \infty$, so that $f \in L^q$ for all $q > p$, then $\|f\|_\infty = \lim_{q\to\infty} \|f\|_q$.

8. Suppose $\mu(X) = 1$ and $f \in L^p$ for some $p > 0$, so that $f \in L^q$ for $0 < q < p$.
a. $\log \|f\|_q \geq \int \log |f|$. (Use Exercise 42d in §3.5, with $F(t) = e^t$.)
b. $(\int |f|^q - 1)/q \geq \log \|f\|_q$, and $(\int |f|^q - 1)/q \to \int \log |f|$ as $q \to 0$.
c. $\lim_{q\to 0} \|f\|_q = \exp(\int \log |f|)$.

9. Suppose $1 \leq p < \infty$. If $\|f_n - f\|_p \to 0$, then $f_n \to f$ in measure, and hence some subsequence converges to f a.e. On the other hand, if $f_n \to f$ in measure and $|f_n| \leq g \in L^p$ for all n, then $\|f_n - f\|_p \to 0$.

10. Suppose $1 \leq p < \infty$. If $f_n, f \in L^p$ and $f_n \to f$ a.e., then $\|f_n - f\|_p \to 0$ iff $\|f_n\|_p \to \|f\|_p$. (Use Exercise 20 in §2.3.)

11. If f is a measurable function on X, define the **essential range** R_f of f to be the set of all $z \in \mathbb{C}$ such that $\{x : |f(x) - z| < \epsilon\}$ has positive measure for all $\epsilon > 0$.
a. R_f is closed.
b. If $f \in L^\infty$, then R_f is compact and $\|f\|_\infty = \max\{|z| : z \in R_f\}$.

12. If $p \neq 2$, the L^p norm does not arise from an inner product on L^p, except in trivial cases when $\dim(L^p) \leq 1$. (Show that the parallelogram law fails.)

13. $L^p(\mathbb{R}^n, m)$ is separable for $1 \leq p < \infty$. However, $L^\infty(\mathbb{R}^n, m)$ is not separable. (There is an uncountable set $\mathcal{F} \subset L^\infty$ such that $\|f - g\|_\infty \geq 1$ for all $f, g \in \mathcal{F}$ with $f \neq g$.)

14. If $g \in L^\infty$, the operator T defined by $Tf = fg$ is bounded on L^p for $1 \leq p \leq \infty$. Its operator norm is at most $\|g\|_\infty$, with equality if μ is semifinite.

15. (**The Vitali Convergence Theorem**) Suppose $1 \leq p < \infty$ and $\{f_n\}_1^\infty \subset L^p$. In order for $\{f_n\}$ to be Cauchy in the L^p norm it is necessary and sufficient for the following three conditions to hold: (i) $\{f_n\}$ is Cauchy in measure; (ii) the sequence $\{|f_n|^p\}$ is uniformly integrable (see Exercise 11 in §3.2); and (iii) for every $\epsilon > 0$ there exists $E \subset X$ such that $\mu(E) < \infty$ and $\int_{E^c} |f_n|^p < \epsilon$ for all n. (To prove the sufficiency: Given $\epsilon > 0$, let E be as in (iii), and let $A_{mn} = \{x \in E : |f_m(x) - f_n(x)| \geq \epsilon\}$. Then the integrals of $|f_n - f_m|^p$ over $E \setminus A_{mn}$, A_{mn}, and E^c are small when m and n are large — for three different reasons.)

16. If $0 < p < 1$, the formula $\rho(f, g) = \int |f - g|^p$ defines a metric on L^p that makes L^p into a complete topological vector space. (The proof of Theorem 6.6 still works

for $p < 1$ if $\|f\|_p$ is replaced by $\int |f|^p$, as it uses only the triangle inequality and not the homogeneity of the norm.)

6.2 THE DUAL OF L^P

Suppose that p and q are conjugate exponents. Hölder's inequality shows that each $g \in L^q$ defines a bounded linear functional ϕ_g on L^p by

$$\phi_g(f) = \int fg,$$

and the operator norm of ϕ_g is at most $\|g\|_q$. (If $p = 2$ and we are thinking of L^2 as a Hilbert space, it is more appropriate to define $\phi_g(f) = \int f\overline{g}$. The same convention can be used for $p \neq 2$ without changing the results below in an essential way.) In fact, the map $g \to \phi_g$ is almost always an isometry from L^q into $(L^p)^*$.

6.13 Proposition. *Suppose that p and q are conjugate exponents and $1 \leq q < \infty$. If $g \in L^q$, then*

$$\|g\|_q = \|\phi_g\| = \sup\left\{\left|\int fg\right| : \|f\|_p = 1\right\}.$$

If μ is semifinite, this result holds also for $q = \infty$.

Proof. Hölder's inequality says that $\|\phi_g\| \leq \|g\|_q$, and equality is trivial if $g = 0$ (a.e.). If $g \neq 0$ and $q < \infty$, let

$$f = \frac{|g|^{q-1}\overline{\operatorname{sgn} g}}{\|g\|_q^{q-1}}.$$

Then

$$\|f\|_p^p = \frac{\int |g|^{(q-1)p}}{\|g\|_q^{(q-1)p}} = \frac{\int |g|^q}{\int |g|^q} = 1,$$

so

$$\|\phi_g\| \geq \int fg = \frac{\int |g|^q}{\|g\|_q^{q-1}} = \|g\|_q.$$

(If $q = 1$, then $f = \overline{\operatorname{sgn} g}$, $\|f\|_\infty = 1$, and $\int fg = \|g\|_1$.) If $q = \infty$, for $\epsilon > 0$ let $A = \{x : |g(x)| > \|g\|_\infty - \epsilon\}$. Then $\mu(A) > 0$, so if μ is semifinite there exists $B \subset A$ with $0 < \mu(B) < \infty$. Let $f = \mu(B)^{-1}\chi_B\overline{\operatorname{sgn} g}$; then $\|f\|_1 = 1$, so

$$\|\phi_g\| \geq \int fg = \frac{1}{\mu(B)}\int_B |g| \geq \|g\|_\infty - \epsilon.$$

Since ϵ is arbitrary, $\|\phi_g\| = \|g\|_\infty$. ∎

Conversely, if $f \to \int fg$ is a bounded linear functional on L^p, then $g \in L^q$ in almost all cases. In fact, we have the following stronger result.

6.14 Theorem. *Let p and q be conjugate exponents. Suppose that g is a measurable function on X such that $fg \in L^1$ for all f in the space Σ of simple functions that vanish outside a set of finite measure, and the quantity*

$$M_q(g) = \sup\left\{\left|\int fg\right| : f \in \Sigma \text{ and } \|f\|_p = 1\right\}$$

is finite. Also, suppose either that $S_g = \{x : g(x) \neq 0\}$ is σ-finite or that μ is semifinite. Then $g \in L^q$ and $M_q(g) = \|g\|_q$.

Proof. First, we remark that if f is a bounded measurable function that vanishes outside a set E of finite measure and $\|f\|_p = 1$, then $|\int fg| \leq M_q(g)$. Indeed, by Theorem 2.10 there is a sequence $\{f_n\}$ of simple functions such that $|f_n| \leq |f|$ (in particular, f_n vanishes outside E) and $f_n \to f$ a.e. Since $|f_n| \leq \|f\|_\infty \chi_E$ and $\chi_E g \in L^1$, by the dominated convergence theorem we have $|\int fg| = \lim |\int f_n g| \leq M_q(g)$.

Now suppose that $q < \infty$. We may assume that S_g is σ-finite, as this condition automatically holds when μ is semifinite; see Exercise 17. Let $\{E_n\}$ be an increasing sequence of sets of finite measure such that $S_g = \bigcup_1^\infty E_n$. Let $\{\phi_n\}$ be a sequence of simple functions such that $\phi_n \to g$ pointwise and $|\phi_n| \leq |g|$, and let $g_n = \phi_n \chi_{E_n}$. Then $g_n \to g$ pointwise, $|g_n| \leq |g|$, and g_n vanishes outside E_n. Let

$$f_n = \frac{|g_n|^{q-1}\overline{\operatorname{sgn} g}}{\|g_n\|_q^{q-1}}.$$

Then as in the proof of Proposition 6.13 we have $\|f_n\|_p = 1$, and by Fatou's lemma,

$$\begin{aligned}\|g\|_q &\leq \liminf \|g_n\|_q = \liminf \int |f_n g_n| \\ &\leq \liminf \int |f_n g| = \liminf \int f_n g \leq M_q(g).\end{aligned}$$

(For the last estimate we used the remark at the beginning of the proof.) On the other hand, Hölder's inequality gives $M_q(g) \leq \|g\|_q$, so the proof is complete for the case $q < \infty$.

Now suppose $q = \infty$. Given $\epsilon > 0$, let $A = \{x : |g(x)| \geq M_\infty(g) + \epsilon\}$. If $\mu(A)$ were positive, we could choose $B \subset A$ with $0 < \mu(B) < \infty$ (either because μ is semifinite or because $A \subset S_g$). Setting $f = \mu(B)^{-1}\chi_B \overline{\operatorname{sgn} g}$, we would then have $\|f\|_1 = 1$, and $\int fg = \mu(B)^{-1} \int_B |g| \geq M_\infty(g) + \epsilon$. But this is impossible by the remark at the beginning of the proof. Hence $\|g\|_\infty \leq M_\infty(g)$, and the reverse inequality is obvious. ∎

The last and deepest part of the description of $(L^p)^*$ is the fact that the map $g \to \phi_g$ is, in almost all cases, a surjection.

6.15 Theorem. *Let p and q be conjugate exponents. If $1 < p < \infty$, for each $\phi \in (L^p)^*$ there exists $g \in L^q$ such that $\phi(f) = \int fg$ for all $f \in L^p$, and hence L^q is isometrically isomorphic to $(L^p)^*$. The same conclusion holds for $p = 1$ provided μ is σ-finite.*

Proof. First let us suppose that μ is finite, so that all simple functions are in L^p. If $\phi \in (L^p)^*$ and E is a measurable set, let $\nu(E) = \phi(\chi_E)$. For any disjoint sequence $\{E_j\}$, if $E = \bigcup_1^\infty E_j$ we have $\chi_E = \sum_1^\infty \chi_{E_j}$ where the series converges in the L^p norm:

$$\Big\|\chi_E - \sum_1^n \chi_{E_j}\Big\|_p = \Big\|\sum_{n+1}^\infty \chi_{E_j}\Big\|_p = \mu\Big(\bigcup_{n+1}^\infty E_j\Big)^{1/p} \to 0 \text{ as } n \to \infty.$$

(It is at this point that we need the assumption that $p < \infty$.) Hence, since ϕ is linear and continuous,

$$\nu(E) = \sum_1^\infty \phi(\chi_{E_j}) = \sum_1^\infty \nu(E_j),$$

so that ν is a complex measure. Also, if $\mu(E) = 0$, then $\chi_E = 0$ as an element of L^p, so $\nu(E) = 0$; that is, $\nu \ll \mu$. By the Radon-Nikodym theorem there exists $g \in L^1(\mu)$ such that $\phi(\chi_E) = \nu(E) = \int_E g\, d\mu$ for all E and hence $\phi(f) = \int fg\, d\mu$ for all simple functions f. Moreover, $|\int fg| \le \|\phi\|\, \|f\|_p$, so $g \in L^q$ by Theorem 6.14. Once we know this, it follows from Proposition 6.7 that $\phi(f) = \int fg$ for all $f \in L^p$.

Now suppose that μ is σ-finite. Let $\{E_n\}$ be an increasing sequence of sets such that $0 < \mu(E_n) < \infty$ and $X = \bigcup_1^\infty E_n$, and let us agree to identify $L^p(E_n)$ and $L^q(E_n)$ with the subspaces of $L^p(X)$ and $L^q(X)$ consisting of functions that vanish outside E_n. The preceding argument shows that for each n there exists $g_n \in L^q(E_n)$ such that $\phi(f) = \int fg_n$ for all $f \in L^p(E_n)$, and $\|g_n\|_q = \|\phi|L^p(E_n)\| \le \|\phi\|$. The function g_n is unique modulo alterations on nullsets, so $g_n = g_m$ a.e. on E_n for $n < m$, and we can define g a.e. on X by setting $g = g_n$ on E_n. By the monotone convergence theorem, $\|g\|_q = \lim \|g_n\|_q \le \|\phi\|$, so $g \in L^q$. Moreover, if $f \in L^p$, then by the dominated convergence theorem, $f\chi_{E_n} \to f$ in the L^p norm and hence $\phi(f) = \lim \phi(f\chi_{E_n}) = \lim \int_{E_n} fg = \int fg$.

Finally, suppose that μ is arbitrary and $p > 1$, so that $q < \infty$. As above, for each σ-finite set $E \subset X$ there is an a.e.-unique $g_E \in L^q(E)$ such that $\phi(f) = \int fg_E$ for all $f \in L^p(E)$ and $\|g_E\|_q \le \|\phi\|$. If F is σ-finite and $F \supset E$, then $g_F = g_E$ a.e. on E, so $\|g_F\|_q \ge \|g_E\|_q$. Let M be the supremum of $\|g_E\|_q$ as E ranges over all σ-finite sets, noting that $M \le \|\phi\|$. Choose a sequence $\{E_n\}$ so that $\|g_{E_n}\|_q \to M$, and set $F = \bigcup_1^\infty E_n$. Then F is σ-finite and $\|g_F\|_q \ge \|g_{E_n}\|_q$ for all n, whence $\|g_F\|_q = M$. Now, if A is a σ-finite set containing F, we have

$$\int |g_F|^q + \int |g_{A\setminus F}|^q = \int |g_A|^q \le M^q = \int |g_F|^q,$$

and thus $g_{A\setminus F} = 0$ and $g_A = g_F$ a.e. (Here we use the fact that $q < \infty$.) But if $f \in L^p$, then $A = F \cup \{x : f(x) \ne 0\}$ is σ-finite, so $\phi(f) = \int fg_A = \int fg_F$. Thus we may take $g = g_F$, and the proof is complete. ∎

6.16 Corollary. *If $1 < p < \infty$, L^p is reflexive.*

We conclude with some remarks on the exceptional cases $p = 1$ and $p = \infty$. For any measure μ, the correspondence $g \mapsto \phi_g$ maps L^∞ into $(L^1)^*$, but in general it is neither injective nor surjective. Injectivity fails when μ is not semifinite. Indeed, if $E \subset X$ is a set of infinite measure that contains no subsets of positive finite measure, and $f \in L^1$, then $\{x : f(x) \neq 0\}$ is σ-finite and hence intersects E in a null set. It follows that $\phi_{\chi_E} = 0$ although $\chi_E \neq 0$ in L^∞. This problem, however, can be remedied by redefining L^∞; see Exercises 23–24. The failure of surjectivity is more subtle and is best illustrated by an example; see also Exercise 25.

Let X be an uncountable set, $\mu =$ counting measure on $(X, \mathcal{P}(X))$, $\mathcal{M} =$ the σ-algebra of countable or co-countable sets, and $\mu_0 =$ the restriction of μ to $\mathcal{M}$. Every $f \in L^1(\mu)$ vanishes outside a countable set, and it follows that $L^1(\mu) = L^1(\mu_0)$. On the other hand, $L^\infty(\mu)$ consists of all bounded functions on X, whereas $L^\infty(\mu_0)$ consists of those bounded functions that are constant except on a countable set. With this in mind, it is easy to see that the dual of $L^1(\mu_0)$ is $L^\infty(\mu)$ and not the smaller space $L^\infty(\mu_0)$.

As for the case $p = \infty$: the map $g \to \phi_g$ is always an isometric injection of L^1 into $(L^\infty)^*$ by Proposition 6.13, but it is almost *never* a surjection. We shall say more about this in §6.6; for the present, we give a specific example. (Another example can be found in Exercise 19.)

Let $X = [0, 1]$, $\mu =$ Lebesgue measure. The map $f \mapsto f(0)$ is a bounded linear functional on $C(X)$, which we regard as a subspace of L^∞. By the Hahn-Banach theorem there exists $\phi \in (L^\infty)^*$ such that $\phi(f) = f(0)$ for all $f \in C(X)$. To see that ϕ cannot be given by integration against an L^1 function, consider the functions $f_n \in C(X)$ defined by $f_n(x) = \max(1-nx, 0)$. Then $\phi(f_n) = f_n(0) = 1$ for all n, but $f_n(x) \to 0$ for all $x > 0$, so by the dominated convergence theorem, $\int f_n g \to 0$ for all $g \in L^1$.

Exercises

17. With notation as in Theorem 6.14, if μ is semifinite, $q < \infty$, and $M_q(g) < \infty$, then $\{x : |g(x)| > \epsilon\}$ has finite measure for all $\epsilon > 0$ and hence S_g is σ-finite.

18. The self-duality of L^2 follows from Hilbert space theory (Theorem 5.25), and this fact can be used to prove the Lebesgue-Radon-Nikodym theorem by the following argument due to von Neumann. Suppose that μ, ν are positive finite measures on $(X, \mathcal{M})$ (the σ-finite case follows easily as in §3.2), and let $\lambda = \mu + \nu$.

a. The map $f \mapsto \int f\, d\nu$ is a bounded linear functional on $L^2(\lambda)$, so $\int f\, d\nu = \int fg\, d\lambda$ for some $g \in L^2(\lambda)$. Equivalently, $\int f(1-g)\, d\nu = \int fg\, d\mu$ for $f \in L^2(\lambda)$.

b. $0 \leq g \leq 1$ λ-a.e., so we may assume $0 \leq g \leq 1$ everywhere.

c. Let $A = \{x : g(x) < 1\}$, $B = \{x : g(x) = 1\}$, and set $\nu_a(E) = \nu(A \cap E)$, $\nu_s(E) = \nu(B \cap E)$. Then $\nu_s \perp \mu$ and $\nu_a \ll \mu$; in fact, $d\nu_a = g(1-g)^{-1}\chi_A\, d\mu$.

19. Define $\phi_n \in (l^\infty)^*$ by $\phi_n(f) = n^{-1}\sum_1^n f(j)$. Then the sequence $\{\phi_n\}$ has a weak* cluster point ϕ, and ϕ is an element of $(l^\infty)^*$ that does not arise from an element of l^1.

20. Suppose $\sup_n \|f_n\|_p < \infty$ and $f_n \to f$ a.e.

a. If $1 < p < \infty$, then $f_n \to f$ weakly in L^p. (Given $g \in L^q$, where q is conjugate to p, and $\epsilon > 0$, there exist (i) $\delta > 0$ such that $\int_E |g|^q < \epsilon$ whenever $\mu(E) < \delta$, (ii) $A \subset X$ such that $\mu(A) < \infty$ and $\int_{X\setminus A} |g|^q < \epsilon$, and (iii) $B \subset A$ such that $\mu(A \setminus B) < \delta$ and $f_n \to f$ uniformly on B.)

b. The result of (a) is false in general for $p = 1$. (Find counterexamples in $L^1(\mathbb{R}, m)$ and l^1.) It is, however, true for $p = \infty$ if μ is σ-finite and weak convergence is replaced by weak* convergence.

21. If $1 < p < \infty$, $f_n \to f$ weakly in $l^p(A)$ iff $\sup_n \|f_n\|_p < \infty$ and $f_n \to f$ pointwise.

22. Let $X = [0, 1]$, with Lebesgue measure.

a. Let $f_n(x) = \cos 2\pi nx$. Then $f_n \to 0$ weakly in L^2 (see Exercise 63 in §5.5), but $f_n \not\to 0$ a.e. or in measure.

b. Let $f_n(x) = n\chi_{(0,1/n)}$. Then $f_n \to 0$ a.e. and in measure, but $f_n \not\to 0$ weakly in L^p for any p.

23. Let $(X, \mathcal{M}, \mu)$ be a measure space. A set $E \in \mathcal{M}$ is called **locally null** if $\mu(E \cap F) = 0$ for every $F \in \mathcal{M}$ such that $\mu(F) < \infty$. If $f : X \to \mathbb{C}$ is a measurable function, define

$$\|f\|_* = \inf\{a : \{x : |f(x)| > a\} \text{ is locally null}\},$$

and let $\mathcal{L}^\infty = \mathcal{L}^\infty(X, \mathcal{M}, \mu)$ be the space of all measurable f such that $\|f\|_* < \infty$. We consider $f, g \in \mathcal{L}^\infty$ to be identical if $\{x : f(x) \neq g(x)\}$ is locally null.

a. If E is locally null, then $\mu(E)$ is either 0 or ∞. If μ is semifinite, then every locally null set is null.

b. $\|\cdot\|_*$ is a norm on $\mathcal{L}^\infty$ that makes $\mathcal{L}^\infty$ into a Banach space. If μ is semifinite, then $\mathcal{L}^\infty = L^\infty$.

24. If $g \in \mathcal{L}^\infty$ (see Exercise 23), then $\|g\|_* = \sup\{|\int fg| : \|f\|_1 = 1\}$, so the map $g \mapsto \phi_g$ is an isometry from $\mathcal{L}^\infty$ into $(L^1)^*$. Conversely, if $M_\infty(g) < \infty$ as in Theorem 6.14, then $g \in \mathcal{L}^\infty$ and $M_\infty(g) = \|g\|_*$.

25. Suppose μ is decomposable (see Exercise 15 in §3.2). Then every $\phi \in (L^1)^*$ is of the form $\phi(f) = \int fg$ for some $g \in \mathcal{L}^\infty$, and hence $(L^1)^* \cong \mathcal{L}^\infty$ (see Exercises 23 and 24). (If $\mathcal{F}$ is a decomposition of μ and $f \in L^1$, there exists $\{E_j\} \subset \mathcal{F}$ such that $f = \sum_1^\infty f\chi_{E_j}$ where the series converges in L^1.)

6.3 SOME USEFUL INEQUALITIES

Estimates and inequalities lie at the heart of the applications of L^p spaces in analysis. The most basic of these are the Hölder and Minkowski inequalities. In this section we present a few additional important results in this area. The first one is almost a triviality, but it is sufficiently useful to warrant special mention.

6.17 Chebyshev's Inequality. *If $f \in L^p$ $(0 < p < \infty)$, then for any $\alpha > 0$,*

$$\mu(\{x : |f(x)| > \alpha\}) \leq \left[\frac{\|f\|_p}{\alpha}\right]^p.$$

Proof. Let $E_\alpha = \{x : |f(x)| > \alpha\}$. Then

$$\|f\|_p^p = \int |f|^p \geq \int_{E_\alpha} |f|^p \geq \alpha^p \int_{E_\alpha} 1 = \alpha^p \mu(E_\alpha).$$

∎

The next result is a rather general theorem about boundedness of integral operators on L^p spaces.

6.18 Theorem. *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces, and let K be an $(\mathcal{M} \otimes \mathcal{N})$-measurable function on $X \times Y$. Suppose that there exists $C > 0$ such that $\int |K(x,y)|\, d\mu(x) \leq C$ for a.e. $y \in Y$ and $\int |K(x,y)|\, d\nu(y) \leq C$ for a.e. $x \in X$, and that $1 \leq p \leq \infty$. If $f \in L^p(\nu)$, the integral*

$$Tf(x) = \int K(x,y) f(y)\, d\nu(y)$$

converges absolutely for a.e. $x \in X$, the function Tf thus defined is in $L^p(\mu)$, and $\|Tf\|_p \leq C\|f\|_p$.

Proof. Suppose that $1 < p < \infty$. Let q be the conjugate exponent to p. By applying Hölder's inequality to the product

$$|K(x,y)f(y)| = |K(x,y)|^{1/q}\left(|K(x,y)|^{1/p}|f(y)|\right)$$

we have

$$\begin{aligned}\int |K(x,y)f(y)|\, d\nu(y) &\leq \left[\int |K(x,y)|\, d\nu(y)\right]^{1/q} \left[\int |K(x,y)|\, |f(y)|^p\, d\nu(y)\right]^{1/p} \\ &\leq C^{1/q}\left[\int |K(x,y)|\, |f(y)|^p\, d\nu(y)\right]^{1/p}\end{aligned}$$

for a.e. $x \in X$. Hence, by Tonelli's theorem,

$$\begin{aligned}\int \left[\int |K(x,y)f(y)|\, d\nu(y)\right]^p d\mu(x) &\leq C^{p/q} \iint |K(x,y)|\, |f(y)|^p\, d\nu(y)\, d\mu(x) \\ &\leq C^{(p/q)+1} \int |f(y)|^p\, d\nu(y).\end{aligned}$$

Since the last integral is finite, Fubini's theorem implies that $K(x,\cdot)f \in L^1(\nu)$ for a.e. x, so that Tf is well defined a.e., and

$$\int |Tf(x)|^p \, d\mu(x) \le C^{(p/q)+1} \|f\|_p^p.$$

Taking pth roots, we are done.

For $p = 1$ the proof is similar but easier and requires only the hypothesis $\int |K(x,y)| \, d\mu(x) \le C$; for $p = \infty$ the proof is trivial and requires only the hypothesis $\int |K(x,y)| \, d\nu(y) \le C$. Details are left to the reader (Exercise 26). ∎

Minkowski's inequality states that the L^p norm of a sum is at most the sum of the L^p norms. There is a generalization of this result in which sums are replaced by integrals:

6.19 Minkowski's Inequality for Integrals. *Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces, and let f be an $(\mathcal{M} \otimes \mathcal{N})$-measurable function on $X \times Y$.*

a. If $f \ge 0$ and $1 \le p < \infty$, then

$$\left[\int \left(\int f(x,y) \, d\nu(y)\right)^p d\mu(x)\right]^{1/p} \le \int \left[\int f(x,y)^p \, d\mu(x)\right]^{1/p} d\nu(y).$$

b. If $1 \le p \le \infty$, $f(\cdot, y) \in L^p(\mu)$ for a.e. y, and the function $y \mapsto \|f(\cdot,y)\|_p$ is in $L^1(\nu)$, then $f(x,\cdot) \in L^1(\nu)$ for a.e. x, the function $x \mapsto \int f(x,y) \, d\nu(y)$ is in $L^p(\mu)$, and

$$\left\| \int f(\cdot, y) \, d\nu(y) \right\|_p \le \int \|f(\cdot,y)\|_p \, d\nu(y).$$

Proof. If $p = 1$, (a) is merely Tonelli's theorem. If $1 < p < \infty$, let q be the conjugate exponent to p and suppose $g \in L^q(\mu)$. Then by Tonelli's theorem and Hölder's inequality,

$$\begin{aligned}\int \left[\int f(x,y) \, d\nu(y)\right] |g(x)| \, d\mu(x) &= \iint f(x,y)|g(x)| \, d\mu(x) \, d\nu(y) \\ &\le \|g\|_q \int \left[\int f(x,y)^p \, d\mu(x)\right]^{1/p} d\nu(y).\end{aligned}$$

Assertion (a) therefore follows from Theorem 6.14. When $p < \infty$, (b) follows from (a) (with f replaced by $|f|$) and Fubini's theorem; when $p = \infty$, it is a simple consequence of the monotonicity of the integral. ∎

Our final result is a theorem concerning integral operators on $(0, \infty)$ with Lebesgue measure.

6.20 Theorem. *Let K be a Lebesgue measurable function on $(0,\infty)\times(0,\infty)$ such that $K(\lambda x,\lambda y)=\lambda^{-1}K(x,y)$ for all $\lambda>0$ and $\int_0^\infty |K(x,1)|x^{-1/p}\,dx = C<\infty$ for some $p\in[1,\infty]$, and let q be the conjugate exponent to p. For $f\in L^p$ and $g\in L^q$, let*

$$Tf(y)=\int_0^\infty K(x,y)f(x)\,dx, \qquad Sg(x)=\int_0^\infty K(x,y)g(y)\,dy.$$

Then Tf and Sg are defined a.e., and $\|Tf\|_p\le C\|f\|_p$ and $\|Sg\|_q\le C\|g\|_q$.

Proof. Setting $z=x/y$, we have

$$\int_0^\infty |K(x,y)f(x)|\,dx=\int_0^\infty |K(yz,y)f(yz)|y\,dz=\int_0^\infty |K(z,1)f_z(y)|\,dz$$

where $f_z(y)=f(yz)$; moreover,

$$\|f_z\|_p=\left[\int_0^\infty |f(yz)|^p\,dy\right]^{1/p}=\left[\int_0^\infty |f(x)|^p z^{-1}\,dx\right]^{1/p}=z^{-1/p}\|f\|_p.$$

Therefore, by Minkowski's inequality for integrals, Tf exists a.e. and

$$\|Tf\|_p\le\int_0^\infty |K(z,1)|\,\|f_z\|_p\,dz=\|f\|_p\int_0^\infty |K(z,1)|z^{-1/p}\,dz=C\|f\|_p.$$

Finally, setting $u=y^{-1}$, we have

$$\begin{aligned}\int_0^\infty |K(1,y)|y^{-1/q}\,dy&=\int_0^\infty |K(y^{-1},1)|y^{-1-(1/q)}\,dy\\&=\int_0^\infty |K(u,1)|u^{-1/p}\,du=C,\end{aligned}$$

so the same reasoning shows that Sg is defined a.e. and that $\|Sg\|_q\le C\|g\|_q$. ∎

6.21 Corollary. *Let*

$$Tf(y)=y^{-1}\int_0^y f(x)\,dx, \qquad Sg(x)=\int_x^\infty y^{-1}g(y)\,dy.$$

Then for $1<p\le\infty$ and $1\le q<\infty$,

$$\|Tf\|_p\le\frac{p}{p-1}\|f\|_p, \qquad \|Sg\|_q\le q\|g\|_q.$$

Proof. Let $K(x,y)=y^{-1}\chi_E(x,y)$ where $E=\{(x,y):x<y\}$. Then $\int_0^\infty |K(x,1)|x^{-1/p}\,dx=\int_0^1 x^{-1/p}\,dx=p/(p-1)=q$, where q is the conjugate exponent to p, so Theorem 6.20 yields the result. ∎

Corollary 6.21 is a special case of **Hardy's inequalities**; the general result is in Exercise 29.

Exercises

26. Complete the proof of Theorem 6.18 for the cases $p = 1$ and $p = \infty$.

27. (**Hilbert's Inequality**) The operator $Tf(x) = \int_0^\infty (x+y)^{-1} f(y)\, dy$ satisfies $\|Tf\|_p \le C_p \|f\|_p$ for $1 < p < \infty$, where $C_p = \int_0^\infty x^{-1/p}(x+1)^{-1}\, dx$. (For those who know about contour integrals: Show that $C_p = \pi \csc(\pi/p)$.)

28. Let I_α be the αth fractional integral operator as in Exercise 61 of §2.6, and let $J_\alpha f(x) = x^{-\alpha} I_\alpha f(x)$.

a. J_α is bounded on $L^p(0,\infty)$ for $1 < p \le \infty$; more precisely,

$$\|J_\alpha f\|_p \le \frac{\Gamma(1-p^{-1})}{\Gamma(\alpha+1-p^{-1})} \|f\|_p.$$

b. There exists $f \in L^1(0,\infty)$ such that $J_1 f \notin L^1(0,\infty)$.

29. Suppose that $1 \le p < \infty$, $r > 0$, and h is a nonnegative measurable function on $(0,\infty)$. Then:

$$\int_0^\infty x^{-r-1} \left[\int_0^x h(y)\, dy\right]^p dx \le \left(\frac{p}{r}\right)^p \int_0^\infty x^{p-r-1} h(x)^p\, dx,$$
$$\int_0^\infty x^{r-1} \left[\int_x^\infty h(y)\, dy\right]^p dx \le \left(\frac{p}{r}\right)^p \int_0^\infty x^{p+r-1} h(x)^p\, dx.$$

(Apply Theorem 6.20 with $K(x,y) = x^{\beta-1} y^{-\beta} \chi_{(0,\infty)}(y-x)$, $f(x) = x^\gamma h(x)$, and $g(x) = x^\delta h(x)$ for suitable β, γ, δ.)

30. Suppose that K is a nonnegative measurable function on $(0,\infty)$ such that $\int_0^\infty K(x) x^{s-1}\, dx = \phi(s) < \infty$ for $0 < s < 1$.

a. If $1 < p < \infty$, $p^{-1} + q^{-1} = 1$, and f, g are nonnegative measurable functions on $(0,\infty)$, then (with $\int = \int_0^\infty$)

$$\iint K(xy) f(x) g(y)\, dx\, dy \le \phi(p^{-1}) \left[\int x^{p-2} f(x)^p\, dx\right]^{1/p} \left[\int g(x)^q\, dx\right]^{1/q}.$$

b. The operator $Tf(x) = \int_0^\infty K(xy) f(y)\, dy$ is bounded on $L^2((0,\infty))$ with norm $\le \phi(\frac{1}{2})$. (Interesting special case: If $K(x) = e^{-x}$, then T is the Laplace transform and $\phi(s) = \Gamma(s)$.)

31. (**A Generalized Hölder Inequality**) Suppose that $1 \le p_j \le \infty$ and $\sum_1^n p_j^{-1} = r^{-1} \le 1$. If $f_j \in L^{p_j}$ for $j = 1, \ldots, n$, then $\prod_1^n f_j \in L^r$ and $\|\prod_1^n f_j\|_r \le \prod_1^n \|f_j\|_{p_j}$. (First do the case $n = 2$.)

32. Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces and $K \in L^2(\mu \times \nu)$. If $f \in L^2(\nu)$, the integral $Tf(x) = \int K(x,y)f(y)\,d\nu(y)$ converges absolutely for a.e. $x \in X$; moreover, $Tf \in L^2(\mu)$ and $\|Tf\|_2 \le \|K\|_2\|f\|_2$.

33. Given $1 < p < \infty$, let $Tf(x) = x^{-1/p}\int_0^x f(t)\,dt$. If $p^{-1} + q^{-1} = 1$, then T is a bounded linear map from $L^q((0,\infty))$ to $C_0((0,\infty))$.

34. If f is absolutely continuous on $[\epsilon, 1]$ for $0 < \epsilon < 1$ and $\int_0^1 x|f'(x)|^p\,dx < \infty$, then $\lim_{x\to 0} f(x)$ exists (and is finite) if $p > 2$, $|f(x)|/|\log x|^{1/2} \to 0$ as $x \to 0$ if $p = 2$, and $|f(x)|/x^{1-(2/p)} \to 0$ as $x \to 0$ if $p < 2$.

6.4 DISTRIBUTION FUNCTIONS AND WEAK L^P

If f is a measurable function on $(X, \mathcal{M}, \mu)$, we define its **distribution function** $\lambda_f : (0,\infty) \to [0,\infty]$ by

$$\lambda_f(\alpha) = \mu(\{x : |f(x)| > \alpha\}).$$

(This is closely related, but not identical, to the "distribution functions" discussed in §1.5 and §10.1.) We compile the basic properties of λ_f in a proposition:

6.22 Proposition.

a. *λ_f is decreasing and right continuous.*

b. *If $|f| \le |g|$, then $\lambda_f \le \lambda_g$.*

c. *If $|f_n|$ increases to $|f|$, then λ_{f_n} increases to λ_f.*

d. *If $f = g + h$, then $\lambda_f(\alpha) \le \lambda_g(\frac{1}{2}\alpha) + \lambda_h(\frac{1}{2}\alpha)$.*

Proof. Let $E(\alpha, f) = \{x : |f(x)| > \alpha\}$. The function λ_f is decreasing since $E(\alpha, f) \supset E(\beta, f)$ if $\alpha < \beta$, and it is right continuous since $E(\alpha, f)$ is the increasing union of $\{E(\alpha + n^{-1}, f)\}_1^\infty$. If $|f| \le |g|$, then $E(\alpha, f) \subset E(\alpha, g)$, so $\lambda_f \le \lambda_g$. If $|f_n|$ increases to $|f|$, then $E(\alpha, f)$ is the increasing union of $\{E(\alpha, f_n)\}$, so λ_{f_n} increases to λ_f. Finally, if $f = g + h$, then $E(\alpha, f) \subset E(\frac{1}{2}\alpha, g) \cup E(\frac{1}{2}\alpha, h)$, which implies that $\lambda_f(\alpha) \le \lambda_g(\frac{1}{2}\alpha) + \lambda_h(\frac{1}{2}\alpha)$. ∎

Suppose that $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$. In view of Proposition 6.22a, λ_f defines a negative Borel measure ν on $(0,\infty)$ such that $\nu((a,b]) = \lambda_f(b) - \lambda_f(a)$ whenever $0 < a < b$. (Our construction of Borel measures on $\mathbb{R}$ in §1.5 works equally well on $(0,\infty)$.) We can therefore consider the Lebesgue-Stieltjes integrals $\int \phi\,d\lambda_f = \int \phi\,d\nu$ of functions ϕ on $(0,\infty)$. The following result shows that the integrals of functions of $|f|$ on X can be reduced to such Lebesgue-Stieltjes integrals.

6.23 Proposition. *If $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$ and ϕ is a nonnegative Borel measurable function on $(0,\infty)$, then*

$$\int_X \phi \circ |f|\,d\mu = -\int_0^\infty \phi(\alpha)d\lambda_f(\alpha).$$

Proof. If ν is the negative measure determined by λ_f, we have

$$\nu((a,b]) = \lambda_f(b) - \lambda_f(a) = -\mu\big(\{x : a < |f(x)| \le b\}\big) = -\mu\big(|f|^{-1}((a,b])\big).$$

It follows that $\nu(E) = -\mu(|f|^{-1}(E))$ for all Borel sets $E \subset (0,\infty)$, by the uniqueness of extensions (Theorem 1.14). But this means that $\int_X \phi \circ |f|\, d\mu = -\int_0^\infty \phi(\alpha)\, d\lambda_f(\alpha)$ when ϕ is the characteristic function of a Borel set, and hence when ϕ is simple. The general case then follows by virtue of Theorem 2.10 and the monotone convergence theorem. ∎

The case of this result in which we are most interested is $\phi(\alpha) = \alpha^p$, which gives

$$\int |f|^p\, d\mu = -\int_0^\infty \alpha^p\, d\lambda_f(\alpha).$$

A more useful form of this equation is obtained by integrating the right side by parts (Theorem 3.36) to obtain $\int |f|^p\, d\mu = p\int_0^\infty \alpha^{p-1}\lambda_f(\alpha)\, d\alpha$. The validity of this calculation is not clear unless we know that $\alpha^p\lambda_f(\alpha) \to 0$ as $\alpha \to 0$ and $\alpha \to \infty$; nonetheless, the conclusion is correct.

6.24 Proposition. *If* $0 < p < \infty$, *then*

$$\int |f|^p\, d\mu = p\int_0^\infty \alpha^{p-1}\lambda_f(\alpha)\, d\alpha.$$

Proof. If $\lambda_f(\alpha) = \infty$ for some $\alpha > 0$, then both integrals are infinite. If not, and f is simple, then λ_f is bounded as $\alpha \to 0$ and vanishes for α sufficiently large, so the integration by parts described above works. (It is also easy to verify the formula directly in this case.) For the general case, let $\{g_n\}$ be a sequence of simple functions that increases to $|f|$; then the desired result is true for g_n, and it follows for f by Proposition 6.22c and the monotone convergence theorem. ∎

A variant of the L^p spaces that turns up rather often is the following. If f is a measurable function on X and $0 < p < \infty$, we define

$$[f]_p = \Big(\sup_{\alpha>0} \alpha^p\lambda_f(\alpha)\Big)^{1/p},$$

and we define **weak** L^p to be the set of all f such that $[f]_p < \infty$. $[\cdot]_p$ is not a norm; it is easily checked that $[cf]_p = |c|[f]_p$, but the triangle inequality fails. However, weak L^p is a topological vector space; see Exercise 35.

The relationship between L^p and weak L^p is as follows. On the one hand,

$$L^p \subset \text{weak } L^p, \quad \text{and} \quad [f]_p \le \|f\|_p.$$

(This is just a restatement of Chebyshev's inequality.) On the other hand, if we replace $\lambda_f(\alpha)$ by $([f]_p/\alpha)^p$ in the integral $p\int_0^\infty \alpha^{p-1}\lambda_f(\alpha)\, d\alpha$, which equals $\|f\|_p^p$, we obtain a constant times $\int_0^\infty \alpha^{-1}\, d\alpha$, which is divergent at both 0 and ∞ — but

just barely. One needs only slightly stronger estimates on λ_f near 0 and ∞ to obtain $f \in L^p$. (See also Exercise 36.) The standard example of a function that is in weak L^p but not in L^p is $f(x) = x^{-1/p}$ on $(0, \infty)$ (with Lebesgue measure).

Frequently it is convenient to express a function as the sum of a "small" part and a "big" part. The following is a way of doing this that gives a simple formula for the distribution functions.

6.25 Proposition. *If f is a measurable function and $A > 0$, let $E(A) = \{x : |f(x)| > A\}$, and set*

$$h_A = f\chi_{X\setminus E(A)} + A(\operatorname{sgn} f)\chi_{E(A)}, \qquad g_A = f - h_A = (\operatorname{sgn} f)(|f| - A)\chi_{E(A)}.$$

Then

$$\lambda_{g_A}(\alpha) = \lambda_f(\alpha + A), \qquad \lambda_{h_A}(\alpha) = \begin{cases} \lambda_f(\alpha) & \text{if } \alpha < A, \\ 0 & \text{if } \alpha \geq A. \end{cases}$$

The proof is left to the reader (Exercise 37).

Exercises

35. For any measurable f and g we have $[cf]_p = |c|[f]_p$ and $[f+g]_p \leq 2([f]_p^p + [g]_p^p)^{1/p}$; hence weak L^p is a vector space. Moreover, the "balls" $\{g : [g-f]_p < r\}$ ($r > 0$, $f \in$ weak L^p) generate a topology on weak L^p that makes weak L^p into a topological vector space.

36. If $f \in$ weak L^p and $\mu(\{x : f(x) \neq 0\}) < \infty$, then $f \in L^q$ for all $q < p$. On the other hand, if $f \in$ (weak L^p) $\cap L^\infty$, then $f \in L^q$ for all $q > p$.

37. Prove Proposition 6.25.

38. $f \in L^p$ iff $\sum_{-\infty}^{\infty} 2^{kp}\lambda_f(2^k) < \infty$.

39. If $f \in L^p$, then $\lim_{\alpha\to 0} \alpha^p\lambda_f(\alpha) = \lim_{\alpha\to\infty} \alpha^p\lambda_f(\alpha) = 0$. (First suppose f is simple.)

40. If f is a measurable function on X, its **decreasing rearrangement** is the function $f^* : (0, \infty) \to [0, \infty]$ defined by

$$f^*(t) = \inf\{\alpha : \lambda_f(\alpha) \leq t\} \quad (\text{where } \inf \varnothing = \infty).$$

a. f^* is decreasing. If $f^*(t) < \infty$ then $\lambda_f(f^*(t)) \leq t$, and if $\lambda_f(\alpha) < \infty$ then $f^*(\lambda_f(\alpha)) \leq \alpha$.
b. $\lambda_f = \lambda_{f^*}$, where λ_{f^*} is defined with respect to Lebesgue measure on $(0, \infty)$.
c. If $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$ and $\lim_{\alpha\to\infty} \lambda_f(\alpha) = 0$ (so that $f^*(t) < \infty$ for all $t > 0$), and ϕ is a nonnegative measurable function on $(0, \infty)$, then $\int_X \phi \circ |f|\, d\mu = \int_0^\infty \phi \circ f^*(t)\, dt$. In particular, $\|f\|_p = \|f^*\|_p$ for $0 < p < \infty$.
d. If $0 < p < \infty$, $[f]_p = \sup_{t>0} t^{1/p} f^*(t)$.
e. The name "rearrangement" for f^* comes from the case where f is a nonnegative function on $(0, \infty)$. To see why it is appropriate, pick a step function on $(0, \infty)$ assuming four or five different values and draw the graphs of f and f^*.

6.5 INTERPOLATION OF L^P SPACES

If $1 \le p < q < r \le \infty$, then $(L^p \cap L^r) \subset L^q \subset (L^p + L^r)$, and it is natural to ask whether a linear operator T on $L^p + L^r$ that is bounded on both L^p and L^r is also bounded on L^q. The answer is affirmative, and this result can be generalized in various ways. The two fundamental theorems on this question are the Riesz-Thorin and Marcinkiewicz interpolation theorems, which we present in this section. We begin with the Riesz-Thorin theorem, whose proof is based on the following result from complex function theory.

6.26 The Three Lines Lemma. *Let ϕ be a bounded continuous function on the strip $0 \le \operatorname{Re} z \le 1$ that is holomorphic on the interior of the strip. If $|\phi(z)| \le M_0$ for $\operatorname{Re} z = 0$ and $|\phi(z)| \le M_1$ for $\operatorname{Re} z = 1$, then $|\phi(z)| \le M_0^{1-t} M_1^t$ for $\operatorname{Re} z = t$, $0 < t < 1$.*

Proof. For $\epsilon > 0$ let $\phi_\epsilon(z) = \phi(z) M_0^{z-1} M_1^{-z} \exp(\epsilon z(z-1))$. Then ϕ_ϵ satisfies the hypotheses of the lemma with M_0 and M_1 replaced by 1, and also $|\phi_\epsilon(z)| \to 0$ as $|\operatorname{Im} z| \to \infty$. Thus $|\phi_\epsilon(z)| \le 1$ on the boundary of the rectangle $0 \le \operatorname{Re} z \le 1$, $-A \le \operatorname{Im} z \le A$ provided that A is large, and the maximum modulus principle therefore implies that $|\phi_\epsilon(z)| \le 1$ on the strip $0 \le \operatorname{Re} z \le 1$. Letting $\epsilon \to 0$, we obtain the desired result:

$$|\phi(z)| M_0^{t-1} M_1^{-t} = \lim_{\epsilon \to 0} |\phi_\epsilon(z)| \le 1 \text{ for } \operatorname{Re} z = t.$$

∎

6.27 The Riesz-Thorin Interpolation Theorem. *Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are measure spaces and $p_0, p_1, q_0, q_1 \in [1, \infty]$. If $q_0 = q_1 = \infty$, suppose also that ν is semifinite. For $0 < t < 1$, define p_t and q_t by*

$$\frac{1}{p_t} = \frac{1-t}{p_0} + \frac{t}{p_1}, \qquad \frac{1}{q_t} = \frac{1-t}{q_0} + \frac{t}{q_1}.$$

If T is a linear map from $L^{p_0}(\mu) + L^{p_1}(\mu)$ into $L^{q_0}(\nu) + L^{q_1}(\nu)$ such that $\|Tf\|_{q_0} \le M_0 \|f\|_{p_0}$ for $f \in L^{p_0}(\mu)$ and $\|Tf\|_{q_1} \le M_1 \|f\|_{p_1}$ for $f \in L^{p_1}(\mu)$, then $\|Tf\|_{q_t} \le M_0^{1-t} M_1^t \|f\|_{p_t}$ for $f \in L^{p_t}(\mu)$, $0 < t < 1$.

Proof. To begin with, we observe that the case $p_0 = p_1$ follows from Proposition 6.10: If $p = p_0 = p_1$, then

$$\|Tf\|_{q_t} \le \|Tf\|_{q_0}^{1-t} \|Tf\|_{q_1}^t \le M_0^{1-t} M_1^t \|f\|_p.$$

Thus we may assume that $p_0 \ne p_1$, and in particular that $p_t < \infty$ for $0 < t < 1$.

Let Σ_X (resp. Σ_Y) be the space of all simple functions on X (resp. Y) that vanish outside sets of finite measure. Then $\Sigma_X \subset L^p(\mu)$ for all p and Σ_X is dense in $L^p(\mu)$ for $p < \infty$, by Proposition 6.7; similarly for Σ_Y. The main part of the proof consists

of showing that $\|Tf\|_{q_t} \le M_0^{1-t} M_1^t \|f\|_{p_t}$ for all $f \in \Sigma_X$. However, by Theorem 6.14,

$$\|Tf\|_{q_t} = \sup\left\{ \left| \int (Tf) g \, d\nu \right| : g \in \Sigma_Y \text{ and } \|g\|_{q_t'} = 1 \right\},$$

where q_t' is the conjugate exponent to q_t. (Note that $Tf \in L^{q_0} \cap L^{q_1}$, so $\{y : Tf(y) \neq 0\}$ must be σ-finite unless $q_0 = q_1 = \infty$; hence the hypotheses of Theorem 6.14 are satisfied.) Moreover, we may assume that $f \neq 0$ and rescale f so that $\|f\|_{p_t} = 1$. We therefore wish to establish the following claim:

- If $f \in \Sigma_X$ and $\|f\|_{p_t} = 1$, then $|\int (Tf) g \, d\nu| \le M_0^{1-t} M_1^t$ for all $g \in \Sigma_Y$ such that $\|g\|_{q_t'} = 1$.

Let $f = \sum_1^m c_j \chi_{E_j}$ and $g = \sum_1^n d_k \chi_{F_k}$ where the E_j's and the F_k's are disjoint in X and Y and the c_j's and d_k's are nonzero. Write c_j and d_k in polar form: $c_j = |c_j| e^{i\theta_j}$, $d_k = |d_k| e^{i\psi_k}$. Also, let

$$\alpha(z) = (1-z)p_0^{-1} + z p_1^{-1}, \qquad \beta(z) = (1-z)q_0^{-1} + z q_1^{-1};$$

thus $\alpha(t) = p_t^{-1}$ and $\beta(t) = q_t^{-1}$ for $0 < t < 1$. Fix $t \in (0,1)$; we have assumed that $p_t < \infty$ and hence $\alpha(t) > 0$, so we may define

$$f_z = \sum_1^m |c_j|^{\alpha(z)/\alpha(t)} e^{i\theta_j} \chi_{E_j}.$$

If $\beta(t) < 1$, we define

$$g_z = \sum_1^n |d_k|^{(1-\beta(z))/(1-\beta(t))} e^{i\psi_k} \chi_{F_k},$$

while if $\beta(t) = 1$ we define $g_z = g$ for all z. (We henceforth assume that $\beta(t) < 1$ and leave the easy modification for $\beta(t) = 1$ to the reader.) Finally, we set

$$\phi(z) = \int (Tf_z) g_z \, d\nu.$$

Thus,

$$\phi(z) = \sum_{j,k} A_{jk} |c_j|^{\alpha(z)/\alpha(t)} |d_k|^{(1-\beta(z))/(1-\beta(t))}$$

where

$$A_{jk} = e^{i(\theta_j + \psi_k)} \int (T\chi_{E_j}) \chi_{F_k} \, d\nu,$$

so that ϕ is an entire holomorphic function of z that is bounded in the strip $0 \le \operatorname{Re} z \le 1$. Since $\int (Tf) g \, d\nu = \phi(t)$, by the three lines lemma it will suffice to show that $|\phi(z)| \le M_0$ for $\operatorname{Re} z = 0$ and $|\phi(z)| \le M_1$ for $\operatorname{Re} z = 1$. However, since

$$\alpha(is) = p_0^{-1} + is(p_1^{-1} - p_0^{-1}), \qquad 1 - \beta(is) = (1 - q_0^{-1}) - is(q_1^{-1} - q_0^{-1})$$

for $s \in \mathbb{R}$, we have

$$|f_{is}| = |f|^{\mathrm{Re}[\alpha(is)/\alpha(t)]} = |f|^{p_t/p_0}, \qquad |g_{is}| = |g|^{\mathrm{Re}[(1-\beta(is))/(1-\beta(t))]} = |g|^{q_t'/q_0'}.$$

Therefore, by Hölder's inequality,

$$|\phi(is)| \leq \|Tf_{is}\|_{q_0}\|g_{is}\|_{q_0'} \leq M_0\|f_{is}\|_{p_0}\|g_{is}\|_{q_0'} = M_0\|f\|_{p_t}\|g\|_{q_t'} = M_0.$$

A similar calculation shows that $|\phi(1+is)| \leq M_1$, so the claim is proved.

We have now shown that $\|Tf\|_{q_t} \leq M_0^{1-t}M_1^t\|f\|_{p_t}$ for $f \in \Sigma_X$, so in view of Proposition 6.7, $T|\Sigma_X$ has a unique extension to $L^{p_t}(\mu)$ satisfying the same estimate there. It remains to show that this extension is T itself, that is, that T satisfies this estimate for all $f \in L^{p_t}(\mu)$. Given such an f, choose a sequence $\{f_n\}$ in Σ_X such that $|f_n| \leq |f|$ and $f_n \to f$ pointwise. Also, let $E = \{x : |f(x)| > 1\}$, $g = f\chi_E$, $g_n = f_n\chi_E$, $h = f - g$, and $h_n = f_n - g_n$. Then if $p_0 < p_1$ (which we may assume, by relabeling the p's), we have $g \in L^{p_0}(\mu)$, $h \in L^{p_1}(\mu)$, and by the dominated convergence theorem, $\|f_n - f\|_{p_t} \to 0$, $\|g_n - g\|_{p_0} \to 0$, and $\|h_n - h\|_{p_1} \to 0$. Hence $\|Tg_n - Tg\|_{q_0} \to 0$ and $\|Th_n - Th\|_{q_1} \to 0$, so by passing to a suitable subsequence we may assume that $Tg_n \to Tg$ a.e. and $Th_n \to Th$ a.e. (Exercise 9). But then $Tf_n \to Tf$ a.e., so by Fatou's lemma,

$$\|Tf\|_{q_t} \leq \liminf \|Tf_n\|_{q_t} \leq \liminf M_0^{1-t}M_1^t\|f_n\|_{p_t} = M_0^{1-t}M_1^t\|f\|_{p_t},$$

and we are done. ∎

The conclusion of the Riesz-Thorin theorem can be restated in a slightly stronger form. Let $M(t)$ be the operator norm of T as a map from $L^{p_t}(\mu)$ to $L^{q_t}(\nu)$. We have shown that $M(t) \leq M_0^{1-t}M_1^t$. It is possible for strict inequality to hold; however, if $0 < s < t < u < 1$ and $t = (1-\tau)s + \tau u$, the theorem may be applied again to show that $M(t) \leq M(s)^{1-\tau}M(u)^\tau$. In short, the conclusion is that $\log M(t)$ is a convex function of t.

We now turn to the Marcinkiewicz theorem, for which we need some more terminology. Let T be a map from some vector space $\mathcal{D}$ of measurable functions on $(X, \mathcal{M}, \mu)$ to the space of all measurable functions on $(Y, \mathcal{N}, \nu)$.

- T is called **sublinear** if $|T(f+g)| \leq |Tf| + |Tg|$ and $|T(cf)| = c|Tf|$ for all $f, g \in \mathcal{D}$ and $c > 0$.

- A sublinear map T is **strong type** (p, q) $(1 \leq p, q \leq \infty)$ if $L^p(\mu) \subset \mathcal{D}$, T maps $L^p(\mu)$ into $L^q(\nu)$, and there exists $C > 0$ such that $\|Tf\|_q \leq C\|f\|_p$ for all $f \in L^p(\mu)$.

- A sublinear map T is **weak type** (p, q) $(1 \leq p \leq \infty,\ 1 \leq q < \infty)$ if $L^p(\mu) \subset \mathcal{D}$, T maps $L^p(\mu)$ into weak $L^q(\nu)$, and there exists $C > 0$ such that $[Tf]_q \leq C\|f\|_p$ for all $f \in L^p(\mu)$. Also, we shall say that T is weak type (p, ∞) iff T is strong type (p, ∞).

6.28 The Marcinkiewicz Interpolation Theorem. *Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are measure spaces; p_0, p_1, q_0, q_1 are elements of $[1, \infty]$ such that $p_0 \le q_0$, $p_1 \le q_1$, and $q_0 \ne q_1$; and*

$$\frac{1}{p} = \frac{1-t}{p_0} + \frac{t}{p_1} \text{ and } \frac{1}{q} = \frac{1-t}{q_0} + \frac{t}{q_1}, \qquad \text{where } 0 < t < 1.$$

If T is a sublinear map from $L^{p_0}(\mu) + L^{p_1}(\mu)$ to the space of measurable functions on Y that is weak types (p_0, q_0) and (p_1, q_1), then T is strong type (p, q). More precisely, if $[Tf]_{q_j} \le C_j \|f\|_{p_j}$ for $j = 0, 1$, then $\|Tf\|_q \le B_p \|f\|_p$ where B_p depends only on p_j, q_j, C_j in addition to p; and for $j = 0, 1$, $B_p|p - p_j|$ (resp. B_p) remains bounded as $p \to p_j$ if $p_j < \infty$ (resp. $p_j = \infty$).

Proof. The case $p_0 = p_1$ is easy and is left to the reader (Exercise 42). Without loss of generality we may therefore assume that $p_0 < p_1$, and for the time being we also assume that $q_0 < \infty$ and $q_1 < \infty$ (whence also $p_0 < p_1 < \infty$). Given $f \in L^p(\mu)$ and $A > 0$, let g_A and h_A be as in Proposition 6.25. Then by Propositions 6.24 and 6.25,

$$\begin{aligned} \int |g_A|^{p_0}\, d\mu &= p_0 \int_0^\infty \beta^{p_0-1} \lambda_{g_A}(\beta)\, d\beta = p_0 \int_0^\infty \beta^{p_0-1} \lambda_f(\beta + A)\, d\beta \\ &= p_0 \int_A^\infty (\beta - A)^{p_0-1} \lambda_f(\beta)\, d\beta \le p_0 \int_A^\infty \beta^{p_0-1} \lambda_f(\beta)\, d\beta, \qquad (6.29) \\ \int |h_A|^{p_1}\, d\mu &= p_1 \int_0^\infty \beta^{p_1-1} \lambda_{h_A}(\beta)\, d\beta = p_1 \int_0^A \beta^{p_1-1} \lambda_f(\beta)\, d\beta. \end{aligned}$$

Likewise,

$$\int |Tf|^q\, d\nu = q \int_0^\infty \alpha^{q-1} \lambda_{Tf}(\alpha)\, d\alpha = 2^q q \int_0^\infty \alpha^{q-1} \lambda_{Tf}(2\alpha)\, d\alpha. \tag{6.30}$$

Since T is sublinear, by Proposition 6.22d we have

$$\lambda_{Tf}(2\alpha) \le \lambda_{Tg_A}(\alpha) + \lambda_{Th_A}(\alpha). \tag{6.31}$$

This is true for all $\alpha > 0$ and $A > 0$, so we may take A to depend on α. We now make a specific choice of A. Namely, it follows from the equations defining p and q that

$$\frac{p_0(q_0 - q)}{q_0(p_0 - p)} = \frac{p^{-1}(q^{-1} - q_0^{-1})}{q^{-1}(p^{-1} - p_0^{-1})} = \frac{p^{-1}(q^{-1} - q_1^{-1})}{q^{-1}(p^{-1} - p_1^{-1})} = \frac{p_1(q_1 - q)}{q_1(p_1 - p)}; \tag{6.32}$$

we denote the common value of these quantities by σ, and we take $A = \alpha^\sigma$. Then by (6.29), (6.30), (6.31), and the weak type estimates on T,

$$
\begin{aligned}
\|Tf\|_q^q &\le 2^q q \int_0^\infty \alpha^{q-1}\Big[\big(C_0\|g_A\|_{p_0}/\alpha\big)^{q_0} + \big(C_1\|h_A\|_{p_1}/\alpha\big)^{q_1}\Big]\,d\alpha \\
&\le 2^q q C_0^{q_0} p_0^{q_0/p_0} \int_0^\infty \alpha^{q-q_0-1}\Big[\int_{\alpha^\sigma}^\infty \beta^{p_0-1}\lambda_f(\beta)\,d\beta\Big]^{q_0/p_0} d\alpha \\
&\quad + 2^q q C_1^{q_1} p_1^{q_1/p_1} \int_0^\infty \alpha^{q-q_1-1}\Big[\int_0^{\alpha^\sigma} \beta^{p_1-1}\lambda_f(\beta)\,d\beta\Big]^{q_1/p_1} d\alpha \\
&= \sum_{j=0}^{1} 2^q q C_j^{p_j} p_j^{q_j/p_j} \int_0^\infty \Big[\int_0^\infty \phi_j(\alpha,\beta)\,d\beta\Big]^{q_j/p_j} d\alpha,
\end{aligned}
\tag{6.33}
$$

where, denoting by χ_0 and χ_1 the characteristic functions of $\{(\alpha,\beta) : \beta > \alpha^\sigma\}$ and $\{(\alpha,\beta) : \beta < \alpha^\sigma\}$,

$$\phi_j(\alpha,\beta) = \chi_j(\alpha,\beta)\alpha^{(q-q_j-1)p_j/q_j}\beta^{p_j-1}\lambda_f(\beta).$$

Since $q_0/p_0 \ge 1$ and $q_1/p_1 \ge 1$, we may apply Minkowski's inequality for integrals to obtain

$$
\begin{aligned}
&\int_0^\infty \Big[\int_0^\infty \phi_j(\alpha,\beta)d\beta\Big]^{q_j/p_j} d\alpha \\
&\qquad \le \Big[\int_0^\infty \Big[\int_0^\infty \phi_j(\alpha,\beta)^{q_j/p_j} d\alpha\Big]^{p_j/q_j} d\beta\Big]^{q_j/p_j}.
\end{aligned}
\tag{6.34}
$$

Let $\tau = 1/\sigma$. If $q_1 > q_0$, then $q - q_0$ and σ are positive and the inequality $\beta > \alpha^\sigma$ is equivalent to $\alpha < \beta^\tau$, so

$$
\begin{aligned}
&\int_0^\infty \Big[\int_0^\infty \phi_0(\alpha,\beta)^{q_0/p_0} d\alpha\Big]^{p_0/q_0} d\beta \\
&\qquad = \int_0^\infty \Big[\int_0^{\beta^\tau} \alpha^{q-q_0-1} d\alpha\Big]^{p_0/q_0} \beta^{p_0-1}\lambda_f(\beta)\,d\beta \\
&\qquad = (q-q_0)^{-p_0/q_0} \int_0^\infty \beta^{p_0-1+p_0(q-q_0)/q_0\sigma}\lambda_f(\beta)\,d\beta \\
&\qquad = (q-q_0)^{-p_0/q_0} \int_0^\infty \beta^{p-1}\lambda_f(\beta)\,d\beta \\
&\qquad = |q-q_0|^{-p_0/q_0} p^{-1}\|f\|_p^p,
\end{aligned}
$$

where we have used (6.32) to simplify the exponent of β. On the other hand, if $q_1 < q_0$, then $q - q_0$ and σ are negative and the inequality $\beta > \alpha^\sigma$ is equivalent to

$\alpha > \beta^\tau$, so as above,

$$\int_0^\infty \left[\int_0^\infty \phi_0(\alpha,\beta)^{q_0/p_0} d\alpha\right]^{p_0/q_0} d\beta = \int_0^\infty \left[\int_{\beta^\tau}^\infty \alpha^{q-q_0-1} d\alpha\right]^{p_0/q_0} \beta^{p_0-1}\lambda_f(\beta)\, d\beta$$
$$= (q_0-q)^{-p_0/q_0} \int_0^\infty \beta^{p-1}\lambda_f(\beta)\, d\beta$$
$$= |q-q_0|^{-p_0/q_0} p^{-1} \|f\|_p^p.$$

A similar calculation shows that

$$\int_0^\infty \left[\int_0^\infty \phi_1(\alpha,\beta)^{q_1/p_1} d\alpha\right]^{p_1/q_1} d\beta = |q-q_1|^{-p_1/q_1} p^{-1} \|f\|_p^p.$$

Combining these results with (6.33) and (6.34), we see that

$$\sup\{\|Tf\|_q : \|f\|_p = 1\} \le B_p = 2q^{1/q}\left[\sum_{j=0}^{1} C_j^{q_j}(p_j/p)^{q_j/p_j}|q-q_j|^{-1}\right]^{1/q}.$$

But since $|T(cf)| = c|Tf|$ for $c > 0$, this implies that $\|Tf\|_q \le B_p\|f\|_p$ for all $f \in L^p(\mu)$, and we are done. (The verification of the asserted properties of B_p is left as an easy exercise.)

It remains to show how to modify this argument to deal with the exceptional cases $q_0 = \infty$ or $q_1 = \infty$. We distinguish three cases.

Case I: $p_1 = q_1 = \infty$ (so $p_0 \le q_0 < \infty$). Instead of taking $A = \alpha^\sigma$ in the decomposition of f, we take $A = \alpha/C_1$. Then $\|Th_A\|_\infty \le C_1\|h_A\|_\infty \le \alpha$, so $\lambda_{Th_A}(\alpha) = 0$, and we obtain (6.33) with $\phi_1 = 0$ and α^σ replaced by α/C_1 in the definition of ϕ_0. The same argument as above then gives

$$\|Tf\|_q \le 2\left[qC_0^{q_0}C_1^{q-q_0}(p_0/p)^{q_0/p_0}|q-q_0|^{-1}\right]^{1/q}\|f\|_p.$$

Case II: $p_0 < p_1 < \infty$, $q_0 < q_1 = \infty$. Again the idea is to choose A so that $\lambda_{Th_A}(\alpha) = 0$, and the proper choice is $A = (\alpha/d)^\sigma$ where $d = C_1\left[p_1\|f\|_p^p/p\right]^{1/p_1}$ and $\sigma = p_1/(p_1-p)$ (the limiting value of the σ defined by (6.32) as $q_1 \to \infty$). Indeed, since $p_1 > p$, we have

$$\|Th_A\|_\infty^{p_1} \le C_1^{p_1}\|h_A\|_{p_1}^{p_1} = C_1^{p_1}p_1\int_0^A \alpha^{p_1-1}\lambda_f(\alpha)\, d\alpha$$
$$\le C_1^{p_1}p_1A^{p_1-p}\int_0^A \alpha^{p-1}\lambda_f(\alpha)\, d\alpha = C_1^{p_1}\frac{p_1}{p}\left[\frac{\alpha}{d}\right]^{p_1}\|f\|_p^p = \alpha^{p_1}.$$

As in Case I, then, we find that $\phi_1 = 0$ in (6.33) and the integral involving ϕ_0 is majorized by a constant B_p when $\|f\|_p = 1$, which yields the desired result.

Case III: $p_0 < p_1 < \infty$, $q_1 < q_0 = \infty$. The argument is essentially the same as in Case II, except that we take $A = (\alpha/d)^\sigma$ with d chosen so that $\lambda_{Tg_A}(\alpha) = 0$. ∎

The lengthy formulas in this proof may seem daunting, but the ideas are reasonably simple. To elucidate them, we recommend the exercise of writing out the proof for two special (but important) cases: (i) $p_0 = q_0 = 1$, $p_1 = q_1 = 2$, and (ii) $p_0 = q_0 = 1$, $p_1 = q_1 = \infty$.

Let us compare our two interpolation theorems. The Marcinkiewicz theorem requires some restrictions on p_j and q_j that are not present in the Riesz-Thorin theorem; these restrictions, however, are satisfied in all the interesting applications. Apart from this, the hypotheses of the Marcinkiewicz theorem are weaker: T is allowed to be sublinear rather than linear, and it needs only to satisfy weak-type estimates at the endpoints. The conclusion in both cases is that T is bounded from $L^p(\mu)$ to $L^q(\nu)$, but the Riesz-Thorin theorem produces a much sharper estimate for the operator norm of T. Thus neither theorem includes the other.

We conclude with two applications of the Marcinkiewicz theorem. The first one concerns the Hardy-Littlewood maximal operator H discussed in §3.4,

$$Hf(x) = \sup_{r>0} \frac{1}{m(B(r,x))} \int_{B(r,x)} |f(y)|\, dy \qquad (f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)).$$

H is obviously sublinear and satisfies $\|Hf\|_\infty \le \|f\|_\infty$ for all $f \in L^\infty$. Moreover, Theorem 3.17 says precisely that H is weak type $(1,1)$. We conclude:

6.35 Corollary. *There is a constant $C > 0$ such that if $1 < p < \infty$ and $f \in L^p(\mathbb{R}^n)$, then*

$$\|Hf\|_p \le C\frac{p}{p-1}\|f\|_p.$$

Our second application is a theorem on integral operators related to Theorem 6.18.

6.36 Theorem. *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces, and $1 < q < \infty$. Let K be a measurable function on $X \times Y$ such that, for some $C > 0$, we have $[K(x,\cdot)]_q \le C$ for a.e. $x \in X$ and $[K(\cdot,y)]_q \le C$ for a.e. $y \in Y$. If $1 \le p < \infty$ and $f \in L^p(\nu)$, the integral*

$$Tf(x) = \int K(x,y)f(y)\, d\nu(y)$$

converges absolutely for a.e. $x \in X$, and the operator T thus defined is weak type $(1,q)$ and strong type (p,r) for all p, r such that $1 < p < r < \infty$ and $p^{-1} + q^{-1} = r^{-1} + 1$. More precisely, there exist constants B_p independent of K such that

$$[Tf]_q \le B_1 C\|f\|_1, \qquad \|Tf\|_r \le B_p C\|f\|_p \quad (p > 1,\ r^{-1} = p^{-1} + q^{-1} - 1 > 0).$$

Proof. Let p', q' be the conjugate exponents to p, q; then

$$r^{-1} = p^{-1} + q^{-1} - 1 = p^{-1} - (q')^{-1} = q^{-1} - (p')^{-1},$$

so $p < q'$ and $q < p'$. Suppose $0 \neq f \in L^p$ $(1 \leq p < q')$; by multiplying f and K by constants, we may assume that $\|f\|_p = C = 1$. Given a positive number A whose value will be fixed later, define

$$E = \{(x,y) : |K(x,y)| > A\}, \quad K_1 = (\operatorname{sgn} K)(|K| - A)\chi_E, \quad K_2 = K - K_1,$$

and let T_1, T_2 be the operators corresponding to K_1, K_2. Then by Propositions 6.24 and 6.25, since $q > 1$ we have

$$\int |K_1(x,y)|\, d\nu(y) = \int_0^\infty \lambda_{K(x,\cdot)}(\alpha + A)\, d\alpha \leq \int_A^\infty \alpha^{-q}\, d\alpha = \frac{A^{1-q}}{q-1},$$

and likewise

$$\int |K_1(x,y)|\, d\mu(x) \leq \frac{A^{1-q}}{q-1}.$$

Hence, by Theorem 6.18, the integral defining $T_1 f(x)$ converges for a.e. x and

$$\|T_1 f\|_p \leq \frac{A^{1-q}}{q-1}\|f\|_p = \frac{A^{1-q}}{q-1}. \tag{6.37}$$

Similarly, since $q < p'$,

$$\begin{aligned}\int |K_2(x,y)|^{p'}\, d\nu(y) &= p' \int_0^A \alpha^{p'-1}\lambda_{K(x,\cdot)}(\alpha)\, d\alpha \\ &\leq p' \int_0^A \alpha^{p'-1-q}\, d\alpha = \frac{p' A^{p'-q}}{p'-q}.\end{aligned}$$

Therefore, by Hölder's inequality, the integral defining $T_2 f(x)$ converges for every x, and

$$\|T_2 f\|_\infty \leq \left[\frac{p' A^{p'-q}}{p'-q}\right]^{1/p'} \|f\|_p = \left[\frac{r}{q}\right]^{1/p'} A^{q/r}. \tag{6.38}$$

We have thus established that $Tf = T_1 f + T_2 f$ is well defined a.e.

Next, given $\alpha > 0$, we wish to estimate $\lambda_{Tf}(\alpha)$. But by Proposition 6.22d,

$$\lambda_{Tf}(\alpha) \leq \lambda_{T_1 f}(\tfrac{1}{2}\alpha) + \lambda_{T_2 f}(\tfrac{1}{2}\alpha),$$

and by (6.38), if we choose

$$A = \left[\frac{\alpha}{2}\right]^{r/q} \left[\frac{q}{r}\right]^{r/qp'}$$

we will have $\|T_2 f\|_\infty \leq \frac{1}{2}\alpha$, so that $\lambda_{T_2 f}(\frac{1}{2}\alpha) = 0$. With this choice of A, then, by (6.37) and Chebyshev's inequality we obtain

$$\begin{aligned}\lambda_{Tf}(\alpha) &\leq \lambda_{T_1 f}(\tfrac{1}{2}\alpha) \leq \left[\frac{2\|T_1 f\|_p}{\alpha}\right]^p \leq \left[\frac{2A^{1-q}}{(q-1)\alpha}\right]^p \\ &= \frac{2^{p-(1-q)pr/q}}{(q-1)^p}\left[\frac{q}{r}\right]^{(1-q)pr/qp'} \alpha^{-p+(1-q)pr/q} = C_p\left[\frac{\|f\|_p}{\alpha}\right]^r,\end{aligned}$$

because $\|f\|_p = 1$ and

$$\frac{(1-q)pr}{q} - p = p\left(\frac{-r}{q'} - 1\right) = -p \cdot \frac{r}{p} = -r.$$

A simple homogeneity argument now yields the estimate $\lambda_{Tf}(\alpha) \le C_p(\|f\|_p/\alpha)^r$ with no restriction on $\|f\|_p$, so we have shown that T is weak type (p, r), and in particular (for $p = 1$) weak type $(1, q)$.

Finally, given $p \in (1, q')$, choose $\widetilde{p} \in (p, q')$ and define $\widetilde{r}$ by $\widetilde{r}^{-1} = \widetilde{p}^{-1} - (q')^{-1}$. Then T is weak types $(1, q)$ and $(\widetilde{p}, \widetilde{r})$, so it follows from the Marcinkiewicz theorem that T is strong type (p, r). ∎

Exercises

41. Suppose $1 < p \le \infty$ and $p^{-1} + q^{-1} = 1$. If T is a bounded operator on L^p such that $\int (Tf)g = \int f(Tg)$ for all $f, g \in L^p \cap L^q$, then T extends uniquely to a bounded operator on L^r for all r in $[p, q]$ (if $p < q$) or $[q, p]$ (if $q < p$).

42. Prove the Marcinkiewicz theorem in the case $p_0 = p_1$. (Setting $p = p_0 = p_1$, we have $\lambda_{Tf}(\alpha) \le (C_0\|f\|_p/\alpha)^{q_0}$ and $\lambda_{Tf}(\alpha) \le (C_1\|f\|_p/\alpha)^{q_1}$. Use whichever estimate is better, depending on α, to majorize $q\int_0^\infty \alpha^{q-1}\lambda_{Tf}(\alpha)\,d\alpha$.)

43. Let H be the Hardy-Littlewood maximal operator on $\mathbb{R}$. Compute $H\chi_{(0,1)}$ explicitly. Show that it is in L^p for all $p > 1$ and in weak L^1 but not in L^1, and that its L^p norm tends to ∞ like $(p-1)^{-1}$ as $p \to 1$, although $\|\chi_{(0,1)}\|_p = 1$ for all p.

44. Let I_α be the fractional integration operator of Exercise 61 in §2.6. If $0 < \alpha < 1$, $1 < p < \alpha^{-1}$, and $r^{-1} = p^{-1} - \alpha$, then I_α is weak type $(1, (1-\alpha)^{-1})$ and strong type (p, r) with respect to Lebesgue measure on $(0, \infty)$.

45. If $0 < \alpha < n$, define an operator T_α on functions on $\mathbb{R}^n$ by

$$T_\alpha f(x) = \int |x-y|^{-\alpha} f(y)\,dy.$$

Then T_α is weak type $(1, (n-\alpha)^{-1})$ and strong type (p, r) with respect to Lebesgue measure on $\mathbb{R}^n$, where $1 < p < n\alpha^{-1}$ and $r^{-1} = p^{-1} - \alpha n^{-1}$. (The case $n = 3$, $\alpha = 1$ is of particular interest in physics: If f represents the density of a mass or charge distribution, $-(4\pi)^{-1}T_1 f$ represents the induced gravitational or electrostatic potential.)

6.6 NOTES AND REFERENCES

The importance of the space $L^2([a, b])$ was recognized soon after the invention of the Lebesgue integral because of its connection with Fourier series and other orthogonal expansions; and one of the early triumphs of the Lebesgue theory was the discovery in 1907 by Fischer [44] and F. Riesz [114] that $L^2([a, b])$ is isomorphic to l^2, or

what amounts to the same thing, that $L^2([a,b])$ is complete. The spaces $L^p([a,b])$ for $1 < p < \infty$ were first investigated by F. Riesz [117], who proved all of the major results in §§6.1–2 for them as well as the weak sequential compactness of the closed unit ball in L^p. The fact that $(L^1)^* = L^\infty$ was first proved by Steinhaus [143].

In some respects it is unfortunate that L^p spaces were not named $L^{1/p}$ spaces, for — as one sees in the conjugacy relation $p^{-1} + q^{-1} = 1$ and in the results of §6.5 — relationships among different L^p spaces usually involve linear equations in p^{-1}.

A discussion of some of the deeper aspects of L^p spaces and their applications in other areas of analysis can be found in Lieb and Loss [93].

§6.1: Hölder's inequality, in the case $p = 2$, is commonly associated with the names of Cauchy (who proved it for finite sums) and Buniakovsky and Schwarz (who proved it, independently, for integrals). For general p it was discovered independently by Hölder and Rogers. Minkowski's original inequality was for finite sums. (See Hardy, Littlewood, and Pólya [66] for references.) A neat proof of Hölder's inequality using complex function theory can be found in Rubel [122].

The relations among the spaces $L^p + L^q$, defined in Exercise 4, are studied in Alvarez [5]. See Romero [120] for a discussion of Exercise 5, including some other conditions for the inclusion $L^p \subset L^q$ to hold, and Miamee [100] for a discussion of the more general relation $L^p(\mu) \subset L^q(\nu)$.

§6.2: A quite different approach to the L^p duality theory for $1 < p < \infty$ can be found in Hewitt and Stromberg [76, §15]. J. Schwartz [130] has found a characterization of $(L^1)^*$ that is valid on arbitrary measure spaces.

The proof of Theorem 6.15 breaks down for $p = \infty$ because the set function $\nu(E) = \phi(\chi_E)$ need not be countably additive. It is, however, a bounded, *finitely* additive complex measure on $(X, \mathcal{M})$ that is absolutely continuous with respect to μ in the sense that $\nu(E) = 0$ whenever $\mu(E) = 0$. Conversely, given a bounded, finitely additive complex measure ν on $(X, \mathcal{M})$, one can define the integral of a bounded measurable function with respect to ν. (One defines $\int f\,d\nu$ in the obvious way when f is simple and then shows that $|\int f\,d\nu| \le C\|f\|_u$, so that the integral extends to all uniform limits of simple functions.) In this way one obtains a representation of $(L^\infty)^*$ as a space of finitely additive complex measures. See Hewitt and Stromberg [76, §20], and for a more general treatment of finitely additive integrals, Dunford and Schwartz [35, Chapter 3]. (The example of a $\phi \in (L^\infty)^* \setminus L^1$ that we presented at the end of §6.2 shows how horrible finitely additive measures can be: If $\nu(E) = \phi(\chi_E)$, then $\nu \ll m$, but ν behaves like the point mass at zero when integrated against any continuous function.)

§6.3: Theorem 6.18 generalizes results of Schur [129] (for the case $p = 2$) and W. H. Young [164] (for the case $K(x,y) = k(x-y)$; see §8.2). Theorem 6.20 is also essentially due to Schur [129].

The reader whose appetite for inequalities is not satisfied by this section can find a feast in Hardy, Littlewood, and Pólya [66].

§6.4: The weak L^p spaces first appeared implicitly in weak-type estimates, instances of which go back to the 1920s; see also the notes for §6.5 below. Decreasing

rearrangements (Exercise 40) were introduced by Hardy and Littlewood [65], who give an entertaining motivation of their principal theorem on rearrangements in terms of cricket averages.

§6.5: The Riesz-Thorin theorem was first proved by M. Riesz (F. Riesz's younger brother) [118] under the assumption that $p_j \leq q_j$ for $j = 0, 1$; the proof in the general case and the idea of using the three lines lemma are due to Thorin [149]. E. M. Stein has proved a very powerful generalization of the Riesz-Thorin theorem. It deals with a family $\{T_z : 0 \leq \operatorname{Re} z \leq 1\}$ of operators that (roughly speaking) depend holomorphically on z and satisfy some mild growth conditions as $|\operatorname{Im} z| \to \infty$, and it asserts that if T_z is bounded from L^{p_j} to L^{q_j} for $\operatorname{Re} z = j$ $(j = 0, 1)$, then T_z is bounded from L^{p_t} to L^{q_t} for $\operatorname{Re} z = t$ $(0 < t < 1)$, where p_t, q_t are defined as in the Riesz-Thorin theorem. The precise statement and proof can be found in Bennett and Sharpley [15, §4.3], Stein and Weiss [142, §V.4], or Zygmund [167, §XII.1]. For a further extension of these ideas, see Coifman et al. [28].

The Marcinkiewicz interpolation theorem was announced by Marcinkiewicz [97] for the case $p_j = q_j$ $(j = 0, 1)$; after his untimely death in World War II, the work was completed by Zygmund [166]. The theorem can be proved under still weaker hypotheses on T; an extra twist to the argument we have given yields the same result under the sole assumption that $|T(f + g)| \leq C(|Tf| + |Tg|)$ for some constant C. See Zygmund [166], [167, §XII.4]. The spaces L^p and weak L^p form part of a two-parameter family $\{L(p, q) : 1 \leq p, q \leq \infty\}$ of function spaces, the so-called Lorentz spaces, such that $L^p = L(p, p)$ and weak $L^p = L(p, \infty)$, and the Marcinkiewicz theorem can be extended to a result about interpolation of operators on the $L(p, q)$ spaces. See Bennett and Sharpley [15, §4.4], or Stein and Weiss [142, §5.3]

There are many other examples of "continuous families" of Banach spaces for which interpolation theorems can be proved — for example, the spaces Λ_α discussed in Exercise 11 in §5.1 and the Sobolev spaces discussed in §9.3. There are also two general techniques for constructing "intermediate spaces" between pairs of Banach spaces, known as the "complex method" and the "real method," which may be regarded as abstract forms of the Riesz-Thorin and Marcinkiewicz theorems. An account of these theories and their applications can be found in Bergh and Löfström [16]; see also Bennett and Sharpley [15] the real method and its applications.

Corollary 6.35 is due to Hardy and Littlewood [65]. Theorem 6.36 appears first in Folland and Stein [51], but the essential idea of the proof was discovered by Stein several years earlier (see, e.g., Stein [140, §5.1]), and the special case discussed in Exercise 44 goes back to Hardy and Littlewood [64].

7

Radon Measures

The subject of this chapter is measure and integration theory on locally compact Hausdorff (LCH) spaces. We have seen in §2.6 that Lebesgue measure on $\mathbb{R}^n$ interacts nicely with the topology on $\mathbb{R}^n$ — measurable sets can be approximated by open or compact sets, and integrable functions can be approximated by continuous functions — and it is of interest to study measures having similar properties on more general spaces. Moreover, it turns out that certain linear functionals on spaces of continuous functions are given by integration against such measures. This fact constitutes an important link between measure theory and functional analysis, and it also provides a powerful tool for constructing measures.

Throughout this chapter, X will denote an LCH space. We continue to employ the terminology developed in Chapter 1 in the context of metric spaces: $\mathcal{B}_X$ will denote the Borel σ-algebra on X, that is, the σ-algebra generated by the open sets; measures on $\mathcal{B}_X$ will be called Borel measures; countable unions (intersections) of closed (open) sets will be called F_σ (G_δ) sets, and so forth.

7.1 POSITIVE LINEAR FUNCTIONALS ON $C_C(X)$

We recall that $C_c(X)$ is the space of continuous functions on X with compact support. A linear functional I on $C_c(X)$ will be called **positive** if $I(f) \geq 0$ whenever $f \geq 0$. In this definition there is no mention of continuity, but it is worth noting that positivity itself implies a rather strong continuity property.

7.1 Proposition. *If I is a positive linear functional on $C_c(X)$, for each compact $K \subset X$ there is a constant C_K such that $|I(f)| \le C_K \|f\|_u$ for all $f \in C_c(X)$ such that* $\operatorname{supp}(f) \subset K$.

Proof. It suffices to consider real-valued f. Given a compact K, choose $\phi \in C_c(X, [0,1])$ such that $\phi = 1$ on K (Urysohn's lemma). Then if $\operatorname{supp}(f) \subset K$, we have $|f| \le \|f\|_u \phi$, that is, $\|f\|_u \phi - f \ge 0$ and $\|f\|_u \phi + f \ge 0$. Thus $\|f\|_u I(\phi) - I(f) \ge 0$ and $\|f\|_u I(\phi) + I(f) \ge 0$, so that $|I(f)| \le I(\phi)\|f\|_u$. ∎

If μ is a Borel measure on X such that $\mu(K) < \infty$ for every compact $K \subset X$, then clearly $C_c(X) \subset L^1(\mu)$, so the map $f \mapsto \int f\, d\mu$ is a positive linear functional on $C_c(X)$. The principal result of this section is that *every* positive linear functional on $C_c(X)$ arises in this fashion; moreover, one can impose some additional regularity conditions on μ, subject to which μ is unique. These conditions are as follows.

Let μ be a Borel measure on X and E a Borel subset of X. The measure μ is called **outer regular** on E if

$$\mu(E) = \inf\{\mu(U) : U \supset E,\ U \text{ open}\}$$

and **inner regular** on E if

$$\mu(E) = \sup\{\mu(K) : K \subset E,\ K \text{ compact}\}.$$

If μ is outer and inner regular on all Borel sets, μ is called **regular**. It turns out that regularity is a bit too much to ask for when X is not σ-compact, so we adopt the following definition. A **Radon measure** on X is a Borel measure that is finite on all compact sets, outer regular on all Borel sets, and inner regular on all open sets. We shall show in §7.2 that Radon measures are also inner regular on all of their σ-finite sets.

One further bit of notation: If U is open in X and $f \in C_c(X)$, we shall write

$$f \prec U$$

to mean that $0 \le f \le 1$ and $\operatorname{supp}(f) \subset U$. (This is slightly stronger than the condition $0 \le f \le \chi_U$, which implies only that $\operatorname{supp}(f) \subset \overline{U}$.)

7.2 The Riesz Representation Theorem. *If I is a positive linear functional on $C_c(X)$, there is a unique Radon measure μ on X such that $I(f) = \int f\, d\mu$ for all $f \in C_c(X)$. Moreover, μ satisfies*

$$\mu(U) = \sup\{I(f) : f \in C_c(X),\ f \prec U\} \text{ for all open } U \subset X \tag{7.3}$$

and

$$\mu(K) = \inf\{I(f) : f \in C_c(X),\ f \ge \chi_K\} \text{ for all compact } K \subset X. \tag{7.4}$$

Proof. Let us begin by establishing uniqueness. If μ is a Radon measure such that $I(f) = \int f\, d\mu$ for all $f \in C_c(X)$, and $U \subset X$ is open, then clearly $I(f) \le \mu(U)$

whenever $f \prec U$. On the other hand, if $K \subset U$ is compact, by Urysohn's lemma there is an $f \in C_c(X)$ such that $f \prec U$ and $f = 1$ on K, whence $\mu(K) \leq \int f\,d\mu = I(f)$. Since μ is inner regular on U, it follows that (7.3) is satisfied. Thus μ is determined by I on open sets, and hence on all Borel sets because of outer regularity.

This argument proves the uniqueness of μ and also suggests how to go about proving existence. Namely, we begin by defining

$$\mu(U) = \sup\{I(f) : f \in C_c(X),\ f \prec U\}$$

for U open, and we then define $\mu^*(E)$ for an arbitrary $E \subset X$ by

$$\mu^*(E) = \inf\{\mu(U) : U \supset E,\ U \text{ open}\}.$$

Clearly $\mu(U) \leq \mu(V)$ if $U \subset V$, and hence $\mu^*(U) = \mu(U)$ if U is open.

The outline of the proof is now as follows. We shall establish that

i. μ^* is an outer measure.

ii. Every open set is μ^*-measurable.

At this point it follows from Carathéodory's theorem that every Borel set is μ^*-measurable and that $\mu = \mu^*|\mathcal{B}_X$ is a Borel measure. (The notation is consistent because $\mu^*(U) = \mu(U)$ for U open.) The measure μ is outer regular and satisfies (7.3) by definition. We next show that

iii. μ satisfies (7.4).

This clearly implies that μ is finite on compact sets, and inner regularity on open sets also follows easily. Indeed, if U is open and $\alpha < \mu(U)$, choose $f \in C_c(X)$ such that $f \prec U$ and $I(f) > \alpha$, and let $K = \operatorname{supp}(f)$. If $g \in C_c(X)$ and $g \geq \chi_K$, then $g - f \geq 0$ and hence $I(g) \geq I(f) > \alpha$. But then $\mu(K) > \alpha$ by (7.4), so μ is inner regular on U. Finally, we prove that

iv. $I(f) = \int f\,d\mu$ for all $f \in C_c(X)$.

With this, the proof of the theorem will be complete.

Proof of (i): It suffices to show that if $\{U_j\}$ is a sequence of open sets and $U = \bigcup_1^\infty U_j$, then $\mu(U) \leq \sum_1^\infty \mu(U_j)$. Indeed, from this it follows that for any $E \subset X$,

$$\mu^*(E) = \inf\Big\{\sum_1^\infty \mu(U_j) : U_j \text{ open}, E \subset \bigcup_1^\infty U_j\Big\},$$

and the expression on the right defines an outer measure by Proposition 1.10. If $U = \bigcup_1^\infty U_j$, $f \in C_c(X)$, and $f \prec U$, let $K = \operatorname{supp}(f)$. Since K is compact, we have $K \subset \bigcup_1^n U_j$ for some finite n, so by Proposition 4.41 there exist $g_1, \ldots, g_n \in C_c(X)$ with $g_j \prec U_j$ and $\sum_1^n g_j = 1$ on K. But then $f = \sum_1^n fg_j$ and $fg_j \prec U_j$, so

$$I(f) = \sum_1^n I(fg_j) \leq \sum_1^n \mu(U_j) \leq \sum_1^\infty \mu(U_j).$$

Since this is true for any $f \prec U$, we conclude that $\mu(U) \leq \sum_1^\infty \mu(U_j)$ as desired.

Proof of (ii): We must show that if U is open and E is any subset of X such that $\mu^*(E) < \infty$, then $\mu^*(E) \geq \mu^*(E \cap U) + \mu^*(E \setminus U)$. First suppose that E is open. Then $E \cap U$ is open, so given $\epsilon > 0$ we can find $f \in C_c(X)$ such that $f \prec E \cap U$ and $I(f) > \mu(E \cap U) - \epsilon$. Also, $E \setminus (\mathrm{supp}(f))$ is open, so we can find $g \in C_c(X)$ such that $g \prec E \setminus \mathrm{supp}(f)$ and $I(g) > \mu(E \setminus \mathrm{supp}(f)) - \epsilon$. But then $f + g \prec E$, so

$$\begin{aligned}\mu(E) &\geq I(f) + I(g) > \mu(E \cap U) + \mu(E \setminus \mathrm{supp}(f)) - 2\epsilon \\ &\geq \mu^*(E \cap U) + \mu^*(E \setminus U) - 2\epsilon.\end{aligned}$$

Letting $\epsilon \to 0$, we obtain the desired inequality. For the general case, if $\mu^*(E) < \infty$ we can find an open $V \supset E$ such that $\mu(V) < \mu^*(E) + \epsilon$, and hence

$$\begin{aligned}\mu^*(E) + \epsilon > \mu(V) &\geq \mu^*(V \cap U) + \mu^*(V \setminus U) \\ &\geq \mu^*(E \cap U) + \mu^*(E \setminus U).\end{aligned}$$

Letting $\epsilon \to 0$, we are done.

Proof of (iii): If K is compact, $f \in C_c(X)$, and $f \geq \chi_K$, let $U_\epsilon = \{x : f(x) > 1 - \epsilon\}$. Then U_ϵ is open, and if $g \prec U_\epsilon$, we have $(1-\epsilon)^{-1} f - g \geq 0$ and so $I(g) \leq (1-\epsilon)^{-1} I(f)$. Thus $\mu(K) \leq \mu(U_\epsilon) \leq (1-\epsilon)^{-1} I(f)$, and letting $\epsilon \to 0$ we see that $\mu(K) \leq I(f)$. On the other hand, for any open $U \supset K$, by Urysohn's lemma there exists $f \in C_c(X)$ such that $f \geq \chi_K$ and $f \prec U$, whence $I(f) \leq \mu(U)$. Since μ is outer regular on K, (7.4) follows.

Proof of (iv): If suffices to show that $I(f) = \int f \, d\mu$ if $f \in C_c(X, [0,1])$, as $C_c(X)$ is the linear span of the latter set. Given $N \in \mathbb{N}$, for $1 \leq j \leq N$ let $K_j = \{x : f(x) \geq jN^{-1}\}$ and let $K_0 = \mathrm{supp}(f)$. Also, define $f_1, \ldots, f_N \in C_c(X)$ by $f_j(x) = 0$ if $x \notin K_{j-1}$, $f_j(x) = f(x) - (j-1)N^{-1}$ if $x \in K_{j-1} \setminus K_j$, and $f_j(x) = N^{-1}$ if $x \in K_j$. In other words,

$$f_j = \min\left\{\max\left\{f - \frac{j-1}{N}, 0\right\}, \frac{1}{N}\right\}.$$

Then $N^{-1}\chi_{K_j} \leq f_j \leq N^{-1}\chi_{K_{j-1}}$, hence

$$\frac{1}{N}\mu(K_j) \leq \int f_j \, d\mu \leq \frac{1}{N}\mu(K_{j-1}).$$

Also, if U is an open set containing K_{j-1} we have $Nf_j \prec U$ and so $I(f_j) \leq N^{-1}\mu(U)$. Hence, by (7.4) and outer regularity,

$$\frac{1}{N}\mu(K_j) \leq I(f_j) \leq \frac{1}{N}\mu(K_{j-1}).$$

Moreover, $f = \sum_1^N f_j$, so that

$$\frac{1}{N}\sum_1^N \mu(K_j) \le \int f\,d\mu \le \frac{1}{N}\sum_0^{N-1} \mu(K_j),$$
$$\frac{1}{N}\sum_1^N \mu(K_j) \le I(f) \le \frac{1}{N}\sum_0^{N-1} \mu(K_j).$$

It follows that

$$\left|I(f) - \int f\,d\mu\right| \le \frac{\mu(K_0) - \mu(K_N)}{N} \le \frac{\mu(\operatorname{supp}(f))}{N}.$$

Since $\mu(\operatorname{supp}(f)) < \infty$ and N is arbitrary, we conclude that $I(f) = \int f\,d\mu$. ∎

The proof of this theorem yields something stronger than the statement: We obtain not just a Borel measure μ but an extension $\overline{\mu}$ of μ to the σ-algebra of μ^*-measurable sets. However, it follows from outer regularity that for any $E \subset X$,

$$\mu^*(E) = \inf\{\mu(B) : B \in \mathcal{B}_X,\ B \supset E\},$$

so μ^* is the outer measure induced by μ in the sense of §1.4. According to Exercise 22 in §1.4, therefore, $\overline{\mu}$ is the completion of μ if μ is σ-finite and is the saturation of the completion of μ in general.

On the other hand, some authors prefer to restrict attention to a smaller σ-algebra than $\mathcal{B}_X$, namely, the σ-algebra $\mathcal{B}_X^0$ generated by $C_c(X)$ (that is, the smallest σ-algebra with respect to which every $f \in C_c(X)$ is measurable). The elements of $\mathcal{B}_X^0$ are called **Baire sets**. For more about Baire sets, see Exercises 4–6.

Exercises

1. Let X be an LCH space, Y a closed subset of X (which is an LCH space in the relative topology), and μ a Radon measure on Y. Then $I(f) = \int (f|Y)\,d\mu$ is a positive linear functional on $C_c(X)$, and the induced Radon measure ν on X is given by $\nu(E) = \mu(E \cap Y)$.

2. Let μ be a Radon measure on X.
a. Let N be the union of all open $U \subset X$ such that $\mu(U) = 0$. Then N is open and $\mu(N) = 0$. The complement of N is called the **support** of μ.
b. $x \in \operatorname{supp}(\mu)$ iff $\int f\,d\mu > 0$ for every $f \in C_c(X, [0,1])$ such that $f(x) > 0$.

3. Let X be the one-point compactification of a set with the discrete topology. If μ is a Radon measure on X, then $\operatorname{supp}(\mu)$ (see Exercise 2) is countable.

4. Let X be an LCH space.
a. If $f \in C_c(X, [0,\infty))$, then $f^{-1}([a,\infty))$ is a compact G_δ set for all $a > 0$.
b. If $K \subset X$ is a compact G_δ set, there exists $f \in C_c(X, [0,1])$ such that $K = f^{-1}(\{1\})$.

c. The σ-algebra $\mathcal{B}_X^0$ of Baire sets is the σ-algebra generated by the compact G_δ sets.

5. Let X be a second countable LCH space.
a. Every compact subset of X is a G_δ set.
b. $\mathcal{B}_X = \mathcal{B}_X^0$.

6. Let X be an uncountable set with the discrete topology, or the one-point compactification of such a set. Then $\mathcal{B}_X \neq \mathcal{B}_X^0$.

7.2 REGULARITY AND APPROXIMATION THEOREMS

In this section we explore the properties of Radon measures in more detail.

7.5 Proposition. *Every Radon measure is inner regular on all of its σ-finite sets.*

Proof. Suppose that μ is Radon and E is σ-finite. If $\mu(E) < \infty$, for any $\epsilon > 0$ we can choose an open $U \supset E$ such that $\mu(U) < \mu(E) + \epsilon$ and a compact $F \subset U$ such that $\mu(F) > \mu(U) - \epsilon$. Since $\mu(U \setminus E) < \epsilon$, we can also choose an open $V \supset U \setminus E$ such that $\mu(V) < \epsilon$. Let $K = F \setminus V$. Then K is compact, $K \subset E$, and

$$\mu(K) = \mu(F) - \mu(F \cap V) > \mu(E) - \epsilon - \mu(V) > \mu(E) - 2\epsilon.$$

Thus μ is inner regular on E. On the other hand, if $\mu(E) = \infty$, E is an increasing union of sets E_j with $\mu(E_j) < \infty$ and $\mu(E_j) \to \infty$. Thus for any $N \in \mathbb{N}$ there exists j such that $\mu(E_j) > N$ and hence, by the preceding argument, a compact $K \subset E_j$ with $\mu(K) > N$. Hence μ is inner regular on E. ∎

7.6 Corollary. *Every σ-finite Radon measure is regular. If X is σ-compact, every Radon measure on X is regular.*

For an example of a nonregular Radon measure, see Exercise 12.

7.7 Proposition. *Suppose that μ is a σ-finite Radon measure on X and E is a Borel set in X.*

a. For every $\epsilon > 0$ there exist an open U and a closed F with $F \subset E \subset U$ and $\mu(U \setminus F) < \epsilon$.

b. There exist an F_σ set A and a G_δ set B such that $A \subset E \subset B$ and $\mu(B \setminus A) = 0$.

Proof. Write $E = \bigcup_1^\infty E_j$ where the E_j's are disjoint and have finite measure. For each j, choose an open $U_j \supset E_j$ with $\mu(U_j) < \mu(E_j) + \epsilon 2^{-j-1}$ and let $U = \bigcup_1^\infty U_j$. Then U is open, $U \supset E$, and $\mu(U \setminus E) \leq \sum_1^\infty \mu(U_j \setminus E_j) < \epsilon/2$. Applying the same reasoning to E^c, we obtain an open $V \supset E^c$ with $\mu(V \setminus E^c) < \epsilon/2$. Let $F = V^c$. Then F is closed, $F \subset E$, and

$$\mu(U \setminus F) = \mu(U \setminus E) + \mu(E \setminus F) = \mu(U \setminus E) + \mu(V \setminus E^c) < \epsilon.$$

This proves (a), and (b) follows easily; details are left to the reader. ∎

7.8 Theorem. *Let X be an LCH space in which every open set is σ-compact (which is the case, for example, if X is second countable). Then every Borel measure on X that is finite on compact sets is regular and hence Radon.*

Proof. If μ is a Borel measure that is finite on compact sets, then $C_c(X) \subset L^1(\mu)$, so the map $I(f) = \int f\,d\mu$ is a positive linear functional on $C_c(X)$. Let ν be the associated Radon measure according to Theorem 7.2. If $U \subset X$ is open, let $U = \bigcup_1^\infty K_j$ where each K_j is compact. Choose $f_1 \in C_c(X)$ so that $f \prec U$ and $f = 1$ on K_1. Proceeding inductively, for $n > 1$ choose $f_n \in C_c(X)$ so that $f_n \prec U$ and $f_n = 1$ on $\bigcup_1^n K_j$ and on $\bigcup_1^{n-1} \operatorname{supp}(f_j)$. Then f_n increases pointwise to χ_U as $n \to \infty$, so

$$\mu(U) = \lim \int f_n\,d\mu = \lim \int f_n\,d\nu = \nu(U)$$

by the monotone convergence theorem. Next, if E is any Borel set and $\epsilon > 0$, by Proposition 7.7 there exist an open $V \supset E$ and a closed $F \subset E$ with $\nu(V \setminus F) < \epsilon$. But $V \setminus F$ is open, so $\mu(V \setminus F) = \nu(V \setminus F) < \epsilon$. In particular, $\mu(V) \le \mu(E) + \epsilon$, so μ is outer regular. Also, $\mu(F) \ge \mu(E) - \epsilon$, and F is σ-compact (since X is), so there exist compact $K_j \subset F$ with $\mu(K_j) \to \mu(F)$, whence μ is inner regular. Thus μ is regular (and equal to ν, by the uniqueness part of Theorem 7.2.) ∎

Examples of non-Radon measures are considered in Exercises 13–15. In particular, Exercise 15 exhibits an example of a finite, non-Radon Borel measure on a compact Hausdorff space.

We now turn to some approximation theorems for measurable functions.

7.9 Proposition. *If μ is a Radon measure on X, $C_c(X)$ is dense in $L^p(\mu)$ for $1 \le p < \infty$.*

Proof. Since the L^p simple functions are dense in L^p (Proposition 6.7), it suffices to show that for any Borel set E with $\mu(E) < \infty$, χ_E can be approximated in the L^p norm by elements of $C_c(X)$. Given $\epsilon > 0$, by Proposition 7.5 we can choose a compact $K \subset E$ and an open $U \supset E$ such that $\mu(U \setminus K) < \epsilon$, and by Urysohn's lemma we can choose $f \in C_c(X)$ such that $\chi_K \le f \le \chi_U$. Then $\|\chi_E - f\|_p \le \mu(U \setminus K)^{1/p} < \epsilon^{1/p}$, so we are done. ∎

7.10 Lusin's Theorem. *Suppose that μ is a Radon measure on X and $f : X \to \mathbb{C}$ is a measurable function that vanishes outside a set of finite measure. Then for any $\epsilon > 0$ there exists $\phi \in C_c(X)$ such that $\phi = f$ except on a set of measure $< \epsilon$. If f is bounded, ϕ can be taken to satisfy $\|\phi\|_u \le \|f\|_u$.*

Proof. Let $E = \{x : f(x) \ne 0\}$, and suppose to begin with that f is bounded. Then $f \in L^1(\mu)$, so by Proposition 7.9 there is a sequence $\{g_n\}$ in $C_c(X)$ that converges to f in L^1, and hence by Corollary 2.32 a subsequence (still denoted by $\{g_n\}$) that converges to f a.e. By Egoroff's theorem there is a set $A \subset E$ such that $\mu(E \setminus A) < \epsilon/3$ and $g_n \to f$ uniformly on A, and there exist a compact $B \subset A$ and

an open $U \supset E$ such that $\mu(A \setminus B) < \epsilon/3$ and $\mu(U \setminus E) < \epsilon/3$. Since $g_n \to f$ uniformly on B, $f|B$ is continuous, so by Theorem 4.34 there exists $h \in C_c(X)$ such that $h = f$ on B and $\operatorname{supp}(h) \subset U$. But then $\{x : f(x) \neq h(x)\}$ is contained in $U \setminus B$, which has measure $< \epsilon$.

To complete the proof for f bounded, define $\beta : \mathbb{C} \to \mathbb{C}$ by $\beta(z) = z$ if $|z| \leq \|f\|_u$ and $\beta(z) = \|f\|_u \operatorname{sgn} z$ if $|z| > \|f\|_u$, and set $\phi = \beta \circ h$. Then $\phi \in C_c(X)$ since β is continuous and $\beta(0) = 0$. Moreover, $\|\phi\|_u \leq \|f\|_u$, and $\phi = f$ on the set where $h = f$, so we are done.

If f is unbounded, let $A_n = \{x : 0 < |f(x)| \leq n\}$. Then A_n increases to E as $n \to \infty$, so $\mu(E \setminus A_n) < \epsilon/2$ for sufficiently large n. By the preceding argument there exists $\phi \in C_c(X)$ such that $\phi = f\chi_{A_n}$ except on a set of measure $< \epsilon/2$, and hence $\phi = f$ except on a set of measure $< \epsilon$. ∎

Our final group of results in this section concerns semicontinuous functions. If X is a topological space, a function $f : X \to (-\infty, \infty]$ is called **lower semicontinuous (LSC)** if $\{x : f(x) > a\}$ is open for all $a \in \mathbb{R}$, and $f : X \to [-\infty, \infty)$ is called **upper semicontinuous (USC)** if $\{x : f(x) < a\}$ is open for all $a \in \mathbb{R}$.

7.11 Proposition. *Let X be a topological space.*

a. If U is open in X, then χ_U is LSC.

b. If f is LSC and $c \in [0, \infty)$, then cf is LSC.

c. If $\mathcal{G}$ is a family of LSC functions and $f(x) = \sup\{g(x) : g \in \mathcal{G}\}$, then f is LSC.

d. If f_1 and f_2 are LSC, so is $f_1 + f_2$.

e. If X is an LCH space and f is LSC and nonnegative, then

$$f(x) = \sup\{g(x) : g \in C_c(X),\ 0 \leq g \leq f\}.$$

Proof. (a) and (b) are obvious, and (c) follows from the observation that

$$f^{-1}((a, \infty]) = \bigcup_{g \in \mathcal{G}} g^{-1}((a, \infty]).$$

As for (d), if $f_1(x_0) + f_2(x_0) > a$, choose $\epsilon > 0$ so that $f_1(x_0) > a - f_2(x_0) + \epsilon$. Then

$$\{x : (f_1 + f_2)(x) > a\} \supset \{x : f_1(x) > a - f_2(x_0) + \epsilon\} \cap \{x : f_2(x) > f_2(x_0) - \epsilon\},$$

which is a neighborhood of x_0. Thus $f_1 + f_2$ is LSC. Finally, if X is LCH, $f(x) > 0$, and $0 < a < f(x)$, then $U = \{y : f(y) > a\}$ is an open set containing x, so by Urysohn's lemma there exists $g \in C_c(X)$ such that $g(x) = a$ and $0 \leq g \leq a\chi_u \leq f$. This establishes (e) when $f(x) > 0$, and (e) is trivial when $f(x) = 0$. ∎

There is, of course, a corresponding set of results for USC functions, whose formulation is left to the reader. The following result is a monotone convergence theorem for nets of LSC functions.

7.12 Proposition. *Let $\mathcal{G}$ be a family of nonnegative LSC functions on an LCH space X that is directed by $\leq$ (that is, for every $g_1, g_2 \in \mathcal{G}$ there exists $g \in \mathcal{G}$ such that $g_1 \leq g$ and $g_2 \leq g$). Let $f = \sup\{g : g \in \mathcal{G}\}$. If μ is any Radon measure on X, then $\int f\, d\mu = \sup\{\int g\, d\mu : g \in \mathcal{G}\}$.*

Proof. By Proposition 7.11c, f is LSC and hence Borel measurable, and clearly $\int f\, d\mu \geq \sup\{\int g\, d\mu\}$. To prove the reverse inequality, consider the sequence ϕ_n of simple functions increasing to f that was constructed in Theorem 2.10:

$$\phi_n = \frac{1}{2^n} \sum_{j=1}^{2^{2n}} \chi_{U_{nj}}, \text{ where } U_{nj} = \{x : f(x) > j2^{-n}\}.$$

By the monotone convergence theorem, given $a < \int f\, d\mu$ we can fix n large enough so that $2^{-n} \sum_j \mu(U_{nj}) = \int \phi_n\, d\mu > a$. Since U_{nj} is open, there exist compact $K_j \subset U_{nj}$ $(1 \leq j \leq 2^{2n})$ such that $2^{-n} \sum_j \mu(K_j) > a$. Let $\psi = 2^{-n} \sum_j \chi_{K_j}$. For each $x \in \bigcup_j K_j$ we have $f(x) > \phi_n(x) \geq \psi(x)$, so we can pick $g_x \in \mathcal{G}$ such that $g_x(x) > \psi(x)$. But $-\chi_{K_j}$ is LSC, so $g_x - \psi$ is LSC by Proposition 7.11d, and hence the set $V_x = \{y : \psi(y) < g_x(y)\}$ is open. Thus $\{V_x : x \in \bigcup_j K_j\}$ is an open cover of $\bigcup_j K_j$, so there is a finite subcover $V_{x_1}, \ldots, V_{x_m}$. Pick $g \in \mathcal{G}$ such that $g_{x_k} \leq g$ for $k = 1, \ldots, m$; then $\psi \leq g$, so $\int g\, d\mu > a$. Since a was any number less than $\int f\, d\mu$, we are done. ∎

7.13 Corollary. *If μ is Radon and f is nonnegative and LSC, then*

$$\int f\, d\mu = \sup\left\{\int g\, d\mu : g \in C_c(X),\ 0 \leq g \leq f\right\}.$$

Proof. Combine Propositions 7.11e and 7.12. ∎

7.14 Proposition. *If μ is a Radon measure and f is a nonnegative Borel measurable function, then*

$$\int f\, d\mu = \inf\left\{\int g\, d\mu : g \geq f \text{ and } g \text{ is LSC}\right\}.$$

If $\{x : f(x) > 0\}$ is σ-finite, then

$$\int f\, d\mu = \sup\left\{\int g\, d\mu : 0 \leq g \leq f \text{ and } g \text{ is USC}\right\}.$$

Proof. Let $\{\phi_n\}$ be a sequence of nonnegative simple functions that increase pointwise to f. Then $f = \phi_1 + \sum_2^\infty (\phi_n - \phi_{n-1})$, and each term in this series is a nonnegative simple function, so we can write $f = \sum_1^\infty a_j \chi_{E_j}$ where $a_j > 0$. Given $\epsilon > 0$, for each j choose an open $U_j \supset E_j$ such that $\mu(U_j) \leq \mu(E_j) + \epsilon/(2^j a_j)$. Then $g = \sum_1^\infty a_j \chi_{U_j}$ is LSC by Proposition 7.11, $g \geq f$, and $\int g\, d\mu \leq \int f\, d\mu + \epsilon$. This establishes the first assertion. For the second, if $a < \int f\, d\mu$, let N be large enough so that $\sum_1^N a_j \mu(E_j) > a$. Since the E_j's are σ-finite, by Proposition 7.5 there are compact sets $K_j \subset E_j$ such that $\sum_1^N a_j \mu(K_j) > a$. Thus if $g = \sum_1^n a_j \chi_{K_j}$, then g is USC, $g \leq f$, and $\int g\, d\mu > a$. ∎

Exercises

7. If μ is a σ-finite Radon measure on X and $A \in \mathcal{B}_X$, the Borel measure μ_A defined by $\mu_A(E) = \mu(E \cap A)$ is a Radon measure. (See also Exercise 13.)

8. Suppose that μ is a Radon measure on X. If $\phi \in L^1(\mu)$ and $\phi \geq 0$, then $\nu(E) = \int_E \phi \, d\mu$ is a Radon measure. (Use Corollary 3.6.)

9. Suppose that μ is a Radon measure on X and $\phi \in C(X, (0, \infty))$. Let $\nu(E) = \int_E \phi \, d\mu$, and let ν' be the Radon measure associated to the functional $f \mapsto \int f\phi \, d\mu$ on $C_c(X)$.

a. If U is open, $\nu(U) = \nu'(U)$. (Apply Corollary 7.13 to $\phi\chi_U$.)

b. ν is outer regular on all Borel sets. (Hint: The open sets $V_k = \{x : 2^k < \phi(x) < 2^{k+2}\}$, $k \in \mathbb{Z}$, cover X.)

c. $\nu = \nu'$, and hence ν is a Radon measure. (See also Exercise 13.)

10. If μ is a Radon measure and $f \in L^1(\mu)$ is real-valued, for every $\epsilon > 0$ there exist an LSC function g and a USC function h such that $h \leq f \leq g$ and $\int (g - h) \, d\mu < \epsilon$.

11. Suppose that μ is a Radon measure on X such that $\mu(\{x\}) = 0$ for all $x \in X$, and $A \in \mathcal{B}_X$ satisfies $0 < \mu(A) < \infty$. Then for any α such that $0 < \alpha < \mu(A)$ there is a Borel set $B \subset A$ such that $\mu(B) = \alpha$.

12. Let $X = \mathbb{R} \times \mathbb{R}_d$, where $\mathbb{R}_d$ denotes $\mathbb{R}$ with the discrete topology. If f is a function on X, let $f^y(x) = f(x, y)$; and if $E \subset X$, let $E^y = \{x : (x, y) \in E\}$.

a. $f \in C_c(X)$ iff $f^y \in C_c(\mathbb{R})$ for all y and $f^y = 0$ for all but finitely many y.

b. Define a positive linear functional on $C_c(X)$ by $I(f) = \sum_{y \in \mathbb{R}} \int f(x, y) \, dx$, and let μ be the associated Radon measure on X. Then $\mu(E) = \infty$ for any E such that $E^y \neq \varnothing$ for uncountably many y.

c. Let $E = \{0\} \times \mathbb{R}_d$. Then $\mu(E) = \infty$ but $\mu(K) = 0$ for all compact $K \subset E$.

13. In the setting of Exercise 12, let $A = (\mathbb{R} \setminus \{0\}) \times \mathbb{R}_d$ and $\phi(x, y) = |x|$. Then the measures $\mu_A(E) = \mu(A \cap E)$ and $\nu(E) = \int_E \phi \, d\mu$ are not Radon. (Thus, the hypotheses that μ be σ-finite in Exercise 7, that $\phi \in L^1(\mu)$ in Exercise 8, and that $\phi > 0$ in Exercise 9, cannot be dropped.)

14. Let μ be a Radon measure on X, and let μ_0 be the semifinite part of μ (see Exercise 15 in §1.3).

a. μ_0 is inner regular on all Borel sets.

b. μ_0 is outer regular on all Borel sets E such that $\mu(E) < \infty$.

c. $\int f \, d\mu = \int f \, d\mu_0$ for all $f \in C_c(X)$.

d. If μ is the measure of Exercise 12 and m is Lebesgue measure on $\mathbb{R}$, then $\mu_0(E) = \sum_{y \in \mathbb{R}} m(E^y)$ for any Borel set E.

15. Let Ω be the set of countable ordinals, ω_1 the first uncountable ordinal, and $\Omega^* = \Omega \cup \{\omega_1\}$. Let Ω^* be endowed with the order topology (see Exercise 9 in §4.1).

a. Ω^* is a compact Hausdorff space. (Hint: Ω^* contains no infinite strictly decreasing sequences.)

b. Ω is an open set in Ω^* that is not σ-compact.
c. A subset E of Ω is uncountable iff for each $x \in \Omega$ there exists $y \in E$ such that $x < y$.
d. If $\{E_n\}$ is a sequence of uncountable closed sets in Ω^*, then $\bigcap_1^\infty E_n$ is uncountable. (If $\{x_j\}$ is an increasing sequence in Ω such that each E_n contains infinitely many x_j's, then $\lim_{j\to\infty} x_j$ exists and is in $\bigcap_1^\infty E_n$.)
e. If $E \subset \mathcal{B}_{\Omega^*}$, then either $E \cup \{\omega_1\}$ or $E^c \cup \{\omega_1\}$ contains an uncountable closed set. (Hint: The set of all E satisfying the latter condition is a σ-algebra.)
f. Define μ on $\mathcal{B}_{\Omega^*}$ by $\mu(E) = 1$ if $E \cup \{\omega_1\}$ contains an uncountable closed set, $\mu(E) = 0$ otherwise. Then μ is a measure, $\mu(\{\omega_1\}) = 0$, but $\mu(U) = 1$ for every open U containing ω_1.
g. If $f \in C(\Omega^*)$, there exists $x \in \Omega$ such that $f(y) = f(\omega_1)$ for $y \geq x$. (If $E_n = \{x : |f(x) - f(\omega_1)| < n^{-1}\}$, then E_n^c is countable.)
h. With μ as in (f), the Radon meausre on Ω^* associated to the functional $f \mapsto \int f\,d\mu$ is the point mass at ω_1.

7.3 THE DUAL OF $C_0(X)$

We recall that for any LCH space X, $C_0(X)$ is the uniform closure of $C_c(X)$ (Proposition 4.35), and hence if μ is a Radon measure on X, the functional $I(f) = \int f\,d\mu$ extends continuously to $C_0(X)$ iff it is bounded with respect to the uniform norm. In view of the equality

$$\mu(X) = \sup\left\{\int f\,d\mu : f \in C_c(X),\ 0 \leq f \leq 1\right\}$$

(a special case of (7.3)) together with the fact that $|\int f\,d\mu| \leq \int |f|\,d\mu$, this happens precisely when $\mu(X) < \infty$, in which case $\mu(X)$ is the operator norm of I.

We have therefore identified the positive bounded linear functionals on $C_0(X)$: they are given by integration against finite Radon measures. Our object in this section is to extend this result to give a complete description of $C_0(X)^*$. The key fact is that real linear functionals on $C_0(X, \mathbb{R})$ have a "Jordan decomposition."

7.15 Lemma. *If $I \in C_0(X, \mathbb{R})^*$, there exist positive functionals $I^\pm \in C_0(X, \mathbb{R})^*$ such that $I = I^+ - I^-$.*

Proof. If $f \in C_0(X, [0, \infty))$, we define

$$I^+(f) = \sup\{I(g) : g \in C_0(X, \mathbb{R}),\ 0 \leq g \leq f\}.$$

Since $|I(g)| \leq \|I\|\,\|g\|_u \leq \|I\|\,\|f\|_u$ for $0 \leq g \leq f$, and $I(0) = 0$, we have $0 \leq I^+(f) \leq \|I\|\,\|f\|_u$. We claim that I^+ is the restriction to $C_0(X, [0, \infty))$ of a linear functional; the proof is much the same as the proof of the linearity of the integral in §2.3.

Obviously $I^+(cf) = cI^+(f)$ if $c \geq 0$. Also, whenever $0 \leq g_1 \leq f_1$ and $0 \leq g_2 \leq f_2$ we have $0 \leq g_1+g_2 \leq f_1+f_2$, so that $I^+(f_1+f_2) \geq I(g_1)+I(g_2)$, and it follows that $I^+(f_1+f_2) \geq I^+(f_1)+I^+(f_2)$. On the other hand, if $0 \leq g \leq f_1+f_2$, let $g_1 = \min(g, f_1)$ and $g_2 = g - g_1$. Then $0 \leq g_1 \leq f_1$ and $0 \leq g_2 \leq f_2$, so $I(g) = I(g_1)+I(g_2) \leq I^+(f_1)+I^+(f_2)$; therefore $I^+(f_1+f_2) \leq I^+(f_1)+I^+(f_2)$. In short, $I^+(f_1 + f_2) = I^+(f_1) + I^+(f_2)$.

Now, if $f \in C_0(X, \mathbb{R})$, then its positive and negative parts f^+ and f^- are in $C_0(X, [0, \infty))$, and we define $I^+(f) = I^+(f^+) - I^+(f^-)$. If also $f = g - h$ where $g, h \geq 0$, then $g + f^- = h + f^+$, whence $I^+(g) + I^+(f^-) = I^+(h) + I^+(f^+)$. Thus $I^+(f) = I^+(g) - I^+(h)$, and it follows easily as in the proof of Proposition 2.21 that I^+ is linear on $C_0(X, \mathbb{R})$. Moreover,

$$|I^+(f)| \leq \max\bigl(I^+(f^+),\, I^+(f^-)\bigr) \leq \|I\| \max\bigl(\|f^+\|_u,\, \|f^-\|_u\bigr) = \|I\|\, \|f\|_u,$$

so that $\|I^+\| \leq \|I\|$.

Finally, let $I^- = I^+ - I$. Then $I^- \in C_0(X, \mathbb{R})^*$, and it is immediate from the definition of I^+ that I^+ and I^- are positive. ∎

Any $I \in C_0(X)^*$ is uniquely determined by its restriction J to $C_0(X, \mathbb{R})$, and we have $J = J_1 + iJ_2$ where J_1, J_2 are real linear functionals. We therefore conclude from Lemma 7.15 and the discussion preceding it that for any $I \in C_0(X)^*$ there are finite Radon measures $\mu_1, \ldots, \mu_4$ such that $I(f) = \int f\, d\mu$ where $\mu = \mu_1 - \mu_2 + i(\mu_3 - \mu_4)$.

At this point we need some more definitions. A **signed Radon measure** is a signed Borel measure whose positive and negative variations are Radon, and a **complex Radon measure** is a complex Borel measure whose real and imaginary parts are signed Radon measures. (It is worth noting that on a second countable LCH space, every complex Borel measure is Radon. This follows from Theorem 7.8 since complex measures are bounded.) We denote that space of complex Radon measures on X by $M(X)$, and for $\mu \in M(X)$ we define

$$\|\mu\| = |\mu|(X),$$

where, of course, $|\mu|$ is the total variation of μ.

7.16 Proposition. *If μ is a complex Borel measure, then μ is Radon iff $|\mu|$ is Radon. Moreover, $M(X)$ is a vector space and $\mu \mapsto \|\mu\|$ is a norm on it.*

Proof. We observe that a finite positive Borel measure ν is Radon iff for every Borel set E and every $\epsilon > 0$ there exist a compact K and an open U such that $K \subset E \subset U$ and $\nu(U \setminus K) < \epsilon$, by Propositions 7.5 and 7.7. The first assertion follows easily from this. Indeed, if $\mu = \mu_1 - \mu_2 + i(\mu_3 - \mu_4)$ and $|\mu|(U \setminus K) < \epsilon$, then $\mu_j(U \setminus K) < \epsilon$ for all j; conversely, if $\mu_j(U_j \setminus K_j) < \epsilon/4$ for all j, then $\mu(U \setminus K) < \epsilon$ where $K = \bigcup_1^4 K_j$ and $U = \bigcap_1^4 U_j$. The same argument shows that $M(X)$ is closed under addition and scalar multiplication. Finally, that $\|\cdot\|$ is a norm on $M(X)$ follows from Proposition 3.14. ∎

7.17 The Riesz Representation Theorem. *Let X be an LCH space, and for $\mu \in M(X)$ and $f \in C_0(X)$ let $I_\mu(f) = \int f\,d\mu$. Then the map $\mu \mapsto I_\mu$ is an isometric isomorphism from $M(X)$ to $C_0(X)^*$.*

Proof. We have already shown that every $I \in C_0(X)^*$ is of the form I_μ. On the other hand, if $\mu \in M(X)$,by Proposition 3.13c we have

$$\left|\int f\,d\mu\right| \le \int |f|\,d|\mu| \le \|f\|_u \|\mu\|,$$

so $I_\mu \in C_0(X)^*$ and $\|I_\mu\| \le \|\mu\|$. Moreover, if $h = d\mu/d|\mu|$, then $|h| = 1$ by Proposition 3.13b, so by Lusin's theorem, for any $\epsilon > 0$ there exists $f \in C_c(X)$ such that $\|f\|_u = 1$ and $f = \overline{h}$ except on a set E with $|\mu|(E) < \epsilon/2$. Then

$$\begin{aligned}\|\mu\| &= \int |h|^2 d|\mu| = \int \overline{h}\,d\mu \le \left|\int f\,d\mu\right| + \left|\int (f - \overline{h})\,d\mu\right| \\ &\le \left|\int f\,d\mu\right| + 2|\mu|(E) < \left|\int f\,d\mu\right| + \epsilon \le \|I_\mu\| + \epsilon.\end{aligned}$$

It follows that $\|\mu\| \le \|I_\mu\|$, so the proof is complete. ∎

7.18 Corollary. *If X is a compact Hausdorff space, then $C(X)^*$ is isometrically isomorphic to $M(X)$.*

Let μ be a fixed positive Radon measure on X. If $f \in L^1(\mu)$, the complex measure $d\nu_f = f\,d\mu$ is easily seen to be Radon (Exercise 8), and $\|\nu_f\| = \int |f|\,d\mu = \|f\|_1$. Thus $f \mapsto \nu_f$ is an isometric embedding of $L^1(\mu)$ into $M(X)$ whose range consists precisely of those $\nu \in M(X)$ such that $\nu \ll \mu$. (The last statement follows from the Radon-Nikodym theorem, which applies even if μ is not σ-finite; see §7.5.) The most important example of this situation is $\mu = m =$ Lebesgue measure on $\mathbb{R}^n$, and we shall identify $L^1(m)$ with a subspace of $M(\mathbb{R}^n)$.

The weak* topology on $M(X) = C_0(X)^*$, in which $\mu_\alpha \to \mu$ iff $\int f\,d\mu_\alpha \to \int f\,d\mu$ for all $f \in C_0(X)$, is of considerable importance in applications; we shall call it the **vague topology** on $M(X)$. (The term "vague" is common in probability theory and has the advantage of forming an adverb more gracefully than "weak*.") The vague topology is sometimes called the weak topology, but this terminology conflicts with ours, since $C_0(X)$ is rarely reflexive (see Exercise 20). Weak convergence arguments for $L^p(\mu)$ generally fail for $p = 1$ because $L^1(\mu)$ is not the dual of $L^\infty(\mu)$, but good substitute results can often be obtained by regarding $L^1(\mu)$ as a subspace of $M(X)$ as in the preceding paragraph and using the vague topology there.

We conclude by presenting a useful criterion for vague convergence in $M(\mathbb{R})$.

7.19 Proposition. *Suppose $\mu, \mu_1, \mu_2, \ldots \in M(\mathbb{R})$, and let $F_n(x) = \mu_n((-\infty, x])$ and $F(x) = \mu((-\infty, x])$.*

a. If $\sup_n \|\mu_n\| < \infty$ and $F_n(x) \to F(x)$ for every x at which F is continuous, then $\mu_n \to \mu$ vaguely.

b. If $\mu_n \to \mu$ vaguely, then $\sup_n \|\mu_n\| < \infty$. If, in addition, the μ_n's are positive, then $F_n(x) \to F(x)$ at every x at which F is continuous.

Proof. (a) Since F is continuous except at countably many points (Theorems 3.27 and 3.29), $F_n \to F$ a.e. with respect to Lebesgue measure. Also, $\|F_n\|_u \le \|\mu_n\|$, so the F_n's are uniformly bounded. If f is continuously differentiable and has compact support, then, integration by parts (Theorem 3.36) and the dominated convergence theorem yield

$$\int f\,d\mu_n = \int f'(x)F_n(x)\,dx \to \int f'(x)F(x)\,dx = \int f\,d\mu.$$

But by Theorem 4.52, the set of all such f's is dense in $C_0(\mathbb{R})$, so $\int f_n\,d\mu \to \int f\,d\mu$ for all $f \in C_0(\mathbb{R})$ by Proposition 5.17. Thus $\mu_n \to \mu$ vaguely.

(b) If $\mu_n \to \mu$ vaguely, then $\sup_n \|\mu_n\| < \infty$ by the uniform boundedness principle. Suppose that $\mu_n \ge 0$, and hence $\mu \ge 0$, and that F is continuous at $x = a$. If $f \in C_c(\mathbb{R})$ is the function that is 1 on $[-N, a]$, 0 on $(-\infty, -N-\epsilon]$ and $[a+\epsilon, \infty)$, and linear in between, we have

$$\begin{aligned} F_n(a) - F_n(-N) = \mu_n((-N,a]) \le \int f\,d\mu_n &\to \int f\,d\mu \\ &\le F(a+\epsilon) - F(-N-\epsilon). \end{aligned}$$

As $N \to \infty$, $F_n(-N)$ and $F(-N-\epsilon)$ tend to zero, so

$$\limsup_{n\to\infty} F_n(a) \le F(a+\epsilon).$$

Similarly, by considering the function that is 1 on $[-N+\epsilon, a-\epsilon]$, 0 on $(-\infty, N]$ and $[a, \infty)$, and linear in between, we see that

$$\liminf_{n\to\infty} F_n(a) \ge F(a-\epsilon).$$

Since ϵ is arbitrary and F is continuous at a, we have $F_n(a) \to F(a)$ as desired. ∎

Exercises

16. Suppose that $I \in C_0(X, \mathbb{R})^*$ and I^+, I^- are the functionals constructed in the proof of Lemma 7.15. If μ is the signed Radon measure associated to I, then the positive and negative variations of μ are the Radon measures associated to I^+ and I^-.

17. If μ is a positive Radon measure on X with $\mu(X) = \infty$, there exists $f \in C_0(X)$ such that $\int f\,d\mu = \infty$. Consequently, every positive linear functional on $C_0(X)$ is bounded.

18. If μ is a σ-finite Radon measure on X and $\nu \in \mathcal{M}(X)$, let $\nu = \nu_1 + \nu_2$ be the Lebesgue decomposition of ν with respect to μ. Then ν_1 and ν_2 are Radon. (Use Exercise 8.)

19. Let X be a completely regular space and $\mathcal{A}$ a completely regular subalgebra of $BC(X)$ (see Exercise 73 in §4.8). Find a description of $\mathcal{A}^*$ as a space of measures.

20. Some examples of nonreflexivity of $C_0(X)$:

a. If $\mu \in M(X)$, let $\Phi(\mu) = \sum_{x \in X} \mu(\{x\})$. This sum is well defined, and $\Phi \in M(X)^*$. If there exists a nonzero $\mu \in M(X)$ such that $\mu(\{x\}) = 0$ for all $x \in X$, then Φ is not in the image of $C_0(X)$ in $M(X)^* \cong C_0(X)^{**}$.

b. At the other extreme, let $X = \mathbb{N}$ with the discrete topology; then $C_0(X)^* \cong l^1$ and $(l^1)^* \cong l^\infty$. (*Note:* $C_0(\mathbb{N})$ is usually denoted by c_0.)

21. Let $\{f_\alpha\}_{\alpha \in A}$ be a subset of $C_0(X)$ and $\{c_\alpha\}_{\alpha \in A}$ a family of complex numbers. If for each finite set $B \subset A$ there exists $\mu_B \in M(X)$ such that $\|\mu_B\| \leq 1$ and $\int f_\alpha \, d\mu_B = c_\alpha$ for $\alpha \in B$, then there exists $\mu \in M(X)$ such that $\|\mu\| \leq 1$ and $\int f_\alpha \, d\mu = c_\alpha$ for all $\alpha \in A$.

22. A sequence $\{f_n\}$ in $C_0(X)$ converges weakly to $f \in C_0(X)$ iff $\sup \|f_n\|_u < \infty$ and $f_n \to f$ pointwise.

23. The hypothesis of positivity in Proposition 7.19b is necessary. (Take μ_n to be the difference of the point masses at n^{-1} and $-n^{-1}$.)

24. Find examples of sequences $\{\mu_n\}$ in $M(\mathbb{R})$ such that

a. $\mu_n \to 0$ vaguely, but $\|\mu_n\| \not\to 0$.

b. $\mu_n \to 0$ vaguely, but $\int f \, d\mu_n \not\to \int f \, d\mu$ for some bounded measurable f with compact support.

c. $\mu_n \geq 0$ and $\mu_n \to 0$ vaguely, but there exists $x \in \mathbb{R}$ such that $F_n(x) \not\to F(x)$ (notation as in Proposition 7.19).

25. Let μ be a Radon measure on X such that every nonempty open set has positive measure (e.g., Lebesgue measure). For each $x \in X$ there is a net $\{f_\alpha\}$ in $L^1(\mu)$ that converges vaguely in $M(X)$ to the point mass at x. If X is first countable, the net can be taken to be a sequence. (Consider functions of the form $\mu(U)^{-1}\chi_U$.)

26. If $\{\mu_n\} \subset M(X)$, $\mu_n \to \mu$ vaguely, and $\|\mu_n\| \to \|\mu\|$, then $\int f \, d\mu_n \to \int f \, d\mu$ for every $f \in BC(X)$. (If $\mu = 0$ the result is trivial. Otherwise, there exists $g \in C_c(X)$ with $\|g\|_u \leq 1$ such that $\int g \, d\mu > \|\mu\| - \epsilon$, and $\int gf \, d\mu_n \to \int gf \, d\mu$ for $f \in BC(X)$.) Moreover, the hypothesis $\|\mu_n\| \to \|\mu\|$ cannot be omitted.

27. Let $C^k([0,1])$ be as in Exercise 9 in §5.1. If $I \in C^k([0,1])^*$, there exist $\mu \in M([0,1])$ and constants $c_0, \ldots, c_{k-1}$, all unique, such that

$$I(f) = \int f^{(k)} \, d\mu + \sum_0^{k-1} c_j f^{(j)}(0).$$

(The functionals $f \mapsto f^{(j)}(0)$ could be replaced by any set of k functionals that separate points in the space of polynomials of degree $< k$.)

7.4 PRODUCTS OF RADON MEASURES

In this section we study Radon measures on product spaces. X and Y will denote LCH spaces, and π_X and π_Y will denote the projections of $X \times Y$ onto X and Y, respectively.

7.20 Theorem.

a. $\mathcal{B}_X \otimes \mathcal{B}_Y \subset \mathcal{B}_{X\times Y}$.

b. If X and Y are second countable, then $\mathcal{B}_X \otimes \mathcal{B}_Y = \mathcal{B}_{X\times Y}$.

c. If X and Y are second countable and μ and ν are Radon measures on X and Y, then $\mu \times \nu$ is a Radon measure on $X \times Y$.

Proof. Parts (a) and (b) are direct generalizations of Proposition 1.5, and the proof is essentially the same. The main tool is Proposition 1.4. It implies, first, that $\mathcal{B}_X \otimes \mathcal{B}_Y$ is generated by the sets $U \times V$ where U is open in X and V is open in Y. Since these sets are open in $X \times Y$, we have $\mathcal{B}_X \otimes \mathcal{B}_Y \subset \mathcal{B}_{X\times Y}$. If the topologies on X and Y have countable bases $\mathcal{E}$ and $\mathcal{F}$, then every open set in X, Y, or $X \times Y$ is a countable union of sets in $\mathcal{E}$, $\mathcal{F}$, or $\{U \times V : U \in \mathcal{E},\ V \in \mathcal{F}\}$. It follows that $\mathcal{B}_X$, $\mathcal{B}_Y$, and $\mathcal{B}_{X\times Y}$ are generated by these families and hence that $\mathcal{B}_X \otimes \mathcal{B}_Y = \mathcal{B}_{X\times Y}$. As for (c), $\mu \times \nu$ is a Borel measure by (b), so by Theorem 7.8 we need only show that $(\mu \times \nu)(K)$ is finite for every compact $K \subset X \times Y$. But this is easy: $\pi_X(K)$ and $\pi_Y(K)$ are compact, and $K \subset \pi_1(K) \times \pi_2(K)$, so $(\mu \times \nu)(K) = \mu(\pi_X(K))\nu(\pi_Y(K)) < \infty$. ∎

When X or Y is not second countable it can happen that $\mathcal{B}_X \otimes \mathcal{B}_Y \neq \mathcal{B}_{X\times Y}$; see Exercises 28 and 29. In this case the product of Radon measures is certainly not a Radon measure. However, there is a natural way of manufacturing a Radon measure from it. To see this, we need a couple of facts about continuous functions. If g and h are functions on X and Y, we define $g \otimes h$ on $X \times Y$ by

$$g \otimes h(x, y) = g(x)h(y).$$

7.21 Proposition. *Let $\mathcal{P}$ be the vector space spanned by the functions $g \otimes h$ with $g \in C_c(X)$, $h \in C_c(Y)$. Then $\mathcal{P}$ is dense in $C_c(X \otimes Y)$ in the uniform norm. More precisely, given $f \in C_c(X \times Y)$, $\epsilon > 0$, and precompact open sets $U \subset X$ and $V \subset Y$ containing $\pi_X(\operatorname{supp}(f))$ and $\pi_Y(\operatorname{supp}(f))$, there exists $F \in \mathcal{P}$ such that $\|F - f\|_u < \epsilon$ and $\operatorname{supp}(F) \subset U \times V$.*

Proof. $\overline{U} \times \overline{V}$ is a compact Hausdorff space. It follows easily from the Stone-Weierstrass theorem that the linear span of $\{g \otimes h : g \in C(\overline{U}),\ h \in C(\overline{V})\}$ is dense in $C(\overline{U} \times \overline{V})$. In particular, there is an element G of this linear span such that $\sup_{\overline{U}\times\overline{V}} |G - f| < \epsilon$. Also, by Urysohn's lemma there exist $\phi \in C_c(U, [0, 1])$ and $\psi \in C_c(V, [0, 1])$ such that $\phi = 1$ on $\pi_X(\operatorname{supp}(f))$ and $\psi = 1$ on $\pi_Y(\operatorname{supp}(f))$. Thus if we define $F = (\phi \otimes \psi)G$ on $\overline{U} \times \overline{V}$ and $F = 0$ elsewhere, we have $F \in \mathcal{P}$, $\operatorname{supp}(F) \subset U \times V$, and $\|F - f\|_u < \epsilon$. ∎

7.22 Proposition. *Every $f \in C_c(X \times Y)$ is $\mathcal{B}_X \otimes \mathcal{B}_Y$-measurable. Moreover, if μ and ν are Radon measures on X and Y, then $C_c(X \times Y) \subset L^1(\mu \times \nu)$, and*

$$\int f\, d(\mu \times \nu) = \iint f\, d\mu\, d\nu = \iint f\, d\nu\, d\mu \qquad \big(f \in C_c(X \times Y)\big).$$

Proof. If $g \in C_c(X)$ and $h \in C_c(Y)$, we have $g \otimes h = (g \circ \pi_X)(h \circ \pi_Y)$. Since π_X and π_Y are measurable from $\mathcal{B}_X \otimes \mathcal{B}_Y$ to $\mathcal{B}_X$ and $\mathcal{B}_Y$ (by definition of $\mathcal{B}_X \otimes \mathcal{B}_Y$) and g and h are continuous, $g \circ \pi_X$ and $h \circ \pi_Y$ are $\mathcal{B}_X \otimes \mathcal{B}_Y$-measurable. Since products, sums, and pointwise limits of measurable functions are measurable, the first assertion follows from Proposition 7.21. Also, every $f \in C_c(X \times Y)$ is bounded and supported in a set of finite $(\mu \times \nu)$-measure, hence is in $L^1(\mu \times \nu)$. Fubini's theorem holds for such f even if μ and ν are not σ-finite because one can replace μ and ν by the finite measures $\mu|\pi_X(\mathrm{supp}(f))$ and $\nu|\pi_Y(\mathrm{supp}(f))$. ∎

It is now clear how to obtain a Radon measure on $X \times Y$ from Radon measures μ and ν on X and Y. Namely, by Proposition 7.22 the formula $I(f) = \int f\, d(\mu \times \nu)$ defines a positive linear functional on $C_c(X \times Y)$, so it determines a Radon measure on $X \times Y$ by the Riesz representation theorem. We call this measure the **Radon product** of μ and ν and denote it by $\mu \mathbin{\widehat{\times}} \nu$. The obvious question is: Does $\mu \mathbin{\widehat{\times}} \nu$ agree with $\mu \times \nu$ on $\mathcal{B}_X \otimes \mathcal{B}_Y$? In general, the answer is no. Indeed, a counterexample may be obtained by taking $X = \mathbb{R}$, $Y = \mathbb{R}_d$ ($\mathbb{R}$ with the discrete topology), $\mu =$ Lebesgue measure, and $\nu =$ counting measure. It is not hard to see that in this case $\mathcal{B}_{X \times Y} = \mathcal{B}_X \otimes \mathcal{B}_Y$, but Exercises 12 and 14 show that $\mu \mathbin{\widehat{\times}} \nu$ is not semifinite and that $\mu \times \nu$ is the semifinite part of $\mu \mathbin{\widehat{\times}} \nu$. However, some results are still available, and in the σ-finite case everything works out beautifully. In what follows, we employ the notation of x-sections and y-sections introduced in §2.5.

7.23 Lemma.

a. If $E \in \mathcal{B}_{X \times Y}$, then $E_x \in \mathcal{B}_Y$ for all $x \in X$ and $E^y \in \mathcal{B}_X$ for all $y \in Y$.

b. If $f : X \times Y \to \mathbb{C}$ is $\mathcal{B}_{X \times Y}$-measurable, then f_x is $\mathcal{B}_Y$-measurable for all $x \in X$ and f^y is $\mathcal{B}_X$-measurable for all $y \in Y$.

Proof. The collection of all $E \subset X \times Y$ such that $E_x \in \mathcal{B}_Y$ and $E^y \in \mathcal{B}_X$ for all x, y is easily seen to be a σ-algebra. It contains all open sets — if E is open, so are E_x and E^y, being inverse images of E under the maps $y' \mapsto (x, y')$ and $x' \mapsto (x', y)$ — and hence it contains $\mathcal{B}_{X \times Y}$. This proves (a), and (b) follows since $(f_x)^{-1}(A) = (f^{-1}(A))_x$ and $(f^y)^{-1}(A) = (f^{-1}(A))^y$. ∎

7.24 Lemma. *If $f \in C_c(X \times Y)$ and μ and ν are Radon measures on X and Y, then the functions $x \mapsto \int f_x\, d\nu$ and $y \mapsto \int f^y\, d\mu$ are continuous.*

Proof. We write out the proof only for f_x. It suffices to show that for any $x_0 \in X$ and $\epsilon > 0$ there is a neighborhood U of x_0 such that $\|f_x - f_{x_0}\|_u < \epsilon$ for $x \in U$, since then

$$\left| \int (f_x - f_{x_0})\, d\nu \right| \le \epsilon \nu\big(\pi_Y(\mathrm{supp}(f))\big).$$

However, for each $y \in \pi_Y(\operatorname{supp}(f))$ there exist neighborhoods U_y, V_y of x_0 and y such that if $(x,z) \in U_y \times V_y$, then $|f(x_0,y) - f(x,z)| < \frac{1}{2}\epsilon$. We may choose a finite subcover $V_{y_1}, \ldots, V_{y_n}$ of $\pi_Y(\operatorname{supp}(f))$ and then take $U = \bigcap_1^m U_{y_j}$; details are left to the reader. ∎

7.25 Proposition. *Let μ and ν be Radon measures on X and Y. If U is open in $X \times Y$, then the functions $x \mapsto \nu(U_x)$ and $y \mapsto \mu(U^y)$ are Borel measurable on X and Y, and*

$$\mu \mathbin{\widehat{\times}} \nu(U) = \int \nu(U_x)\, d\mu(x) = \int \mu(U^y)\, d\nu(y).$$

Proof. Let $\mathcal{F} = \{f \in C_c(X \times Y) : 0 \le f \le \chi_U\}$. By Proposition 7.11 we have $\chi_U = \sup\{f : f \in \mathcal{F}\}$ and hence $\chi_{U_x} = \sup\{f_x : f \in \mathcal{F}\}$ and $\chi_{U^y} = \sup\{f^y : f \in \mathcal{F}\}$. Thus by Proposition 7.12,

$$\mu \mathbin{\widehat{\times}} \nu(U) = \sup\left\{\int f\, d(\mu \mathbin{\widehat{\times}} \nu) : f \in \mathcal{F}\right\},$$

$$\nu(U_x) = \sup\left\{\int f_x\, d\nu : f \in \mathcal{F}\right\}, \qquad \mu(U^y) = \sup\left\{\int f^y\, d\mu : f \in \mathcal{F}\right\}.$$

From Lemma 7.24 and Proposition 7.11 it follows that $x \mapsto \nu(U_x)$ and $y \mapsto \mu(U^y)$ are LSC and hence Borel measurable. Another application of Proposition 7.12, together with Proposition 7.22, yields

$$\begin{aligned}\mu \mathbin{\widehat{\times}} \nu(U) &= \sup\left\{\iint f_x\, d\nu\, d\mu(x) : f \in \mathcal{F}\right\} \\ &= \int \left[\sup\left\{\int f_x\, d\nu : f \in \mathcal{F}\right\}\right] d\mu(x) = \int \nu(U_x)\, d\mu(x),\end{aligned}$$

and likewise $\mu \mathbin{\widehat{\times}} \nu(U) = \int \mu(U^y)\, d\nu(y)$. ∎

7.26 Theorem. *Suppose that μ and ν are σ-finite Radon measures on X and Y. If $E \in \mathcal{B}_{X\times Y}$, then the functions $x \mapsto \nu(E_x)$ and $y \mapsto \mu(E^y)$ (which make sense by Lemma 7.23) are Borel measurable on X and Y, and*

$$\mu \mathbin{\widehat{\times}} \nu(E) = \int \nu(E_x)\, d\mu(x) = \int \mu(E^y)\, d\nu(y).$$

Moreover, the restriction of $\mu \mathbin{\widehat{\times}} \nu$ to $\mathcal{B}_X \otimes \mathcal{B}_Y$ is $\mu \times \nu$.

Proof. For the moment, let us fix open sets $U \subset X$ and $V \subset Y$ with $\mu(U)$ and $\nu(V)$ finite, and let $W = U \times V$. Let $\mathcal{M}$ be the collection of all sets $E \in \mathcal{B}_{X\times Y}$ such that $E \cap W$ satisfies the conclusions of the theorem. We then have

i. $\mathcal{M}$ contains all open sets, by Proposition 7.25.

ii. If $E, F \in \mathcal{M}$ and $F \subset E$, then $E \setminus F \in \mathcal{M}$; in particular, if $F \in \mathcal{M}$, then $F^c = X \setminus F \in \mathcal{M}$. Indeed, we have

$$\mu \mathbin{\widehat{\times}} \nu(E \cap W) = \mu \mathbin{\widehat{\times}} \nu(F \cap W) + \mu \mathbin{\widehat{\times}} \nu((E \setminus F) \cap W),$$

and likewise for $\nu((E \cap W)_x)$ and $\mu((E \cap W)^y)$. Since the conclusions are true for $E \cap W$ and $F \cap W$ and all the sets involved have finite measure (this is why we introduced W), we can subtract to obtain the conclusions for $(E \setminus F) \cap W$.

iii. $\mathcal{M}$ is closed under finite disjoint unions. (This is simply the additivity of the measures.)

iv. $\mathcal{M}$ is closed under countable increasing unions, and hence (by (ii)) under countable decreasing intersections. (This follows from the monotone convergence theorem.)

Now, let $\mathcal{E} = \{A \setminus B : A, B \text{ open in } X \times Y\}$, and let $\mathcal{A}$ be the collection of finite disjoint unions of sets in $\mathcal{E}$. Since

$$(A_1 \setminus B_1) \cap (A_2 \setminus B_2) = (A_1 \cap A_2) \setminus (B_1 \cup B_2),$$
$$(A \setminus B)^c = \big[(X \times Y) \setminus A\big] \cup \big[(A \cap B) \setminus \varnothing\big],$$

$\mathcal{E}$ is an elementary family, so by Proposition 1.7, $\mathcal{A}$ is an algebra. By Lemma 2.35, the monotone class generated by $\mathcal{A}$ coincides with the σ-algebra generated by $\mathcal{A}$, which is clearly $\mathcal{B}_{X \times Y}$. But by (i)–(iv) (since $A \setminus B = A \setminus (A \cap B)$), $\mathcal{M}$ contains this monotone class, so $\mathcal{M} = \mathcal{B}_{X \times Y}$.

Next, since μ and ν are σ-finite and outer regular, we have $X = \bigcup_1^\infty U_n$ and $Y = \bigcup_1^\infty V_n$ where U_n and V_n are open and have finite measure, and we may assume that the sequences $\{U_n\}$ and $\{V_n\}$ are increasing. If $E \in \mathcal{B}_{X \times Y}$, the preceding argument shows that $E \cap (U_n \times V_n)$ satisfies the conclusions of the theorem for all n, and the monotone convergence theorem then implies that E satifies the conclusions too.

Finally, if $E \in \mathcal{B}_X \times \mathcal{B}_Y$, by Tonelli's theorem we have

$$\mu \times \nu(E) = \int \nu(E_x)\, d\mu(x) = \mu \mathbin{\widehat{\times}} \nu(E),$$

and the proof is complete. ∎

7.27 The Fubini-Tonelli Theorem for Radon Products. *Let μ and ν be σ-finite Radon measures on X and Y, and let f be a Borel measurable function on $X \times Y$. Then f_x and f^y are Borel measurable for every x and y. If $f \geq 0$, then $x \mapsto \int f_x\, d\nu$ and $y \mapsto \int f^y\, d\mu$ are Borel measurable on X and Y. If $f \in L^1(\mu \mathbin{\widehat{\times}} \nu)$, then $f_x \in L^1(\nu)$ for a.e. x, $f^y \in L^1(\mu)$ for a.e. y, and $x \mapsto \int f_x\, d\nu$ and $y \mapsto \int f^y\, d\mu$ are in $L^1(\mu)$ and $L^1(\nu)$. In both cases, we have*

$$\int f\, d(\mu \mathbin{\widehat{\times}} \nu) = \iint f\, d\mu\, d\nu = \iint f\, d\nu\, d\mu.$$

Proof. The measurability of f_x and f^y was established in Lemma 7.23. The rest of the proof is identical to the proof of the ordinary Fubini-Tonelli theorem, except that Theorem 7.26 is used in place of Theorem 2.36. ∎

The extension of the notion of Radon products to any finite number of factors is straightforward. More interestingly, the theory can be extended to infinitely many factors provided that the spaces in question are compact and the measures on them are normalized to have total mass 1.

To be precise, suppose that $\{X_\alpha\}_{\alpha\in A}$ is a family of compact Hausdorff spaces and, for each α, μ_α is a Radon measure on X_α such that $\mu_\alpha(X_\alpha) = 1$. Let $X = \prod_{\alpha\in A} X_\alpha$, a compact Hausdorff space by Tychonoff's theorem. We would like to define a Radon measure μ on X such that if E_α is a Borel set in X_α for each α and $E_\alpha = X_\alpha$ for all but finitely many α, then $\mu(\prod_{\alpha\in A} E_\alpha) = \prod_{\alpha\in A} \mu_\alpha(E_\alpha)$. (The product on the right is well defined since all but finitely many factors are equal to 1.) A bit of notation will be helpful: Given $\alpha_1, \dots, \alpha_n \in A$, let $\pi_{(\alpha_1,\dots,\alpha_n)}$ be the natural projection from X onto $\prod_1^n X_{\alpha_j}$,

$$\pi_{(\alpha_1,\dots,\alpha_n)}(x) = (x_{\alpha_1}, \dots, x_{\alpha_n}).$$

Thus $\pi^{-1}_{(\alpha_1,\dots,\alpha_n)}(E_{\alpha_1} \times \cdots \times E_{\alpha_n}) = \prod_{\alpha\in A} E_\alpha$ where $E_\alpha = X_\alpha$ for $\alpha \neq \alpha_1, \dots, \alpha_n$.

7.28 Theorem. *Suppose that, for each $\alpha \in A$, μ_α is a Radon measure on the compact Hausdorff space X_α such that $\mu_\alpha(X_\alpha) = 1$. Then there is a unique Radon measure μ on $X = \prod_{\alpha\in A} X_\alpha$ such that for any $\alpha_1, \dots, \alpha_n \in A$ and any Borel set E in $\prod_1^n X_{\alpha_j}$,*

$$\mu(\pi^{-1}_{(\alpha_1,\dots,\alpha_n)}(E)) = (\mu_{\alpha_1} \mathbin{\widehat{\times}} \cdots \mathbin{\widehat{\times}} \mu_{\alpha_n})(E).$$

Proof. Let $C_F(X)$ be the set of all $f \in C(X)$ that depend on only finitely many coordinates, that is, all f of the form $f = g \circ \pi_{(\alpha_1,\dots,\alpha_n)}$ for some $\alpha_1, \dots, \alpha_n \in A$ and $g \in C(\prod_1^n X_{\alpha_j})$. If f is such a function, we define

$$I(f) = \int g \, d(\mu_{\alpha_1} \mathbin{\widehat{\times}} \cdots \mathbin{\widehat{\times}} \mu_{\alpha_n}).$$

Adding on some extra coordinates to the set $\alpha_1, \dots, \alpha_n$ has no effect on this formula since $\mu_\alpha(X_\alpha) = 1$ for all α. Thus $I(f)$ is a well-defined positive linear functional on $C_F(X)$, and $|I(f)| \le \|f\|_u$ with equality when f is constant.

Now, $C_F(X)$ is clearly an algebra that separates points, contains constant functions, and is closed under complex conjugation, so by the Stone-Weierstrass theorem it is dense in $C(X)$. Hence the functional I extends uniquely to a positive linear functional of norm 1 on $C(X)$, and the Riesz representation theorem therefore yields a unique Radon measure μ on X such that $I(f) = \int f \, d\mu$ for all $f \in C_F(X)$.

Given $\alpha_1, \dots, \alpha_n \in A$, let $\mu_{(\alpha_1,\dots,\alpha_n)} = \mu \circ \pi^{-1}_{(\alpha_1,\dots,\alpha_n)}$. Then $\mu_{(\alpha_1,\dots,\alpha_n)}$ is a Borel measure on $\prod_1^n X_{\alpha_j}$ that satisfies

$$\int g \, d\mu_{(\alpha_1,\dots,\alpha_n)} = \int g \circ \pi_{(\alpha_1,\dots,\alpha_n)} \, d\mu$$

when g is the characteristic function of a Borel set, and hence (by the usual linearity and approximation arguments) when g is any bounded Borel function. In particular, from the definition of μ, for all $g \in C(\prod_1^n X_{\alpha_j})$ we have

$$\int g\, d\mu_{(\alpha_1,\ldots,\alpha_n)} = \int g\, d(\mu_{\alpha_1} \mathbin{\widehat{\times}} \cdots \mathbin{\widehat{\times}} \mu_{\alpha_n}).$$

If we can show that $\mu_{(\alpha_1,\ldots,\alpha_n)}$ is Radon, the uniqueness of the Riesz representation will imply that $\mu_{(\alpha_1,\ldots,\alpha_n)} = \mu_{\alpha_1} \mathbin{\widehat{\times}} \cdots \mathbin{\widehat{\times}} \mu_{\alpha_n}$, which will complete the proof.

Let E be a Borel set in $\prod_1^n X_{\alpha_j}$, and write $\pi = \pi_{(\alpha_1,\ldots,\alpha_n)}$ for short. Since μ is regular, for any $\epsilon > 0$ there is a compact $K \subset \pi^{-1}(E)$ such that $\mu(K) > \mu(\pi^{-1}(E)) - \epsilon$. Then $K' = \pi(K)$ is a compact subset of E, and $\mu_{(\alpha_1,\ldots,\alpha_n)}(K') = \mu(\pi^{-1}(K')) \geq \mu(K)$ since $K \subset \pi^{-1}(K')$, so $\mu_{(\alpha_1,\ldots,\alpha_n)}(K') > \mu(\pi^{-1}(E)) - \epsilon = \mu_{(\alpha_1,\ldots,\alpha_n)}(E) - \epsilon$. Thus $\mu_{(\alpha_1,\ldots,\alpha_n)}$ is inner regular, and the same argument applied to E^c shows that it is outer regular. Thus $\mu_{(\alpha_1,\ldots,\alpha_n)}$ is Radon, and we are done. ∎

Exercises

28. If X is the set of ordinals less than or equal to the first uncountable ordinal ω_1, with the order topology, then $\mathcal{B}_{X\times X} \neq \mathcal{B}_X \otimes \mathcal{B}_X$. In fact, $\{(x,y) : x < y < \omega_1\}$ is open but not in $\mathcal{B}_X \otimes \mathcal{B}_X$. (Reexamine Exercise 47 in §2.5 in the light of Exercise 15 in §7.2.)

29. If X is a set of cardinality $> \mathfrak{c}$ with the discrete topology, then $\mathcal{B}_{X\times X} \neq \mathcal{B}_X \otimes \mathcal{B}_X$. In fact, $D = \{(x,y) : x = y\}$ is closed but not in $\mathcal{B}_X \otimes \mathcal{B}_X$. (Use Exercise 5 in §1.2 and Proposition 1.23. If $D \in \mathcal{B}_X \otimes \mathcal{B}_X$, then $D \in \mathcal{M}$ where $\mathcal{M}$ is a sub-σ-algebra of $\mathcal{B}_X \otimes \mathcal{B}_X$ generated by a countable family of rectangles, hence $\mathcal{D} \in \mathcal{N} \otimes \mathcal{N}$ where $\mathcal{N}$ is a countably generated sub-σ-algebra of $\mathcal{B}_X$. Then $\{x\} = D_x \in \mathcal{N}$ for all x, but $\operatorname{card}(\mathcal{N}) \leq \mathfrak{c}$.) The same reasoning applies if X is replaced by its one-point compactification.

30. Let μ and ν be Radon measures on X and Y, not necessarily σ-finite. If f is a nonnegative LSC function on $X \times Y$, then $x \mapsto \int f_x\, d\nu$ and $y \mapsto \int f^y\, d\mu$ are Borel measurable and $\int f\, d(\mu \mathbin{\widehat{\times}} \nu) = \iint f\, d\mu\, d\nu = \iint f\, d\nu\, d\mu$.

31. Some results concerning Baire sets on product spaces:

a. $\mathcal{B}^0_{X\times Y} \subset \mathcal{B}^0_X \otimes \mathcal{B}^0_Y$. (Hint: Proposition 7.22 remains true if $\mathcal{B}$ is replaced by $\mathcal{B}^0$.)

b. If X and Y are either compact or second countable, then $\mathcal{B}^0_{X\times Y} = \mathcal{B}^0_X \otimes \mathcal{B}^0_Y$.

c. If X is an uncountable set with the discrete topology, then $\mathcal{B}^0_{X\times X} \neq \mathcal{B}^0_X \otimes \mathcal{B}^0_X$.

7.5 NOTES AND REFERENCES

§7.1: The Riesz representation theorem is actually the work of many hands. F. Riesz [116] first proved it for the case $X = [a,b] \subset \mathbb{R}$; he formulated the result in terms of

Riemann-Stieltjes integrals and used no measure theory. It was extended to compact subsets of $\mathbb{R}^n$ by Radon [111], to compact metric spaces by Banach (see Saks [128]), and to compact Hausdorff spaces by Kakutani [80]. For the noncompact case, the first general results were obtained by Markov [98], who characterized positive linear functionals on $BC(X)$ for a normal space X in terms of certain finitely additive set functions. A theorem essentially equivalent to Theorem 7.2 was apparently known to Bourbaki about 1940 (see Weil [158, §6]), but his treatment of integration was not published until 1952, by which time several others had obtained similar results. For more detailed references, see Dunford and Schwartz [35, §IV.16] and Hewitt and Ross [75, §11]. See also König [86] for a generalization of the Riesz representation theorem to spaces that are not locally compact.

Our use of the term "Radon measure," which derives from Radon's seminal paper [111], is common but not entirely standard. Some authors refer to such measures as "regular Borel measures"; others use the term "Radon measure" to mean a positive linear functional on $C_c(X)$, and still others define Radon measures to be inner regular rather than outer regular on all Borel sets. It should also be noted that some older texts define the Borel σ-algebra to be the σ-algebra generated by the compact sets, which is in general smaller than our $\mathcal{B}_X$.

If μ is a Radon measure, let $\overline{\mu}$ denote the complete, saturated extension of μ discussed at the end of §7.1. It is a significant fact that $\overline{\mu}$ is always decomposable in the sense of Exercise 15 in §3.2; see Hewitt and Stromberg [76, Theorem 19.30]. Consequently, the extension of the Radon-Nikodym theorem in that exercise and the fact that $L^1(\overline{\mu}) \cong \mathcal{L}^\infty(\overline{\mu})$ (Exercise 25 in §6.2) are available. In this connection one should note that $L^p(\overline{\mu})$ is essentially identical to $L^p(\mu)$ for $p < \infty$, by Propositions 2.12 and 2.20.

§7.2 See Cohn [27, Proposition 7.2.3] for a proof of Theorem 7.8 that does not use the Riesz representation theorem.

Propositions 7.12 and 7.14 suggest an alternative way of constructing the Radon measure μ associated to a positive linear functional I on $C_c(X)$ in the spirit of the Daniell integral (see §2.8). Namely, one first extends I to nonnegative LSC functions g by setting

$$I(g) = \sup\{I(f) : f \in C_c(X),\ 0 \le f \le g\}$$

and then extends I to arbitrary nonnegative functions h by setting

$$I(h) = \inf\{I(g) : g \text{ LSC}, g \ge h\}.$$

It is then not difficult to verify that if $E \subset X$, $I(\chi_E) = \mu^*(E)$ where μ^* is the outer measure in the proof of Theorem 7.2. For details, see Hewitt and Ross [75] or Hewitt and Stromberg [76].

Kupka and Prikry [88] contains a readable discussion of some of the more advanced topics in the theory of measures on LCH spaces.

§7.3: Theorem 7.17 is frequently stated only for the case where X is compact; however, the more general formulation follows easily from the compact case by considering the one-point compactification of X. An interesting proof of the Baire measure version of this result, quite different from ours, can be found in Hartig [67].

§7.4: The Fubini-Tonelli theorem for Radon products, as presented here, is essentially due to deLeeuw [31]; see also Cohn [27, §7.6]. Another variant of this theorem, which includes some further results for the non-σ-finite case, can be found in Hewitt and Ross [75, §13].

Theorem 7.28 is essentially due to Nelson [103]. There is also a purely measure-theoretic version of this result: If $\{(X_\alpha, \mathcal{M}_\alpha, \mu_\alpha)\}_{\alpha \in A}$ is a family of measure spaces with $\mu_\alpha(X_\alpha) = 1$ for all α, one can define a product measure $\prod_\alpha \mu_\alpha$ on $(\prod_\alpha X_\alpha, \bigotimes_\alpha \mathcal{M}_\alpha)$. See Saeki [127], Halmos [62, §38], or Hewitt and Stromberg [76, §22]. The hypotheses of Theorem 7.28 are more restrictive, but the conclusion is stronger; in particular, the domain of the measure μ in Theorem 7.28 is the Borel σ-algebra on $\prod_\alpha X_\alpha$, which is much larger than $\bigotimes_\alpha \mathcal{B}_{X_\alpha}$ when the index set A is uncountable.

8

Elements of Fourier Analysis

It is easy to say that Fourier analysis, or harmonic analysis, originated in the work of Euler, Fourier, and others on trigonometric series; it is much harder to describe succinctly what the subject comprises today, for it is a meeting ground for ideas from many parts of analysis and has applications in such diverse areas as partial differential equations and algebraic number theory. Two of the central ingredients of harmonic analysis, however, are convolution operators and the Fourier transform, which we study in this chapter.

8.1 PRELIMINARIES

We begin by making some notational conventions. Throughout this chapter we shall be working on $\mathbb{R}^n$, and n will always refer to the dimension. In any measure-theoretic considerations we always have Lebesgue measure in mind unless we specify otherwise. Thus, if E is a measurable set in $\mathbb{R}^n$, we shall denote $L^p(E, m)$ by $L^p(E)$. If U is open in $\mathbb{R}^n$ and $k \in \mathbb{N}$, we denote by $C^k(U)$ the space of all functions on U whose partial derivatives of order $\leq k$ all exist and are continuous, and we set $C^\infty(U) = \bigcap_1^\infty C^k(U)$. Furthermore, for any $E \subset \mathbb{R}^n$ we denote by $C_c^\infty(E)$ the space of all C^∞ functions on $\mathbb{R}^n$ whose support is compact and contained in E. If $E = \mathbb{R}^n$ or $U = \mathbb{R}^n$, we shall usually omit it in naming function spaces: thus, $L^p = L^p(\mathbb{R}^n)$, $C^k = C^k(\mathbb{R}^n)$, $C_c^\infty = C_c^\infty(\mathbb{R}^n)$. If $x, y \in \mathbb{R}^n$, we set

$$x \cdot y = \sum_1^n x_j y_j, \qquad |x| = \sqrt{x \cdot x}.$$

It will be convenient to have a compact notation for partial derivatives. We shall write

$$\partial_j = \frac{\partial}{\partial x_j},$$

and for higher-order derivatives we use multi-index notation. A **multi-index** is an ordered n-tuple of nonnegative integers. If $\alpha = (\alpha_1, \dots, \alpha_n)$ is a multi-index, we set

$$|\alpha| = \sum_1^n \alpha_j, \qquad \alpha! = \prod_1^n \alpha_j!, \qquad \partial^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n},$$

and if $x = (x_1, \dots, x_n) \in \mathbb{R}^n$,

$$x^\alpha = \prod_1^n x_j^{\alpha_j}.$$

(The notation $|\alpha| = \sum \alpha_j$ is inconsistent with the notation $|x| = (\sum x_j^2)^{1/2}$, but the meaning will always be clear from the context.) Thus, for example, Taylor's formula for functions $f \in C^k$ reads

$$f(x) = \sum_{|\alpha| \le k} (\partial^\alpha f)(x_0) \frac{(x - x_0)^\alpha}{\alpha!} + R_k(x), \qquad \lim_{x \to x_0} \frac{|R_k(x)|}{|x - x_0|^k} = 0,$$

and the product rule for derivatives becomes

$$\partial^\alpha (fg) = \sum_{\beta + \gamma = \alpha} \frac{\alpha!}{\beta! \gamma!} (\partial^\beta f)(\partial^\gamma g)$$

(Exercise 1).

We shall often avail ourselves of the sloppy but handy device of using the same notation for a function and its value at a point. Thus, "x^α" may be used to denote the function whose value at any point x is x^α.

Two spaces of C^∞ functions on $\mathbb{R}^n$ will be of particular importance for us. The first is the space C_c^∞ of C^∞ functions with compact support. The existence of nonzero functions in C_c^∞ is not quite obvious; the standard construction is based on the fact that the function $\eta(t) = e^{-1/t}\chi_{(0,\infty)}(t)$ is C^∞ even at the origin (Exercise 3). If we set

$$\psi(x) = \eta(1 - |x|^2) = \begin{cases} \exp[(|x|^2 - 1)^{-1}] & \text{if } |x| < 1, \\ 0 & \text{if } |x| \ge 1, \end{cases} \tag{8.1}$$

it follows that $\psi \in C^\infty$, and $\operatorname{supp}(\psi)$ is the closed unit ball. In the next section we shall use this single function to manufacture elements of C_c^∞ in great profusion; see Propositions 8.17 and 8.18.

The other space of C^∞ functions we shall need is the **Schwartz space** $\mathcal{S}$ consisting of those C^∞ functions which, together with all their derivatives, vanish at infinity

faster than any power of $|x|$. More precisely, for any nonnegative integer N and any multi-index α we define

$$\|f\|_{(N,\alpha)} = \sup_{x\in\mathbb{R}^n} (1+|x|)^N |\partial^\alpha f(x)|;$$

then

$$\mathcal{S} = \{f \in C^\infty : \|f\|_{(N,\alpha)} < \infty \text{ for all } N, \alpha\}.$$

Examples of functions in $\mathcal{S}$ are easy to find: for instance, $f_\alpha(x) = x^\alpha e^{-|x|^2}$ where α is any multi-index. Also, clearly $C_c^\infty \subset \mathcal{S}$.

It is an important observation that if $f \in \mathcal{S}$, then $\partial^\alpha f \in L^p$ for all α and all $p \in [1,\infty]$. Indeed, $|\partial^\alpha f(x)| \le C_N(1+|x|)^{-N}$ for all N, and $(1+|x|)^{-N} \in L^p$ for $N > n/p$ by Corollary 2.52.

8.2 Proposition. *$\mathcal{S}$ is a Fréchet space with the topology defined by the norms $\|\cdot\|_{(N,\alpha)}$.*

Proof. The only nontrivial point is completeness. If $\{f_k\}$ is a Cauchy sequence in $\mathcal{S}$, then $\|f_j - f_k\|_{(N,\alpha)} \to 0$ for all N, α. In particular, for each α the sequence $\{\partial^\alpha f_k\}$ converges uniformly to a function g_α. Denoting by e_j the vector $(0,\dots,1,\dots,0)$ with the 1 in the jth position, we have

$$f_k(x+te_j) - f_k(x) = \int_0^t \partial_j f_k(x+se_j)\,ds.$$

Letting $k \to \infty$, we obtain

$$g_0(x+te_j) - g_0(x) = \int_0^t g_{e_j}(x+se_j)\,ds.$$

The fundamental theorem of calculus implies that $g_{e_j} = \partial_j g_0$, and an induction on $|\alpha|$ then yields $g_\alpha = \partial^\alpha g_0$ for all α. It is then easy to check that $\|f_k - g_0\|_{(N,\alpha)} \to 0$ for all α. ∎

Another useful characterization of $\mathcal{S}$ is the following.

8.3 Proposition. *If $f \in C^\infty$, then $f \in \mathcal{S}$ iff $x^\beta \partial^\alpha f$ is bounded for all multi-indices α, β iff $\partial^\alpha(x^\beta f)$ is bounded for all multi-indices α, β.*

Proof. Obviously $|x^\beta| \le (1+|x|)^N$ for $|\beta| \le N$. On the other hand, $\sum_1^n |x_j|^N$ is strictly positive on the unit sphere $|x| = 1$, so it has a positive minimum δ there. It follows that $\sum_1^n |x_j|^N \ge \delta|x|^N$ for all x since both sides are homogeneous of degree N, and hence

$$(1+|x|)^N \le 2^N(1+|x|^N) \le 2^N\Big[1+\delta^{-1}\sum_1^n |x_j^N|\Big] \le 2^N\delta^{-1}\sum_{|\beta|\le N} |x^\beta|.$$

This establishes the first equivalence. The second one follows from the fact that each $\partial^\alpha(x^\beta f)$ is a linear combination of terms of the form $x^\gamma \partial^\delta f$ and vice versa, by the product rule (Exercise 1). ∎

We next investigate the continuity of translations on various function spaces. The following notation for translations will be used throughout this chapter and the next one: If f is a function on $\mathbb{R}^n$ and $y \in \mathbb{R}^n$,

$$\tau_y f(x) = f(x - y).$$

We observe that $\|\tau_y f\|_p = \|f\|_p$ for $1 \le p \le \infty$ and that $\|\tau_y f\|_u = \|f\|_u$. A function f is called **uniformly continuous** if $\|\tau_y f - f\|_u \to 0$ as $y \to 0$. (The reader should pause to check that this is equivalent to the usual ϵ-δ definition of uniform continuity.)

8.4 Lemma. *If $f \in C_c(\mathbb{R}^n)$, then f is uniformly continuous.*

Proof. Given $\epsilon > 0$, for each $x \in \operatorname{supp}(f)$ there exists $\delta_x > 0$ such that $|f(x-y) - f(x)| < \frac{1}{2}\epsilon$ if $|y| < \delta_x$. Since $\operatorname{supp}(f)$ is compact, there exist $x_1, \ldots, x_N$ such that the balls of radius $\frac{1}{2}\delta_{x_j}$ about x_j cover $\operatorname{supp}(f)$. If $\delta = \frac{1}{2}\min\{\delta_{x_j}\}$, then, one easily sees that $\|\tau_y f - f\|_u < \epsilon$ whenever $|y| < \delta$. ∎

8.5 Proposition. *If $1 \le p < \infty$, translation is continuous in the L^p norm; that is, if $f \in L^p$ and $z \in \mathbb{R}^n$, then $\lim_{y\to 0} \|\tau_{y+z} f - \tau_z f\|_p = 0$.*

Proof. Since $\tau_{y+z} = \tau_y \tau_z$, by replacing f by $\tau_z f$ it suffices to assume that $z = 0$. First, if $g \in C_c$, for $|y| \le 1$ the functions $\tau_y g$ are all supported in a common compact set K, so by Lemma 8.4,

$$\int |\tau_y g - g|^p \le \|\tau_y g - g\|_u^p \, m(K) \to 0 \text{ as } y \to 0.$$

Now suppose $f \in L^p$. If $\epsilon > 0$, by Proposition 7.9 there exists $g \in C_c$ with $\|g - f\|_p < \epsilon/3$, so

$$\|\tau_y f - f\|_p \le \|\tau_y(f - g)\|_p + \|\tau_y g - g\|_p + \|g - f\|_p < \tfrac{2}{3}\epsilon + \|\tau_y g - g\|_p,$$

and $\|\tau_y g - g\|_p < \epsilon/3$ if y is sufficiently small. ∎

Proposition 8.5 is false for $p = \infty$, as one should expect since the L^∞ norm is closely related to the uniform norm; see Exercise 4.

Some of our results will concern multiply periodic functions in $\mathbb{R}^n$, and for simplicity we shall take the fundamental period in each variable to be 1. That is, we define a function f on $\mathbb{R}^n$ to be **periodic** if $f(x + k) = f(x)$ for all $x \in \mathbb{R}^n$ and $k \in \mathbb{Z}^n$. Every periodic function is thus completely determined by its values on the unit cube

$$Q = \left[-\tfrac{1}{2}, \tfrac{1}{2}\right)^n.$$

Periodic functions may be regarded as functions on the space $\mathbb{R}^n/\mathbb{Z}^n \cong (\mathbb{R}/\mathbb{Z})^n$ of cosets of $\mathbb{Z}^n$, which we call the **n-dimensional torus** and denote by $\mathbb{T}^n$. (When $n = 1$ we write $\mathbb{T}$ rather than $\mathbb{T}^1$.) $\mathbb{T}^n$ is a compact Hausdorff space; it may be

identified with the set of all $z = (z_1, \ldots, z_n) \in \mathbb{C}^n$ such that $|z_j| = 1$ for all j, via the map

$$(x_1, \ldots, x_n) \mapsto (e^{2\pi i x_1}, \ldots, e^{2\pi i x_n}).$$

On the other hand, for measure-theoretic purposes we identify $\mathbb{T}^n$ with the unit cube Q, and when we speak of Lebesgue measure on $\mathbb{T}^n$ we mean the measure induced on $\mathbb{T}^n$ by Lebesgue measure on Q. In particular, $m(\mathbb{T}^n) = 1$. Functions on $\mathbb{T}^n$ may be considered as periodic functions on $\mathbb{R}^n$ or as functions on Q; the point of view will be clear from the context when it matters.

Exercises

1. Prove the product rule for partial derivatives as stated in the text. Deduce that

$$\partial^\alpha(x^\beta f) = x^\beta \partial^\alpha f + \sum c_{\gamma\delta} x^\delta \partial^\gamma f, \qquad x^\beta \partial^\alpha f = \partial^\alpha(x^\beta f) + \sum c'_{\gamma\delta} \partial^\gamma(x^\delta f)$$

for some constants $c_{\gamma\delta}$ and $c'_{\gamma\delta}$ with $c_{\gamma\delta} = c'_{\gamma\delta} = 0$ unless $|\gamma| < |\alpha|$ and $|\delta| < |\beta|$.

2. Observe that the binomial theorem can be written as follows:

$$(x_1 + x_2)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} x^\alpha \qquad (x = (x_1, x_2),\ \alpha = (\alpha_1, \alpha_2)).$$

Prove the following generalizations:

a. The multinomial theorem: If $x \in \mathbb{R}^n$,

$$(x_1 + \cdots + x_n)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} x^\alpha.$$

b. The n-dimensional binomial theorem: If $x, y \in \mathbb{R}^n$,

$$(x + y)^\alpha = \sum_{\beta+\gamma=\alpha} \frac{\alpha!}{\beta!\gamma!} x^\beta y^\gamma.$$

3. Let $\eta(t) = e^{-1/t}$ for $t > 0$, $\eta(t) = 0$ for $t \le 0$.

a. For $k \in \mathbb{N}$ and $t > 0$, $\eta^{(k)}(t) = P_k(1/t)e^{-1/t}$ where P_k is a polynomial of degree $2k$.

b. $\eta^{(k)}(0)$ exists and equals zero for all $k \in \mathbb{N}$.

4. If $f \in L^\infty$ and $\|\tau_y f - f\|_\infty \to 0$ as $y \to 0$, then f agrees a.e. with a uniformly continuous function. (Let $A_r f$ be as in Theorem 3.18. Then $A_r f$ is uniformly continuous for $r > 0$ and uniformly Cauchy as $r \to 0$.)

8.2 CONVOLUTIONS

Let f and g be measurable functions on $\mathbb{R}^n$. The **convolution** of f and g is the function $f * g$ defined by

$$f * g(x) = \int f(x - y) g(y)\, dy$$

for all x such that the integral exists. Various conditions can be imposed on f and g to guarantee that $f * g$ is defined at least almost everywhere. For example, if f is bounded and compactly supported, g can be any locally integrable function; see also Propositions 8.7–8.9 below.

In what follows, we shall need the fact that if f is a measurable function on $\mathbb{R}^n$, then the function $K(x, y) = f(x-y)$ is measurable on $\mathbb{R}^n \times \mathbb{R}^n$. We have $K = f \circ s$ where $s(x, y) = x - y$; since s is continuous, K is Borel measurable if f is Borel measurable. This can always be assumed without affecting the definition of $f * g$, by Proposition 2.12. However, the Lebesgue measurability of K also follows from the Lebesgue measurability of f; see Exercise 5.

The elementary properties of convolutions are summarized in the following proposition.

8.6 Proposition. *Assuming that all integrals in question exist, we have*

a. $f * g = g * f$,

b. $(f * g) * h = f * (g * h)$.

c. For $z \in \mathbb{R}^n$, $\tau_z(f * g) = (\tau_z f) * g = f * (\tau_z g)$.

d. If A *is the closure of* $\{x+y : x \in \operatorname{supp}(f),\ y \in \operatorname{supp}(g)\}$, *then* $\operatorname{supp}(f*g) \subset A$.

Proof. (a) is proved by the substitution $z = x - y$:

$$f * g(x) = \int f(x-y)g(y)\,dy = \int f(z)g(x-z)\,dz = g * f(x).$$

(b) follows from (a) and Fubini's theorem:

$$\begin{aligned}\int (f * g) * h(x) &= \iint f(y)g(x-z-y)h(z)\,dy\,dz \\ &= \iint f(y)g(x-y-z)h(z)\,dz\,dy = f * (g * h)(x).\end{aligned}$$

As for (c),

$$\tau_z(f * g)(x) = \int f(x-z-y)g(y)\,dy = \int \tau_z f(x-y)g(y)\,dy = (\tau_z f) * g(x),$$

and by (a),

$$\tau_z(f * g) = \tau_z(g * f) = (\tau_z g) * f = f * (\tau_z g).$$

For (d), we observe that if $x \notin A$, then for any $y \in \operatorname{supp}(g)$ we have $x - y \notin \operatorname{supp}(f)$; hence $f(x-y)g(y) = 0$ for all y, so $f * g(x) = 0$. ∎

The following two propositions contain the basic facts about convolutions of L^p functions.

8.7 Young's Inequality. *If* $f \in L^1$ *and* $g \in L^p$ $(1 \le p \le \infty)$, *then* $f * g(x)$ *exists for almost every* x, $f * g \in L^p$, *and* $\|f * g\|_p \le \|f\|_1 \|g\|_p$.

Proof. This is a special case of Theorem 6.18, with $K(x,y) = f(x-y)$. Alternatively, one can use Minkowski's inequality for integrals:

$$\|f * g\|_p = \left\| \int f(y)g(\cdot - y)\, dy \right\|_p \le \int |f(y)|\, \|\tau_y g\|_p\, dy = \|f\|_1 \|g\|_p.$$

∎

8.8 Proposition. *If p and q are conjugate exponents, $f \in L^p$, and $g \in L^q$, then $f * g(x)$ exists for every x, $f * g$ is bounded and uniformly continuous, and $\|f * g\|_u \le \|f\|_p \|g\|_q$. If $1 < p < \infty$ (so that $1 < q < \infty$ also), then $f * g \in C_0(\mathbb{R}^n)$.*

Proof. The existence of $f * g$ and the estimate for $\|f * g\|_u$ follow immediately from Hölder's inequality. In view of Propositions 8.5 and 8.6, so does the uniform continuity of $f * g$: If $1 \le p < \infty$,

$$\|\tau_y(f * g) - f * g\|_u = \|(\tau_y f - f) * g\|_\infty \le \|\tau_y f - f\|_p \|g\|_q \to 0 \text{ as } y \to 0.$$

(If $p = \infty$, interchange the roles of f and g.) Finally, if $1 < p, q < \infty$, choose sequences $\{f_n\}$ and $\{g_n\}$ of functions with compact support such that $\|f_n - f\|_p \to 0$ and $\|g_n - g\|_q \to 0$. By Proposition 8.6d and what we have just proved, $f_n * g_n \in C_c$. But

$$\|f_n * g_n - f * g\|_u \le \|f_n - f\|_p \|g_n\|_q + \|f\|_p \|g_n - g\|_q \to 0,$$

so $f * g \in C_0$ by Proposition 4.35. ∎

The preceding results are all we shall use, but for the sake of completeness we state also the following generalization.

8.9 Proposition. *Suppose $1 \le p, q, r \le \infty$ and $p^{-1} + q^{-1} = r^{-1} + 1$.*

a. **(Young's Inequality, General Form)** *If $f \in L^p$ and $g \in L^q$, then $f * g \in L^r$ and $\|f * g\|_r \le \|f\|_p \|g\|_q$.*

b. Suppose also that $p > 1$, $q > 1$, and $r < \infty$. If $f \in L^p$ and $g \in$ weak L^q, then $f * g \in L^r$ and $\|f * g\|_r \le C_{pq} \|f\|_p [g]_q$ where C_{pq} is independent of f and g.

c. Suppose that $p = 1$ and $r = q > 1$. If $f \in L^1$ and $g \in$ weak L^q, then $f * g \in$ weak L^q and $[f * g]_q \le C_q \|f\|_1$, where C_q is independent of f and g.

Proof. To prove (a), let q be fixed. The special cases $p = 1$, $r = q$ and $p = q/(q-1)$, $r = \infty$ are Propositions 8.7 and 8.8. The general case then follows from the Riesz-Thorin interpolation theorem. (See also Exercise 6 for a direct proof.) (b) and (c) are special cases of Theorem 6.36. ∎

One of the most important properties of convolution is that, roughly speaking, $f * g$ is at least as smooth as either f or g, because formally we have

$$\partial^\alpha (f * g)(x) = \partial^\alpha \int f(x-y)g(y)\, dy = \int \partial^\alpha f(x-y) g(y)\, dy = (\partial^\alpha f) * g(x),$$

and similarly $\partial^\alpha(f * g) = f * (\partial^\alpha g)$. To make this precise, one needs only to impose conditions on f and g so that differentiation under the integral sign is legitimate. One such result is the following; see also Exercises 7 and 8.

8.10 Proposition. *If $f \in L^1$, $g \in C^k$, and $\partial^\alpha g$ is bounded for $|\alpha| \le k$, then $f * g \in C^k$ and $\partial^\alpha(f * g) = f * (\partial^\alpha g)$ for $|\alpha| \le k$.*

Proof. This is clear from Theorem 2.27. ∎

8.11 Proposition. *If $f, g \in \mathcal{S}$, then $f * g \in \mathcal{S}$.*

Proof. First, $f * g \in C^\infty$ by Proposition 8.10. Since

$$1 + |x| \le 1 + |x - y| + |y| \le (1 + |x - y|)(1 + |y|), \tag{8.12}$$

we have

$$\begin{aligned}(1 + |x|)^N |\partial^\alpha(f * g)(x)| &\le \int (1 + |x - y|)^N |\partial^\alpha f(x - y)| (1 + |y|)^N |g(y)|\, dy \\ &\le \|f\|_{(N,\alpha)} \|g\|_{(N+n+1,\alpha)} \int (1 + |y|)^{-n-1}\, dy,\end{aligned}$$

which is finite by Corollary 2.52. ∎

Convolutions of functions on the torus $\mathbb{T}^n$ are defined just as for functions on $\mathbb{R}^n$. (If one regards functions on $\mathbb{T}^n$ as periodic functions on $\mathbb{R}^n$, of course, the integration is to be extended over the unit cube rather than $\mathbb{R}^n$.) All of the preceding results remain valid, with the same proofs.

The following theorem underlies many of the important applications of convolutions on $\mathbb{R}^n$. We introduce a bit of notation that will be used frequently hereafter: If ϕ is any function on $\mathbb{R}^n$ and $t > 0$, we set

$$\phi_t(x) = t^{-n}\phi(t^{-1}x). \tag{8.13}$$

We observe that if $\phi \in L^1$, then $\int \phi_t$ is independent of t, by Theorem 2.44:

$$\int \phi_t = \int \phi(t^{-1}x)t^{-n}\, dx = \int \phi(y)\, dy = \int \phi.$$

Moreover, the "mass" of ϕ_t becomes concentrated at the origin as $t \to 0$. (Draw a picture if this isn't clear.)

8.14 Theorem. *Suppose $\phi \in L^1$ and $\int \phi(x)\, dx = a$.*

a. If $f \in L^p$ $(1 \le p < \infty)$, then $f * \phi_t \to af$ in the L^p norm as $t \to 0$.

b. If f is bounded and uniformly continuous, then $f * \phi_t \to af$ uniformly as $t \to 0$.

c. If $f \in L^\infty$ and f is continuous on an open set U, then $f * \phi_t \to af$ uniformly on compact subsets of U as $t \to 0$.

Proof. Setting $y = tz$, we have

$$\begin{aligned} f * \phi_t(x) - af(x) &= \int [f(x-y) - f(x)]\phi_t(y)\,dy \\ &= \int [f(x-tz) - f(x)]\phi(z)\,dz \\ &= \int [\tau_{tz} f(x) - f(x)]\phi(z)\,dz. \end{aligned}$$

Apply Minkowski's inequality for integrals:

$$\|f * \phi_t - af\|_p \le \int \|\tau_{tz} f - f\|_p |\phi(z)|\,dz.$$

Now, $\|\tau_{tz} f - f\|_p$ is bounded by $2\|f\|_p$ and tends to 0 as $t \to 0$ for each z, by Proposition 8.5. Assertion (a) therefore follows from the dominated convergence theorem.

The proof of (b) is exactly the same, with $\|\cdot\|_p$ replaced by $\|\cdot\|_u$. The estimate for $\|f * \phi_t - af\|_u$ is obvious, and $\|\tau_{tz} f - f\|_u \to 0$ as $t \to 0$ by the uniform continuity of f.

As for (c), given $\epsilon > 0$ let us choose a compact $E \subset \mathbb{R}^n$ such that $\int_{E^c} |\phi| < \epsilon$. Also, let K be a compact subset of U. If t is sufficiently small, then, we will have $x - tz \in U$ for all $x \in K$ and $z \in E$, so from the compactness of K it follows as in Lemma 8.4 that

$$\sup_{x \in K,\ z \in E} |f(x - tz) - f(x)| < \epsilon$$

for small t. But then

$$\begin{aligned} \sup_{x \in K} |f * \phi_t(x) - af(x)| &\le \sup_{x \in K} \left[\int_E + \int_{E^c} \right] |f(x-tz) - f(x)|\,|\phi(z)|\,dz \\ &\le \epsilon \int |\phi| + 2\|f\|_\infty \epsilon, \end{aligned}$$

from which (c) follows. ∎

If we impose slightly stronger conditions on ϕ, we can also show that $f * \phi_t \to af$ almost everywhere for $f \in L^p$. The device in the following proof of breaking up an integral into pieces corresponding to the dyadic intervals $[2^k, 2^{k+1}]$ and estimating each piece separately is a standard trick of the trade in Fourier analysis.

8.15 Theorem. *Suppose $|\phi(x)| \le C(1+|x|)^{-n-\epsilon}$ for some $C, \epsilon > 0$ (which implies that $\phi \in L^1$ by Corollary 2.52), and $\int \phi(x)\,dx = a$. If $f \in L^p$ ($1 \le p \le \infty$), then $f * \phi_t(x) \to af(x)$ as $t \to 0$ for every x in the Lebesgue set of f — in particular, for almost every x, and for every x at which f is continuous.*

Proof. If x is in the Lebesgue set of f, for any $\delta > 0$ there exists $\eta > 0$ such that

$$\int_{|y|<r} |f(x-y) - f(x)|\,dy \le \delta r^n \text{ for } r \le \eta. \tag{8.16}$$

Let us set

$$I_1 = \int_{|y|<\eta} |f(x-y) - f(x)|\,|\phi_t(y)|\,dy,$$
$$I_2 = \int_{|y|\geq\eta} |f(x-y) - f(x)|\,|\phi_t(y)|\,dy.$$

We claim that I_1 is bounded by $A\delta$ where A is independent of t, whereas $I_2 \to 0$ as $t \to 0$. Since

$$|f * \phi_t(x) - af(x)| \leq I_1 + I_2,$$

we will have

$$\limsup_{t\to 0} |f * \phi_t(x) - af(x)| \leq A\delta,$$

and since δ is arbitrary, this will complete the proof.

To estimate I_1, let K be the integer such that $2^K \leq \eta/t < 2^{K+1}$ if $\eta/t \geq 1$, and $K = 0$ if $\eta/t < 1$. We view the ball $|y| < \eta$ as the union of the annuli $2^{-k}\eta \leq |y| < 2^{1-k}\eta$ $(1 \leq k \leq K)$ and the ball $|y| < 2^{-K}\eta$. On the kth annulus we use the estimate

$$|\phi_t(y)| \leq Ct^{-n}\left|\frac{y}{t}\right|^{-n-\epsilon} \leq Ct^{-n}\left[\frac{2^{-k}\eta}{t}\right]^{-n-\epsilon},$$

and on the ball $|y| < 2^{-K}\eta$ we use the estimate $|\phi_t(y)| \leq Ct^{-n}$. Thus

$$I_1 \leq \sum_1^K Ct^{-n}\left[\frac{2^{-k}\eta}{t}\right]^{-n-\epsilon} \int_{2^{-k}\eta\leq|y|<2^{1-k}\eta} |f(x-y) - f(x)|\,dy$$
$$+Ct^{-n}\int_{|y|<2^{-K}\eta} |f(x-y) - f(x)|\,dy.$$

Therefore, by (8.16) and the fact that $2^K \leq \eta/t < 2^{K+1}$,

$$\begin{aligned} I_1 &\leq C\delta\sum_1^K (2^{1-k}\eta)^n t^{-n}\left[\frac{2^{-k}\eta}{t}\right]^{-n-\epsilon} + C\delta t^{-n}(2^{-K}\eta)^n \\ &= 2^n C\delta\left[\frac{\eta}{t}\right]^{-\epsilon}\sum_1^K 2^{k\epsilon} + C\delta\left[\frac{2^{-K}\eta}{t}\right]^n \\ &= 2^n C\delta\left[\frac{\eta}{t}\right]^{-\epsilon}\frac{2^{(K+1)\epsilon} - 2^\epsilon}{2^\epsilon - 1} + C\delta\left[\frac{2^{-K}\eta}{t}\right]^n \\ &\leq 2^n C\left[2^\epsilon(2^\epsilon - 1)^{-1} + 1\right]\delta. \end{aligned}$$

As for I_2, if p' is the conjugate exponent to p and χ is the characteristic function of $\{y : |y| \geq \eta\}$, by Hölder's inequality we have

$$\begin{aligned} I_2 &\leq \int_{|y|\geq\eta} \big(|f(x-y)| + |f(x)|\big)|\phi_t(y)|\,dy \\ &\leq \|f\|_p\|\chi\phi_t\|_{p'} + |f(x)|\,\|\chi\phi_t\|_1, \end{aligned}$$

so it suffices to show that for $1 \le q \le \infty$, and in particular for $q = 1$ and $q = p'$, $\|\chi\phi_t\|_q \to 0$ as $t \to 0$. If $q = \infty$, this is obvious:

$$\|\chi\phi_t\|_\infty \le Ct^{-n}\left[1 + (\eta/t)\right]^{-n-\epsilon} = Ct^\epsilon(t+\eta)^{-n-\epsilon} \le C\eta^{-n-\epsilon}t^\epsilon.$$

If $q < \infty$, by Corollary 2.51 we have

$$\begin{aligned}\|\chi\phi_t\|_q^q &= \int_{|y|\ge\eta} t^{-nq}|\phi(t^{-1}y)|^q\,dy = t^{n(1-q)}\int_{|z|\ge\eta/t}|\phi(z)|^q\,dz \\ &\le C_1 t^{n(1-q)}\int_{\eta/t}^\infty r^{n-1-(n+\epsilon)q}\,dr = C_2 t^{n(1-q)}\left[\frac{\eta}{t}\right]^{n-(n+\epsilon)q} = C_3 t^{\epsilon q}.\end{aligned}$$

In either case, $\|\chi\phi_t\|_q$ is dominated by t^ϵ, so we are done. ∎

In most of the applications of the preceding two theorems one has $a = 1$, although the case $a = 0$ is also useful. If $a = 1$, $\{\phi_t\}_{t>0}$ is called an **approximate identity**, as it furnishes an approximation to the identity operator on L^p by convolution operators. This construction is useful for approximating L^p functions by functions having specified regularity properties. For example, we have the following two important results:

8.17 Proposition. *C_c^∞ (and hence also $\mathcal{S}$) is dense in L^p ($1 \le p < \infty$) and in C_0.*

Proof. Given $f \in L^p$ and $\epsilon > 0$, there exists $g \in C_c$ with $\|f - g\|_p < \epsilon/2$, by Proposition 7.9. Let ϕ be a function in C_c^∞ such that $\int \phi = 1$ — for example, take $\phi = (\int \psi)^{-1}\psi$ where ψ is as in (8.1). Then $g * \phi_t \in C_c^\infty$ by Propositions 8.6d and 8.10, and $\|g * \phi_t - g\|_p < \epsilon/2$ for sufficiently small t by Theorem 8.14. The same argument applies if L^p is replaced by C_0, $\|\cdot\|_p$ by $\|\cdot\|_u$, and Proposition 7.9 by Proposition 4.35. ∎

8.18 The C^∞ Urysohn Lemma. *If $K \subset \mathbb{R}^n$ is compact and U is an open set containing K, there exists $f \in C_c^\infty$ such that $0 \le f \le 1$, $f = 1$ on K, and* $\operatorname{supp}(f) \subset U$.

Proof. Let $\delta = \rho(K, U^c)$ (the distance from K to U^c, which is positive since K is compact), and let $V = \{x : \rho(x, K) < \delta/3\}$. Choose a nonnegative $\phi \in C_c^\infty$ such that $\int \phi = 1$ and $\phi(x) = 0$ for $|x| \ge \delta/3$ (for example, $(\int \psi)^{-1}\psi_{\delta/3}$ with ψ as in (8.1)), and set $f = \chi_V * \phi$. Then $f \in C_c^\infty$ by Propositions 8.6d and 8.10, and it is easily checked that $0 \le f \le 1$, $f = 1$ on K, and $\operatorname{supp}(f) \subset \{x : \rho(x, K) \le 2\delta/3\} \subset U$. ∎

Exercises

5. If $s : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is defined by $s(x, y) = x - y$, then $s^{-1}(E)$ is Lebesgue measurable whenever E is Lebesgue measurable. (For $n = 1$, draw a picture of $s^{-1}(E) \subset \mathbb{R}^2$. It should be clear that after rotation through an angle $\pi/4$, $s^{-1}(E)$

becomes $F \times \mathbb{R}$ where $F = \{x : \sqrt{2}\,x \in E\}$, and Theorem 2.44 can be applied. The same idea works in higher dimensions.)

6. Prove Theorem 8.9a by using Exercise 31 in §6.3 to show that

$$|f * g(x)|^r \leq \|f\|_p^{r-p} \|g\|_q^{r-q} \int |f(y)|^p |g(x-y)|^q \, dy.$$

7. If f is locally integrable on $\mathbb{R}^n$ and $g \in C^k$ has compact support, then $f * g \in C^k$.

8. Suppose that $f \in L^p(\mathbb{R})$. If there exists $h \in L^p(\mathbb{R})$ such that

$$\lim_{y \to 0} \left\| y^{-1}(\tau_{-y} f - f) - h \right\|_p = 0,$$

we call h the **(strong) L^p derivative** of f. If $f \in L^p(\mathbb{R}^n)$, L^p partial derivatives of f are defined similarly. Suppose that p and q are conjugate exponents, $f \in L^p$, $g \in L^q$, and the L^p derivative $\partial_j f$ exists. Then $\partial_j(f * g)$ exists (in the ordinary sense) and equals $(\partial_j f) * g$.

9. If $f \in L^p(\mathbb{R})$, the L^p derivative of f (call it h; see Exercise 8) exists iff f is absolutely continuous on every bounded interval (perhaps after modification on a null set) and its pointwise derivative f' is in L^p, in which case $h = f'$ a.e. (For "only if," use Exercise 8: If $g \in C_c$ with $\int g = 1$, then $f * g_t \to f$ and $(f * g_t)' \to h$ as $t \to 0$. For "if," write

$$\frac{f(x+y) - f(x)}{y} - f'(x) = \frac{1}{y} \int_0^y [f'(x+t) - f'(x)] \, dt$$

and use Minkowski's inequality for integrals.)

10. Let ϕ satisfy the hypotheses of Theorem 8.15. If $f \in L^p$ $(1 \leq p \leq \infty)$, define the **ϕ-maximal function** of f to be $M_\phi f(x) = \sup_{t>0} |f * \phi_t(x)|$. (Observe that the Hardy-Littlewood maximal function Hf is $M_\phi |f|$ where ϕ is the characteristic function of the unit ball divided by the volume of the ball.) Show that there is a constant C, independent of f, such that $M_\phi f \leq C \cdot Hf$. (Break up the integral $\int f(x-y)\phi_t(y)\, dy$ as the sum of the integrals over $|y| \leq t$ and over $2^k t < |y| \leq 2^{k+1} t$ $(k = 0, 1, 2, \ldots)$, and estimate ϕ_t on each region.) It follows from Theorem 3.17 that M_ϕ is weak type (1,1), and the proof of Theorem 3.18 can then be adapted to give an alternate demonstration that $f * \phi_t \to (\int \phi) f$ a.e.

11. Young's inequality shows that L^1 is a Banach algebra, the product being convolution.

a. If $\mathfrak{I}$ is an ideal in the algebra L^1, so is its closure in L^1.

b. If $f \in L^1$, the smallest closed ideal in L^1 containing f is the smallest closed subspace of L^1 containing all translates of f. (If $g \in C_c$, $f * g(x)$ can be approximated by sums $\sum f(x - y_j) g(y_j) \Delta y_j$. On the other hand, if $\{\phi_t\}$ is an approximate identity, $f * \tau_y(\phi_t) \to \tau_y f$ as $t \to 0$.)

8.3 THE FOURIER TRANSFORM

One of the fundamental principles of harmonic analysis is the exploitation of symmetry. To be more specific, if one is doing analysis on a space on which a group acts, it is a good idea to study functions (or other analytic objects) that transform in simple ways under the group action, and then try to decompose arbitrary functions as sums or integrals of these basic functions.

The spaces we are studying are $\mathbb{R}^n$ and $\mathbb{T}^n$, which are Abelian groups under addition and act on themselves by translation. The building blocks of harmonic analysis on these spaces are the functions that transform under translation by multiplication by a factor of absolute value one, that is, functions f such that for each x there is a number $\phi(x)$ with $|\phi(x)| = 1$ such that $f(y+x) = \phi(x)f(y)$. If f and ϕ have this property, then $f(x) = \phi(x)f(0)$, so f is completely determined by ϕ once $f(0)$ is given; moreover,

$$\phi(x)\phi(y)f(0) = \phi(x)f(y) = f(x+y) = \phi(x+y)f(0),$$

so that (unless $f = 0$) $\phi(x+y) = \phi(x)\phi(y)$. In short, to find all f's that transform as described above, it suffices to find all ϕ's of absolute value one that satisfy the functional equation $\phi(x+y) = \phi(x)\phi(y)$. Upon imposing the natural requirement that ϕ should be measurable, we have a complete solution to this problem.

8.19 Theorem. *If ϕ is a measurable function on $\mathbb{R}^n$ (resp. $\mathbb{T}^n$) such that $\phi(x+y) = \phi(x)\phi(y)$ and $|\phi| = 1$, there exists $\xi \in \mathbb{R}^n$ (resp. $\xi \in \mathbb{T}^n$) such that $\phi(x) = e^{2\pi i\xi\cdot x}$.*

Proof. We first prove this assertion on $\mathbb{R}$. Let $a \in \mathbb{R}$ be such that $\int_0^a \phi(t)\,dt \neq 0$; such an a surely exists, for otherwise the Lebesgue differentiation theorem would imply that $\phi = 0$ a.e. Setting $A = (\int_0^a \phi(t)\,dt)^{-1}$, then, we have

$$\phi(x) = A\int_0^a \phi(x)\phi(t)\,dt = A\int_0^a \phi(x+t)\,dt = A\int_x^{x+a} \phi(t)\,dt.$$

Thus ϕ, being the indefinite integral of a locally integrable function, is continuous; and then, being the integral of a continuous function, it is C^1. Moreover,

$$\phi'(x) = A[\phi(x+a) - \phi(x)] = B\phi(x), \text{ where } B = A[\phi(a) - 1].$$

It follows that $(d/dx)(e^{-Bx}\phi(x)) = 0$, so that $e^{-Bx}\phi(x)$ is constant. Since $\phi(0) = 1$, we have $\phi(x) = e^{Bx}$, and since $|\phi| = 1$, B is purely imaginary, so $B = 2\pi i\xi$ for some $\xi \in \mathbb{R}$. This completes the proof for $\mathbb{R}$; as for $\mathbb{T}$, the ϕ we have been considering will be periodic (with period 1) iff $e^{2\pi i\xi} = 1$ iff $\xi \in \mathbb{Z}$.

The n-dimensional case follows easily, for if $e_1, \ldots, e_n$ is the standard basis for $\mathbb{R}^n$, the functions $\psi_j(t) = \phi(te_j)$ satisfy $\psi_j(t+s) = \psi_j(t)\psi_j(s)$ on $\mathbb{R}$, so that $\psi_j(t) = e^{2\pi i\xi_j t}$, and hence

$$\phi(x) = \phi\Big(\sum_1^n x_j e_j\Big) = \prod_1^n \psi_j(x_j) = e^{2\pi i\xi\cdot x}.$$

∎

The idea now is to decompose more or less arbitrary functions on $\mathbb{T}^n$ or $\mathbb{R}^n$ in terms of the exponentials $e^{2\pi i\xi\cdot x}$. In the case of $\mathbb{T}^n$ this works out very simply for L^2 functions:

8.20 Theorem. *Let $E_\kappa(x) = e^{2\pi i\kappa\cdot x}$. Then $\{E_\kappa : \kappa \in \mathbb{Z}^n\}$ is an orthonormal basis of $L^2(\mathbb{T}^n)$.*

Proof. Verification of orthonormality is an easy exercise in calculus; by Fubini's theorem it boils down to the fact that $\int_0^1 e^{2\pi ikt}\,dt$ equals 1 if $k = 0$ and equals 0 otherwise. Next, since $E_\kappa E_\lambda = E_{\kappa+\lambda}$, the set of finite linear combinations of the E_κ's is an algebra. It clearly separates points on $\mathbb{T}^n$; also, $E_0 = 1$ and $\overline{E}_\kappa = E_{-\kappa}$. Since $\mathbb{T}^n$ is compact, the Stone-Weierstrass theorem implies that this algebra is dense in $C(\mathbb{T}^n)$ in the uniform norm and hence in the L^2 norm, and $C(\mathbb{T}^n)$ is itself dense in $L^2(\mathbb{T}^n)$ by Proposition 7.9. It follows that $\{E_\kappa\}$ is a basis. ∎

To restate this result: If $f \in L^2(\mathbb{T}^n)$, we define its **Fourier transform** $\widehat{f}$, a function on $\mathbb{Z}^n$, by

$$\widehat{f}(\kappa) = \langle f, E_\kappa\rangle = \int_{\mathbb{T}^n} f(x)e^{-2\pi i\kappa\cdot x}\,dx.$$

and we call the series

$$\sum_{\kappa\in\mathbb{Z}^n} \widehat{f}(\kappa)E_\kappa$$

the **Fourier series** of f. The term "Fourier transform" is also used to mean the map $f \mapsto \widehat{f}$. Theorem 8.20 then says that the Fourier transform maps $L^2(\mathbb{T}^n)$ onto $l^2(\mathbb{Z}^n)$, that $\|\widehat{f}\|_2 = \|f\|_2$ (Parseval's identity), and that the Fourier series of f converges to f in the L^2 norm. We shall consider the question of pointwise convergence in the next two sections.

Actually, the definition of $\widehat{f}(\kappa)$ makes sense if f is merely in $L^1(\mathbb{T}^n)$, and $|\widehat{f}(\kappa)| \le \|f\|_1$, so the Fourier transform extends to a norm-decreasing map from $L^1(\mathbb{T}^n)$ to $l^\infty(\mathbb{Z}^n)$. (The Fourier series of an L^1 function may be quite badly behaved, but there are still methods for recovering f from $\widehat{f}$ when $f \in L^1$, as we shall see in the next section.) Interpolating between L^1 and L^2, we have the following result.

8.21 The Hausdorff-Young Inequality. *Suppose that $1 \le p \le 2$ and q is the conjugate exponent to p. If $f \in L^p(\mathbb{T}^n)$, then $\widehat{f} \in l^q(\mathbb{Z}^n)$ and $\|\widehat{f}\|_q \le \|f\|_p$.*

Proof. Since $\|\widehat{f}\|_\infty \le \|f\|_1$ and $\|\widehat{f}\|_2 = \|f\|_2$ for $f \in L^1$ or $f \in L^2$, the assertion follows from the Riesz-Thorin interpolation theorem. ∎

The situation on $\mathbb{R}^n$ is more delicate. The formal analogue of Theorem 8.20 should be

$$f(x) = \int_{\mathbb{R}^n} \widehat{f}(\xi)e^{2\pi i\xi\cdot x}\,d\xi, \text{ where } \widehat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-2\pi i\xi\cdot x}\,dx.$$

These relations turn out to be valid when suitably interpreted, but some care is needed. In the first place, the integral defining $\widehat{f}(\xi)$ is likely to diverge if $f \in L^2$. However, it certainly converges if $f \in L^1$. We therefore begin by defining the **Fourier transform** of $f \in L^1(\mathbb{R}^n)$ by

$$\mathcal{F}f(\xi) = \widehat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-2\pi i\xi\cdot x}\,dx.$$

(We use the notation $\mathcal{F}$ for the Fourier transform only in certain situations where it is needed for clarity.) Clearly $\|\widehat{f}\|_u \le \|f\|_1$, and $\widehat{f}$ is continuous by Theorem 2.27; thus

$$\mathcal{F} : L^1(\mathbb{R}^n) \to BC(\mathbb{R}^n).$$

We summarize the elementary properties of $\mathcal{F}$ in a theorem.

8.22 Theorem. *Suppose $f, g \in L^1(\mathbb{R}^n)$.*

a. $(\tau_y f)\widehat{\ }(\xi) = e^{-2\pi i\xi\cdot y}\widehat{f}(\xi)$ *and* $\tau_\eta(\widehat{f}) = \widehat{h}$ *where* $h(x) = e^{2\pi i\eta\cdot x}f(x)$.

b. If T is an invertible linear transformation of $\mathbb{R}^n$ and $S = (T^*)^{-1}$ is its inverse transpose, then $(f\circ T)\widehat{\ } = |\det T|^{-1}\widehat{f}\circ S$. *In particular, if T is a rotation, then* $(f\circ T)\widehat{\ } = \widehat{f}\circ T$; *and if* $Tx = t^{-1}x$ $(t > 0)$, *then* $(f\circ T)\widehat{\ }(\xi) = t^n\widehat{f}(t\xi)$, *so that* $(f_t)\widehat{\ }(\xi) = \widehat{f}(t\xi)$ *in the notation of (8.13).*

c. $(f * g)\widehat{\ } = \widehat{f}\widehat{g}$.

d. If $x^\alpha f \in L^1$ *for* $|\alpha| \le k$, *then* $\widehat{f} \in C^k$ *and* $\partial^\alpha\widehat{f} = [(-2\pi i x)^\alpha f]\widehat{\ }$.

e. If $f \in C^k$, $\partial^\alpha f \in L^1$ *for* $|\alpha| \le k$, *and* $\partial^\alpha f \in C_0$ *for* $|\alpha| \le k-1$, *then* $(\partial^\alpha f)\widehat{\ }(\xi) = (2\pi i\xi)^\alpha\widehat{f}(\xi)$.

f. **(The Riemann-Lebesgue Lemma)** $\mathcal{F}(L^1(\mathbb{R}^n)) \subset C_0(\mathbb{R}^n)$.

Proof. a. We have

$$(\tau_y f)\widehat{\ }(\xi) = \int f(x-y)e^{-2\pi i\xi\cdot x}\,dx = \int f(x)e^{-2\pi i\xi\cdot(x+y)}\,dx = e^{-2\pi i\xi\cdot y}\widehat{f}(\xi),$$

and similarly for the other formula.

b. By Theorem 2.44,

$$\begin{aligned}(f\circ T)\widehat{\ }(\xi) &= \int f(Tx)e^{-2\pi i\xi\cdot x}\,dx = |\det T|^{-1}\int f(x)e^{-2\pi i\xi\cdot T^{-1}x}\,dx\\ &= |\det T|^{-1}\int f(x)e^{-2\pi i S\xi\cdot x}\,dx = |\det T|^{-1}\widehat{f}(S\xi).\end{aligned}$$

c. By Fubini's theorem,

$$\begin{aligned}(f * g)\widehat{\ }(\xi) &= \iint f(x-y)g(y)e^{-2\pi i\xi\cdot x}\,dy\,dx \\ &= \iint f(x-y)e^{-2\pi i\xi\cdot(x-y)}g(y)e^{-2\pi i\xi\cdot y}\,dx\,dy \\ &= \widehat{f}(\xi)\int g(y)e^{-2\pi i\xi\cdot y}\,dy \\ &= \widehat{f}(\xi)\widehat{g}(\xi).\end{aligned}$$

d. By Theorem 2.27 and induction on $|\alpha|$,

$$\partial^\alpha \widehat{f}(\xi) = \partial_\xi^\alpha \int f(x)e^{-2\pi i\xi\cdot x}\,dx = \int f(x)(-2\pi i x)^\alpha e^{-2\pi i\xi\cdot x}\,dx.$$

e. First assume $n = |\alpha| = 1$. Since $f \in C_0$, we can integrate by parts:

$$\begin{aligned}\int f'(x)e^{-2\pi i\xi\cdot x}\,dx &= f(x)e^{-2\pi i\xi\cdot x}\Big|_{-\infty}^{\infty} - \int f(x)(-2\pi i\xi)e^{-2\pi i\xi\cdot x}\,dx \\ &= 2\pi i\xi\widehat{f}(\xi).\end{aligned}$$

The argument for $n > 1$, $|\alpha| = 1$ is the same — to compute $(\partial_j f)\widehat{\ }$, integrate by parts in the jth variable — and the general case follows by induction on $|\alpha|$.

f. By (e), if $f \in C^1 \cap C_c$, then $|\xi|\widehat{f}(\xi)$ is bounded and hence $\widehat{f} \in C_0$. But the set of all such f's is dense in L^1 by Proposition 8.17, and $\widehat{f_n} \to \widehat{f}$ uniformly whenever $f_n \to f$ in L^1. Since C_0 is closed in the uniform norm, the result follows. ∎

Parts (d) and (e) of Theorem 8.22 point to a fundamental property of the Fourier transform: Smoothness properties of f are reflected in the rate of decay of $\widehat{f}$ at infinity, and vice versa. Parts (a), (c), (e), and (f) of this theorem are valid also on $\mathbb{T}^n$, as is (b) provided that T leaves the lattice $\mathbb{Z}^n$ invariant (Exercise 12).

8.23 Corollary. *$\mathcal{F}$ maps the Schwartz class $\mathcal{S}$ continuously into itself.*

Proof. If $f \in \mathcal{S}$, then $x^\alpha\partial^\beta f \in L^1 \cap C_0$ for all α, β, so by Theorem 8.22d,e, $\widehat{f}$ is C^∞ and

$$(x^\alpha\partial^\beta f)\widehat{\ } = (-1)^{|\alpha|}(2\pi i)^{|\beta|-|\alpha|}\partial^\alpha(\xi^\beta\widehat{f}).$$

Thus $\partial^\alpha(\xi^\beta\widehat{f})$ is bounded for all α, β, whence $\widehat{f} \in \mathcal{S}$ by Proposition 8.3. Moreover, since $\int(1+|x|)^{-n-1}\,dx < \infty$,

$$\|(x^\alpha\partial^\beta f)\widehat{\ }\|_u \le \|x^\alpha\partial^\beta f\|_1 \le C\|(1+|x|)^{n+1}x^\alpha\partial^\beta f\|_u.$$

It then follows that $\|\widehat{f}\|_{(N,\beta)} \le C_{N,\beta}\sum_{|\gamma|\le|\beta|}\|f\|_{(N+n+1,\gamma)}$ by the proof of Proposition 8.3, so the Fourier transform is continuous on $\mathcal{S}$. ∎

At this point we need to compute an important specific Fourier transform.

8.24 Proposition. *If $f(x) = e^{-\pi a|x|^2}$ where $a > 0$, then $\widehat{f}(\xi) = a^{-n/2}e^{-\pi|\xi|^2/a}$.*

Proof. First consider the case $n = 1$. Since the derivative of $e^{-\pi ax^2}$ is $-2\pi a e^{-\pi ax^2}$, by Theorem 8.22d,e we have

$$(\widehat{f})'(\xi) = (-2\pi i x e^{-\pi ax^2})\widehat{\ }(\xi) = \frac{i}{a}(f')\widehat{\ }(\xi) = \frac{i}{a}(2\pi i\xi)\widehat{f}(\xi) = -\frac{2\pi}{a}\xi\widehat{f}(\xi).$$

It follows that $(d/d\xi)(e^{\pi\xi^2/a}\widehat{f}(\xi)) = 0$, so that $e^{\pi\xi^2/a}\widehat{f}(\xi)$ is constant. To evaluate the constant, set $\xi = 0$ and use Proposition 2.53:

$$\widehat{f}(0) = \int e^{-\pi ax^2}\,dx = a^{-1/2}.$$

The n-dimensional case follows by Fubini's theorem, since $|x|^2 = \sum_1^n x_j^2$:

$$\begin{aligned}\widehat{f}(\xi) &= \prod_1^n \int \exp(-\pi a x_j^2 - 2\pi i\xi_j x_j)\,dx_j \\ &= \prod_1^n \left[a^{-1/2}\exp(-\pi\xi_j^2/a)\right] = a^{-n/2}\exp(-\pi|\xi|^2/a).\end{aligned}$$

∎

We are now ready to invert the Fourier transform. If $f \in L^1$, we define

$$f^\vee(x) = \widehat{f}(-x) = \int f(\xi)e^{2\pi i\xi\cdot x}\,d\xi,$$

and we claim that if $f \in L^1$ and $\widehat{f} \in L^1$ then $(\widehat{f})^\vee = f$. A simple appeal to Fubini's theorem fails because the integrand in

$$(\widehat{f})^\vee(x) = \iint f(y)\,e^{-2\pi i\xi\cdot y}e^{2\pi i\xi\cdot x}\,dy\,d\xi$$

is not in $L^1(\mathbb{R}^n \times \mathbb{R}^n)$. The trick is to introduce a convergence factor and then pass to the limit, using Fubini's theorem via the following lemma:

8.25 Lemma. *If $f, g \in L^1$ then $\int \widehat{f}g = \int f\widehat{g}$.*

Proof. Both integrals are equal to $\iint f(x)g(\xi)e^{-2\pi i\xi\cdot x}\,dx\,d\xi$. ∎

8.26 The Fourier Inversion Theorem. *If $f \in L^1$ and $\widehat{f} \in L^1$, then f agrees almost everywhere with a continuous function f_0, and $(\widehat{f})^\vee = (f^\vee)\widehat{\ } = f_0$.*

Proof. Given $t > 0$ and $x \in \mathbb{R}^n$, set

$$\phi(\xi) = \exp(2\pi i\xi \cdot x - \pi t^2|\xi|^2).$$

By Theorem 8.22a and Proposition 8.24,

$$\widehat{\phi}(y) = t^{-n}\exp(-\pi|x-y|^2/t^2) = g_t(x-y),$$

where $g(x) = e^{-\pi|x|^2}$ and the subscript t has the meaning in (8.13). By Lemma 8.25, then,

$$\int e^{-\pi t^2|\xi|^2} e^{2\pi i\xi\cdot x}\widehat{f}(\xi)\,d\xi = \int \widehat{f}\phi = \int f\widehat{\phi} = f * g_t(x).$$

Since $\int e^{-\pi|x|^2}\,dx = 1$, by Theorem 8.14 we have $f * g_t \to f$ in the L^1 norm as $t \to 0$. On the other hand, since $\widehat{f} \in L^1$ the dominated convergence theorem yields

$$\lim_{t\to 0}\int e^{-\pi t^2|\xi|^2} e^{2\pi i\xi\cdot x}\widehat{f}(\xi)\,d\xi = \int e^{2\pi i\xi\cdot x}\widehat{f}(\xi)\,d\xi = (\widehat{f})^\vee(x).$$

It follows that $f = (\widehat{f})^\vee$ a.e., and similarly $(f^\vee)\widehat{\ } = f$ a.e. Since $(\widehat{f})^\vee$ and $(f^\vee)\widehat{\ }$ are continuous, being Fourier transforms of L^1 functions, the proof is complete. ∎

8.27 Corollary. *If $f \in L^1$ and $\widehat{f} = 0$, then $f = 0$ a.e.*

8.28 Corollary. *$\mathcal{F}$ is an isomorphism of $\mathcal{S}$ onto itself.*

Proof. By Corollary 8.23, $\mathcal{F}$ maps $\mathcal{S}$ continuously into itself, and hence so does $f \mapsto f^\vee$, since $f^\vee(x) = \widehat{f}(-x)$. By the Fourier inversion theorem, these maps are inverse to each other. ∎

At last we are in a position to derive the analogue of Theorem 8.20 on $\mathbb{R}^n$.

8.29 The Plancherel Theorem. *If $f \in L^1 \cap L^2$, then $\widehat{f} \in L^2$; and $\mathcal{F}|(L^1 \cap L^2)$ extends uniquely to a unitary isomorphism on L^2.*

Proof. Let $\mathcal{X} = \{f \in L^1 : \widehat{f} \in L^1\}$. Since $\widehat{f} \in L^1$ implies $f \in L^\infty$, we have $\mathcal{X} \subset L^2$ by Proposition 6.10, and $\mathcal{X}$ is dense in L^2 because $\mathcal{S} \subset \mathcal{X}$ and $\mathcal{S}$ is dense in L^2 by Proposition 8.17. Given $f, g \in \mathcal{X}$, let $h = \overline{\widehat{g}}$. By the inversion theorem,

$$\widehat{h}(\xi) = \int e^{-2\pi i\xi\cdot x}\overline{\widehat{g}(x)}\,dx = \int \overline{e^{2\pi i\xi\cdot x}\widehat{g}(x)}\,dx = \overline{g(\xi)}.$$

Hence, by Lemma 8.25,

$$\int f\overline{g} = \int f\widehat{h} = \int \widehat{f}h = \widehat{f}\overline{\widehat{g}}.$$

Thus $\mathcal{F}|\mathcal{X}$ preserves the L^2 inner product; in particular, by taking $g = f$, we obtain $\|\widehat{f}\|_2 = \|f\|_2$. Since $\mathcal{F}(\mathcal{X}) = \mathcal{X}$ by the inversion theorem, $\mathcal{F}|\mathcal{X}$ extends by continuity to a unitary isomorphism on L^2.

It remains only to show that this extension agrees with $\mathcal{F}$ on all of $L^1 \cap L^2$. But if $f \in L^1 \cap L^2$ and $g(x) = e^{-\pi|x|^2}$ as in the proof of the inversion theorem, we have $f * g_t \in L^1$ by Young's inequality and $(f * g_t)\widehat{} \in L^1$ because $(f * g_t)\widehat{}(\xi) = e^{-\pi t^2|\xi|^2}\widehat{f}(\xi)$ and $\widehat{f}$ is bounded. Hence $f * g_t \in \mathcal{X}$; moreover, by Theorem 8.14, $f * g_t \to f$ in both the L^1 and L^2 norms. Therefore $(f * g_t)\widehat{} \to \widehat{f}$ both uniformly and in the L^2 norm, and we are done. ∎

We have thus extended the domain of the Fourier transform from L^1 to $L^1 + L^2$. Just as on $\mathbb{T}^n$, the Riesz-Thorin theorem yields the following result for the intermediate L^p spaces:

8.30 The Hausdorff-Young Inequality. *Suppose that $1 \le p \le 2$ and q is the conjugate exponent to p. If $f \in L^p(\mathbb{R}^n)$, then $\widehat{f} \in L^q(\mathbb{R}^n)$ and $\|\widehat{f}\|_q \le \|f\|_p$.*

If $f \in L^1$ and $\widehat{f} \in L^1$, the inversion formula

$$f(x) = \int \widehat{f}(\xi) e^{2\pi i \xi \cdot x}\, d\xi$$

exhibits f as a superposition of the basic functions $e^{2\pi i \xi \cdot x}$; it is often called the **Fourier integral** representation of f. This formula remains valid in spirit for all $f \in L^2$, although the integral (as well as the integral defining $\widehat{f}$) may not converge pointwise. The interpretation of the inversion formula will be studied further in the next section.

We conclude this section with a beautiful theorem that involves an interplay of Fourier series and Fourier integrals. To motivate it, consider the following problem: Given a function $f \in L^1(\mathbb{R}^n)$, how can one manufacture a periodic function (that is, a function on $\mathbb{T}^n$) from it? Two possible answers suggest themselves. One way is to "average" f over all periods, producing the series $\sum_{k \in \mathbb{Z}^n} f(x - k)$. This series, if it converges, will surely define a periodic function. The other way is to restrict f to the lattice $\mathbb{Z}^n$ and use it to form a Fourier series $\sum_{\kappa \in \mathbb{Z}^n} \widehat{f}(\kappa) e^{2\pi i \kappa \cdot x}$. The content of the following theorem is that these methods both work and both give the same answer.

8.31 Theorem. *If $f \in L^1(\mathbb{R}^n)$, the series $\sum_{k \in \mathbb{Z}^n} \tau_k f$ converges pointwise a.e. and in $L^1(\mathbb{T}^n)$ to a function Pf such that $\|Pf\|_1 \le \|f\|_1$. Moreover, for $\kappa \in \mathbb{Z}^n$, $(Pf)\widehat{}(\kappa)$ (Fourier transform on $\mathbb{T}^n$) equals $\widehat{f}(\kappa)$ (Fourier transform on $\mathbb{R}^n$).*

Proof. Let $Q = [-\frac{1}{2}, \frac{1}{2})^n$. Then $\mathbb{R}^n$ is the disjoint union of the cubes $Q + k = \{x + k : x \in Q\}$, $k \in \mathbb{Z}^n$, so

$$\int_Q \sum_{k \in \mathbb{Z}^n} |f(x - k)|\, dx = \sum_{k \in \mathbb{Z}^n} \int_{Q+k} |f(x)|\, dx = \int_{\mathbb{R}^n} |f(x)|\, dx.$$

Now apply Theorem 2.25. First, it shows that the series $\sum \tau_k f$ converges a.e. and in $L^1(\mathbb{T}^n)$ to a function $Pf \in L^1(\mathbb{T}^n)$ such that $\|Pf\|_1 \leq \|f\|_1$, since $\mathbb{T}^n$ is measure-theoretically identical to Q. Second, it yields

$$\begin{aligned}(Pf)\widehat{\ }(\kappa) &= \int_Q \sum_{k\in\mathbb{Z}^n} f(x-k)e^{-2\pi i\kappa\cdot x}\,dx = \sum_{k\in\mathbb{Z}^n}\int_{Q+k} f(x)e^{-2\pi i\kappa\cdot(x+k)}\,dx \\ &= \sum_{k\in\mathbb{Z}^N}\int_{Q+k} f(x)e^{-2\pi i\kappa\cdot x}\,dx = \int_{\mathbb{R}^n} f(x)e^{-2\pi i\kappa\cdot x}\,dx = \widehat{f}(\kappa).\end{aligned}$$

∎

If we impose conditions on f to guarantee that the series in question converge absolutely, we obtain a more refined result.

8.32 The Poisson Summation Formula. *Suppose $f \in C(\mathbb{R}^n)$ satisfies $|f(x)| \leq C(1+|x|)^{-n-\epsilon}$ and $|\widehat{f}(\xi)| \leq C(1+|\xi|)^{-n-\epsilon}$ for some $C, \epsilon > 0$. Then*

$$\sum_{k\in\mathbb{Z}^n} f(x+k) = \sum_{\kappa\in\mathbb{Z}^n} \widehat{f}(\kappa)e^{2\pi i\kappa\cdot x},$$

where both series converge absolutely and uniformly on $\mathbb{T}^n$. In particular, taking $x = 0$,

$$\sum_{k\in\mathbb{Z}^n} f(k) = \sum_{\kappa\in\mathbb{Z}^n} \widehat{f}(\kappa).$$

Proof. The absolute and uniform convergence of the series follows from the fact that $\sum_{k\in\mathbb{Z}^n}(1+|k|)^{-n-\epsilon} < \infty$, which can be seen by comparing the latter series to the convergent integral $\int(1+|x|)^{-n-\epsilon}\,dx$. Thus the function $Pf = \sum_k \tau_k f$ is in $C(\mathbb{T}^n)$ and hence in $L^2(\mathbb{T}^n)$, so Theorem 8.35 implies that the series $\sum \widehat{f}(\kappa)e^{2\pi i\kappa\cdot x}$ converges in $L^2(\mathbb{T}^n)$ to Pf. Since it also converges uniformly, its sum equals Pf pointwise. (The replacement of k by $-k$ in the formula for Pf is immaterial since the sum is over all $k \in \mathbb{Z}^n$.) ∎

Exercises

12. Work out the analogue of Theorem 8.22 for the Fourier transform on $\mathbb{T}^n$.

13. Let $f(x) = \frac{1}{2} - x$ on the interval $[0,1)$, and extend f to be periodic on $\mathbb{R}$.
- **a.** $\widehat{f}(0) = 0$, and $\widehat{f}(\kappa) = (2\pi i\kappa)^{-1}$ if $\kappa \neq 0$.
- **b.** $\sum_1^\infty k^{-2} = \pi^2/6$. (Use the Parseval identity.)

14. **(Wirtinger's Inequality)** If $f \in C^1([a,b])$ and $f(a) = f(b) = 0$, then

$$\int_a^b |f(x)|^2\,dx \leq \left(\frac{b-a}{\pi}\right)^2 \int_a^b |f'(x)|^2\,dx.$$

(By a change of variable it suffices to assume $a = 0$, $b = \frac{1}{2}$. Extend f to $[-\frac{1}{2}, \frac{1}{2}]$ by setting $f(-x) = -f(x)$, and then extend f to be periodic on $\mathbb{R}$. Check that f, thus extended, is in $C^1(\mathbb{T})$ and apply the Parseval identity.)

15. Let $\operatorname{sinc} x = (\sin \pi x)/\pi x$ ($\operatorname{sinc} 0 = 1$).

a. If $a > 0$, $\widehat{\chi}_{[-a,a]}(x) = \chi^{\vee}_{[a,a]}(x) = 2a \operatorname{sinc} 2ax$.

b. Let $\mathcal{H}_a = \{f \in L^2 : \widehat{f}(\xi) = 0 \text{ (a.e.) for } |\xi| > a\}$. Then $\mathcal{H}$ is a Hilbert space and $\{\sqrt{2a} \operatorname{sinc}(2ax - k) : k \in \mathbb{Z}\}$ is an orthonormal basis for $\mathcal{H}$.

c. (**The Sampling Theorem**) If $f \in \mathcal{H}_a$, then $f \in C_0$ (after modification on a null set), and $f(x) = \sum_{-\infty}^{\infty} f(k/2a) \operatorname{sinc}(2ax - k)$, where the series converges both uniformly and in L^2. (In the terminology of signal analysis, a signal of bandwidth $2a$ is completely determined by sampling its values at a sequence of points $\{k/2a\}$ whose spacing is the reciprocal of the bandwidth.)

16. Let $f_k = \chi_{[-1,1]} * \chi_{[-k,k]}$.

a. Compute $f_k(x)$ explicitly and show that $\|f\|_u = 2$.

b. $f_k^{\vee}(x) = (\pi x)^{-2} \sin 2\pi kx \sin 2\pi x$, and $\|f_k^{\vee}\|_1 \to \infty$ as $k \to \infty$. (Use Exercise 15a, and substitute $y = 2\pi kx$ in the integral defining $\|f_k^{\vee}\|_1$.)

c. $\mathcal{F}(L^1)$ is a proper subset of C_0. (Consider $g_k = f_k^{\vee}$ and use the open mapping theorem.)

17. Given $a > 0$, let $f(x) = e^{-2\pi x} x^{a-1}$ for $x > 0$ and $f(x) = 0$ for $x \le 0$.

a. $f \in L^1$, and $f \in L^2$ if $a > \frac{1}{2}$.

b. $\widehat{f}(\xi) = \Gamma(a)[(2\pi)(1 + i\xi)]^{-a}$. (Here we are using the branch of z^a in the right half plane that is positive when z is positive. Cauchy's theorem may be used to justify the complex substitution $y = (1 + i\xi)x$ in the integral defining $\widehat{f}$.)

c. If $a, b > \frac{1}{2}$ then

$$\int_{-\infty}^{\infty} (1 - ix)^{-a}(1 + ix)^{-b}\, dx = \frac{2^{2-a-b}\pi\Gamma(a + b - 1)}{\Gamma(a)\Gamma(b)}.$$

18. Suppose $f \in L^2(\mathbb{R})$.

a. The L^2 derivative f' (in the sense of Exercises 8 and 9) exists iff $\xi \widehat{f} \in L^2$, in which case $\widehat{f'}(\xi) = 2\pi i \xi \widehat{f}(\xi)$.

b. If the L^2 derivative f' exists, then

$$\left[\int |f(x)|^2\, dx\right] \le 4 \int |xf(x)|^2\, dx \int |f'(x)|^2\, dx.$$

(If the integrals on the right are finite, one can integrate by parts to obtain $\int |f|^2 = -2 \operatorname{Re} \int x\overline{f}f'$.)

c. (**Heisenberg's Inequality**) For any $b, \beta \in \mathbb{R}$,

$$\int (x - b)^2 |f(x)|^2\, dx \int (\xi - \beta)^2 |\widehat{f}(\xi)|^2\, d\xi \ge \frac{\|f\|_2^4}{16\pi^2}.$$

(The inequality is trivial if either integral on the right is infinite; if not, reduce to the case $b = \beta = 0$ by considering $g(x) = e^{-2\pi i\beta x} f(x+b)$.) This inequality, a form of the quantum uncertainty principle, says that f and $\widehat{f}$ cannot both be sharply localized about single points b and β.

19. (A variation on the theme of Exercise 18) If $f \in L^2(\mathbb{R}^n)$ and the set $S = \{x : f(x) \neq 0\}$ has finite measure, then for any measurable $E \subset \mathbb{R}^n$, $\int_E |\widehat{f}|^2 \leq \|f\|_2^2 m(S) m(E)$.

20. If $f \in L^1(\mathbb{R}^{n+m})$, define $Pf(x) = \int f(x,y)\,dy$. (Here $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$.) Then $Pf \in L^1(\mathbb{R}^n)$, $\|Pf\|_1 \leq \|f\|_1$, and $(Pf)\widehat{\ }(\xi) = \widehat{f}(\xi, 0)$.

21. State and prove a result that encompasses both Theorem 8.31 and Exercise 20, in the setting of Fourier transforms on closed subgroups and quotient groups of $\mathbb{R}^n$.

22. Since $\mathcal{F}$ commutes with rotations, the Fourier transform of a radial function is radial; that is, if $F \in L^1(\mathbb{R}^n)$ and $F(x) = f(|x|)$, then $\widehat{F}(\xi) = g(|\xi|)$, where f and g are related as follows.

a. Let $J(\xi) = \int_S e^{ix\xi} d\sigma(x)$ where σ is surface measure on the unit sphere S in $\mathbb{R}^n$ (Theorem 2.49). Then J is radial — say, $J(\xi) = j(|\xi|)$ — and $g(\rho) = \int_0^\infty j(2\pi r\rho) f(r) r^{n-1}\,dr$.

b. J satisfies $\sum_1^n \partial_k^2 J + J = 0$.

c. j satisfies $\rho j''(\rho) + (n-1)j'(\rho) + \rho j(\rho) = 0$. (This equation is a variant of Bessel's equation. The function j is completely determined by the fact that it is a solution of this equation, is smooth at $\rho = 0$, and satisfies $j(0) = \sigma(S) = 2\pi^{n/2}/\Gamma(n/2)$. In fact, $j(\rho) = (2\pi)^{n/2}\rho^{(2-n)/2} J_{(n-2)/2}(\rho)$ where J_α is the Bessel function of the first kind of order α.)

d. If $n = 3$, $j(\rho) = 4\pi\rho^{-1}\sin\rho$. (Set $f(\rho) = \rho j(\rho)$ and use (c) to show that $f'' + f = 0$. Alternatively, use spherical coordinates to compute the integral defining $J(0,0,\rho)$ directly.)

23. In this exercise we develop the theory of Hermite functions.

a. Define operators T, T^* on $\mathcal{S}(\mathbb{R})$ by $Tf(x) = 2^{-1/2}[xf(x) - f'(x)]$ and $T^*f(x) = 2^{-1/2}[xf(x) + f'(x)]$. Then $\int (Tf)\overline{g} = \int f(\overline{T^*g})$ and $T^*T^k - T^kT^* = kT^{k-1}$.

b. Let $h_0(x) = \pi^{-1/4}e^{-x^2/2}$, and for $k \geq 1$ let $h_k = (k!)^{-1/2}T^k h_0$. (h_k is the kth **normalized Hermite function**.) We have $Th_k = \sqrt{k+1}\,h_{k+1}$ and $T^*h_k = \sqrt{k}\,h_{k-1}$, and hence $TT^*h_k = kh_k$.

c. Let $S = 2TT^* + I$. Then $Sf(x) = x^2 f(x) - f''(x)$ and $Sh_k = (2k+1)h_k$. (S is called the **Hermite operator**.)

d. $\{h_k\}_0^\infty$ is an orthonormal set in $L^2(\mathbb{R})$. (Check directly that $\|h_0\|_2 = 1$, then observe that for $k > 0$, $\int h_k \overline{h}_m = k^{-1}\int (TT^*h_k)\overline{h}_m$ and use (a) and (b).)

e. We have

$$T^k f(x) = (-1)^k 2^{-k/2} e^{x^2/2} \left(\frac{d}{dx}\right)^k \left[e^{-x^2/2} f(x)\right]$$

(use induction on k), and in particular,

$$h_k(x) = \frac{(-1)^k}{[\pi^{1/2}2^k k!]^{1/2}} e^{x^2/2} \left(\frac{d}{dx}\right)^k e^{-x^2}.$$

f. Let $H_k(x) = e^{x^2/2}h_k(x)$. Then H_k is a polynomial of degree k, called the kth **normalized Hermite polynomial**. The linear span of $H_0, \dots, H_m$ is the set of all polynomials of degree $\leq m$. (The kth Hermite polynomial as usually defined is $[\pi^{1/2}2^k k!]^{1/2}H_k$.)

g. $\{h_k\}_0^\infty$ is an orthonormal basis for $L^2(\mathbb{R})$. (Suppose $f \perp h_k$ for all k, and let $g(x) = f(x)e^{-x^2/2}$. Show that $\widehat{g} = 0$ by expanding $e^{-2\pi i\xi\cdot x}$ in its Maclaurin series and using (f).)

h. Define $A : L^2 \to L^2$ by $Af(x) = (2\pi)^{1/4}f(x\sqrt{2\pi})$, and define $\widetilde{f} = A^{-1}\mathcal{F}Af$ for $f \in L^2$. Then A is unitary and $\widetilde{f}(\xi) = (2\pi)^{-1/2}\int f(x)e^{-i\xi x}\,dx$. Moreover, $\widetilde{Tf} = -iT(\widetilde{f})$ for $f \in \mathcal{S}$, and $\widetilde{h}_0 = h_0$; hence $\widetilde{h}_k = (-i)^k h_k$. Therefore, if $\phi_k = Ah_k$, $\{\phi_k\}_0^\infty$ is an orthonormal basis for L^2 consisting of eigenfunctions for $\mathcal{F}$; namely, $\widehat{\phi}_k = (-i)^k\phi_k$.

8.4 SUMMATION OF FOURIER INTEGRALS AND SERIES

The Fourier inversion theorem shows how to express a function f on $\mathbb{R}^n$ in terms of $\widehat{f}$ provided that f and $\widehat{f}$ are in L^1. The same result holds for periodic functions. Namely, if $f \in L^1(\mathbb{T}^n)$ and $\widehat{f} \in l^1(\mathbb{Z}^n)$, then the Fourier series $\sum_\kappa \widehat{f}(\kappa)e^{2\pi i\kappa\cdot x}$ converges absolutely and uniformly to a function g. Since $l^1 \subset l^2$, it follows from Theorem 8.20 that $f \in L^2$ and that the series converges to f in the L^2 norm. Hence $f = g$ a.e., and $f = g$ everywhere if f is assumed continuous at the outset.

Two questions therefore arise. What conditions on f will guarantee that $\widehat{f}$ is integrable? And how can f be recovered from $\widehat{f}$ if $\widehat{f}$ is not integrable?

As for the first question, since $\widehat{f}$ is bounded for $f \in L^1$, the issue is the decay of $\widehat{f}$ at infinity, and this is related to the smoothness properties of f. For example, by Theorem 8.22e, if $f \in C^{n+1}(\mathbb{R}^n)$ and $\partial^\alpha f \in L^1 \cap C_0$ for $|\alpha| \leq n+1$, then $|\widehat{f}(\xi)| \leq C(1+|\xi|)^{-n-1}$ and hence $\widehat{f} \in L^1(\mathbb{R}^n)$ by Corollary 2.52. The same result holds for periodic functions, for the same reason: If $f \in C^{n+1}(\mathbb{T}^n)$, then $|\widehat{f}(\kappa)| \leq C(1+|\kappa|)^{-n-1}$ and hence $\widehat{f} \in l^1(\mathbb{Z}^n)$.

To obtain sharper results when $n > 1$ requires a generalized notion of partial derivatives, so we shall postpone this task until §9.3. (See Theorem 9.17.) However, for $n = 1$ we can easily obtain a better theorem that covers the useful case of functions that are continuous and piecewise C^1. We state it for periodic functions and leave the nonperiodic case to the reader (Exercise 24).

8.33 Theorem. *Suppose that f is periodic and absolutely continuous on $\mathbb{R}$, and that $f' \in L^p(\mathbb{T})$ for some $p > 1$. Then $\widehat{f} \in l^1(\mathbb{Z})$.*

Proof. Since $p > 1$, we have $C_p = \sum_1^\infty \kappa^{-p} < \infty$; and since $L^p(\mathbb{T}) \subset L^2(\mathbb{T})$ for $p > 2$, we may assume that $p \le 2$. Integration by parts (Theorem 3.36) shows that $\widehat{(f')}(\kappa) = 2\pi i\kappa\widehat{f}(\kappa)$. Hence, by the inequalities of Hölder and Hausdorff-Young, if q is the conjugate exponent to p,

$$\begin{aligned}\sum_{\kappa\neq 0}|\widehat{f}(\kappa)| &\le \Big[\sum_{\kappa\neq 0}(2\pi|\kappa|)^{-p}\Big]^{1/p}\Big[\sum_{\kappa\neq 0}(2\pi|\kappa\widehat{f}(\kappa)|)^q\Big]^{1/q}\\ &= \frac{(2C_p)^{1/p}}{2\pi}\|\widehat{(f')}\|_q \le \frac{(2C_p)^{1/p}}{2\pi}\|f'\|_p.\end{aligned}$$

Adding $|\widehat{f}(0)|$ to both sides, we see that $\|\widehat{f}\|_1 < \infty$. ∎

We now turn to the problem of recovering f from $\widehat{f}$ under minimal hypotheses on f, and we consider first the case of $\mathbb{R}^n$. The proof of the Fourier inversion theorem contains the essential idea: Replace the divergent integral $\int \widehat{f}(\xi)e^{2\pi i\xi\cdot x}\,d\xi$ by $\int \widehat{f}(\xi)\Phi(t\xi)e^{2\pi i\xi\cdot x}\,d\xi$ where Φ is a continuous function that vanishes rapidly enough at infinity to make the integral converge. If we choose Φ to satisfy $\Phi(0) = 1$, then $\Phi(t\xi) \to 1$ as $t \to 0$, and with any luck the corresponding integral will converge to f in some sense. One Φ that works is the function $\Phi(\xi) = e^{-\pi|\xi|^2}$ used in the proof of the inversion theorem, but we shall see below that there are others of independent interest. We therefore formulate a fairly general theorem, for which we need the following lemma that complements Theorem 8.22c.

8.34 Lemma. *If $f, g \in L^2(\mathbb{R}^n)$, then $(\widehat{f}\widehat{g})^\vee = f * g$.*

Proof. $\widehat{f}\widehat{g} \in L^1$ by Plancherel's theorem and Hölder's inequality, so $(\widehat{f}\widehat{g})^\vee$ makes sense. Given $x \in \mathbb{R}^n$, let $h(y) = \overline{g(x-y)}$. It is easily verified that $\widehat{h}(\xi) = \overline{\widehat{g}(\xi)}e^{-2\pi i\xi\cdot x}$, so since $\mathcal{F}$ is unitary on L^2,

$$f * g(x) = \int f\overline{h} = \int \widehat{f}\overline{\widehat{h}} = \int \widehat{f}(\xi)\widehat{g}(\xi)e^{2\pi i\xi\cdot x}\,d\xi = (\widehat{f}\widehat{g})^\vee(x).$$

∎

8.35 Theorem. *Suppose that $\Phi \in L^1 \cap C_0$, $\Phi(0) = 1$, and $\phi = \Phi^\vee \in L^1$. Given $f \in L^1 + L^2$, for $t > 0$ set*

$$f^t(x) = \int \widehat{f}(\xi)\Phi(t\xi)e^{2\pi i\xi\cdot x}\,d\xi.$$

a. If $f \in L^p$ $(1 \le p < \infty)$, then $f^t \in L^p$ and $\|f^t - f\|_p \to 0$ as $t \to 0$.

b. If f is bounded and uniformly continuous, then so is f^t, and $f^t \to f$ uniformly as $t \to 0$.

c. Suppose also that $|\phi(x)| \le C(1+|x|)^{-n-\epsilon}$ for some $C, \epsilon > 0$. Then $f^t(x) \to f(x)$ for every x in the Lebesgue set of f.

Proof. We have $f = f_1 + f_2$ where $f_1 \in L^1$ and $f_2 \in L^2$. Since $\widehat{f_1} \in L^\infty$, $\widehat{f_2} \in L_2$, and $\Phi \in (L^1 \cap C_0) \subset (L^1 \cap L^2)$, the integral defining f^t converges absolutely for every x. Moreover, if $\phi_t(x) = t^{-n}\phi(t^{-1}x)$, we have $\Phi(t\xi) = (\phi_t)\widehat{\ }(\xi)$ by the inversion theorem and Theorem 8.22b, and $\int \phi(x)\,dx = \Phi(0) = 1$. Since $\phi, \Phi \in L^1$ we have $f_1 * \phi \in L^1$ and $\widehat{f_1}\Phi \in L^1$, so by Theorem 8.22c and the inversion formula,

$$\int \widehat{f_1}(\xi)\Phi(t\xi)e^{2\pi i\xi\cdot x}\,d\xi = f_1 * \phi_t(x).$$

Also, $\phi \in L^2$ by the Plancherel theorem, so by Lemma 8.34,

$$\int \widehat{f_2}(\xi)\Phi(t\xi)e^{2\pi i\xi\cdot x}\,d\xi = f_2 * \phi_t(\xi).$$

In short, $f^t = f * \phi_t$, so the assertions follow from Theorems 8.14 and 8.15. ∎

By combining this theorem with the Poisson summation formula, we obtain a corresponding result for periodic functions.

8.36 Theorem. *Suppose that $\Phi \in C(\mathbb{R}^n)$ satisfies $|\Phi(\xi)| \le C(1 + |\xi|)^{-n-\epsilon}$, $|\Phi^\vee(x)| \le C(1 + |x|)^{-n-\epsilon}$, and $\Phi(0) = 1$. Given $f \in L^1(\mathbb{T}^n)$, for $t > 0$ set*

$$f^t(x) = \sum_{\kappa\in\mathbb{Z}^n} \widehat{f}(\kappa)\Phi(t\kappa)e^{2\pi i\kappa\cdot x}$$

(which converges absolutely since $\sum_\kappa |\Phi(t\kappa)| < \infty$).

a. If $f \in L^p(\mathbb{T}^n)$ $(1 \le p < \infty)$, then $\|f^t - f\|_p \to 0$ as $t \to 0$, and if $f \in C(\mathbb{T}^n)$, then $f^t \to f$ uniformly as $t \to 0$.

b. $f^t(x) \to f(x)$ for every x in the Lebesgue set of f.

Proof. Let $\phi = \Phi^\vee$ and $\phi_t(x) = t^{-n}\phi(t^{-1}x)$. Then $(\phi_t)\widehat{\ }(\xi) = \Phi(t\xi)$, and ϕ_t satisfies the hypotheses of the Poisson summation formula, so

$$\sum_{k\in\mathbb{Z}^n} \phi_t(x - k) = \sum_{k\in\mathbb{Z}^n} \Phi(t\kappa)e^{2\pi i\kappa\cdot x}.$$

Let us denote the common value of these sums by $\psi_t(x)$. Then

$$(f * \psi_t)\widehat{\ }(\kappa) = \widehat{f}(\kappa)\widehat{\psi_t}(\kappa) = \widehat{f}(\kappa)\Phi(t\kappa) = (f^t)\widehat{\ }(\kappa),$$

so $f^t = f * \psi_t$. Hence, by Young's inequality and Theorem 8.31 we have

$$\|f^t\|_p \le \|f\|_p\|\psi_t\|_1 \le \|f\|_p\|\phi_t\|_1 = \|f\|_p\|\phi\|_1,$$

so the operators $f \mapsto f^t$ are uniformly bounded on L^p, $1 \le p \le \infty$.

Now, since Φ is continuous and $\Phi(0) = 1$, we clearly have $f^t \to f$ uniformly (and hence in $L^p(\mathbb{T}^n)$) if f is a trigonometric polynomial — that is, if $\widehat{f}(\kappa) = 0$ for all but finitely many κ. But the trigonometric polynomials are dense in $C(\mathbb{T}^n)$ in the

uniform norm by the Stone-Weierstrass theorem, and hence also dense in $L^p(\mathbb{T}^n)$ in the L^p norm for $p < \infty$. Assertion (a) therefore follows from Proposition 5.17.

To prove (b), suppose that x is in the Lebesgue set of f; by translating f we may assume that $x = 0$, which simplifies the notation. With $Q = [-\frac{1}{2}, \frac{1}{2})^n$, we have

$$\begin{aligned} f^t(0) = f * \psi_t(0) &= \int_Q f(x)\psi_t(-x)\,dx \\ &= \int_Q f(x)\phi_t(-x)\,dx + \sum_{k\neq 0}\int_Q f(x)\phi_t(-x+k)\,dx. \end{aligned}$$

Since

$$|\phi_t(x)| \le Ct^{-n}(1+t^{-1}|x|)^{-n-\epsilon} \le Ct^{\epsilon}|x|^{-n-\epsilon},$$

for $x \in Q$ and $k \neq 0$ we have $|\phi_t(-x+k)| \le C2^{n+\epsilon}t^{\epsilon}|k|^{-n-\epsilon}$, and hence

$$\sum_{k\neq 0}\int_Q |f(x)\phi_t(-x+k)|\,dx \le \Big[C2^{n+\epsilon}\|f\|_1 \sum_{k\neq 0}|k|^{-n-\epsilon}\Big]t^{\epsilon},$$

which vanishes as $t \to 0$. On the other hand, if we define $g = f\chi_Q \in L^1(\mathbb{R}^n)$, then 0 is in the Lebesgue set of g (because 0 is in the interior of Q, and the condition that 0 be in the Lebesgue set of g depends only on the behavior of g near 0), so by Theorem 8.15,

$$\lim_{t\to 0}\int_Q f(x)\phi_t(-x)\,dx = \lim_{t\to 0} g * \phi_t(0) = g(0) = f(0).$$

∎

Let us examine some specific examples of functions Φ that can be used in Theorems 8.35 and 8.36. The first is the one already used in the proof of the inversion theorem,

$$\Phi(\xi) = e^{-\pi|\xi|^2}, \qquad \phi(x) = \Phi^{\vee}(x) = e^{-\pi|x|^2}.$$

This ϕ is called the **Gauss kernel** or **Weierstrass kernel**. It is important for a number of reasons, including its connection with the heat equation that we shall explain in §8.7. When $n = 1$, its periodized version

$$\psi_t(x) = \frac{1}{t}\sum_{k\in\mathbb{Z}} e^{-\pi|x-k|^2/t^2} = \sum_{\kappa\in\mathbb{Z}} e^{-\pi t^2\kappa^2}e^{2\pi i\kappa\cdot x},$$

in terms of which the f^t in Theorem 8.36 is given by $f^t = f * \psi_t$, is essentially one of the Jacobi theta functions, which are connected with elliptic functions and have applications in number theory.

The second example is $\Phi(\xi) = e^{-2\pi|\xi|}$, whose inverse Fourier transform ϕ is called the **Poisson kernel** on $\mathbb{R}^n$. When $n = 1$, we have

$$\begin{aligned} \phi(x) &= \int_{-\infty}^{0} e^{2\pi(1+ix)\xi}\,d\xi + \int_0^{\infty} e^{2\pi(-1+ix)\xi}\,d\xi \\ &= \frac{1}{2\pi}\left[\frac{1}{1+ix} + \frac{1}{1-ix}\right] = \frac{1}{\pi(1+x^2)}. \end{aligned} \tag{8.37}$$

The formula for ϕ in higher dimensions is worked out in Exercise 26; it turns out that $\phi(x)$ is a constant multiple of $(1+|x|^2)^{-(n+1)/2}$. Like the Gauss kernel, the Poisson kernel has an interpretation in terms of partial differential equations that we shall explain in §8.7.

If we take $n = 1$ and $\Phi(\xi) = e^{-2\pi|\xi|}$ in Theorem 8.36, make the substitution $r = e^{-2\pi t}$, and write $A_r f$ in place of f^t, we obtain

$$\begin{aligned} A_r f(x) &= \sum_{\kappa\in\mathbb{Z}} r^{|\kappa|}\widehat{f}(\kappa)e^{-2\pi i\kappa x} \\ &= \widehat{f}(0) + \sum_{k=1}^{\infty} r^k\big[\widehat{f}(k)e^{2\pi ikx} + \widehat{f}(-k)e^{-2\pi ikx}\big]. \end{aligned} \tag{8.38}$$

This formula is a special case of one of the classical methods for summing a (possibly) divergent series. Namely, if $\sum_0^\infty a_k$ is a series of complex numbers, for $0 < r < 1$ its **rth Abel mean** is the series $\sum_0^\infty r^k a_k$. If the latter series converges for $r < 1$ to the sum $S(r)$ and the limit $S = \lim_{r\nearrow 1} S(r)$ exists, the series $\sum_0^\infty a_k$ is said to be **Abel summable** to S. If $\sum_0^\infty a_k$ converges to the sum S, then it is also Abel summable to S (Exercise 27), but the Abel sum may exist even when the series diverges.

In (8.38), $A_r f(x)$ is the rth Abel mean of the Fourier series of f, in which the kth and $(-k)$th terms are grouped together to make a series indexed by the nonnegative integers. It has the following complex-variable interpretation: If we set $z = re^{2\pi ix}$, then

$$A_r f(x) = \sum_0^\infty \widehat{f}(k)z^k + \sum_1^\infty \widehat{f}(-k)\bar{z}^k.$$

The two series on the right define, respectively, a holomorphic and an antiholomorphic function on the unit disc $|z| < 1$. In particular, $A_r f(x)$ is a harmonic function on the unit disc, and the fact that $A_r f \to f$ as $r \to 1$ means that f is the boundary value of this function on the unit circle. See also Exercise 28.

Our final example is the function $\Phi(\xi) = \max(1-|\xi|, 0)$ with $n = 1$. Its inverse Fourier transform is

$$\begin{aligned} \phi(x) &= \int_{-1}^{0}(1+\xi)e^{2\pi i\xi\cdot x}\,d\xi + \int_0^1 (1-\xi)e^{2\pi i\xi\cdot x}\,d\xi \\ &= \frac{e^{2\pi ix} + e^{-2\pi ix} - 2}{(2\pi ix)^2} = \left(\frac{\sin \pi x}{\pi x}\right)^2. \end{aligned}$$

If we use this Φ in Theorem 8.36, take $t = (m+1)^{-1}$ $(m = 0, 1, 2, \ldots)$, and write $\sigma_m f(x)$ for $f^{1/(m+1)}(x)$, we obtain

$$\begin{aligned} \sigma_m f(x) &= \sum_{\kappa=-m}^{m} \frac{m+1-|\kappa|}{m+1}\widehat{f}(\kappa)e^{2\pi i\kappa x} \\ &= \widehat{f}(0) + \sum_{k=1}^{m}\frac{m+1-k}{m+1}\big[\widehat{f}(k)e^{2\pi ikx} + \widehat{f}(-k)e^{-2\pi ikx}\big]. \end{aligned} \tag{8.39}$$

This is an instance of another classical method for summing divergent series. Namely, if $\sum_0^\infty a_k$ is a series of complex numbers, its ***m*th Cesàro mean** is the average of its first $m+1$ partial sums, $(m+1)^{-1}\sum_0^m S_n$, where $S_n = \sum_0^n a_k$. If the sequence of Cesàro means converges as $m \to \infty$ to a limit S, the series is said to be **Cesàro summable** to S. It is easily verified that if $\sum_0^\infty a_k$ converges to S, then it is Cesàro summable to S (but perhaps not conversely), and that $\sigma_m f(x)$ is the mth Cesàro mean of the Fourier series of f with the kth and $(-k)$th terms grouped together. See Exercise 29, and also Exercise 33 in the next section.

Exercises

24. State and prove an analogue of Theorem 8.33 for functions on $\mathbb{R}$. (In addition to the hypotheses that f be locally absolutely continuous and that $f' \in L^p$ for some $p > 1$, you will need some further conditions f and/or f' at infinity to make the argument work. Make them as mild as possible.)

25. For $0 < \alpha \le 1$, let $\Lambda_\alpha(\mathbb{T})$ be the space of Hölder continuous functions on $\mathbb{T}$ of exponent α as in Exercise 11 in §5.1. Suppose $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$.

a. If f satisfies the hypotheses of Theorem 8.33, then $f \in \Lambda_{1/q}(\mathbb{T})$, but f need not lie in $\Lambda_\alpha(\mathbb{T})$ for any $\alpha > 1/q$. (Hint: $f(b) - f(a) = \int_a^b f'(t)\,dt$.)

b. If $\alpha < 1$, $\Lambda_\alpha(\mathbb{T})$ contains functions that are not of bounded variation and hence are not absolutely continuous. (But cf. Exercise 37 in §3.5.)

26. The aim of this exercise is to show that the inverse Fourier transform of $e^{-2\pi|\xi|}$ on $\mathbb{R}^n$ is

$$\phi(x) = \frac{\Gamma(\frac{1}{2}(n+1))}{\pi^{(n+1)/2}}\left(1 + |x|^2\right)^{-(n+1)/2}.$$

a. If $\beta \ge 0$, $e^{-\beta} = \pi^{-1}\int_{-\infty}^\infty (1+t^2)^{-1}e^{-i\beta t}\,dt$. (Use (8.37).)

b. If $\beta \ge 0$, $e^{-\beta} = \int_0^\infty (\pi s)^{-1/2}e^{-s}e^{-\beta^2/4s}\,ds$. (Use (a), Proposition 8.24, and the formula $(1+t^2)^{-1} = \int_0^\infty e^{-(1+t^2)s}\,ds$.)

c. Let $\beta = 2\pi|\xi|$ where $\xi \in \mathbb{R}^n$; then the formula in (b) expresses $e^{-2\pi|\xi|}$ as a superposition of dilated Gauss kernels. Use Proposition 8.24 again to derive the asserted formula for ϕ.

27. Suppose that the numerical series $\sum_0^\infty a_k$ is convergent.

a. Let $S_m^n = \sum_m^n a_k$. Then $\sum_m^n r^k a_k = \sum_m^{n-1} S_m^j(r^j - r^{j+1}) + S_m^n r^n$ for $0 \le r \le 1$ ("summation by parts").

b. $|\sum_m^n r^k a_k| \le \sup_{j \ge m} |S_m^j|$.

c. The series $\sum_0^\infty r^k a_k$ is uniformly convergent for $0 \le r \le 1$, and hence its sum $S(r)$ is continuous there. In particular, $\sum_0^\infty a_k = \lim_{r \nearrow 1} S(r)$.

28. Suppose that $f \in L^1(\mathbb{T})$, and let $A_r f$ be given by (8.38).

a. $A_r f = f * P_r$ where $P_r(x) = \sum_{-\infty}^\infty r^{|\kappa|}e^{2\pi i\kappa x}$ is the **Poisson kernel** for $\mathbb{T}$.

b. $P_r(x) = (1-r^2)/(1+r^2-2r\cos 2\pi x)$.

29. Given $\{a_k\}_0^\infty \subset \mathbb{C}$, let $S_n = \sum_0^n a_k$ and $\sigma_m = (m+1)^{-1}\sum_0^m S_n$.

a. $\sigma_m = (m+1)^{-1}\sum_0^m (m+1-k)a_k$.
b. If $\lim_{n\to\infty} S_n = \sum_0^\infty a_k$ exists, then so does $\lim_{m\to\infty}\sigma_m$, and the two limits are equal.
c. The series $\sum_0^\infty (-1)^k$ diverges but is Abel and Cesàro summable to $\frac{1}{2}$.

30. If $f \in L^1(\mathbb{R}^n)$, f is continuous at 0, and $\widehat{f} \geq 0$, then $\widehat{f} \in L^1$. (Use Theorem 8.35c and Fatou's lemma.)

31. Suppose $a > 0$. Use (8.37) to show that

$$\sum_{-\infty}^{\infty} \frac{1}{k^2+a^2} = \frac{\pi}{a}\frac{1+e^{-2\pi a}}{1-e^{-2\pi a}}.$$

Then subtract a^{-2} from both sides and let $a \to 0$ to show that $\sum_1^\infty k^{-2} = \pi^2/6$.

32. A C^∞ function f on $\mathbb{R}$ is **real-analytic** if for every $x \in \mathbb{R}$, f is the sum of its Taylor series based at x in some neighborhood of x. If f is periodic and we regard f as a function on $S = \{z \in \mathbb{C} : |z| = 1\}$, this condition is equivalent to the condition that f be the restriction to S of a holomorphic function on some neighborhood of S. Show that $f \in C^\infty(\mathbb{T})$ is real-analytic iff $|\widehat{f}(\kappa)| \leq Ce^{-\epsilon|\kappa|}$ for some $C, \epsilon > 0$. (See the discussion of the Abel means $A_r f$ in the text, and note that $\overline{z} = z^{-1}$ when $|z| = 1$.)

8.5 POINTWISE CONVERGENCE OF FOURIER SERIES

The techniques and results of the previous two sections, involving such things as L^p norms and summability methods, are relatively modern; they were preceded historically by the study of pointwise convergence of one-dimensional Fourier series. Although the latter is one of the oldest parts of Fourier analysis, it is also one of the most difficult — unfortunately for the mathematicians who developed it, but fortunately for us who are the beneficiaries of the ideas and techniques they invented in doing so. A thorough study of this issue is beyond the scope of this book, but we would be remiss not to present a few of the classic results.

To set the stage, suppose $f \in L^1(\mathbb{T})$. We denote by $S_m f$ the mth symmetric partial sum of the Fourier series of f:

$$S_m f(x) = \sum_{-m}^{m} \widehat{f}(k)e^{2\pi ikx}.$$

From the definition of $\widehat{f}(k)$, we have

$$S_m f(x) = \sum_{-m}^{m} \int_0^1 f(y)e^{2\pi ik(x-y)}\,dy = f * D_m(x),$$

where D_m is the mth **Dirichlet kernel**:

$$D_m(x) = \sum_{-m}^{m} e^{2\pi ikx}.$$

The terms in this sum form a geometric progression, so

$$D_m(x) = e^{-2\pi imx} \sum_{0}^{2m} e^{2\pi ikx} = e^{-2\pi imx} \frac{e^{2\pi(2m+1)x} - 1}{e^{2\pi ix} - 1}.$$

Multiplying top and bottom by $e^{-\pi ix}$ yields the standard closed formula for D_m:

$$D_m(x) = \frac{e^{(2m+1)\pi ix} - e^{-(2m+1)\pi ix}}{e^{\pi ix} - e^{-\pi ix}} = \frac{\sin(2m+1)\pi x}{\sin \pi x}. \tag{8.40}$$

The difficulty with the partial sums $S_m f$, as opposed to (for example) the Abel or Cesàro means, can be summed up in a nutshell as follows. $S_m f$ can be regarded as a special case of the construction in Theorem 8.36; in fact, with the notation used there, $S_m f = f^{1/m}$ if we take $\Phi = \chi_{[-1,1]}$. *But* $\chi_{[-1,1]}$ *does not satisfy the hypotheses of Theorem 8.36,* because its inverse Fourier transform $(\pi x)^{-1} \sin 2\pi x$ (Exercise 15a) is not in $L^1(\mathbb{R})$. On the level of periodic functions, this is reflected in the fact that although $D_m \in L^1(\mathbb{T})$ for all m, $\|D_m\|_1 \to \infty$ as $m \to \infty$ (Exercise 34).

Among the consequences of this is that the Fourier series of a continuous function f need not converge pointwise, much less uniformly, to f; see Exercise 35. (This does not contradict the fact that trigonometric polynomials are dense in $C(\mathbb{T})$! It just means that if one wants to approximate a function $f \in C(\mathbb{T})$ uniformly by trigonometric polynomials, one should not count on the partial sums $S_m f$ to do the job; the Cesàro means defined by (8.39) work much better in general.) To obtain positive results for pointwise convergence, one must look in other directions.

The first really general theorem about pointwise convergence of Fourier series was obtained in 1829 by Dirichlet, who showed that $S_m f(x) \to \frac{1}{2}[f(x+) + f(x-)]$ for every x provided that f is piecewise continuous and piecewise monotone. Later refinements of the argument showed that what is really needed is for f to be of bounded variation. We now prove this theorem, for which we need two lemmas. The first one is a slight generalization of one of the more arcane theorems of elementary calculus, the "second mean value theorem for integrals."

8.41 Lemma. *Let ϕ and ψ be real-valued functions on $[a, b]$. Suppose that ϕ is monotone and right continuous on $[a, b]$ and ψ is continuous on $[a, b]$. Then there exists $\eta \in [a, b]$ such that*

$$\int_a^b \phi(x)\psi(x)\,dx = \phi(a) \int_a^\eta \psi(x)\,dx + \phi(b) \int_\eta^b \psi(x)\,dx.$$

Proof. Adding a constant c to ϕ changes both sides of the equation by the amount $c \int_a^b \psi(x)\,dx$, so we may assume that $\phi(a) = 0$. We may also assume that ϕ

is increasing; otherwise replace ϕ by $-\phi$. Let $\Psi(x) = \int_x^b \psi(t)\,dt$ (so that $\Psi' = -\psi$) and apply Theorem 3.36:

$$\int_a^b \phi(x)\psi(x)\,dx = -\phi(x)\Psi(x)\Big|_a^b + \int_{(a,b]} \Psi(x)\,d\phi(x).$$

The endpoint evaluations vanish since $\phi(a) = \Psi(b) = 0$. Since ϕ is increasing and $\int_{(a,b]} d\phi = \phi(b) - \phi(a) = \phi(b)$, if m and M are the minimum and maximum values of Ψ on $[a,b]$ we have $m\phi(b) \le \int_{(a,b]} \Psi\,d\phi \le M\phi(b)$. By the intermediate value theorem, then, there exists $\eta \in [a,b]$ such that $\int_{(a,b]} \Psi\,d\phi = \Psi(\eta)\phi(b)$, which is the desired result. ∎

8.42 Lemma. *There is a constant $C < \infty$ such that for every $m \ge 0$ and every $[a,b] \subset [-\frac{1}{2}, \frac{1}{2}]$,*

$$\left|\int_a^b D_m(x)\,dx\right| \le C.$$

Moreover, $\int_{-1/2}^0 D_m(x)\,dx = \int_0^{1/2} D_m(x)\,dx = \frac{1}{2}$ for all m.

Proof. By (8.40),

$$\int_a^b D_m(x)\,dx = \int_a^b \frac{\sin(2m+1)\pi x}{\pi x}\,dx + \int_a^b \sin(2m+1)\pi x\left[\frac{1}{\sin \pi x} - \frac{1}{\pi x}\right]dx.$$

Since $(\sin \pi x)^{-1} - (\pi x)^{-1}$ is bounded on $[-\frac{1}{2}, \frac{1}{2}]$ and $|\sin(2m+1)\pi x| \le 1$, the second integral on the right is bounded in absolute value by a constant. With the substitution $y = (2m+1)\pi x$, the first one becomes

$$\int_{(2m+1)\pi a}^{(2m+1)\pi b} \frac{\sin y}{\pi y}\,dy = \frac{\mathrm{Si}[(2m+1)\pi b] - \mathrm{Si}[(2m+1)\pi a]}{\pi}$$

where $\mathrm{Si}(x) = \int_0^x y^{-1}\sin y\,dy$. But $\mathrm{Si}(x)$ is continuous and approaches the finite limits $\pm\frac{1}{2}\pi$ as $x \to \pm\infty$ (see Exercise 59b in §2.6), so $\mathrm{Si}(x)$ is bounded. This proves the first assertion. As for the second one,

$$\int_{-1/2}^{1/2} D_m(x)\,dx = \sum_{-m}^{m} \int_{-1/2}^{1/2} e^{2\pi i k x}\,dx = 1$$

(only the term with $k = 0$ is nonzero), so since D_m is even,

$$\int_{-1/2}^{0} D_m(x)\,dx = \int_0^{\frac{1}{2}} D_m(x)\,dx = \tfrac{1}{2}.$$

∎

8.43 Theorem. *If $f \in BV(\mathbb{T})$ — that is, if f is periodic on $\mathbb{R}$ and of bounded variation on $[-\frac{1}{2}, \frac{1}{2}]$ — then*

$$\lim_{m\to\infty} S_m f(x) = \tfrac{1}{2}\big[f(x+) + f(x-)\big] \text{ for every } x.$$

In particular, $\lim_{m\to\infty} S_m f(x) = f(x)$ at every x at which f is continuous.

Proof. We begin by making some reductions. In examining the convergence of $S_m f(x)$, we may assume that $x = 0$ (by replacing f with the translated function $\tau_{-x} f$), that f is real-valued (by considering the real and imaginary parts separately), and that f is right continuous (since replacing $f(t)$ by $f(t+)$ affects neither $S_m f$ nor $\frac{1}{2}[f(0+) + f(0-)]$). In this case, by Theorem 3.27b, on the interval $[-\frac{1}{2}, \frac{1}{2})$ we can write f as the difference of two right continuous increasing functions g and h. If these functions are extended to $\mathbb{R}$ by periodicity, they are again of bounded variation, and it is enough to show that $S_m g(0) \to \frac{1}{2}[g(0+) + g(0-)]$ and likewise for h.

In short, it suffices to consider the case where $x = 0$ and f is increasing and right continuous on $[-\frac{1}{2}, \frac{1}{2})$. Since D_m is even, we have $S_m f(0) = f * D_m(0) = \int_{-1/2}^{1/2} f(x) D_m(x)\, dx$, so by Lemma 8.42,

$$\begin{aligned} &S_m f(0) - \tfrac{1}{2}\big[f(0+) + f(0-)\big] \\ &\qquad = \int_0^{1/2} \big[f(x) - f(0+)\big] D_m(x)\, dx + \int_{-1/2}^{0} \big[f(x) - f(0-)\big] D_m(x)\, dx. \end{aligned}$$

We shall show that the first integral on the right tends to zero as $m \to \infty$; a similar argument shows that the second integral also tends to zero, thereby completing the proof.

Given $\epsilon > 0$, choose $\delta > 0$ small enough so that $f(\delta) - f(0+) < \epsilon/C$ where C is as in Lemma 8.42. Then by Lemma 8.41, for some $\eta \in [0, \delta]$,

$$\left|\int_0^{\delta} \big[f(x) - f(0+)\big] D_m(x)\, dx\right| = \big[f(\delta) - f(0+)\big]\left|\int_{\eta}^{\delta} D_m(x)\, dx\right|,$$

which is less than ϵ. On the other hand, by (8.40),

$$\int_{\delta}^{1/2} \big[f(x) - f(0+)\big] D_m(x)\, dx = \widehat{g}_+(-m) - \widehat{g}_-(m),$$

where $g_\pm$ is the periodic function given on the interval $[-\frac{1}{2}, \frac{1}{2})$ by

$$g_\pm(x) = \frac{\big[f(x) - f(0+)\big] e^{\pm \pi i x}}{2i \sin \pi x} \chi_{[\delta, 1/2)}(x).$$

But $g_\pm \in L^1(\mathbb{T})$, so $\widehat{g}_\pm(\mp m) \to 0$ as $m \to \infty$ by the Riemann-Lebesgue lemma (the periodic analogue of Theorem 8.22f). Therefore,

$$\limsup_{m\to\infty} \left|\int_0^{1/2} \big[f(x) - f(0+)\big] D_m(x)\, dx\right| < \epsilon$$

for every $\epsilon > 0$, and we are done. ∎

One of the less attractive features of Fourier series is that bad behavior of a function at one point affects the behavior of its Fourier series at all points. For example, if f has even one jump discontinuity, then $\widehat{f}$ cannot be in $l^1(\mathbb{Z})$ and so the series $\sum \widehat{f}(k)e^{2\pi ikx}$ cannot converge absolutely at any point. However, to a limited extent the convergence of the series at a point x depends only on the behavior of f near x, as explained in the following localization theorem.

8.44 Theorem. *If f and g are in $L^1(\mathbb{T})$ and $f = g$ on an open interval I, then $S_m f - S_m g \to 0$ uniformly on compact subsets of I.*

Proof. It is enough to assume that $g = 0$ (consider $f - g$), and by translating f we may assume that I is centered at 0, say $I = (-c, c)$ where $c \le \frac{1}{2}$. Fix $\delta < c$; we shall show that if $f = 0$ on I then $S_m f \to 0$ uniformly on $[-\delta, \delta]$.

The first step is to show that $S_m f \to 0$ pointwise on $[-\delta, \delta]$, and the argument is similar to the preceding proof. Namely, by (8.40) we have

$$S_m f(x) = \int_{-1/2}^{1/2} f(x-y)D_m(y)\,dy = \widehat{g}_{x,+}(-m) - \widehat{g}_{x,-}(m),$$

where

$$g_{x,\pm}(y) = \frac{f(x-y)e^{\pm\pi iy}}{2i\sin\pi y}.$$

Since $f(x-y) = 0$ on a neighborhood of the zeros of $\sin\pi y$, the functions $g_{x,\pm}$ are in $L^1(\mathbb{T})$, so $\widehat{g}_{x,\pm}(\mp m) \to 0$ by the Riemann-Lebesgue lemma.

The next step is to show that if $x_1, x_2 \in [-\delta, \delta]$, then $S_m f(x_1) - S_m f(x_2)$ vanishes as $x_1 - x_2 \to 0$, uniformly in m. By (8.40) again,

$$S_m f(x_1) - S_m f(x_2) = \int_{-1/2}^{1/2} \frac{\sin(2m+1)\pi y}{\sin\pi y}\big[f(x_1-y) - f(x_2-y)\big]\,dy.$$

But $f(x_1 - y) - f(x_2 - y) = 0$ for $|y| < c - \delta$, and for $c - \delta \le |y| \le \frac{1}{2}$ we have

$$\left|\frac{\sin(2m+1)\pi y}{\sin\pi y}\right| \le \frac{1}{\sin\pi(c-\delta)} = A,$$

where A is independent of m. Hence

$$|S_m f(x_1) - S_m f(x_2)| \le A\int_{-1/2}^{1/2} |f(x_1-y) - f(x_2-y)|\,dy = A\|\tau_{x_1}f - \tau_{x_2}f\|_1,$$

which vanishes as $x_1 - x_2 \to 0$ by (the periodic analogue of) Proposition 8.5.

Now, given $\epsilon > 0$, we can choose η small enough so that if $x_1, x_2 \in [-\delta, \delta]$ and $|x_1 - x_2| < \eta$, then $|S_m f(x_1) - S_m f(x_2)| < \epsilon/2$. Choose $x_1, \ldots, x_k \in [-\delta, \delta]$ so that the intervals $|x - x_j| < \eta$ cover $[-\delta, \delta]$. Since $S_m f(x_j) \to 0$ for each j, we can choose M large enough so that $|S_m f(x_j)| < \epsilon/2$ for $m > M$ and $1 \le j \le k$. If $|x| \le \delta$, then, we have $|x - x_j| < \eta$ for some j, so

$$|S_m f(x)| \le |S_m f(x) - S_m f(x_j)| + |S_m f(x_j)| < \epsilon$$

for $m > M$, and we are done. ∎

8.45 Corollary. *Suppose that $f \in L^1(\mathbb{T})$ and I is an open interval of length ≤ 1.*

a. If f agrees on I with a function g such that $\widehat{g} \in l^1(\mathbb{Z})$, then $S_m f \to f$ uniformly on compact subsets of I.

b. If f is absolutely continuous on I and $f' \in L^p(I)$ for some $p > 1$, then $S_m f \to f$ uniformly on compact subsets of I.

Proof. If $f = g$ on I, then $S_m f - f = S_m f - g = (S_m f - S_m g) + (S_m g - g)$ on I, and if $\widehat{g} \in l^1(\mathbb{Z})$, then $S_m g \to g$ uniformly on $\mathbb{R}$; (a) follows. As for (b), given $[a_0, b_0] \subset I$, pick $a < a_0$ and $b > b_0$ so that $[a, b] \subset I$, and let g be the continuous periodic function that equals f on $[a, b]$ and is linear on $[b, a+1]$ (which is unique since $g(b) = f(b)$ and $g(a+1) = g(a) = f(a)$). Under the hypotheses of (b), g is absolutely continuous on $\mathbb{R}$ and $g' \in L^p(\mathbb{T})$, so $\widehat{g} \in l^1(\mathbb{Z})$ by Theorem 8.33. Thus $S_m f \to f$ uniformly on $[a_0, b_0]$ by (a). ∎

Finally, we discuss the behavior of $S_m f$ near a jump discontinuity of f. Let us first consider a simple example: Let

$$\phi(x) = \tfrac{1}{2} - x - [x] \qquad ([x] = \text{ greatest integer } \leq x). \tag{8.46}$$

Then ϕ is periodic and is C^∞ except for jump discontinuities at the integers, where $\phi(j+) - \phi(j-) = 1$. It is easy to check that $\widehat{\phi}(0) = 0$ and $\widehat{\phi}(k) = (2\pi i k)^{-1}$ for $k \neq 0$ (Exercise 13a), so that

$$S_m \phi(x) = \sum_{0<|k|\leq m} \frac{e^{2\pi i k x}}{2\pi i k} = \sum_1^m \frac{\sin 2\pi k x}{\pi k}.$$

From Corollary 8.45 it follows that $S_m \phi \to \phi$ uniformly on any compact set not containing an integer, and it is obvious that $S_m \phi(x) = 0$ when x is an integer. But near the integers a peculiar thing happens: $S_m \phi$ contains a sequence of spikes that overshoot and undershoot ϕ, as shown in Figure 8.1, and as $m \to \infty$ the spikes tend to zero in width but *not* in height. In fact, when m is large the value of $S_m \phi$ at its first maximum to the right of 0 is about 0.5895, about 18% greater than $\phi(0+) = \frac{1}{2}$. This is known as the **Gibbs phenomenon**; the precise statement and proof are given in Exercise 37.

Now suppose that f is any periodic function on $\mathbb{R}$ having a jump discontinuity at $x = a$ (that is, $f(a+)$ and $f(a-)$ exist and are unequal). Then the function

$$g(x) = f(x) - \big[f(a+) - f(a-)\big]\phi(x - a)$$

is continuous at every point where f is, and also at $x = a$ provided that we (re)define $g(a)$ to be $\frac{1}{2}[f(a+) + f(a-)]$, as the jumps in f and ϕ cancel out. If g satisfies one of the hypotheses of Corollary 8.45 on an interval I containing a, the Fourier series of g will converge uniformly near a, and hence the Fourier series of f will exhibit the same Gibbs phenomenon as that of ϕ.

Finally, suppose that f is periodic and continuous except at finitely many points $a_1, \ldots, a_k \in \mathbb{T}$, where f has jump discontinuities. We can then subtract off all the

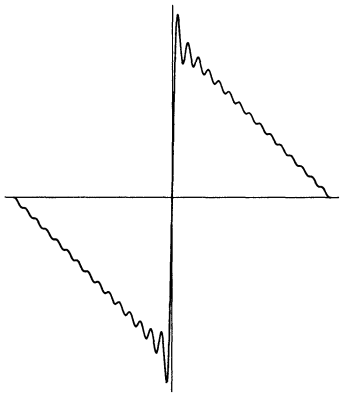

Fig. 8.1 The Gibbs phenomenon: the graph $y = \sum_1^{30} (\pi k)^{-1} \sin 2\pi kx$, $-\frac{1}{2} \le x \le \frac{1}{2}$.

jumps to form a continuous function g:

$$g(x) = f(x) - \sum \big[f(a_j+) - f(a_j-)\big]\phi(x - a_j)$$

If f satisfies some mild smoothness conditions — for example, if f is absolutely continuous on any interval not containing any a_j and $f' \in L^p$ for some $p > 1$ — then $\widehat{g}$ will be in $l^1(\mathbb{Z})$. Conclusion: $S_m f \to f$ uniformly on any interval not containing any a_j, $S_m(a_j) \to \frac{1}{2}[f(a_j+) + f(a_j-)]$, and $S_m f$ exhibits the Gibbs phenomenon near every a_j.

Exercises

33. Let $\sigma_m f$ be the Cesàro means of the Fourier series of f given by (8.39).

a. $\sigma_m f = f * F_m$ where $F_m = (m+1)^{-1} \sum_0^m D_k$ and D_k is the kth Dirichlet kernel. (See Exercise 29a.) F_m is called the mth **Fejér kernel**.

b. $F_m(x) = \sin^2(m+1)\pi x/(m+1)\sin^2 \pi x$. (Use (8.40) and the fact that $\sin(2k+1)\pi x = \operatorname{Im} e^{(2k+1)\pi i x}$.)

34. If D_m is the mth Dirichlet kernel, $\|D_m\|_1 \to \infty$ as $m \to \infty$. (Make the substitution $y = (2m+1)\pi x$ and use Exercise 59a in §2.6.)

35. The purpose of this exercise is to show that the Fourier series of "most" continuous functions on $\mathbb{T}$ do not converge pointwise.

a. Define $\phi_m(f) = S_m f(0)$. Then $\phi \in C(\mathbb{T})^*$ and $\|\phi\| = \|D_m\|_1$.

b. The set of all $f \in C(\mathbb{T})$ such that the sequence $\{S_m f(0)\}$ converges is meager in $C(\mathbb{T})$. (Use Exercise 34 and the uniform boundedness principle.)

c. There exist $f \in C(\mathbb{T})$ (in fact, a residual set of such f's) such that $\{S_m f(x)\}$ diverges for every x in a dense subset of $\mathbb{T}$. (The result of (b) holds if the point 0 is replaced by any other point in $\mathbb{T}$. Apply Exercise 40 in §5.3.)

36. The Fourier transform is not surjective from $L^1(\mathbb{T})$ to $C_0(\mathbb{Z})$. (Use Exercise 34, and cf. Exercise 16c.)

37. Let ϕ be given by (8.46) and let $\Delta_m = S_m\phi - \phi$.

a. $(d/dx)\Delta_m(x) = D_m(x)$ for $x \notin \mathbb{Z}$.

b. The first maximum of Δ_m to the right of 0 occurs at $x = (2m+1)^{-1}$, and

$$\lim_{m\to\infty} \Delta_m\left(\frac{1}{2m+1}\right) = \frac{1}{\pi}\int_0^\pi \frac{\sin t}{t}\,dt - \frac{1}{2} \cong 0.0895.$$

(Use (8.40) and the fact that $\Delta_m(x) = \int_0^x \Delta_m'(t)\,dt - \frac{1}{2}$.)

c. More generally, the jth critical point of Δ_m to the right of 0 occurs at $x = j/(2m+1)$ $(j = 1, \dots, 2m)$, and

$$\lim_{m\to\infty} \Delta_m\left(\frac{j}{2m+1}\right) = \frac{1}{\pi}\int_0^{j\pi} \frac{\sin t}{t}\,dt - \frac{1}{2}.$$

These numbers are positive for j odd and negative for j even. (See Exercise 59b in §2.6.)

8.6 FOURIER ANALYSIS OF MEASURES

We recall that $M(\mathbb{R}^n)$ is the space of complex Borel measures on $\mathbb{R}^n$ (which are automatically Radon measures by Theorem 7.8), and we embed $L^1(\mathbb{R}^n)$ into $M(\mathbb{R}^n)$ by identifying $f \in L^1$ with the measure $d\mu = f\,dm$. We shall need to define products of complex measures on Cartesian product spaces, which can easily be done in terms of products of positive measures by using Radon-Nikodym derivatives. Namely, if $\mu, \nu \in M(\mathbb{R}^n)$, we define $\mu \times \nu \in M(\mathbb{R}^n \times \mathbb{R}^n)$ by

$$d(\mu \times \nu)(x,y) = \frac{d\mu}{d|\mu|}(x)\frac{d\nu}{d|\nu|}(y)\,d\big(|\mu| \times |\nu|\big)(x,y).$$

If $\mu, \nu \in M(\mathbb{R}^n)$, we define their convolution $\mu * \nu \in M(\mathbb{R}^n)$ by $\mu * \nu(E) = \mu \times \nu(\alpha^{-1}(E))$ where $\alpha : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is addition, $\alpha(x,y) = x + y$. In other words,

$$\mu \times \nu(E) = \iint \chi_E(x+y)\,d\mu(x)\,d\nu(y). \tag{8.47}$$

8.48 Proposition.

a. Convolution of measures is commutative and associative.

b. For any bounded Borel measurable function h,

$$\int h\,d(\mu * \nu) = \iint h(x+y)\,d\mu(x)\,d\nu(y).$$

c. $\|\mu * \nu\| \le \|\mu\|\,\|\nu\|$.

d. *If* $d\mu = f\,dm$ *and* $d\nu = g\,dm$, *then* $d(\mu * \nu) = (f * g)\,dm$; *that is, on* L^1 *the new and old definitions of convolution coincide.*

Proof. Commutativity is obvious from Fubini's theorem, as is associativity, for $\lambda * \mu * \nu$ is unambiguously defined by the formula

$$\lambda * \mu * \nu(E) = \iiint \chi_E(x + y + z)\,d\lambda(x)\,d\mu(y)\,d\nu(z).$$

Assertion (b) follows from (8.47) by the usual linearity and approximation arguments. In particular, taking $h = d|\mu * \nu|/d(\mu * \nu)$, since $|h| = 1$ we obtain

$$\|\mu * \nu\| = \int h\,d(\mu * \nu) \le \iint |h|\,d|\mu|\,d|\nu| = \|\mu\|\,\|\nu\|,$$

which proves (c). Finally, if $d\mu = f\,dm$ and $d\nu = g\,dm$, for any bounded measurable h we have

$$\begin{aligned}\int h\,d(\mu * \nu) &= \iint h(x + y)f(x)g(y)\,dx\,dy \\ &= \iint h(x)f(x - y)g(y)\,dx\,dy = \int h(x)(f * g)(x)\,dx,\end{aligned}$$

whence $d(\mu * \nu) = (f * g)\,dm$. ∎

We can also define convolutions of measures with functions in $L^p(\mathbb{R}^n, m)$, which we implicitly assume to be Borel measurable. (By Proposition 2.12, this is no restriction.)

8.49 Proposition. *If* $f \in L^p(\mathbb{R}^n)$ $(1 \le p \le \infty)$ *and* $\mu \in M(\mathbb{R}^n)$, *then the integral* $f * \mu(x) = \int f(x - y)\,d\mu(y)$ *exists for a.e.* x, $f * \mu \in L^p$, *and* $\|f * \mu\|_p \le \|f\|_p\|\mu\|$. *(Here "$L^p$" and "a.e." refer to Lebesgue measure.)*

Proof. If f and μ are nonnegative, then $f * \mu(x)$ exists (possibly being equal to ∞) for every x, and by Minkowski's inequality for integrals,

$$\|f * \mu\|_p \le \int \|f(\cdot - y)\|_p\,d\mu(y) = \|f\|_p\|\mu\|.$$

In particular, $f * \mu(x) < \infty$ for a.e. x. In the general case this argument applies to $|f|$ and $|\mu|$, and the result follows easily. ∎

In the case $p = 1$, the definition of $f * \mu$ in Proposition 8.49 coincides with the definition given earlier in which f is identified with $f\,dm$, for

$$\int_E f * \mu(x)\,dx = \iint \chi_E(x)f(x - y)\,d\mu(y)\,dx = \iint \chi_E(x + y)f(x)\,dx\,d\mu(y)$$

for any Borel set E. Thus $L^1(\mathbb{R}^n)$ is not merely a subalgebra of $M(\mathbb{R}^n)$ with respect to convolution but an ideal.

We extend the Fourier transform from $L^1(\mathbb{R}^n)$ to $M(\mathbb{R}^n)$ in the obvious way: If $\mu \in M(\mathbb{R}^n)$, $\widehat{\mu}$ is the function defined by

$$\widehat{\mu}(\xi) = \int e^{-2\pi i \xi \cdot x}\, d\mu(x).$$

(The Fourier transform on measures is sometimes called the **Fourier-Stieltjes transform.**) Since $e^{-2\pi i \xi \cdot x}$ is uniformly continuous in x, it is clear that $\widehat{\mu}$ is a bounded continuous function and that $\|\widehat{\mu}\|_u \leq \|\mu\|$. Moreover, by taking $h(x) = e^{-2\pi i \xi \cdot x}$ in Proposition 8.48b, one sees immediately that $\widehat{(\mu * \nu)} = \widehat{\mu}\widehat{\nu}$.

We conclude by giving a useful criterion for vague convergence of measures in terms of Fourier transforms.

8.50 Proposition. *Suppose that $\mu_1, \mu_2, \ldots$, and μ are in $M(\mathbb{R}^n)$. If $\|\mu_k\| \leq C < \infty$ for all k and $\widehat{\mu}_k \to \widehat{\mu}$ pointwise, then $\mu_k \to \mu$ vaguely.*

Proof. If $f \in \mathcal{S}$,then $f^\vee \in \mathcal{S}$ (Corollary 8.23), so by the Fourier inversion theorem,

$$\int f\, d\mu_k = \iint f^\vee(y) e^{-2\pi i y \cdot x}\, dy\, d\mu_k(x) = \int f^\vee(y) \widehat{\mu}_k(y)\, dy.$$

Since $f^\vee \in L^1$ and $\|\widehat{\mu}_k\|_u \leq C$, the dominated convergence theorem implies that $\int f\, d\mu_k \to \int f\, d\mu$. But $\mathcal{S}$ is dense in $C_0(\mathbb{R}^n)$ (Proposition 8.17), so by Proposition 5.17, $\int f\, d\mu_k \to \int f\, d\mu$ for all $f \in C_0(\mathbb{R}^n)$, that is, $\mu_k \to \mu$ vaguely. ∎

This result has a partial converse: If $\mu_k \to \mu$ vaguely and $\|\mu_k\| \to \|\mu\|$, then $\widehat{\mu}_k \to \widehat{\mu}$ pointwise. This follows from Exercise 26 in §7.3.

Exercises

38. Work out the analogues of the results in this section for measures on the torus $\mathbb{T}^n$.

39. If μ is a positive Borel measure on $\mathbb{T}$ with $\mu(\mathbb{T}) = 1$, then $|\widehat{\mu}(k)| < 1$ for all $k \neq 0$ unless μ is a linear combination, with positive coefficients, of the point masses at $0, \frac{1}{m}, \ldots, \frac{m-1}{m}$ for some $m \in \mathbb{N}$, in which case $\widehat{\mu}(jm) = 1$ for all $j \in \mathbb{Z}$.

40. $L^1(\mathbb{R}^n)$ is vaguely dense in $M(\mathbb{R}^n)$. (If $\mu \in M(\mathbb{R}^n)$, consider $\phi_t * \mu$ where $\{\phi_t\}_{t>0}$ is an approximate identity.)

41. Let Δ be the set of finite linear combinations of the point masses δ_x, $x \in \mathbb{R}^n$. Then Δ is vaguely dense in $M(\mathbb{R}^n)$. (If f is in the dense subset $C_c(\mathbb{R}^n)$ of $L^1(\mathbb{R}^n)$ and $g \in C_0(\mathbb{R}^n)$, approximate $\int fg$ by Riemann sums. Then use Exercise 40.)

42. A function ϕ on $\mathbb{R}^n$ that satisfies $\sum_{j,k=1}^m z_j \overline{z}_k \phi(x_j - x_k) \geq 0$ for all $z_1, \ldots, z_m \in \mathbb{C}$ and all $x_1, \ldots, x_m \in \mathbb{R}^n$, for any $m \in \mathbb{N}$, is called **positive definite**. If $\mu \in M(\mathbb{R}^n)$ is positive, then $\widehat{\mu}$ is positive definite.

8.7 APPLICATIONS TO PARTIAL DIFFERENTIAL EQUATIONS

In this section we present a few of the many applications of Fourier analysis to the theory of partial differential equations; others will be found in Chapter 9. We shall use the term **differential operator** to mean a linear partial differential operator with smooth coefficients, that is, an operator L of the form

$$Lf(x) = \sum_{|\alpha|\le m} a_\alpha(x)\partial^\alpha f(x), \qquad a_\alpha \in C^\infty.$$

If the a_α's are constants, we call L a **constant-coefficient** operator. In this case, if for all sufficiently well-behaved functions f (for example, $f \in \mathcal{S}$) we have

$$(Lf)\widehat{\ }(\xi) = \sum_{|\alpha|\le m} a_\alpha(2\pi i\xi)^\alpha \widehat{f}(\xi).$$

It is therefore convenient to write L in a slightly different form: We set $b_\alpha = (2\pi i)^{|\alpha|}a_\alpha$ and introduce the operators

$$D^\alpha = (2\pi i)^{-|\alpha|}\partial^\alpha,$$

so that

$$L = \sum_{|\alpha|\le m} b_\alpha D^\alpha, \qquad (Lf)\widehat{\ } = \sum_{|\alpha|\le m} b_\alpha \xi^\alpha \widehat{f}.$$

Thus, if P is any polynomial in n complex variables, say $P(\xi) = \sum_{|\alpha|\le m} b_\alpha\xi^\alpha$, we can form the constant-coefficient operator $P(D) = \sum_{|\alpha|\le m} b_\alpha D^\alpha$, and we then have $[P(D)f]\widehat{\ } = P\widehat{f}$. The polynomial P is called the **symbol** of the operator $P(D)$.

Clearly, one potential application of the Fourier transform is in finding solutions of the differential equation $P(D)u = f$. Indeed, application of the Fourier transform to both sides yields $\widehat{u} = P^{-1}\widehat{f}$, whence $u = (P^{-1}\widehat{f})^\vee$. Moreover, if P^{-1} is the Fourier transform of a function ϕ, we can express u directly in terms of f as $u = f * \phi$. For these calculations to make sense, however, the functions f and $P^{-1}\widehat{f}$ (or P^{-1}) must be ones to which the Fourier transform can be applied, which is a serious limitation within the theory we have developed so far. The full power of this method becomes available only when the the domain of the Fourier transform is substantially extended. We shall do this in §9.2; for the time being, we invite the reader to work out a fairly simple example in Exercise 43. (It must also be pointed out that even when this method works, $u = (P^{-1}\widehat{f})^\vee$ is far from being the only solution of $P(D)u = f$; there are others that grow too fast at infinity to be within the scope even of the extended Fourier transform.)

Let us turn to some more concrete problems. The most important of all partial differential operators is the **Laplacian**

$$\Delta = \sum_1^n \frac{\partial^2}{\partial x_j^2} = -4\pi^2 \sum_1^n D_j^2 = P(D) \text{ where } P(\xi) = -4\pi^2|\xi|^2.$$

The reason for this is that Δ is essentially the only (scalar) differential operator that is invariant under translations and rotations. (If one considers operators on vector-valued functions, there are others, such as the familiar grad, curl, and div of 3-dimensional vector analysis.) More precisely, we have:

8.51 Theorem. *A differential operator L satisfies $L(f \circ T) = (Lf) \circ T$ for all translations and rotations T iff there is a polynomial P in one variable such that $L = P(\Delta)$.*

Proof. Clearly L is translation-invariant iff L has constant coefficients, in which case $L = Q(D)$ for some polynomial Q in n variables. Moreover, since $\widehat{(Lf)} = Q\widehat{f}$ and the Fourier transform commutes with rotations, L commutes with rotations iff Q is rotation-invariant. Let $Q = \sum_0^m Q_j$ where Q_j is homogeneous of degree j; then it is easy to see that Q is rotation-invariant iff each Q_j is rotation-invariant. (Use induction on j and the fact that $Q_j(\xi) = \lim_{r\to 0} r^{-j} \sum_j^m Q_i(r\xi)$.) But this means that $Q_j(\xi)$ depends only on $|\xi|$, so $Q_j(\xi) = c_j|\xi|^j$ by homogeneity. Moreover, $|\xi|^j$ is a polynomial precisely when j is even, so $c_j = 0$ for j odd. Setting $b_k = (-4\pi^2)^{-k}c_{2k}$, then, we have $Q(\xi) = \sum b_k(-4\pi^2|\xi|^2)^k$, that is, $L = \sum b_k\Delta^k$. ∎

One of the basic boundary value problems for the Laplacian is the **Dirichlet problem**: Given an open set $\Omega \subset \mathbb{R}^n$ and a function f on its boundary $\partial\Omega$, find a function u on $\overline{\Omega}$ such that $\Delta u = 0$ on Ω and $u|\partial\Omega = f$. (This statement of the problem is deliberately a bit imprecise.) We shall solve the Dirichlet problem when Ω is a half-space.

For this purpose it will be convenient to replace n by $n+1$ and to denote the coordinates on $\mathbb{R}^{n+1}$ by $x_1, \ldots, x_n, t$. We continue to use the symbol Δ to denote the Laplacian on $\mathbb{R}^n$, and we set

$$\partial_t = \frac{\partial}{\partial t},$$

so the Laplacian on $\mathbb{R}^{n+1}$ is $\Delta + \partial_t^2$. We take the half-space Ω to be $\mathbb{R}^n \times (0, \infty)$. Thus, given a function f on $\mathbb{R}^n$, satisfying conditions to be made more precise below, we wish to find a function u on $\mathbb{R}^n \times [0, \infty)$ such that $(\Delta + \partial_t^2)u = 0$ and $u(x, 0) = f(x)$.

The idea is to apply the Fourier transform on $\mathbb{R}^n$, thus converting the partial differential equation $(\Delta + \partial_t^2)u = 0$ into the simple ordinary differential equation $(-4\pi^2|\xi|^2 + \partial_t^2)\widehat{u} = 0$. The general solution of this equation is

$$\widehat{u}(\xi, t) = c_1(\xi)e^{-2\pi t|\xi|} + c_2(\xi)e^{2\pi t|\xi|}, \tag{8.52}$$

and we require that $\widehat{u}(\xi, 0) = \widehat{f}(\xi)$. We therefore obtain a solution to our problem by taking $c_1(\xi) = \widehat{f}(\xi)$, $c_2(\xi) = 0$ (more about the reasons for this choice below); this gives $\widehat{u}(\xi, t) = \widehat{f}(\xi)e^{-2\pi t|\xi|}$, or $u(x, t) = (f * P_t)(x)$ where $P_t = (e^{-2\pi t|\xi|})^\vee$ is the Poisson kernel introduced in §8.4. As we calculated in Exercise 26,

$$P_t(x) = \frac{\Gamma(\frac{1}{2}(n+1))}{\pi^{(n+1)/2}} \frac{t}{\left(t^2 + |x|^2\right)^{-(n+1)/2}}.$$

So far this is all formal, since we have not specified conditions on f to ensure that these manipulations are justified. We now give a precise result.

8.53 Theorem. *Suppose $f \in L^p(\mathbb{R}^n)$ $(1 \le p \le \infty)$. Then the function $u(x,t) = (f * P_t)(x)$ satisfies $(\Delta + \partial_t^2)u = 0$ on $\mathbb{R}^n \times (0,\infty)$, and $\lim_{t\to 0} u(x,t) = f(x)$ for a.e. x and for every x at which f is continuous. Moreover, $\lim_{t\to 0} \|u(\cdot,t) - f\|_p = 0$ provided $p < \infty$.*

Proof. P_t and all of its derivatives are in $L^q(\mathbb{R}^n)$ for $1 \le q \le \infty$, since a rough calculation shows that $|\partial_x^\alpha P_t(x)| \le C_\alpha |x|^{-n-1-|\alpha|}$ and $|\partial_t^j P_t(x)| \le C_j |x|^{-n-1}$ for large x. Also, $(\Delta + \partial_t^2)P_t(x) = 0$, as can be verified by direct calculation or (more easily) by taking the Fourier transform. Hence $f * P_t$ is well defined and

$$(\Delta + \partial_t^2)(f * P_t) = f * (\Delta + \partial_t^2)P_t = 0.$$

Since $P_t(x) = t^{-n}P_1(t^{-1}x)$ and $\int P_1(x)\,dx = \widehat{P}_1(0) = 1$, the remaining assertions follow from Theorems 8.14 and 8.15. ∎

The function $u(x,t) = (f * P_t)(x)$ is not the only one satisfying the conclusions of Theorem 8.53; for example, $v(x,t) = u(x,t) + ct$ also works, for any $c \in \mathbb{C}$. For $f \in L^1$, we could also obtain a large family of solutions by taking c_2 in (8.52) to be an arbitrary function in C_c^∞ and $c_1 = \widehat{f} - c_2$. (But there is no nice convolution formula for the resulting function u, because $e^{2\pi t|\xi|}$ is not the Fourier transform of a function or even a distribution.) The solution $u(x,t) = (f * P_t)(x)$ is distinguished, however, by its regularity at infinity; for example, it can be shown that if $f \in BC(\mathbb{R}^n)$, then u is the unique solution in $BC(\mathbb{R}^n \times [0,\infty))$.

The same idea can be used to solve the **heat equation**

$$(\partial_t - \Delta)u = 0$$

on $\mathbb{R}^n \times (0,\infty)$ subject to the initial condition $u(x,0) = f(x)$. (Physical interpretation: $u(x,t)$ represents the temperature at position x and time t in a homogeneous isotropic medium, given that the temperature at time 0 is $f(x)$.) Indeed, Fourier transformation leads to the ordinary differential equation $(\partial_t + 4\pi^2|\xi|^2)\widehat{u} = 0$ with initial condition $\widehat{u}(\xi,0) = \widehat{f}(\xi)$. The unique solution of the latter problem is $\widehat{u}(\xi,t) = \widehat{f}(\xi)e^{-4\pi^2 t|\xi|^2}$. In view of Proposition 8.24, this yields

$$u(x,t) = f * G_t(x), \qquad G_t(x) = (4\pi t)^{-n/2}e^{-|x|^2/4t}.$$

Here we have $G_t(x) = t^{-n/2}G_1(t^{-1/2}x)$, so after the change of variable $s = \sqrt{t}$, Theorems 8.14 and 8.15 apply again, and we obtain an exact analogue of Theorem 8.53 for the initial value problem $(\partial_t - \Delta)u = 0$, $u(x,0) = f(x)$. Actually, in the present case the hypotheses on f can be relaxed considerably because $G_t \in \mathcal{S}$; see Exercise 44.

Another fundamental equation of mathematical physics is the **wave equation**

$$(\partial_t^2 - \Delta)u = 0.$$

(Physical interpretation: $u(x,t)$ is the amplitude at position x and time t of a wave traveling in a homogeneous isotropic medium, with units chosen so that the speed of propagation is 1.) Here it is appropriate to specify both $u(x,0)$ and $\partial_t u(x,0)$:

$$(8.54) \qquad (\partial_t^2 - \Delta)u = 0, \qquad u(x,0) = f(x), \qquad \partial_t u(x,0) = g(x).$$

After applying the Fourier transform, we obtain

$$(\partial_t^2 + 4\pi^2|\xi|^2)\widehat{u}(\xi,t) = 0, \qquad \widehat{u}(\xi,0) = \widehat{f}(\xi), \qquad \partial_t\widehat{u}(\xi,0) = \widehat{g}(\xi),$$

the solution to which is

$$(8.55) \qquad \widehat{u}(\xi,t) = (\cos 2\pi t|\xi|)\widehat{f}(\xi) + \frac{\sin 2\pi t|\xi|}{2\pi|\xi|}\widehat{g}(\xi).$$

Since

$$\cos 2\pi t|\xi| = \frac{\partial}{\partial t}\left[\frac{\sin 2\pi t|\xi|}{2\pi|\xi|}\right],$$

it follows that

$$u(x,t) = f * \partial_t W_t(x) + g * W_t(x), \text{ where } W_t = \left[\frac{\sin 2\pi t|\xi|}{2\pi|\xi|}\right]^\vee.$$

But here there is a problem: $(2\pi|\xi|)^{-1}\sin 2\pi t|\xi|$ is the Fourier transform of a function only when $n \le 2$ and the Fourier transform of a measure only when $n \le 3$; for these cases the resulting solution of the wave equation is worked out in Exercises 45–47. To carry out this analysis in higher dimensions requires the theory of distributions, which we shall examine in Chapter 9. (We shall not, however, derive the explicit formula for W_t, which becomes increasingly complicated as n increases.)

Exercises

43. Let $\phi(x) = e^{-|x|/2}$ on $\mathbb{R}$. Use the Fourier transform to derive the solution $u = f * \phi$ of the differential equation $u - u'' = f$, and then check directly that it works. What hypotheses are needed on f?

44. Let $G_t(x) = (4\pi t)^{-n/2}e^{-|x|^2/4t}$, and suppose that $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ satisfies $|f(x)| \le C_\epsilon e^{\epsilon|x|^2}$ for every $\epsilon > 0$. Then $u(x,t) = f * G_t(x)$ is well defined for all $x \in \mathbb{R}^n$ and $t > 0$; $(\partial_t - \Delta)u = 0$ on $\mathbb{R}^n \times (0,\infty)$; and $\lim_{t\to 0} u(x,t) = f(x)$ for a.e. x and for every x at which f is continuous. (To show $u(x,t) \to f(x)$ a.e. on a bounded open set V, write $f = \phi f + (1-\phi)f$ where $\phi \in C_c$ and $\phi = 1$ on V, and show that $[(1-\phi)f] * G_t \to 0$ on V.)

45. Let $n = 1$. Use (8.55) and Exercise 15a to derive d'Alembert's solution to the initial value problem (8.54):

$$u(x,t) = \frac{1}{2}[f(x+t) + f(x-t)] + \frac{1}{2}\int_{x-t}^{x+t} g(s)\,ds.$$

Under what conditions on f and g does this formula actually give a solution?

46. Let $n = 3$, and let σ_t denote surface measure on the sphere $|x| = t$. Then

$$\frac{\sin 2\pi t|\xi|}{2\pi|\xi|} = (4\pi t)^{-1}\widehat{\sigma}_t(\xi).$$

(See Exercise 22d.) What is the resulting solution of the initial value problem (8.54), expressed in terms of convolutions? What conditions on f and g ensure its validity?

47. Let $n = 2$. If $\xi \in \mathbb{R}^2$, let $\widetilde{\xi} = (\xi, 0) \in \mathbb{R}^3$. Rewrite the result of Exercise 46,

$$\frac{\sin 2\pi t|\widetilde{\xi}|}{2\pi|\widetilde{\xi}|} = \frac{1}{4\pi t}\int_{|x|=t} e^{-2\pi i\widetilde{\xi}\cdot x}\, d\sigma_t(x),$$

in terms of an integral over the disc $D_t = \{y : |y| \le t\}$ in $\mathbb{R}^2$ by projecting the upper and lower hemispheres of the sphere $|x| = t$ in $\mathbb{R}^3$ onto the equatorial plane. Conclude that $(2\pi|\xi|)^{-1}\sin 2\pi t|\xi|$ is the Fourier transform of

$$W_t(x) = (2\pi)^{-1}(t^2 - |x|^2)^{-1/2}\chi_{D_t}(x),$$

and write out the resulting solution of the initial value problem (8.54).

48. Solve the following initial value problems in terms of Fourier series, where f, g, and $u(\cdot, t)$ are periodic functions on $\mathbb{R}$:

a. $(\partial_t^2 + \partial_x^2)u = 0$, $u(x, 0) = f(x)$. (Cf. the discussion of Abel means in §8.4.)
b. $(\partial_t - \partial_x^2)u = 0$, $u(x, 0) = f(x)$.
c. $(\partial_t^2 - \partial_x^2)u = 0$, $u(x, 0) = f(x)$, $\partial_t u(x, 0) = g(x)$.

49. In this exercise we discuss heat flow on an interval.

a. Solve $(\partial_t - \partial_x^2)u = 0$ on $(a, b) \times (0, \infty)$ with boundary conditions $u(x, 0) = f(x)$ for $x \in (a, b)$, $u(a, t) = u(b, t) = 0$ for $t > 0$, in terms of Fourier series. (This describes heat flow on (a, b) when the endpoints are held at a constant temperature. It suffices to assume $a = 0$, $b = \frac{1}{2}$; extend f to $\mathbb{R}$ by requiring f to be odd and periodic, and use Exercise 48b.)

b. Solve the same problem with the condition $u(a, t) = u(b, t) = 0$ replaced by $\partial_x u(a, t) = \partial_x u(b, t) = 0$. (This describes heat flow on (a, b) when the endpoints are insulated. This time, extend f to be even and periodic.)

50. Solve $(\partial_t^2 - \partial_x^2)u = 0$ on $(a, b) \times (0, \infty)$ with boundary conditions $u(x, 0) = f(x)$ and $\partial_t u(x, 0) = g(x)$ for $x \in (a, b)$, $u(a, t) = u(b, t) = 0$ for $t > 0$, in terms of Fourier series by the method of Exercise 49a. (This problem describes the motion of a vibrating string that is fixed at the endpoints. It can also be solved by extending f to be odd and periodic and using Exercise 45. That form of the solution tells you what you see when you look at a vibrating string; this one tells you what you hear when you listen to it.)

8.8 NOTES AND REFERENCES

The scope of Fourier analysis is much wider than we have been able to indicate in this chapter. Dym and McKean [36] gives a more comprehensive treatment with many interesting applications. Also recommended are Körner's delightful book [87], which discusses various aspects of classical Fourier analysis and their role in science, and the excellent collection of expository articles edited by Ash [7], which gives a broader view of the mathematical ramifications of the subject. On the more advanced level, the reader should consult Zygmund [167] for the classical theory and Stein [140], [141] and Stein and Weiss [142] for some of the more recent developments.

§8.1: The formulas given in most calculus books for the remainder term $R_k(x) = f(x) - P_k(x)$ in Taylor's formula (where P_k is the Taylor polynomial of degree k) require f to possess derivatives of order $k+1$, but this is not really necessary. The version of Taylor's theorem stated in the text is derived in Folland [45].

§8.3: Trigonometric series and integrals have a very long history, but modern Fourier analysis only became possible after the invention of the Lebesgue integral. When that tool became available, the L^2 theory was quickly established: the Riesz-Fischer theorem [44], [114] for Fourier series (essentially Theorem 8.20), and the Plancherel theorem [110] for Fourier integrals. Since then the subject has developed in many directions.

There is no universal agreement on where to put the factors of 2π in the definition of the Fourier transform. Other common conventions are

$$\mathcal{F}_1 f(\xi) = \int e^{-i\xi\cdot x} f(x)\,dx, \qquad \mathcal{F}_2 f(\xi) = (2\pi)^{-n/2} \int e^{-i\xi\cdot x}\,dx,$$

whose inverse transforms are

$$\mathcal{F}_1^{-1} g(x) = (2\pi)^{-n} \int e^{i\xi\cdot x} g(\xi)\,d\xi, \qquad \mathcal{F}_2^{-1} g(\xi) = (2\pi)^{-n/2} \int e^{i\xi\cdot x} g(\xi)\,d\xi.$$

$\mathcal{F}_1$ has the disadvantage of not being unitary ($\|\mathcal{F}_1 f\|_2 = (2\pi)^{n/2}\|f\|_2$), whereas $\mathcal{F}_2$ is unitary but does not convert convolution into multiplication ($\mathcal{F}_2(f * g) = (2\pi)^{n/2}(\mathcal{F}_2 f)(\mathcal{F}_2 g)$). To make both L^2 norms and convolutions come out right, one can either put the 2π's in the exponent, as we have done, or omit them from the exponent but replace Lebesgue measure dx by $(2\pi)^{-n/2}\,dx$ in defining both the Fourier transform (as in $\mathcal{F}_2$) and convolutions.

The Hausdorff-Young inequality $\|\widehat{f}\|_q \le \|f\|_p$ ($1 \le p \le 2$, $p^{-1} + q^{-1} = 1$) is sharp on $\mathbb{T}^n$, since equality holds when f is a constant function; but on $\mathbb{R}^n$ the optimal result, a deep theorem of Beckner [14], is that $\|\widehat{f}\|_q \le p^{n/2p} q^{-n/2q} \|f\|_p$.

One of the fundamental qualitative features of the Fourier transform is the fact that, roughly speaking, a nonzero function and its Fourier transform cannot both be sharply localized, that is, they cannot both be negligibly small outside of small sets. This general principle has a number of different precise formulations, two of which are derived in Exercises 18 and 19; see Folland and Sitaram [50] for a comprehensive discussion.

A nice complex-variable proof of the fact that the Fourier transform is injective on L^1 can be found in Newman [106].

§§8.4–5: The theory of convergence of one-dimensional Fourier series really began (as mentioned in the text) with Dirichlet's theorem in 1829. The first construction of a continuous function whose Fourier series does not converge pointwise was obtained by du Bois Reymond in 1876, and the fact that the Fourier series of a continuous function f is uniformly Cesàro summable to f was proved by Fejér in 1904. In 1926 Kolmogorov produced an $f \in L^1(\mathbb{T})$ such that $\{S_m f(x)\}$ diverges at *every* x; on the other hand, in 1927 M. Riesz proved that for $1 < p < \infty$, $\|S_m f - f\|_p \to 0$ for every $f \in L^p(\mathbb{T})$. The culmination of this subject is the theorem of Carleson (1966, for $p = 2$) and Hunt (1967, for general p) that if $f \in L^p(\mathbb{T})$ where $p > 1$, then $S_m f \to f$ almost everywhere.

For more information, see Zygmund [167], the articles by Zygmund and Hunt in Ash [7], and Fefferman [42]. Also see Hewitt and Hewitt [74] for an interesting historical discussion of the Gibbs phenomenon.

Convergence of Fourier series in n variables is an even trickier subject. In the first place, one must decide what one means by a partial sum of a series indexed by $\mathbb{Z}^n$. It is a straightforward consequence of the Riesz and Carleson-Hunt theorems that if $f \in L^p(\mathbb{T}^n)$ with $p > 1$, the "cubical partial sums"

$$S_m^c f(x) = \sum_{\|\kappa\| \le m} \widehat{f}(\kappa) e^{2\pi i \kappa \cdot x} \qquad (\|\kappa\| = \max(|\kappa_1|, \dots, |\kappa_n|))$$

converge to f a.e. and (if $p < \infty$) in the L^p norm. On the other hand, C. Fefferman proved the rather shocking result that for the "spherical partial sums"

$$S_r f(x) = \sum_{|\kappa| < r} \widehat{f}(\kappa) e^{2\pi i \kappa \cdot x} \qquad \Big(|\kappa|^2 = \sum_1^n \kappa_j^2\Big),$$

the convergence $\lim_{r\to\infty} \|S_r f - f\|_p = 0$ holds for all $f \in L^p$ *only* when $p = 2$, if $n > 1$. Of course, one can consider modifications of the spherical partial sums in the hope of obtaining positive results; the most intensively studied of these are the **Bochner-Riesz means**

$$\sigma_r^\alpha f(x) = \sum_{|\kappa| < r} \big(1 - |r^{-1}\kappa|^2\big)^\alpha \widehat{f}(\kappa) e^{2\pi i \kappa \cdot x}$$

obtained by taking $\Phi(\xi) = \big[\max(1 - |\xi|^2), 0\big]^\alpha$ in Theorem 8.36. (When $n = 1$, $\sigma_{m+1}^1 f$ is essentially equivalent to the Cesàro mean $\sigma_m f$.) These Φ's satisfy the hypotheses of Theorem 8.36 when $\alpha > \frac{1}{2}(n-1)$, and some positive results are also known for smaller values of α.

Davis and Chang [30] is a good source for all of this material; see also Stein and Weiss [142] and Ash [7].

§8.7: The solution of the initial value problem for the wave equation in arbitrary dimensions can be found in Folland [48]; see also Folland [49]. Further applications of Fourier analysis to differential equations can be found in Folland [46], [48], Körner [87], and Taylor [147].

9

Elements of Distribution Theory

At least as far back as Heaviside in the 1890s, engineers and physicists have found it convenient to consider mathematical objects which, roughly speaking, resemble functions but are more singular than functions. Despite their evident efficacy, such objects were at first received with disdain and perplexity by the pure mathematicians, and one of the most important conceptual advances in modern analysis is the development of methods for dealing with them in a rigorous and systematic way. The method that has proved to be most generally useful is Laurent Schwartz's theory of distributions, based on the idea of linear functionals on test functions. For some purposes, however, it is preferable to use a theory more closely tied to L^2 on which the power of Hilbert space methods and the Plancherel theorem can be brought to bear, namely, the (L^2) Sobolev spaces. In this chapter we present the fundamentals of these theories and some of their applications.

9.1 DISTRIBUTIONS

In order to find a fruitful generalization of the notion of function on $\mathbb{R}^n$, it is necessary to get away from the classical definition of function as a map that assigns to each point of $\mathbb{R}^n$ a numerical value. We have already done this to some extent in the theory of L^p spaces: If $f \in L^p$, the pointwise values $f(x)$ are of little significance for the behavior of f as an element of L^p, as f can be modified on any set of measure zero without affecting the latter. What is more to the point is the family of integrals $\int f\phi$ as ϕ ranges over the dual space L^q. Indeed, we know that f is completely determined by its action as a linear functional on L^q; on the other hand, if we take

$\phi = \phi_r = m(B_r)^{-1}\chi_{B_r}$ where B_r is the ball of radius r about x, by the Lebesgue differentiation theorem we can recover the pointwise value $f(x)$, for almost every x, as $\lim_{r\to 0}\int f\phi_r$. Thus, we lose nothing by thinking of f as a linear map from $L^q(\mathbb{R}^n)$ to $\mathbb{C}$ rather than as a map from $\mathbb{R}^n$ to $\mathbb{C}$.

Let us modify this idea by allowing f to be merely locally integrable on $\mathbb{R}^n$ but requiring ϕ to lie in C_c^∞. Again the map $\phi \mapsto \int f\phi$ is a well-defined linear functional on C_c^∞, and again the pointwise values of f can be recovered a.e. from it, by an easy extension of Theorem 8.15. But there are many linear functionals on C_c^∞ that are *not* of the form $\phi \mapsto \int f\phi$, and these — subject to a mild continuity condition to be specified below — will be our "generalized functions."

Recall that for $E \subset \mathbb{R}^n$ we have defined $C_c^\infty(E)$ to be the set of all C^∞ functions whose support is compact and contained in E. If $U \subset \mathbb{R}^n$ is open, $C_c^\infty(U)$ is the union of the spaces $C_c^\infty(K)$ as K ranges over all compact subsets of U. Each of the latter is a Fréchet space with the topology defined by the norms

$$\phi \mapsto \|\partial^\alpha \phi\|_u \qquad (\alpha \in \{0, 1, 2, \ldots\}^n),$$

in which a sequence $\{\phi_j\}$ converges to ϕ iff $\partial^\alpha \phi_j \to \partial^\alpha \phi$ uniformly for all α. (The completeness of $C_c^\infty(K)$ is easily proved by the argument in Exercise 9 in §5.1.) With this in mind, we make the following definitions, in which U is an open subset of $\mathbb{R}^n$:

i. A sequence $\{\phi_j\}$ in $C_c^\infty(U)$ **converges in C_c^∞** to ϕ if $\{\phi_j\} \subset C_c^\infty(K)$ for some compact set $K \subset U$ and $\phi_j \to \phi$ in the topology of $C_c^\infty(K)$, that is, $\partial^\alpha \phi_j \to \partial^\alpha \phi$ uniformly for all α.

ii. If $\mathcal{X}$ is a locally convex topological vector space and $T : C_c^\infty(U) \to \mathcal{X}$ is a linear map, T is **continuous** if $T|C_c^\infty(K)$ is continuous for each compact $K \subset U$, that is, if $T\phi_j \to T\phi$ whenever $\phi_j \to \phi$ in $C_c^\infty(K)$ and $K \subset U$ is compact.

iii. A linear map $T : C_c^\infty(U) \to C_c^\infty(U')$ is **continuous** if for each compact $K \subset U$ there is a compact $K' \subset U'$ such that $T(C_c^\infty(K)) \subset C_c^\infty(K')$, and T is continuous from $C_c^\infty(K)$ to $C_c^\infty(K')$.

iv. A **distribution** on U is a continuous linear functional on $C_c^\infty(U)$. The space of all distributions on U is denoted by $\mathcal{D}'(U)$, and we set $\mathcal{D}' = \mathcal{D}'(\mathbb{R}^n)$. We impose the weak* topology on $\mathcal{D}'(U)$, that is, the topology of pointwise convergence on $C_c^\infty(U)$.

Two remarks: First, the standard notation $\mathcal{D}'$ for the space of distributions comes from Schwartz's notation $\mathcal{D}$ for C_c^∞, which is also quite common. Second, there is a locally convex topology on C_c^∞ with respect to which sequential convergence in C_c^∞ is given by (i) and continuity of linear maps $T : C_c^\infty \to \mathcal{X}$ and $T : C_c^\infty \to C_c^\infty$ is given by (ii) and (iii). However, its definition is rather complicated and of little importance for the elementary theory of distributions, so we shall omit it.

Here are some examples of distributions; more will be presented below.

- Every $f \in L^1_{\text{loc}}(U)$ — that is, every function f on U such that $\int_K |f| < \infty$ for every compact $K \subset U$ — defines a distribution on U, namely, the functional $\phi \to \int f\phi$, and two functions define the same distribution precisely when they are equal a.e.

- Every Radon measure μ on U defines a distribution by $\phi \mapsto \int \phi \, d\mu$.

- If $x_0 \in U$ and α is a multi-index, the map $\phi \mapsto \partial^\alpha \phi(x_0)$ is a distribution that does not arise from a function; it arises from a measure μ precisely when $\alpha = 0$, in which case μ is the point mass at x_0.

If $f \in L^1_{\text{loc}}(U)$, we denote the distribution $\phi \mapsto \int f\phi$ also by f, thereby identifying $L^1_{\text{loc}}(U)$ with a subspace of $\mathcal{D}'(U)$. In order to avoid notational confusion between $f(x)$ and $f(\phi) = \int f\phi$, we adopt a different notation for the pairing between $C_c^\infty(U)$ and $\mathcal{D}'(U)$. Namely, if $F \in \mathcal{D}'(U)$ and $\phi \in C_c^\infty(U)$, the value of F at ϕ will be denoted by $\langle F, \phi \rangle$. Observe that the pairing $\langle \cdot, \cdot \rangle$ between $\mathcal{D}'(U)$ and $C_c^\infty(U)$ is linear in each variable; this conflicts with our earlier notation for inner products but will cause no serious confusion. If μ is a measure, we shall also identify μ with the distribution $\phi \mapsto \int \phi \, d\mu$

Sometimes it is convenient to pretend that a distribution F is a function even when it really is not, and to write $\int F(x)\phi(x)\, dx$ instead of $\langle F, \phi \rangle$. This is the case especially when the explicit presence of the variable x is notationally helpful.

At this point we set forth two pieces of notation that will be used consistently throughout this chapter. First, we shall use a tilde to denote the reflection of a function in the origin:

$$\widetilde{\phi}(x) = \phi(-x).$$

Second, we denote the point mass at the origin, which plays a central role in distribution theory, by δ:

$$\langle \delta, \phi \rangle = \phi(0).$$

As an illustration of the role of δ and the notion of convergence in $\mathcal{D}'$, we record the following important corollary of Theorem 8.14:

9.1 Proposition. *Suppose that $f \in L^1(\mathbb{R}^n)$ and $\int f = a$, and for $t > 0$ let $f_t(x) = t^{-n} f(t^{-1}x)$. Then $f_t \to a\delta$ in $\mathcal{D}'$ as $t \to 0$.*

Proof. If $\phi \in C_c^\infty$, by Theorem 8.14 we have

$$\langle f_t, \phi \rangle = \int f_t \phi = f_t * \widetilde{\phi}(0) \to a\widetilde{\phi}(0) = a\phi(0) = a\langle \delta, \phi \rangle.$$

∎

Although it does not make sense to say that two distributions F and G in $\mathcal{D}'(U)$ agree at a single point, it does make sense to say that they agree on an open set $V \subset U$; namely, $F = G$ on V iff $\langle F, \phi \rangle = \langle G, \phi \rangle$ for all $\phi \in C_c^\infty(V)$. (Clearly, if F and G are continuous functions, this condition is equivalent to the pointwise equality of F and G on V; if F and G are merely locally integrable, it means that $F = G$ a.e.

on V.) Since a function in $C_c^\infty(V_1 \cup V_2)$ need not be supported in either V_1 or V_2, it is not immediately obvious that if $F = G$ on V_1 and on V_2 then $F = G$ on $V_1 \cup V_2$. However, it is true:

9.2 Proposition. *Let $\{V_\alpha\}$ be a collection of open subsets of U and let $V = \bigcup_\alpha V_\alpha$. If $F, G \in \mathcal{D}'(U)$ and $F = G$ on each V_α, then $F = G$ on V.*

Proof. If $\phi \in C_c^\infty(V)$, there exist $\alpha_1, \dots \alpha_m$ such that $\operatorname{supp}\phi \subset \bigcup_1^m V_{\alpha_j}$. Pick $\psi_1, \dots, \psi_m \in C_c^\infty$ such that $\operatorname{supp}(\psi_j) \subset V_{\alpha_j}$ and $\sum_1^m \psi_j = 1$ on $\operatorname{supp}(\phi)$. (That this can be done is the C^∞ analogue of Proposition 4.41, proved in the same way as that result by using the C^∞ Urysohn lemma.) Then $\langle F, \phi\rangle = \sum\langle F, \psi_j\phi\rangle = \sum\langle G, \psi_j\phi\rangle = \langle G, \phi\rangle$. ∎

According to Proposition 9.2, if $F \in \mathcal{D}'(U)$, there is a maximal open subset of U on which $F = 0$, namely the union of all the open subsets on which $F = 0$. Its complement in U is called the **support** of F.

There is a general procedure for extending various linear operations from functions to distributions. Suppose that U and V are open sets in $\mathbb{R}^n$, and T is a linear map from some subspace $\mathcal{X}$ of $L^1_{\text{loc}}(U)$ into $L^1_{\text{loc}}(V)$. Suppose that there is another linear map $T' : C_c^\infty(V) \to C_c^\infty(U)$ such that

$$\int (Tf)\phi = \int f(T'\phi) \qquad \big(f \in \mathcal{X},\ \phi \in C_c^\infty(V)\big).$$

Suppose also that T' is continuous in the sense defined above. Then T can be extended to a map from $\mathcal{D}'(U)$ to $\mathcal{D}'(V)$, still denoted by T, by

$$\langle TF, \phi\rangle = \langle F, T'\phi\rangle \qquad \big(F \in \mathcal{D}'(U),\ \phi \in C_c^\infty(V)\big).$$

The intervention of the continuous map T' guarantees that the original T, as well as its extension to distributions, is continuous with respect to the weak* topology on distributions: If $F_\alpha \to F \in \mathcal{D}'(U)$, then $TF_\alpha \to TF$ in $\mathcal{D}'(V)$.

Here are the most important instances of this procedure. In each of them, U is an open set in $\mathbb{R}^n$, and the continuity of T' is an easy exercise that we leave to the reader.

i. (Differentiation) Let $Tf = \partial^\alpha f$, defined on $C^{|\alpha|}(U)$. If $\phi \in C_c^\infty(U)$, integration by parts gives $\int(\partial^\alpha f)\phi = (-1)^{|\alpha|}\int f(\partial^\alpha\phi)$; there are no boundary terms since ϕ has compact support. Hence $T' = (-1)^{|\alpha|}T|C_c^\infty(U)$, and we can define the **derivative** $\partial^\alpha F \in \mathcal{D}'(U)$ of any $F \in \mathcal{D}'(U)$ by

$$\langle \partial^\alpha F, \phi\rangle = (-1)^{|\alpha|}\langle F, \partial^\alpha\phi\rangle.$$

Notice, in particular, that by this procedure we can define derivatives of arbitrary locally integrable functions even when they are not differentiable in the classical sense; this is one of the main reasons for the power of distribution theory. We shall discuss this matter in more detail below.

ii. (Multiplication by Smooth Functions) Given $\psi \in C^\infty(U)$, define $Tf = \psi f$. Then $T' = T|C_c^\infty(U)$, so we can define the product $\psi F \in \mathcal{D}'(U)$ for $F \in \mathcal{D}'(U)$ by

$$\langle \psi F, \phi \rangle = \langle F, \psi\phi \rangle.$$

Moreover, if $\psi \in C_c^\infty(U)$, this formula makes sense for any $\phi \in C_c^\infty(\mathbb{R}^n)$ and defines ψF as a distribution on $\mathbb{R}^n$.

iii. (Translation) Given $y \in \mathbb{R}^n$, let $V = U + y = \{x + y : x \in U\}$ and let $T = \tau_y$. (Recall that we have defined $\tau_y f(x) = f(x - y)$.) Since $\int f(x-y)\phi(x)\,dx = \int f(x)\phi(x+y)\,dx$, we have $T' = \tau_{-y}|C_c^\infty(U+y)$. For $F \in \mathcal{D}'(U)$, then, we define the translated distribution $\tau_y F \in \mathcal{D}'(U+y)$ by

$$\langle \tau_y F, \phi \rangle = \langle F, \tau_{-y}\phi \rangle.$$

For example, the point mass at y is $\tau_y \delta$.

iv. (Composition with Linear Maps) Given an invertible linear transformation S of $\mathbb{R}^n$, let $V = S^{-1}(U)$ and let $Tf = f \circ S$. Then $T'\phi = |\det S|^{-1}\phi \circ S^{-1}$ by Theorem 2.44, so for $F \in \mathcal{D}'(U)$ we define $F \circ S \in \mathcal{D}'(S^{-1}(U))$ by

$$\langle F \circ S, \phi \rangle = |\det S|^{-1}\langle F, \phi \circ S^{-1} \rangle.$$

In particular, for $Sx = -x$ we have $f \circ S = \widetilde{f}$, $S^{-1} = S$, and $|\det S| = 1$, so we define the reflection of a distribution in the origin by

$$\langle \widetilde{F}, \phi \rangle = \langle F, \widetilde{\phi} \rangle.$$

v. (Convolution, First Method) Given $\psi \in C_c^\infty$, let

$$V = \{x : x - y \in U \text{ for } y \in \operatorname{supp}(\psi)\}.$$

(V is open but may be empty.) If $f \in L^1_{\text{loc}}(U)$, the integral

$$f * \psi(x) = \int f(x-y)\psi(y)\,dy = \int f(y)\psi(x-y)\,dy = \int f(\tau_x \widetilde{\psi})$$

is well defined for all $x \in V$. The same definition works for $F \in \mathcal{D}'(U)$: the **convolution** $F * \psi$ is the function defined on V by

$$F * \psi(x) = \langle F, \tau_x \widetilde{\psi} \rangle.$$

Since $\tau_x \widetilde{\psi} \to \tau_{x_0}\widetilde{\psi}$ in C_c^∞ as $x \to x_0$, $F * \psi$ is a continuous function (actually C^∞, as we shall soon see) on V. As an example, for any $\psi \in C_c^\infty$ we have

$$\delta * \psi(x) = \langle \delta, \tau_x \widetilde{\psi} \rangle = \tau_x \widetilde{\psi}(0) = \psi(x),$$

so δ is the multiplicative identity for convolution.

vi. (Convolution, Second Method) Let ψ, $\widetilde{\psi}$, and V be as in (v). If $f \in L^1_{\text{loc}}(U)$ and $\phi \in C_c^\infty(V)$, we have

$$\int (f * \psi)\phi = \iint f(y)\psi(x-y)\phi(y)\,dy\,dx = \int f(\phi * \widetilde{\psi}).$$

That is, if $Tf = f * \psi$, then T maps $L^1_{\text{loc}}(U)$ into $L^1_{\text{loc}}(V)$ and $T'\phi = \phi * \widetilde{\psi}$. For $F \in \mathcal{D}'(U)$, we can therefore define $F * \psi$ as a distribution on V by

$$\langle F * \psi, \phi\rangle = \langle F, \phi * \widetilde{\psi}\rangle.$$

Again, we have $\delta * \psi = \psi$, for

$$\langle \delta * \psi, \phi\rangle = \langle \delta, \phi * \widetilde{\psi}\rangle = \phi * \widetilde{\psi}(0) = \int \phi(x)\psi(x)\,dx = \langle \psi, \phi\rangle.$$

The definitions of convolution in (v) and (vi) are actually equivalent, as we shall now show.

9.3 Proposition. *Suppose that U is open in $\mathbb{R}^n$ and $\psi \in C_c^\infty$. Let $V = \{x : x - y \in U$ for $y \in \operatorname{supp}(\psi)\}$. For $F \in \mathcal{D}'(U)$ and $x \in V$ let $F * \psi(x) = \langle F, \tau_x\widetilde{\psi}\rangle$. Then*

a. $F * \psi \in C^\infty(V)$.

b. $\partial^\alpha(F * \psi) = (\partial^\alpha F) * \psi = F * (\partial^\alpha \psi)$.

c. For any $\phi \in C_c^\infty(V)$, $\int (F * \psi)\phi = \langle F, \phi * \widetilde{\psi}\rangle$.

Proof. Let $e_1, \ldots, e_n$ be the standard basis for $\mathbb{R}^n$. If $x \in V$, there exists $t_0 > 0$ such that $x + te_j \in U$ for $|t| < t_0$, and it is easily verified that

$$t^{-1}(\tau_{x+te_j}\widetilde{\psi} - \tau_x\widetilde{\psi}) \to \tau_x\widetilde{\partial_j\psi} \text{ in } C_c^\infty(U) \text{ as } t \to 0.$$

It follows that $\partial_j(F * \psi)(x)$ exists and equals $F * \partial_j\psi(x)$, so by induction, $F * \psi \in C^\infty(V)$ and $\partial^\alpha(F * \psi) = F * \partial^\alpha\psi$. Moreover, since $\partial^\alpha\widetilde{\psi} = (-1)^{|\alpha|}\widetilde{\partial^\alpha\psi}$ and $\partial^\alpha\tau_x = \tau_x\partial^\alpha$, we have

$$(\partial^\alpha F) * \psi(x) = \langle \partial^\alpha F, \tau_x\widetilde{\psi}\rangle = (-1)^{|\alpha|}\langle F, \partial^\alpha\tau_x\widetilde{\psi}\rangle = \langle F, \tau_x\widetilde{\partial^\alpha\psi}\rangle = F * (\partial^\alpha\psi)(x).$$

Next, if $\phi \in C_c^\infty(V)$, we have

$$\phi * \widetilde{\psi}(x) = \int \phi(y)\psi(y-x)\,dy = \int \phi(y)\tau_y\widetilde{\psi}(x)\,dy.$$

The integrand here is continuous and supported in a compact subset of U, so the integral can be approximated by Riemann sums. That is, for each (large) $m \in \mathbb{N}$ we can approximate $\operatorname{supp}(\phi)$ by a union of cubes of side length 2^{-m} (and volume 2^{-nm}) centered at points $y_1^m, \ldots, y_{k(m)}^m \in \operatorname{supp}(\phi)$; then the corresponding Riemann sums $S^m = 2^{-nm}\sum_j \phi(y_j^m)\tau_{y_j^m}\widetilde{\psi}$ are supported in a common

compact subset of U and converge uniformly to $\phi * \widetilde{\psi}$ as $m \to \infty$. Likewise, $\partial^\alpha S^m = 2^{-nm} \sum_j \phi(y_j^m) \tau_{y_j^m} \partial^\alpha \widetilde{\psi}$ converges uniformly to $\phi * \partial^\alpha \widetilde{\psi} = \partial^\alpha(\phi * \widetilde{\psi})$, so $S^m \to \phi * \widetilde{\psi}$ in $C_c^\infty(U)$. Hence,

$$\begin{aligned}\langle F, \phi * \widetilde{\psi}\rangle &= \lim_{m\to\infty} \langle F, S^m\rangle = \lim_{m\to\infty} 2^{-nm} \sum_j \phi(y_j^m) \langle F, \tau_{y_j^m} \widetilde{\psi}\rangle \\ &= \int \phi(y) \langle F, \tau_y \widetilde{\psi}\rangle \, dy = \int \phi(y) F * \psi(y) \, dy.\end{aligned}$$

∎

Next we show that although distributions may be highly singular objects, they can all be approximated in the (weak*) topology of distributions by smooth functions, even by compactly supported ones.

9.4 Lemma. *Suppose that $\phi \in C_c^\infty$, $\psi \in C_c^\infty$, and $\int \psi = 1$, and let $\psi_t(x) = t^{-n}\psi(t^{-1}x)$.*

a. Given any neighborhood U of $\operatorname{supp}(\phi)$, *we have* $\operatorname{supp}(\phi * \psi_t) \subset U$ *for t sufficiently small.*

b. $\phi * \psi_t \to \phi$ in C_c^∞ as $t \to 0$.

Proof. If $\operatorname{supp}(\psi) \subset \{x : |x| \le R\}$ then $\operatorname{supp}(\phi * \psi_t)$ is contained in the set of points whose distance from $\operatorname{supp}(\phi)$ is at most tR; this is included in a fixed compact set if $t \le 1$ and is included in U if t is small. Moreover, $\partial^\alpha(\phi * \psi_t) = (\partial^\alpha \phi) * \psi_t \to \partial^\alpha \phi$ uniformly as $t \to 0$, by Theorem 8.14. The result follows. ∎

9.5 Proposition. *For any open $U \subset \mathbb{R}^n$, $C_c^\infty(U)$ is dense in $\mathcal{D}'(U)$ in the topology of $\mathcal{D}'(U)$.*

Proof. Suppose $F \in \mathcal{D}'(U)$. We shall first approximate F by distributions supported in compact subsets of U, then approximate the latter by functions in $C_c^\infty(U)$.

Let $\{V_j\}$ be an increasing sequence of precompact open subsets of U whose union is U, as in Proposition 4.39. For each j, by the C^∞ Urysohn lemma we can pick $\zeta_j \in C_c^\infty(U)$ such that $\zeta_j = 1$ on $\overline{V}_j$. Given $\phi \in C_c^\infty(U)$, for j sufficiently large we have $\operatorname{supp}(\phi) \subset V_j$ and hence $\langle F, \phi\rangle = \langle F, \zeta_j \phi\rangle = \langle \zeta_j F, \phi\rangle$. Therefore $\zeta_j F \to F$ as $j \to \infty$.

Now, as we noted in defining products of smooth functions and distributions, since $\operatorname{supp}(\zeta_j)$ is compact, $\zeta_j F$ can be regarded as a distribution on $\mathbb{R}^n$. Let ψ, ψ_t be as in Lemma 9.4, and $\widetilde{\psi}(x) = \psi(-x)$. Then $\int \widetilde{\psi} = 1$ also, so given $\phi \in C_c^\infty$, we have $\phi * \widetilde{\psi}_t \to \phi$ in C_c^∞ by Lemma 9.4. But then by Proposition 9.3, we have $(\zeta_j F) * \psi_t \in C^\infty$ and $\langle (\zeta_j F) * \psi_t, \phi\rangle = \langle \zeta_j F, \phi * \widetilde{\psi}_t\rangle \to \langle \zeta_j F, \phi\rangle$, so $(\zeta_j F) * \psi_t \to \zeta_j F$ in $\mathcal{D}'$. In short, every neighborhood of F in $\mathcal{D}'(U)$ contains the C^∞ functions $(\zeta_j F) * \psi_t$ for j large and t small.

Finally, we observe that $\operatorname{supp}(\zeta_j) \subset V_k$ for some k. If $\operatorname{supp}(\phi) \cap \overline{V}_k = \varnothing$, then for sufficiently small t we have $\operatorname{supp}(\phi * \widetilde{\psi}_t) \cap \overline{V}_k = \varnothing$ (Lemma 9.4 again) and hence $\langle (\zeta_j F) * \psi_t, \phi \rangle = \langle F, \zeta_j(\phi * \widetilde{\psi}_t) \rangle = 0$. In other words, $\operatorname{supp}((\zeta_j F) * \psi_t) \subset \overline{V}_k \subset U$, so we are done. ∎

We conclude this section with some further remarks and examples concerning differentiation of distributions. To restate the basic facts: *Every $F \in \mathcal{D}'(U)$ possesses derivatives $\partial^\alpha F \in \mathcal{D}'(U)$ of all orders; moreover, ∂^α is a continuous linear map of $\mathcal{D}'(U)$ into itself.* Let us examine a couple of one-dimensional examples to see what sort of things arise by taking distribution derivatives of functions that are not classically differentiable.

First, differentiating functions with jump discontinuities leads to "delta-functions," that is, distributions given by measures that are point masses. The simplest example is the Heaviside step function $H = \chi_{(0,\infty)}$, for which we have

$$\langle H', \phi \rangle = -\langle H, \phi' \rangle = -\int_0^\infty \phi'(x)\, dx = \phi(0) = \langle \delta, \phi \rangle,$$

so $H' = \delta$. See Exercises 5 and 7 for generalizations.

Second, distribution derivatives can be used to extract "finite parts" from divergent integrals. For example, let $f(x) = x^{-1}\chi_{(0,\infty)}(x)$. f is locally integrable on $\mathbb{R} \setminus \{0\}$ and so defines a distribution there, but $\int f\phi$ diverges whenever $\phi(0) \neq 0$. Nonetheless, there is a distribution on $\mathbb{R}$ that agrees with f on $\mathbb{R} \setminus \{0\}$, namely, the distribution derivative of the locally integrable function $L(x) = (\log x)\chi_{(0,\infty)}(x)$. One way of seeing what is going on here is to consider the functions $L_\epsilon(x) = (\log x)\chi_{(\epsilon,\infty)}(x)$. By the dominated convergence theorem we have $\int L\phi = \lim_{\epsilon \to 0} \int L_\epsilon \phi$ for any $\phi \in C_c^\infty$, that is, $L_\epsilon \to L$ in $\mathcal{D}'$; it follows that $L'_\epsilon \to L'$ in $\mathcal{D}'$. But

$$\langle L'_\epsilon, \phi \rangle = -\langle L_\epsilon, \phi' \rangle = -\int_\epsilon^\infty \phi'(x) \log x\, dx = \int_\epsilon^\infty \frac{\phi(x)}{x}\, dx + \phi(\epsilon) \log \epsilon.$$

As $\epsilon \to 0$, this last sum converges even though the two terms individually do not. Formally, passage to the limit gives $\langle L', \phi \rangle = \int f\phi + (\log 0)\phi(0)$; that is, L' is obtained from f by subtracting an infinite multiple of δ. (This process is akin to the "renormalizations" used by physicists to remove the divergences from quantum field theory.)

Another way to analyze this situation is to consider smooth approximations to L, such as $L^\epsilon(x) = L(x)\psi(\epsilon x)$ where ψ is a smooth function such that $\psi(x) = 0$ for $x \leq 1$ and $\psi(x) = 1$ for $x \geq 2$. The reader is invited to sketch the graphs of L^ϵ and $(L^\epsilon)'$; the latter will look like the graph of f together with a large negative spike near the origin, which turns into "$-\infty \cdot \delta$" as $\epsilon \to 0$. See also Exercises 10 and 12.

Finally, we remark that one of the bugbears of advanced calculus, that equality of mixed partials need not hold for functions whose derivatives are not continuous, disappears in the setting of distributions: $\partial_j \partial_k = \partial_k \partial_j$ on C_c^∞; therefore $\partial_j \partial_k = \partial_k \partial_j$ on $\mathcal{D}'$! In the standard counterexample, $f(x,y) = xy(x^2 - y^2)(x^2 + y^2)^{-1}$ (with $f(0,0) = 0$), $\partial_x \partial_y f$ and $\partial_y \partial_x f$ are locally integrable functions that agree everywhere except at the origin; hence they are identical as distributions.

Exercises

1. Suppose that $f_1, f_2, \ldots$, and f are in $L^1_{\text{loc}}(U)$. The conditions in (a) and (b) below imply that $f_n \to f$ in $\mathcal{D}'(U)$, but the condition in (c) does not.

a. $f_n \in L^p(U)$ $(1 \le p \le \infty)$ and $f_n \to f$ in the L^p norm or weakly in L^p.

b. For all n, $|f_n| \le g$ for some $g \in L^1_{\text{loc}}(U)$, and $f_n \to f$ a.e.

c. $f_n \to f$ pointwise.

2. The product rule for derivatives is valid for products of smooth functions and distributions.

3. On $\mathbb{R}$, if $\psi \in C^\infty$ then $\psi\delta^{(k)} = \sum_0^k (-1)^j \binom{k}{j} \psi^{(j)}(0)\delta^{(k-j)}$, where the superscripts denote derivatives.

4. Suppose that U and V are open in $\mathbb{R}^n$ and $\Phi : V \to U$ is a C^∞ diffeomorphism. Explain how to define $F \circ \Phi \in \mathcal{D}'(U)$ for any $F \in \mathcal{D}'(V)$.

5. Suppose that f is continuously differentiable on $\mathbb{R}$ except at $x_1, \ldots, x_m$, where f has jump discontinuities, and that its pointwise derivative df/dx (defined except at the x_j's) is in $L^1_{\text{loc}}(\mathbb{R})$. Then the distribution derivative f' of f is given by $f' = (df/dx) + \sum_1^m [f(x_j+) - f(x_j-)]\tau_{x_j}\delta$.

6. If f is absolutely continuous on compact subsets of an interval $U \subset \mathbb{R}$, the distribution derivative $f' \in \mathcal{D}'(U)$ coincides with the pointwise (a.e.-defined) derivative of f.

7. Suppose $f \in L^1_{\text{loc}}(\mathbb{R})$. Then the distribution derivative f' is a complex measure on $\mathbb{R}$ iff f agrees a.e. with a function $F \in NBV$, in which case $\langle f', \phi\rangle = \int \phi\, dF$.

8. Suppose $f \in L^p(\mathbb{R}^n)$. If the strong L^p derivatives $\partial_j f$ exist in the sense of Exercise 8 in §8.2, they coincide with the partial derivatives of f in the sense of distributions.

9. A distribution F on $\mathbb{R}^n$ is called **homogeneous** of degree λ if $F \circ S_r = r^\lambda F$ for all $r > 0$, where $S_r(x) = rx$.

a. δ is homogeneous of degree $-n$.

b. If F is homogeneous of degree λ, then $\partial^\alpha F$ is homogeneous of degree $\lambda - |\alpha|$.

c. The distribution $(d/dx)[\chi_{(0,\infty)}(x)\log x]$ discussed in the text is not homogeneous, although it agrees on $\mathbb{R} \setminus \{0\}$ with a function that is homogeneous of degree -1.

10. Let f be a continuous function on $\mathbb{R}^n \setminus \{0\}$ that is homogeneous of degree $-n$ (i.e., $f(rx) = r^{-n}f(x)$) and has mean zero on the unit sphere (i.e., $\int f\, d\sigma = 0$ where σ is surface measure on the sphere). Then f is not locally integrable near the origin (unless $f = 0$), but the formula

$$\langle PV(f), \phi\rangle = \lim_{\epsilon \to 0} \int_{|x|>\epsilon} f(x)\phi(x)\, dx \qquad (\phi \in C_c^\infty)$$

defines a distribution $PV(f)$ — "PV" stands for "principal value" — that agrees with f on $\mathbb{R}^n \setminus \{0\}$ and is homogeneous of degree $-n$ in the sense of Exercise 9. (Hint: For any $a > 0$, the indicated limit equals

$$\int_{|x|\le a} f(x)\big[\phi(x) - \phi(0)\big]\,dx + \int_{|x|>a} f(x)\phi(x)\,dx,$$

and these integrals converge absolutely.)

11. Let F be a distribution on $\mathbb{R}^n$ such that $\operatorname{supp}(F) = \{0\}$.

a. There exist $N \in \mathbb{N}$, $C > 0$ such that for all $\phi \in C_c^\infty$,

$$|\langle F, \phi\rangle| \le C \sum_{|\alpha|\le N} \sup_{|x|\le 1} |\partial^\alpha \phi(x)|.$$

b. Fix $\psi \in C_c^\infty$ with $\psi(x) = 1$ for $|x| \le 1$ and $\psi(x) = 0$ for $|x| \ge 2$. If $\phi \in C_c^\infty$, let $\phi_k(x) = \phi(x)[1 - \psi(kx)]$. If $\partial^\alpha\phi(0) = 0$ for $|\alpha| \le N$, then $\partial^\alpha\phi_k \to \partial^\alpha\phi$ uniformly as $k \to \infty$ for $|\alpha| \le N$. (Hint: By Taylor's theorem, $|\partial^\alpha\phi(x)| \le C|x|^{N+1-|\alpha|}$ for $|\alpha| \le N$.)

c. If $\phi \in C_c^\infty$ and $\partial^\alpha\phi(0) = 0$ for $|\alpha| \le N$, then $\langle F, \phi\rangle = 0$.

d. There exist constants c_α ($|\alpha| \le N$) such that $F = \sum_{|\alpha|\le N} c_\alpha \partial^\alpha \delta$.

12. Suppose $\lambda > n$; then the function $x \mapsto |x|^{-\lambda}$ on $\mathbb{R}^n$ is not locally integrable near the origin. Here are some ways to make it into a distribution:

a. If $\phi \in C_c^\infty$, let P_ϕ^k be the Taylor polynomial of ϕ about $x = 0$ of degree k. Given $k > \lambda - n - 1$ and $a > 0$, define

$$\langle F_a^k, \phi\rangle = \int_{|x|\le a} \big[\phi(x) - P_\phi^k(x)\big]|x|^{-\lambda}\,dx + \int_{|x|>a} \phi(x)|x|^{-\lambda}\,dx.$$

Then F_a^k is a distribution on $\mathbb{R}^n$ that agrees with $|x|^{-\lambda}$ on $\mathbb{R}^n \setminus \{0\}$.

b. If $\lambda \notin \mathbb{Z}$ and we take k to be the greatest integer $\le \lambda - n$, we can let $a \to \infty$ in (a) to obtain another distribution F that agrees with $|x|^{-\lambda}$ on $\mathbb{R}^n \setminus \{0\}$:

$$\langle F, \phi\rangle = \int \big[\phi(x) - P_\phi^k(x)\big]|x|^{-\lambda}\,dx.$$

c. Let $n = 1$ and let k be the greatest integer $\le \lambda$. Let

$$f(x) = \begin{cases} [(k-\lambda)\cdots(1-\lambda)]^{-1}(\operatorname{sgn} x)^k|x|^{k-\lambda} & \text{if } \lambda > k, \\ (-1)^{k-1}[(k-1)!]^{-1}(\operatorname{sgn} x)^k \log|x| & \text{if } \lambda = k. \end{cases}$$

Then $f \in L^1_{\mathrm{loc}}(\mathbb{R})$, and the distribution derivative $f^{(k)}$ agrees with $|x|^{-\lambda}$ on $\mathbb{R} \setminus \{0\}$.

d. According to Exercise 11, the difference between any two of the distributions constructed in (a)–(c) is a linear combination of δ and its derivatives. Which one?

13. If $F \in \mathcal{D}'$ and $\partial_j F = 0$ for $j = 1, \ldots, n$, then F is a constant function. (Consider $f * \psi_t$ where ψ_t is an approximate identity in C_c^∞.)

14. For $n \geq 3$, define $F, F^\epsilon \in L^1_{\text{loc}}(\mathbb{R}^n)$ by

$$F(x) = \frac{|x|^{2-n}}{\omega_n(2-n)}, \qquad F^\epsilon(x) = \frac{(|x|^2 + \epsilon^2)^{(2-n)/2}}{\omega_n(2-n)},$$

where $\omega_n = 2\pi^{n/2}/\Gamma(n/2)$ is the volume of the unit sphere, and let Δ be the Laplacian.

a. $\Delta F^\epsilon(x) = \epsilon^{-n} g(\epsilon^{-1}x)$ where $g(x) = n\omega_n^{-1}(|x|^2+1)^{-(n+2)/2}$.

b. $\int g = 1$. (Use polar coordinates and set $s = r^2/(r^2+1)$.)

c. $\Delta F = \delta$. ($F^\epsilon \to F$ in $\mathcal{D}'$; use Proposition 9.1.)

d. If $\phi \in C_c^\infty$, the function $f = F * \phi$ satisfies $\Delta f = \phi$.

e. The results of (c) and (d) hold also for $n = 1$ but can be proved more simply there. For $n = 2$, they hold provided F, F^ϵ are defined by $F(x) = (2\pi)^{-1}\log|x|$ and $F^\epsilon = (4\pi)^{-1}\log(|x|^2+\epsilon^2)$.

15. Define G on $\mathbb{R}^n \times \mathbb{R}$ by $G(x,t) = (4\pi t)^{-n/2} e^{-|x|^2/4t}\chi_{(0,\infty)}(t)$.

a. $(\partial_t - \Delta)G = \delta$, where Δ is the Laplacian on $\mathbb{R}^n$. (Let $G^\epsilon(x,t) = G(x,t)\chi_{(\epsilon,\infty)}(t)$; then $G^\epsilon \to G$ in $\mathcal{D}'$. Compute $\langle(\partial_t - \Delta)G^\epsilon, \phi\rangle$ for $\phi \in C_c^\infty$, recalling the discussion of the heat equation in §8.7.)

b. If $\phi \in C_c^\infty(\mathbb{R}^n \times \mathbb{R})$, the function $f = G * \phi$ satisfies $(\partial_t - \Delta)f = \phi$.

9.2 COMPACTLY SUPPORTED, TEMPERED, AND PERIODIC DISTRIBUTIONS

If U is an open set in $\mathbb{R}^n$, the space of all distributions on U whose support is a compact subset of U is denoted by $\mathcal{E}'(U)$; as usual, we set $\mathcal{E}' = \mathcal{E}'(\mathbb{R}^n)$. $\mathcal{E}'(U)$ turns out to be a dual space in its own right, as we shall now show.

The space $C^\infty(U)$ of C^∞ functions on U is a Fréchet space with the C^∞ topology — that is, the topology of uniform convergence of functions, together with all their derivatives, on compact subsets of U. This topology can be defined by a countable family of seminorms as follows. Let $\{V_m\}_1^\infty$ be an increasing sequence of precompact open subsets of U whose union is U, as in Proposition 4.39; then for each $m \in \mathbb{N}$ and each multi-index α we have the seminorm

$$\|f\|_{[m,\alpha]} = \sup_{x \in \overline{V}_m} |\partial^\alpha f(x)|. \tag{9.6}$$

Clearly $\partial^\alpha f_j \to \partial^\alpha f$ uniformly on compact sets for all α iff $\|f_j - f\|_{[m,\alpha]} \to 0$ for all m, α; a different choice of sets V_m would yield an equivalent family of seminorms.

9.7 Proposition. *$C_c^\infty(U)$ is dense in $C^\infty(U)$.*

Proof. Let $\{V_m\}_1^\infty$ be as in (9.6). For each m, by the C^∞ Urysohn lemma we can pick $\psi_m \in C_c^\infty(U)$ with $\psi_m = 1$ on $\overline{V}_m$. If $\phi \in C^\infty(U)$, clearly $\|\psi_m\phi - \phi\|_{[m_0,\alpha]} = 0$ provided $m \geq m_0$; thus $\psi_m\phi \to \phi$ in the C^∞ topology. ∎

9.8 Theorem. *$\mathcal{E}'(U)$ is the dual space of $C^\infty(U)$. More precisely: If $F \in \mathcal{E}'(U)$, then F extends uniquely to a continuous linear functional on $C^\infty(U)$; and if G is a continuous linear functional on $C^\infty(U)$, then $G|C_c^\infty(U) \in \mathcal{E}'(U)$.*

Proof. If $F \in \mathcal{E}'(U)$, choose $\psi \in C_c^\infty(U)$ with $\psi = 1$ on $\operatorname{supp}(F)$, and define the linear functional G on $C^\infty(U)$ by $\langle G, \phi\rangle = \langle F, \psi\phi\rangle$. Since F is continuous on $C_c^\infty(\operatorname{supp}(\psi))$, and the topology of the latter is defined by the norms $\phi \mapsto \|\partial^\alpha\phi\|_u$, by Proposition 5.15 there exist $N \in \mathbb{N}$ and $C > 0$ such that $|\langle G, \phi\rangle| \leq C\sum_{|\alpha|\leq N}\|\partial^\alpha(\psi\phi)\|_u$ for $\phi \in C^\infty(U)$. By the product rule, if we choose m large enough so that $\operatorname{supp}(\psi) \subset V_m$, this implies that

$$|\langle G, \phi\rangle| \leq C' \sum_{|\alpha|\leq N} \sup_{x\in\operatorname{supp}(\psi)} |\partial^\alpha\phi(x)| \leq C' \sum_{|\alpha|\leq N} \|\phi\|_{[m,\alpha]},$$

so that G is continuous on $C^\infty(U)$. That G is the unique continuous extension of F follows from Proposition 9.7.

On the other hand, if G is a continuous linear functional on $C^\infty(U)$, by Proposition 5.15 there exist C, m, N such that $|\langle G, \phi\rangle| \leq C\sum_{|\alpha|\leq N}\|\phi\|_{[m,\alpha]}$ for all $\phi \in C^\infty(U)$. Since $\|\phi\|_{[m,\alpha]} \leq \|\partial^\alpha\phi\|_u$, this implies that G is continuous on $C_c^\infty(K)$ for each compact $K \subset U$, so $G|C_c^\infty(U) \in \mathcal{D}'(U)$. Moreover, if $[\operatorname{supp}(\phi)] \cap \overline{V}_m = \varnothing$, then $\langle G, \phi\rangle = 0$; hence $\operatorname{supp}(G) \subset \overline{V}_m$ and $G|C_c^\infty(U) \in \mathcal{E}'(U)$. ∎

The operations of differentiation, multiplication by C^∞ functions, translation, and composition by linear maps discussed in §9.1 all preserve the class $\mathcal{E}'$. As for convolution, there is more to be said.

First, if $F \in \mathcal{E}'$ and $\phi \in C_c^\infty$ then $F * \phi \in C_c^\infty$, as Proposition 8.6d remains valid in this setting. Second, if $F \in \mathcal{E}'$ and $\psi \in C^\infty$, $F * \psi$ can be defined as a C^∞ function or as a distribution just as before:

$$F * \psi(x) = \langle F, \tau_x\widetilde{\psi}\rangle, \qquad \langle F * \psi, \phi\rangle = \langle F, \phi * \widetilde{\psi}\rangle \quad (\phi \in C_c^\infty)$$

(see Exercise 16). Finally, a further dualization allows us to define convolutions of arbitrary distributions with compactly supported distributions. To wit, if $F \in \mathcal{D}'$ and $G \in \mathcal{E}'$, we can define $F * G \in \mathcal{D}'$ and $G * F \in \mathcal{D}'$ as follows:

$$\langle F * G, \phi\rangle = \langle F, \widetilde{G} * \phi\rangle, \quad \langle G * F, \phi\rangle = \langle G, \widetilde{F} * \phi\rangle \qquad (\phi \in C_c^\infty),$$

and likewise for $\widetilde{F}$. The proof that $F * G$ and $G * F$ are indeed distributions (i.e., that they are continuous on C_c^∞) and that $F * G = G * F$ requires a closer examination of the continuity of the maps involved. We shall not pursue this matter here; however, see Exercises 20 and 21.

A notable omission from our list of operations that can be extended from functions to distributions is the Fourier transform $\mathcal{F}$. The trouble is that $\mathcal{F}$ does not map C_c^∞

into itself; in fact, if $\phi \in C_c^\infty$, then $\widehat{\phi}$ cannot vanish on any nonempty open set unless $\phi = 0$. To see this, suppose $\widehat{\phi} = 0$ on a neighborhood of ξ_0. Replacing ϕ by $e^{-2\pi i \xi_0 \cdot x}\phi$, we may assume that $\xi_0 = 0$. Since ϕ has compact support, we can expand $e^{-2\pi i \xi \cdot x}$ in its Maclaurin series and integrate term by term to obtain

$$\widehat{\phi}(\xi) = \sum_{k=0}^{\infty} \frac{1}{k!} \int (-2\pi i \xi \cdot x)^k \phi(x)\, dx = \sum_{\alpha} \frac{1}{\alpha!} \xi^\alpha \int (-2\pi i x)^\alpha \phi(x)\, dx$$

(see Exercise 2a in §8.1). But $\int (-2\pi i x)^\alpha \phi(x)\, dx = \partial^\alpha \widehat{\phi}(0)$ for all α by Theorem 8.22d. These derivatives all vanish by assumption, so $\widehat{\phi} = 0$ and hence $\phi = 0$.

However, we do have available a slightly larger space of smooth functions that is mapped into itself by $\mathcal{F}$, namely, the Schwartz class $\mathcal{S}$. We recall that $\mathcal{S}$ is a Fréchet space with the topology defined by the norms

$$\|\phi\|_{(N,\alpha)} = \sup_{x \in \mathbb{R}^n} (1 + |x|)^N |\partial^\alpha \phi(x)|.$$

9.9 Proposition. *Suppose $\psi \in C_c^\infty$ and $\psi(0) = 1$, and let $\psi^\epsilon(x) = \psi(\epsilon x)$. Then for any $\phi \in \mathcal{S}$, $\psi^\epsilon \phi \to \phi$ in $\mathcal{S}$ as $\epsilon \to 0$. In particular, C_c^∞ is dense in $\mathcal{S}$.*

Proof. Given $N \in \mathbb{N}$, for any $\eta > 0$ we can choose a compact set K such that $(1 + |x|)^N |\phi(x)| < \eta$ for $x \notin K$. Since $\psi(\epsilon x) \to 1$ uniformly for $x \in K$ as $\epsilon \to 0$, it follows easily that $\|\psi^\epsilon \phi - \phi\|_{(N,0)} \to 0$ for every N. For the norms involving derivatives, we observe that by the product rule,

$$(1 + |x|)^N \partial^\alpha (\psi^\epsilon \phi - \phi) = (1 + |x|)^N (\psi^\epsilon \partial^\alpha \phi - \partial^\alpha \phi) + E_\epsilon(x),$$

where E_ϵ is a sum of terms involving derivative of ψ^ϵ. Since

$$|\partial^\beta \psi^\epsilon(x)| = \epsilon^{|\beta|} |\partial^\beta \psi(\epsilon x)| \le C_\beta \epsilon^{|\beta|},$$

we have $\|E_\epsilon\|_u \le C\epsilon \to 0$ as $\epsilon \to 0$. The preceding argument then shows that $\|\psi^\epsilon \phi - \phi\|_{(N,\alpha)} \to 0$. ∎

A **tempered distribution** is a continuous linear functional on $\mathcal{S}$. The space of tempered distributions is denoted by $\mathcal{S}'$; it comes equipped with the weak* topology, that is, the topology of pointwise convergence on $\mathcal{S}$. If $F \in \mathcal{S}'$, then $F|C_c^\infty$ is clearly a distribution, since convergence in C_c^∞ implies convergence in $\mathcal{S}$, and $F|C_c^\infty$ determines F uniquely by Proposition 9.9. Thus we may, and shall, identify $\mathcal{S}'$ with the set of distributions that extend continuously from C_c^∞ to $\mathcal{S}$. We say that a locally integrable function is **tempered** if it is tempered as a distribution.

The condition that a distribution be tempered means, roughly speaking, that it does not grow too fast at infinity. Here are a few examples:

- Every compactly supported distribution is tempered.
- If $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ and $\int (1 + |x|)^N |f(x)|\, dx < \infty$ for some N, then f is tempered, for $|\int f\phi| \le C\|\phi\|_{(0,N)}$.

- The function $f(x) = e^{ax}$ on $\mathbb{R}$ is tempered iff a is purely imaginary. Indeed, suppose $a = b + ic$ with b, c real. If $b = 0$, then f is bounded and hence tempered by (ii). If $b \neq 0$, choose a function $\psi \in C_c^\infty$ such that $\int \psi = 1$, and let $\phi_j(x) = e^{-ax}\psi(x - j)$. It is easily verified that $\phi_j \to 0$ in $\mathcal{S}$ as $j \to +\infty$ (if $b > 0$) or $j \to -\infty$ (if $b < 0$), but $\int f\phi_j = \int \psi = 1$ for all j.

- On the other hand, the function $f(x) = e^x \cos e^x$ on $\mathbb{R}$ is tempered, because it is the derivative of the bounded function $\sin e^x$. Indeed, if $\phi \in \mathcal{S}$, integration by parts yields

$$\left|\int f\phi\right| = \left|-\int \phi'(x) \sin e^x \, dx\right| \leq C\|\phi\|_{(2,1)}.$$

  Intuitively, $f(x)$ is not too large "on average" when x is large, because of its rapid oscillations.

We turn to the consideration of the basic linear operations on tempered distributions. The operations of differentiation, translation, and composition with linear transformations work just the same way for tempered distributions as for plain distributions; these operations all map $\mathcal{S}$ and $\mathcal{S}'$ into themselves. The same is not true of multiplication by arbitrary smooth functions, however. The proper requirement on $\psi \in C^\infty$ in order for the map $F \to \psi F$ to preserve $\mathcal{S}$ and $\mathcal{S}'$ is that ψ and all its derivatives should have at most polynomial growth at infinity:

$$|\partial^\alpha \psi(x)| \leq C_\alpha (1 + |x|)^{N(\alpha)} \text{ for all } \alpha.$$

Such C^∞ functions are called **slowly increasing**. For example, every polynomial is slowly increasing; so are the functions $(1 + |x|^2)^s$ ($s \in \mathbb{R}$), which will play an important role in the next section.

As for convolutions, for any $F \in \mathcal{S}'$ and $\psi \in \mathcal{S}$ we can define the convolution $F * \psi$ by $F * \psi(x) = \langle F, \tau_x \widetilde{\psi} \rangle$, as before, and we have an analogue of Proposition 9.3:

9.10 Proposition. *If $F \in \mathcal{S}'$ and $\psi \in \mathcal{S}$, then $F * \psi$ is a slowly increasing C^∞ function, and for any $\phi \in \mathcal{S}$ we have $\int (F * \psi)\phi = \langle F, \phi * \widetilde{\psi} \rangle$.*

Proof. That $F * \psi \in C^\infty$ is established as in Proposition 9.3. By Proposition 5.15, the continuity of F implies that there exist m, N, C such that

$$|\langle F, \phi \rangle| \leq C \sum_{|\alpha| \leq N} \|\phi\|_{(m,\alpha)} \qquad (\phi \in \mathcal{S}),$$

and hence by (8.12),

$$\begin{aligned}|F * \psi(x)| &\le C \sum_{|\alpha|\le N} \sup_y (1+|y|)^m |\partial^\alpha \psi(x-y)| \\ &\le C(1+|x|)^m \sum_{|\alpha|\le N} \sup_y (1+|x-y|)^m |\partial^\alpha \psi(x-y)| \\ &\le C(1+|x|)^m \sum_{|\alpha|\le N} \|\psi\|_{(m,\alpha)}.\end{aligned}$$

The same reasoning applies with ψ replaced by $\partial^\beta \psi$, so $F * \psi$ is slowly increasing. Next, by Proposition 9.3 we know that the equation $\int (F*\psi)\phi = \langle F, \phi * \widetilde{\psi}\rangle$ holds when $\phi, \psi \in C_c^\infty$. By Proposition 9.9, if $\phi, \psi \in \mathcal{S}$ we can find sequences $\{\phi_j\}$ and $\{\psi_j\}$ in C_c^∞ that converge to ϕ and ψ in $\mathcal{S}$. Then $\phi_j * \widetilde{\psi}_j \to \phi * \widetilde{\psi}$ in $\mathcal{S}$ by (the proof of) Proposition 8.11, so $\langle F, \phi_j * \widetilde{\psi}_j\rangle \to \langle F, \phi * \widetilde{\psi}\rangle$. On the other hand, the preceding estimates show that $|F * \psi_j(x)| \le C(1+|x|)^m$ with C and m independent of j, and likewise $|\phi_j(x)| \le C(1+|x|)^{-m-n-1}$, so $\int (F*\psi_j)\phi_j \to \int (F*\psi)\phi$ by the dominated convergence theorem. ∎

Finally, we come to the principal *raison d'être* of tempered distributions, the Fourier transform. We recall (Corollary 8.23) that the Fourier transform maps $\mathcal{S}$ continuously into itself, and that for $f, g \in L^1$ (in particular, for $f, g \in \mathcal{S}$) we have

$$\int \widehat{f}(y) g(y)\, dy = \iint f(x) g(y) e^{-2\pi i x\cdot y}\, dx\, dy = \int f(x) \widehat{g}(x)\, dx.$$

We can therefore extend the Fourier transform to a continuous linear map from $\mathcal{S}'$ to itself by defining

$$\langle \widehat{F}, \phi\rangle = \langle F, \widehat{\phi}\rangle \qquad (F \in \mathcal{S}',\ \phi \in \mathcal{S}).$$

This definition clearly agrees with the one in Chapter 8 when $F \in L^1 + L^2$.

The basic properties of the Fourier transform in Theorem 8.22 continue to hold in this setting. To wit,

$$\begin{aligned} (\tau_y F)\widehat{\ } &= e^{-2\pi i \xi\cdot y}\widehat{F}, & \tau_\eta \widehat{F} &= [e^{2\pi i \eta\cdot x} F]\widehat{\ }, \\ \partial^\alpha \widehat{F} &= [(-2\pi i x)^\alpha F]\widehat{\ }, & (\partial^\alpha F)\widehat{\ } &= (2\pi i \xi)^\alpha \widehat{F}, \\ (f\circ T)\widehat{\ } &= |\det T|^{-1} \widehat{f} \circ (T^*)^{-1} & &\big(T \in GL(n,\mathbb{R})\big), \\ (F * \psi)\widehat{\ } &= \widehat{\psi}\widehat{F} & &(\psi \in \mathcal{S}). \end{aligned}$$

(The first four of these formulas involve products of slowly increasing C^∞ functions, specified by their values at a general point x or ξ, and tempered distributions.) The easy verifications of these facts are left to the reader (Exercise 17).

Moreover, we can define the inverse transform in the same way:

$$\langle F^\vee, \phi\rangle = \langle F, \phi^\vee\rangle.$$

The Fourier inversion theorem formula $\phi = (\widehat{\phi})^\vee = (\phi^\vee)\widehat{\ }$ then extends to $\mathcal{S}'$:

$$\langle (\widehat{F})^\vee, \phi\rangle = \langle \widehat{F}, \phi^\vee\rangle = \langle F, (\phi^\vee)\widehat{\ }\rangle = \langle F, \phi\rangle,$$

so that $(\widehat{F})^\vee = F$, and likewise $(F^\vee)\widehat{\ } = F$. Thus the Fourier transform is an isomorphism on $\mathcal{S}'$.

If $F \in \mathcal{E}'$, there is an alternative way to define $\widehat{F}$. Indeed, $\langle F, \phi\rangle$ makes sense for any $\phi \in C^\infty$, and if we take $\phi(x) = e^{-2\pi i\xi\cdot x}$, we obtain a function of ξ that has a strong claim to be called $\widehat{F}(\xi)$. In fact, the two definitions are equivalent:

9.11 Proposition. *If $F \in \mathcal{E}'$, then $\widehat{F}$ is a slowly increasing C^∞ function, and it is given by $\widehat{F}(\xi) = \langle F, E_{-\xi}\rangle$ where $E_\xi(x) = e^{2\pi i\xi\cdot x}$.*

Proof. Let $g(\xi) = \langle F, E_{-\xi}\rangle$. Consideration of difference quotients of g, as in the proof of Proposition 9.3, shows that g is a C^∞ function with derivatives given by $\partial^\alpha g(\xi) = \langle F, \partial_\xi^\alpha E_{-\xi}\rangle = (-2\pi i)^{|\alpha|}\langle F, x^\alpha E_{-\xi}\rangle$. Moreover, by Theorem 9.8 and Proposition 5.15, there exist m, N, C such that

$$|\partial^\alpha g(\xi)| \le C \sum_{|\beta|\le N} \sup_{|x|\le m} \left|\partial^\beta[x^\alpha E_{-\xi}(x)]\right| \le C'(1+m)^{|\alpha|}(1+|\xi|)^N,$$

so g is slowly increasing.

It remains to show that $g = \widehat{F}$, and by Proposition 9.9 it suffices to show that $\int g\phi = \langle F, \widehat{\phi}\rangle$ for $\phi \in C_c^\infty$. In this case $g\phi \in C_c^\infty$, so $\int g\phi$ can be approximated by Riemann sums as in the proof of Proposition 9.3, say $\sum g(\xi_j)\phi(\xi_j)\,\Delta\xi_j$. The corresponding sums $\sum \phi(\xi_j)e^{-2\pi i\xi_j\cdot x}\,\Delta\xi_j$ and their derivatives in x converge uniformly, for x in any compact set, to $\widehat{\phi}(x)$ and its derivatives. Therefore, since F is a continuous functional on C^∞,

$$\int g\phi = \lim \sum \langle F, E_{-\xi_j}\rangle \phi(\xi_j)\,\Delta\xi_j = \lim \Big\langle F, \sum \phi(x_j)E_{-\xi_j}\,\Delta\xi_j\Big\rangle = \langle F, \widehat{\phi}\rangle.$$

∎

It is time for some examples. First and foremost, the Fourier transform of the point mass at 0 is the constant function 1: $\langle \delta, E_{-\xi}\rangle = E_{-\xi}(0) = 1$. More generally, for point masses at other points and their derivatives, we have

$$\begin{aligned}(\partial^\alpha \tau_y\delta)\widehat{\ }(\xi) &= (-1)^{|\alpha|}\langle \delta, \tau_{-y}\partial^\alpha E_{-\xi}\rangle = (-1)^{|\alpha|}\partial_x^\alpha(e^{-2\pi i\xi\cdot(x+y)})\big|_{x=0}\\ &= (2\pi i\xi)^\alpha e^{-2\pi i\xi\cdot y}.\end{aligned}$$

In particular:

9.12 Proposition. *The Fourier transforms of the linear combinations of δ and its derivatives are precisely the polynomials.*

The Fourier inversion theorem then yields the formulas for the Fourier transforms of polynomials and imaginary exponentials:

$$(9.13)\qquad (x^\alpha)\widehat{\ } = [(-x)^\alpha]^\vee = (-2\pi i)^{-|\alpha|}\partial^\alpha\delta, \qquad \widehat{E}_y = (E_{-y})^\vee = \tau_y\delta.$$

As an illustration of the heuristics associated to these results, consider the formula

$$\int e^{2\pi i\xi\cdot x}\,d\xi = \delta(x).$$

Although this is nonsensical as a pointwise equality, it is valid when viewed from the right angle. One the one hand, it expresses the fact that the Fourier transform of the constant function 1 is δ. More interestingly, it is a concise statement of the Fourier inversion theorem. Indeed, if we replace x by $x-y$, integrate both sides against $\phi \in \mathcal{S}$, and reverse the order of integration on the left, we obtain

$$\iint \phi(y)e^{2\pi i\xi\cdot(x-y)}\,dy\,dx = \int \delta(x-y)\phi(y)\,dy.$$

The integral on the left is $(\widehat{\phi})^\vee(x)$, and the integral on the right equals $\phi(x)$!

It is an important fact that every distibution is, at least locally, a linear combination of derivatives of continuous functions. The Fourier transform yields an easy proof of this:

9.14 Proposition.

a. If $F \in \mathcal{E}'$, there exist $N \in \mathbb{N}$, constants c_α ($|\alpha| \le N$), and $f \in C_0(\mathbb{R}^n)$ such that $F = \sum_{|\alpha|\le N} c_\alpha\partial^\alpha f$.

b. If $F \in \mathcal{D}'(U)$ and V is a precompact open set with $\overline{V} \subset U$, there exist N, c_α, f as above such that $F = \sum_{|\alpha|\le N} c_\alpha\partial^\alpha f$ on V.

Proof. By Proposition 9.11, if $F \in \mathcal{E}'$ then $\widehat{F}$ is slowly increasing, so the function $g(\xi) = (1+|\xi|^2)^{-M}\widehat{F}(\xi)$ will be in L^1 if the integer M is chosen sufficiently large. Let $f = \widehat{g}$; then $f \in C_0$ and $\widehat{F} = (1+|\xi|^2)^M\widehat{f}$, so $F = (I - (4\pi^2)^{-1}\sum_1^n \partial_j^2)^M f$. This proves (a); for (b), choose $\psi \in C_c^\infty(U)$ such that $\psi = 1$ on V, and apply (a) to ψF. ∎

We conclude this section with a sketch of the theory of periodic distributions; some of the details are fleshed out in Exercises 22–24.

The space $C^\infty(\mathbb{T}^n)$ of smooth periodic functions is a Fréchet space with the topology defined by the seminorms $\phi \mapsto \|\partial^\alpha\phi\|_u$, and a distribution on $\mathbb{T}^n$ is a continuous linear functional on this space; the space of distributions on $\mathbb{T}^n$ is denoted by $\mathcal{D}'(\mathbb{T}^n)$. If $F \in \mathcal{D}'(\mathbb{T}^n)$, its Fourier transform is the function $\widehat{F}$ on $\mathbb{Z}^n$ defined by $\widehat{f}(\kappa) = \langle F, E_{-\kappa}\rangle$ where $E_\kappa(x) = e^{2\pi i\kappa\cdot x}$. Since F satisfies an estimate of the form $|\langle F,\phi\rangle| \le C\sum_{|\alpha|\le N}\|\partial^\alpha\phi\|_u$, there exist C, N such that

$$(9.15)\qquad |\widehat{F}(\kappa)| \le C(1+|\kappa|)^N,$$

and the Fourier transform is an isomorphism from $\mathcal{D}'(\mathbb{T}^n)$ to the space of all functions on $\mathbb{Z}^n$ satisfying such an estimate. Moreover, if $F \in \mathcal{D}'(\mathbb{T}^n)$, the Fourier series $\sum_\kappa \widehat{F}(\kappa)E_\kappa$ converges in $\mathcal{D}'(\mathbb{T}^n)$ to F.

Instead of defining periodic distributions as distributions on $\mathbb{T}^n$ (linear functionals on $C^\infty(\mathbb{T}^n)$), one can define them as distributions on $\mathbb{R}^n$ (linear functionals on $C_c^\infty(\mathbb{R}^n)$) that are invariant under the translations τ_κ, $\kappa \in \mathbb{Z}^n$. Accordingly, let

$$\mathcal{D}'(\mathbb{R}^n)_{\text{per}} = \{F \in \mathcal{D}'(\mathbb{R}^n) : \tau_\kappa F = F \text{ for } \kappa \in \mathbb{Z}\}.$$

The periodization map $P\phi = \sum_{\kappa\in\mathbb{Z}^n} \tau_\kappa\phi$ used in Theorem 8.31 is easily seen to map $C_c^\infty(\mathbb{R}^n)$ continuously into $C^\infty(\mathbb{T}^n)$, so it induces a map $P' : \mathcal{D}'(\mathbb{T}^n) \to \mathcal{D}'(\mathbb{R}^n)$ given by $\langle P'F, \phi\rangle = \langle F, P\phi\rangle$. Since $P \circ \tau_\kappa = P$ for $\kappa \in \mathbb{Z}^n$, we have $\tau_\kappa \circ P' = P'$, that is, the range of P' lies in $\mathcal{D}'(\mathbb{R}^n)_{\text{per}}$. In fact, $P' : \mathcal{D}'(\mathbb{T}^n) \to \mathcal{D}'(\mathbb{R}^n)_{\text{per}}$ is a bijection. (The proof is nontrivial; see Exercise 24.) Moreover, if $f \in L^1(\mathbb{T}^n)$, then f and $P'f$ coincide as periodic functions on $\mathbb{R}$, for if $\phi \in C_c^\infty(\mathbb{R}^n)$,

$$\begin{aligned}\langle P'f, \phi\rangle = \langle f, P\phi\rangle &= \int_{[0,1)^n} f(x) \sum \phi(x-\kappa)\, dx \\ &= \sum \int_{[0,1)^n+\kappa} f(x)\phi(x)\, dx = \int_{\mathbb{R}^n} f(x)\phi(x)\, dx = \langle f, \phi\rangle.\end{aligned}$$

Thus the two descriptions of periodic distributions are equivalent.

If $F \in \mathcal{D}'(\mathbb{T}^n)$, the Fourier series $\sum \widehat{F}(\kappa)E_\kappa$ converges in $\mathcal{D}'(\mathbb{T}^n)$ to F; on the other hand, it follows easily from (9.15) that it also converges in $\mathcal{S}'(\mathbb{R}^n)$, and its sum there is $P'f$. Thus $\mathcal{D}'(\mathbb{R}^n)_{\text{per}} \subset \mathcal{S}'(\mathbb{R}^n)$, and by (9.13) we have

$$(P'F)\widehat{\ } = \sum \widehat{F}(\kappa)\widehat{E}_\kappa = \sum \widehat{F}(\kappa)\tau_\kappa\delta,$$

giving the relation between the $\mathbb{R}^n$- and $\mathbb{T}^n$-Fourier transforms for periodic distributions. In particular, if $F = \delta_{\mathbb{T}^n}$, the point mass at the origin in $\mathbb{T}^n$, then $\widehat{F}(\kappa) = 1$ for all κ; hence $P'F$ and $(P'F)\widehat{\ }$ are both equal to $\sum \tau_\kappa\delta$ — a restatement of the Poisson summation formula.

Exercises

16. Suppose $F \in \mathcal{E}'$ and $\psi \in C^\infty$. Show that for any $\phi \in C_c^\infty$, $\int \langle F, \tau_x\widetilde{\psi}\rangle\phi(x)\, dx = \langle F, \phi * \widetilde{\psi}\rangle$. (The result can be reduced to Proposition 9.3; given F and ϕ, the indicated expressions depend only on the values of ψ in a compact set.)

17. Suppose that $F \in \mathcal{S}'$. Show that

a. $(\tau_y F)\widehat{\ } = e^{-2\pi i\xi\cdot y}\widehat{F}$, $\tau_\eta\widehat{F} = [e^{2\pi i\eta\cdot x}F]\widehat{\ }$.

b. $\partial^\alpha\widehat{F} = [(-2\pi ix)^\alpha F]\widehat{\ }$, $(\partial^\alpha F)\widehat{\ } = (2\pi i\xi)^\alpha\widehat{F}$.

c. $(F \circ T)\widehat{\ } = |\det T|^{-1}\widehat{F} \circ (T^*)^{-1}$ for $T \in GL(n, \mathbb{R})$.

d. $(F * \psi)\widehat{\ } = \widehat{\psi}\widehat{F}$ for $\psi \in \mathcal{S}$.

18. If $n = l + m$, let us write $x \in \mathbb{R}^n$ as (y, z) with $y \in \mathbb{R}^l$ and $z \in \mathbb{R}^m$. Let $\mathcal{F}$ denote the Fourier transform on $\mathbb{R}^n$ and $\mathcal{F}_1$, $\mathcal{F}_2$ the partial Fourier transforms in the first and second sets of variables — i.e., $\mathcal{F}_1 f(\eta, z) = \int f(y, z) e^{-2\pi i \eta \cdot y}\, dy$ and likewise for $\mathcal{F}_2$. Then $\mathcal{F}_1$ and $\mathcal{F}_2$ are isomorphisms on $\mathcal{S}(\mathbb{R}^n)$ and $\mathcal{S}'(\mathbb{R}^n)$, and $\mathcal{F} = \mathcal{F}_1\mathcal{F}_2 = \mathcal{F}_2\mathcal{F}_1$.

19. On $\mathbb{R}$, let $F_0 = PV(1/x)$ as defined in Exercise 10. Also, for $\epsilon > 0$ let $F_\epsilon(x) = x(x^2 + \epsilon^2)^{-1}$, $G_\epsilon^{\pm}(x) = (x \pm i\epsilon)^{-1}$, and $S_\epsilon(x) = e^{-2\pi\epsilon|x|} \operatorname{sgn} x$.

a. $\lim_{\epsilon \to 0} F_\epsilon = F_0$ in the weak* topology of $\mathcal{S}'$. (Theorem 8.14, with $a = 0$, may be useful.)

b. $\lim_{\epsilon \to 0} G_\epsilon = F_0 \mp \pi i \delta$. (Hint: $(x \pm i\epsilon)^{-1} = (x \mp i\epsilon)(x^2 + \epsilon^2)^{-1}$.)

c. $\widehat{S}_\epsilon = (\pi i)^{-1} F_\epsilon$ and hence $\widehat{\operatorname{sgn}} = (\pi i)^{-1} F_0$.

d. From (c) it follows that $\widehat{F}_0 = -\pi i \operatorname{sgn}$. Prove this directly by showing that $F_0 = \lim_{\epsilon \to 0,\, N \to \infty} H_{\epsilon,N}$, where $H_{\epsilon,N}(x) = x^{-1}$ if $\epsilon < |x| < N$ and $H_{\epsilon,N}(x) = 0$ otherwise, and using Exercise 59b in §2.6.

e. Compute $\widehat{\chi}_{(0,\infty)}$ (i) by writing $\chi_{(0,\infty)} = \frac{1}{2}\operatorname{sgn} + \frac{1}{2}$ and using (c), (ii) by writing $\chi_{(0,\infty)}(x) = \lim e^{-\epsilon x} \chi_{(0,\infty)}(x)$ and using (b).

20. Suppose that $F \in \mathcal{S}'$ and $G \in \mathcal{E}'$.

a. $\widehat{F}\widehat{G}$ is well-defined element of $\mathcal{S}'$.

b. If $\psi \in \mathcal{S}$, then $G * \psi \in \mathcal{S}$.

c. Let $F * G$ (or $G * F$) be the tempered distribution such that $(F * G)\widehat{\ } = \widehat{F}\widehat{G}$. Then $\langle F * G, \psi \rangle = \langle F, \widetilde{G} * \psi \rangle = \langle G, \widetilde{F} * \psi \rangle$ for $\psi \in \mathcal{S}$.

21. Suppose that $F, G, H \in \mathcal{S}'$.

a. If at most one of F, G, H has noncompact support, then $(F * G) * H = F * (G * H)$, where the convolutions are defined as in Exercise 20.

b. On $\mathbb{R}$, let F be the constant function 1, $G = d\delta/dx$, and $H = \chi_{(0,\infty)}$. Then $(F * G) * H$ and $F * (G * H)$ are well defined in $\mathcal{S}'$ but are unequal.

22. Let $E_\kappa(x) = e^{2\pi i \kappa \cdot x}$. If $g : \mathbb{Z}^n \to \mathbb{C}$ satisfies $|g(\kappa)| \le C(1 + |\kappa|)^N$ for some $C, N > 0$, then the series $\sum_{\kappa \in \mathbb{Z}^n} g(\kappa) E_\kappa$ converges in $\mathcal{D}'(\mathbb{T}^n)$ to a distribution F that satisfies $\widehat{F} = g$. It also converges in $\mathcal{S}'(\mathbb{R}^n)$ to a tempered distribution G $(= P'F)$ such that $\tau_\kappa G = G$ for all κ.

23. Suppose that $F, G \in \mathcal{D}'(\mathbb{T}^n)$.

a. There is a unique $F * G \in \mathcal{D}'(\mathbb{T}^n)$ such that $(F * G)\widehat{\ } = \widehat{F}\widehat{G}$. (Use Exercise 22.)

b. If $G \in C^\infty(\mathbb{T}^n)$, then $F * G \in C^\infty(\mathbb{T}^n)$ and $F * G(x) = \langle F, \tau_x \widetilde{G} \rangle$ as on $\mathbb{R}^n$.

24. Let P be the periodization map, $P\phi = \sum_{\kappa \in \mathbb{Z}^n} \tau_\kappa \phi$.

a. P is a continuous linear map from $C_c^\infty(\mathbb{R}^n)$ to $C^\infty(\mathbb{T}^n)$. (Note that for $\phi \in C_c^\infty$ and x in a compact set, only finitely many terms of the series $\sum \tau_\kappa \phi(x)$ are nonzero.)

b. Choose $\gamma \in C_c^\infty$ with $\int \gamma = 1$, and let $\omega = \gamma * \chi_{[0,1)^n}$. Then $\omega \in C_c^\infty$ and $P\omega = 1$.

c. If $\psi \in C^\infty(\mathbb{T}^n)$, then $\psi = P(\omega\psi)$ (where ψ is regarded as a function on $\mathbb{T}^n$ on the left and as a function on $\mathbb{R}^n$ on the right). Consequently, $P : C_c^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{T}^n)$ is surjective and the dual map $P' : \mathcal{D}'(\mathbb{T}^n) \to \mathcal{D}'(\mathbb{R}^n)_{\text{per}}$ is injective.
d. Given $G \in \mathcal{D}'(\mathbb{R}^n)_{\text{per}}$, define $F \in \mathcal{D}'(\mathbb{T}^n)$ by $\langle F, \psi\rangle = \langle G, \omega\psi\rangle$ (with the same understanding as in part (c)). Then $P'F = G$, so P' maps $\mathcal{D}'(\mathbb{T}^n)$ onto $\mathcal{D}'(\mathbb{R}^n)_{\text{per}}$.

25. Suppose that P is a polynomial in n variables such that only zero of $P(\xi)$ in $\mathbb{R}^n$ is $\xi = 0$, and let $P(D)$ be as in §8.7.
a. Every tempered distribution F that satifies $P(D)F = 0$ is a polynomial. (Use Proposition 9.12 and Exercise 11.)
b. Every bounded function f that satisfies $P(D)f = 0$ is a constant. (This result, for the special cases where $P(D)$ is the Laplacian or the Cauchy-Riemann operator $\partial_x + i\partial_y$ on $\mathbb{R}^2$, is known as **Liouville's theorem**.)

26. On $\mathbb{R}^n \times \mathbb{R}$, let $G(x,t) = (4\pi t)^{-n/2} e^{-|x|^2/4t}\chi_{(0,\infty)}(t)$.
a. $\widehat{G}$ is the tempered function $\widehat{G}(\xi,\tau) = (2\pi i\tau + 4\pi^2|\xi|^2)^{-1}$. (Use Proposition 8.24 and Exercise 18.)
b. Deduce that $(\partial_t - \Delta)G = \delta$. (Cf. Exercise 15.)

27. Suppose that $0 < \operatorname{Re}\alpha < n$.
a. For any $\phi \in \mathcal{S}$,

$$\frac{\Gamma((n-\alpha)/2)}{\pi^{(n-\alpha)/2}} \int |x|^{\alpha-n}\widehat{\phi}(x)\,dx = \frac{\Gamma(\alpha/2)}{\pi^{\alpha/2}} \int |\xi|^{-\alpha}\phi(\xi)\,d\xi.$$

(Hint: By Proposition 8.24 and Lemma 8.25, if $t > 0$ we have

$$\int e^{-\pi t|x|^2}\widehat{\phi}(x)\,dx = t^{-n/2}\int e^{-\pi|\xi|^2/t}\phi(\xi)\,d\xi.$$

Multiply both sides by $t^{-1+(n-\alpha)/2}\,dt$ and integrate from 0 to ∞.)
b. Let $R_\alpha(x) = \Gamma((n-\alpha)/2)[\Gamma(\alpha/2)2^\alpha\pi^{n/2}]^{-1}|x|^{\alpha-n}$. Then R_α is a tempered function and $\widehat{R}_\alpha$ is the tempered function $\widehat{R}_\alpha(\xi) = (2\pi|\xi|)^{-\alpha}$.
c. If $n > 2$, then $\Delta R_2 = -\delta$. (Cf. Exercise 14. See the next exercise for the case $n = 2$.)

28. Suppose $n = 2$. For $0 < \operatorname{Re}\alpha < 2$, let $c_\alpha = \Gamma((2-\alpha)/2)[\Gamma(\alpha/2)2^\alpha\pi]^{-1}$ and $Q_\alpha(x) = c_\alpha(|\xi|^{\alpha-2} - 1)$. (Note that Q_α differs by a constant from the R_α in Exercise 27.)
a. $\lim_{\alpha\to 2} Q_\alpha(x) = -(2\pi)^{-1}\log|x|$, pointwise and in $\mathcal{S}'$.
b. By (a), $\lim_{\alpha\to 2}\widehat{Q}_\alpha$ exists in $\mathcal{S}'$, and by Exercise 27b, $\widehat{Q}_\alpha(\xi) = (2\pi|\xi|)^{-\alpha} - c_\alpha\delta$. Noting that $(2\pi|\xi|)^{-2}$ is not integrable near the origin and that $\lim_{\alpha\to 2} c_\alpha = \infty$, find an explicit formula for $\lim_{\alpha\to 2}\widehat{Q}_\alpha$. (Exercise 12 may help.)

29. For $1 \le p < \infty$, let $\mathcal{C}_p$ be the set of all $F \in \mathcal{S}'$ for which there exists $C \ge 0$ such that $\|F * \phi\|_p \le C\|\phi\|_p$ for all $\phi \in \mathcal{S}$, so that the map $\phi \mapsto F * \phi$ extends to a bounded operator on L^p.

a. $\mathcal{C}_1 = M(\mathbb{R}^n)$. (If $F \in \mathcal{C}_1$, consider $F * \phi_t$ where $\{\phi_t\}$ is an approximate identity, and apply Alaoglu's theorem.)
b. $\mathcal{C}_2 = \{F \in \mathcal{S}' : \widehat{F} \in L^\infty\}$. (Use the Plancherel theorem.)
c. If p and q are conjugate exponents, then $\mathcal{C}_p = \mathcal{C}_q$. (Hint: $\langle F * \phi, \psi\rangle = \langle F * \widetilde{\psi}, \widetilde{\phi}\rangle$.)
d. If $1 \le p \le 2$ and q is the conjugate exponent to p, then $\mathcal{C}_p \subset \mathcal{C}_r$ for all $r \in (p, q)$. (Use the Riesz-Thorin theorem.)
e. $\mathcal{C}_1 \subset \mathcal{C}_p \subset \mathcal{C}_2$ for all $p \in (1, \infty)$.

9.3 SOBOLEV SPACES

One of the most satisfactory ways of measuring smoothness properties of functions and distributions is in terms of L^2 norms. There are two reasons for this: L^2 has the advantage of being a Hilbert space, and the Fourier transform, which converts differentiation into multiplication by the coordinate functions, is an isometry on L^2.

As a first step, suppose $k \in \mathbb{N}$ and let H_k be the space of all functions $f \in L^2(\mathbb{R}^n)$ whose distribution derivatives $\partial^\alpha f$ are L^2 functions for $|\alpha| \le k$. One can make H_k into a Hilbert space by imposing the inner product

$$(f, g) \mapsto \sum_{|\alpha| \le k} \int (\partial^\alpha f)(\overline{\partial^\alpha g}).$$

However, it is more convenient to use an equivalent inner product defined in terms of the Fourier transform. Theorem 8.22e and the Plancherel theorem imply that $f \in H_k$ iff $\xi^\alpha \widehat{f} \in L^2$ for $|\alpha| \le k$. A simple modification of the argument in the proof of Proposition 8.3 shows that there exist $C_1, C_2 > 0$ such that

$$C_1\big(1 + |\xi|^2\big)^k \le \sum_{|\alpha| \le k} |\xi^\alpha|^2 \le C_2\big(1 + |\xi|^2\big)^k,$$

from which it follows that $f \in H_k$ iff $(1 + |\xi|^2)^{k/2}\widehat{f} \in L^2$ and that the norms

$$f \mapsto \Big(\sum_{|\alpha| \le k} \|\partial^\alpha f\|_2^2\Big)^{1/2} \quad \text{and} \quad f \mapsto \big\|\big(1 + |\xi|^2\big)^{k/2}\widehat{f}\big\|_2$$

are equivalent. The latter norm, however, makes sense for any $k \in \mathbb{R}$, and we can use it to extend the definition of H_k to all real k.

We proceed to the formal definitions. For any $s \in \mathbb{R}$ the function $\xi \mapsto (1+|\xi|^2)^{s/2}$ is C^∞ and slowly increasing (Exercise 30), so the map Λ_s defined by

$$\Lambda_s f = \big[\big(1 + |\xi|^2\big)^{s/2}\widehat{f}\big]^\vee$$

is a continuous linear operator on $\mathcal{S}'$ — actually an isomorphism, since $\Lambda_s^{-1} = \Lambda_{-s}$. If $s \in \mathbb{R}$, we define the **Sobolev space** H_s to be

$$H_s = \big\{f \in \mathcal{S}' : \Lambda_s f \in L^2\big\},$$

and we define an inner product and norm on H_s by

$$\langle f, g\rangle_{(s)} = \int (\Lambda_s f)(\overline{\Lambda_s g}) = \int \widehat{f}(\xi)(1+|\xi|^2)^s \overline{\widehat{g}(\xi)}\, d\xi,$$
$$\|f\|_{(s)} = \|\Lambda_s f\|_2 = \left[\int |\widehat{f}(\xi)|^2 (1+|\xi|^2)^s\, d\xi\right]^{1/2}.$$

(The equality of the two formulas for $\langle f, g\rangle_{(s)}$ and for $\|f\|_{(s)}$ follows from the Plancherel theorem.) Note that the inner products $\langle\cdot,\cdot\rangle_{(s)}$ are conjugate linear in the second variable, but we are continuing to use the notation $\langle\cdot,\cdot\rangle$ for the *bilinear* pairing between $\mathcal{S}'$ and $\mathcal{S}$. This will cause no confusion, since we shall not be using the inner products $\langle\cdot,\cdot\rangle_{(s)}$ explicitly.

The following properties of Sobolev spaces are simple consequences of the definitions and the preceding discussion:

i. The Fourier transform is a unitary isomorphism from H_s to $L^2(\mathbb{R}^n, \mu_s)$ where $d\mu_s(\xi) = (1+|\xi|^2)^s\, d\xi$. In particular, H_s is a Hilbert space.

ii. $\mathcal{S}$ is a dense subspace of H_s for all $s \in \mathbb{R}$. (This follows easily from (i) and Proposition 8.17.)

iii. If $t < s$, H_s is a dense subspace of H_t in the topology of H_t, and $\|\cdot\|_{(t)} \le \|\cdot\|_{(s)}$.

iv. Λ_t is a unitary isomorphism from H_s to H_{s-t} for all $s, t \in \mathbb{R}$.

v. $H_0 = L^2$ and $\|\cdot\|_{(0)} = \|\cdot\|_2$ (by Plancherel).

vi. ∂^α is a bounded linear map from H_s to $H_{s-|\alpha|}$ for all s, α (because $|\xi^\alpha| \le (1+|\xi|^2)^{|\alpha|/2}$).

By (iii) and (v), for $s \ge 0$ the distributions in H_s are L^2 functions. For $s < 0$ the elements of H_s are generally not functions. For example, the point mass δ is in H_s iff $s < -\frac{1}{2}n$, for $\widehat{\delta}$ is the constant function 1, and $\int (1+|\xi|^2)^s\, d\xi < \infty$ iff $s < -\frac{1}{2}n$. Another example: The distribution W_t whose Fourier transform is $(2\pi|\xi|)^{-1}\sin 2\pi t|\xi|$, which arose in the discussion of the wave equation in §8.7, is in H_s iff $s < 1 - \frac{1}{2}n$; it is in $L^1 \cap L^2$ when $n = 1$ and in $L^1 \setminus L^2$ for $n = 2$, but is not a function for $n \ge 3$.

9.16 Proposition. *If $s \in \mathbb{R}$, the duality between $\mathcal{S}'$ and $\mathcal{S}$ induces a unitary isomorphism from H_{-s} to $(H_s)^*$. More precisely, if $f \in H_{-s}$, the functional $\phi \mapsto \langle f, \phi\rangle$ on $\mathcal{S}$ extends to a continuous linear functional on H_s with operator norm equal to $\|f\|_{(-s)}$, and every element of $(H_s)^*$ arises in this fashion.*

Proof. If $f \in H_{-s}$ and $\phi \in \mathcal{S}$,

$$\langle f, \phi\rangle = \langle f^\vee, \widehat{\phi}\rangle = \int f^\vee(\xi)\widehat{\phi}(\xi)\, d\xi$$

since $f^\vee(\xi) = \widehat{f}(-\xi)$ is a tempered function. Thus by the Schwarz inequality,

$$\begin{aligned}|\langle f, \phi\rangle| &\le \Big[\int |f^\vee(\xi)|^2 (1+|\xi|^2)^{-s}\, d\xi\Big]^{1/2} \Big[\int |\widehat{\phi}(\xi)|^2 (1+|\xi|^2)^{s}\, d\xi\Big]^{1/2} \\ &= \|f\|_{(-s)} \|\phi\|_{(s)},\end{aligned}$$

so the functional $\phi \mapsto \langle f, \phi\rangle$ extends continuously to H_s, with norm at most $\|f\|_{(-s)}$. In fact, its norm equals $\|f\|_{(-s)}$, since if $g \in \mathcal{S}'$ is the distribution whose Fourier transform is $\widehat{g}(\xi) = (1+|\xi|^2)^{-s}\overline{\widehat{f}(\xi)}$, we have $g \in H_s$ and

$$\langle f, g\rangle = \int |\widehat{f}(\xi)|^2 (1+|\xi|^2)^{s}\, d\xi = \|f\|_{(-s)}^2 = \|f\|_{(-s)} \|g\|_{(s)}.$$

Finally, if $G \in (H_s)^*$, then $G \circ \mathcal{F}^{-1}$ is a bounded linear functional on $L^2(\mu_s)$ where $d\mu_s(\xi) = (1+|\xi|^2)^s\, d\xi$, so there exists $g \in L^2(\mu_s)$ such that

$$G(\phi) = \int \widehat{\phi}(\xi) g(\xi) (1+|\xi|^2)^{s}\, d\xi.$$

But then $G(\phi) = \langle f, \phi\rangle$ where $f^\vee(\xi) = (1+|\xi|^2)^s g(\xi)$, and $f \in H_{-s}$ since

$$\|f\|_{(-s)}^2 = \int |\widehat{f}(\xi)|^2 (1+|\xi|^2)^{-s}\, d\xi = \int |g(\xi)|^2 (1+|\xi|^2)^{s}\, d\xi.$$

∎

For $s > 0$, the elements of H_s are L^2 functions that are "L^2-differentiable up to order s," and it is natural to ask what is the relationship between this notion of smoothness and ordinary differentiability. Of course, if one thinks of elements of H_s as distributions or elements of L^2, there is no distinction among functions that agree almost everywhere; from this perspective, when one says that a function in H_s is of class C^k, one means that it agrees a.e. with a C^k function. With this understanding, the question just posed has a simple and elegant answer. We introduce the notation

$$C_0^k = \{f \in C^k(\mathbb{R}^n) : \partial^\alpha f \in C_0 \text{ for } |\alpha| \le k\}.$$

C_0^k is a Banach space with the C^k norm $f \mapsto \sum_{|\alpha|\le k} \|\partial^\alpha f\|_u$.

9.17 The Sobolev Embedding Theorem. *Suppose $s > k + \frac{1}{2}n$.*

a. If $f \in H_s$, then $(\partial^\alpha f)\widehat{\ } \in L^1$ and $\|(\partial^\alpha f)\widehat{\ }\|_1 \le C\|f\|_{(s)}$ for $|\alpha| \le k$, where C depends only on $k - s$.

b. $H_s \subset C_0^k$, and the inclusion map is continuous.

Proof. By the Schwarz inequality,

$$\begin{aligned}(2\pi)^{-|\alpha|} \int |(\partial^\alpha f)\widehat{\ }(\xi)|\, d\xi &= \int |\xi^\alpha \widehat{f}(\xi)|\, d\xi \le \int (1+|\xi|^2)^{k/2} |\widehat{f}(\xi)|\, d\xi \\ &\le \Big[\int (1+|\xi|^2)^s |\widehat{f}(\xi)|^2 d\xi\Big]^{1/2} \Big[\int (1+|\xi|^2)^{k-s} d\xi\Big]^{1/2}.\end{aligned}$$

The first factor on the right is $\|f\|_{(s)}$, and the second one is finite by Corollary 2.52 since $2(k-s) < -n$. This proves (a), and (b) follows by the Fourier inversion theorem and the Riemann-Lebesgue lemma. ∎

9.18 Corollary. *If $f \in H_s$ for all s, then $f \in C^\infty$.*

An example may help to elucidate this theorem. Let $f_\lambda(x) = \phi(x)|x|^\lambda$, where $\lambda \in \mathbb{R}$ and $\phi \in C_c^\infty$ with $\phi = 1$ on a neighborhood of 0. Then the (classical) derivative $\partial^\alpha f_\lambda$ is C^∞ except at 0 and is homogeneous of degree $\lambda - |\alpha|$ near 0, so that $|\partial^\alpha f_\lambda| \le C_{\alpha,\lambda}|x|^{\lambda-|\alpha|}$, and in particular $\partial^\alpha f_\lambda \in L^1$ provided $\lambda - |\alpha| > -n$. In this case $\partial^\alpha f_\lambda$, as an L^1 function, is also the distribution derivative of f_λ. (To see this, replace f_λ by the C^∞ function $\phi(x)(|x|^2+\epsilon^2)^{\lambda/2}$ and consider the limit as $\epsilon \to 0$.) Moreover, $\partial^\alpha f_\lambda \in L^2$ iff $\lambda - |\alpha| > -\frac{1}{2}n$, so $f \in H_k$ $(k = 0, 1, 2, \ldots)$ iff $\lambda > k - \frac{1}{2}n$, whereas $f_\lambda \in C_0^k$ iff $\lambda > k$. See also Exercises 33–35 for some related results.

Next, we show that multiplication by suitably smooth functions preserves the H_s spaces. We need a lemma:

9.19 Lemma. *For all $\xi, \eta \in \mathbb{R}^n$ and $s \in \mathbb{R}$,*

$$\left(1+|\xi|^2\right)^s\left(1+|\eta|^2\right)^{-s} \le 2^{|s|}\left(1+|\xi-\eta|^2\right)^{|s|}.$$

Proof. Since $|\xi| \le |\xi-\eta| + |\eta|$, we have $|\xi|^2 \le 2(|\xi-\eta|^2 + |\eta|^2)$ and hence

$$1+|\xi|^2 \le 2\left(1+|\xi-\eta|^2\right)\left(1+|\eta|^2\right).$$

If $s \ge 0$, we have merely to raise both sides to the sth power. If $s < 0$, we interchange ξ and η and replace s by $-s$, obtaining

$$\left(1+|\eta|^2\right)^{-s} \le 2^{-s}\left(1+|\xi|^2\right)^{-s}\left(1+|\xi-\eta|^2\right)^{-s},$$

which is again the desired result. ∎

9.20 Theorem. *Suppose that $\phi \in C_0(\mathbb{R}^n)$ and that $\widehat{\phi}$ is a function that satisfies*

$$\int \left(1+|\xi|^2\right)^{a/2}|\widehat{\phi}(\xi)|\,d\xi = C < \infty$$

for some $a > 0$. Then the map $M_\phi(f) = \phi f$ is a bounded operator on H_s for $|s| \le a$.

Proof. Since Λ_s is a unitary map from H_s to $H_0 = L^2$, it is equivalent to show that $\Lambda_s M_\phi \Lambda_{-s}$ is a bounded operator on L^2. But

$$(\Lambda_s M_\phi \Lambda_{-s} f)\widehat{\ }(\xi) = \left(1+|\xi|^2\right)^{s/2}\left[\widehat{\phi} * (\Lambda_{-s} f)\widehat{\ }\right](\xi) = \int K(\xi,\eta)\widehat{f}(\eta)\,d\eta,$$

where

$$K(\xi,\eta) = \left(1+|\xi|^2\right)^{s/2}\left(1+|\eta|^2\right)^{-s/2}\widehat{\phi}(\xi-\eta).$$

By Lemma 9.19,

$$|K(\xi,\eta)| \le 2^{|s|/2}\left(1+|\xi-\eta|^2\right)^{|s|/2}|\widehat{\phi}(\xi-\eta)|,$$

so if $|s| \le a$, then $\int |K(\xi,\eta)|\,d\xi$ and $\int |K(\xi,\eta)|\,d\eta$ are bounded by $2^{a/2}C$. That $\Lambda_s M_\phi \Lambda_{-s}$ is bounded on L^2 therefore follows from the Plancherel theorem and Theorem 6.18. ∎

9.21 Corollary. *If $\phi \in \mathcal{S}$, then M_ϕ is a bounded operator on H_s for all $s \in \mathbb{R}$.*

Our next result is a compactness theorem that is of great importance in the applications of Sobolev spaces.

9.22 Rellich's Theorem. *Suppose that $\{f_k\}$ is a sequence of distributions in H_s that are all supported in a fixed compact set K and satisfy $\sup_k \|f_k\|_{(s)} < \infty$. Then there is a subsequence $\{f_{k_j}\}$ that converges in H_t for all $t < s$.*

Proof. First we observe that by Proposition 9.11, $\widehat{f_k}$ is a slowly increasing C^∞ function. Pick $\phi \in C_c^\infty$ such that $\phi = 1$ on a neighborhood of K. Then $f_k = \phi f_k$, so $\widehat{f_k} = \widehat{\phi} * \widehat{f_k}$ where the convolution is defined pointwise by an absolutely convergent integral. By Lemma 9.19 and the Schwarz inequality,

$$\begin{aligned}
&\left(1+|\xi|^2\right)^{s/2}|\widehat{f_k}(\xi)| \\
&\le 2^{|s|/2}\int |\widehat{\phi}(\xi-\eta)|\left(1+|\xi-\eta|^2\right)^{|s|/2}|\widehat{f_k}(\eta)|\left(1+|\eta|^2\right)^{s/2}d\eta \\
&\le 2^{|s|/2}\|\phi\|_{(|s|)}\|f_k\|_{(s)} \le \text{constant}.
\end{aligned}$$

Likewise, since $\partial_j(\widehat{\phi} * \widehat{f_k}) = (\partial_j\widehat{\phi}) * \widehat{f_k}$, we see that $(1+|\xi|^2)^{s/2}|\partial_j\widehat{f_k}(\xi)|$ is bounded by a constant independent of ξ, j, and k. In particular, the $\widehat{f_k}$'s and their first derivatives are uniformly bounded on compact sets, so by the mean value theorem and the Arzelà-Ascoli theorem there is a subsequence $\{\widehat{f_{k_j}}\}$ that converges uniformly on compact sets.

We claim that $\{f_{k_j}\}$ is Cauchy in H_t for all $t < s$. Indeed, for any $R > 0$ we can write the integral

$$\|f_{k_i} - f_{k_j}\|_{(t)}^2 = \int \left(1+|\xi|^2\right)^t |\widehat{f_{k_i}} - \widehat{f_{k_j}}|^2(\xi)\,d\xi$$

as the sum of the integrals over the regions $|\xi| \le R$ and $|\xi| > R$. For $|\xi| \le R$ we use the estimate

$$\left(1+|\xi|^2\right)^t \le \left(1+R^2\right)^{\max(t,0)},$$

and for $|\xi| > R$ we use the estimate

$$(1+|\xi|^2)^t \le (1+R^2)^{t-s}(1+|\xi|^2)^s,$$

which yield

$$\|f_{k_i} - f_{k_j}\|_{(t)}^2 \le CR^n(1+R^2)^{\max(t,0)} \sup_{|\xi|\le R} |\widehat{f}_{k_i} - \widehat{f}_{k_j}|^2(\xi)$$
$$+(1+R^2)^{t-s}\|f_{k_i} - f_{k_j}\|_{(s)}^2.$$

Given $\epsilon > 0$, the second term will be less than $\frac{1}{2}\epsilon$ provided R is chosen sufficiently large, since $t-s<0$; once such an R is fixed, the first term will less than $\frac{1}{2}\epsilon$ provided i and j are sufficiently large. The proof is therefore complete. ∎

Although the definition of Sobolev spaces in terms of the Fourier transform entails their elements being defined on all of $\mathbb{R}^n$, these spaces can also be used in the study of local smoothness properties of functions. The key definition is as follows: If U is an open set in $\mathbb{R}^n$, the **localized Sobolev space** $H_s^{\text{loc}}(U)$ is the set of all distributions $f \in \mathcal{D}'(U)$ such that for every precompact open set V with $\overline{V} \subset U$ there exists $g \in H_s$ such that $g = f$ on V.

9.23 Proposition. *A distribution $f \in \mathcal{D}'(U)$ is in $H_s^{\text{loc}}(U)$ iff $\phi f \in H_s$ for every $\phi \in C_c^\infty(U)$.*

Proof. If $f \in H_s^{\text{loc}}(U)$ and $\phi \in C_c^\infty(U)$, then f agrees with some $g \in H_s$ on a neighborhood of $\text{supp}(\phi)$; hence $\phi f = \phi g \in H_s$ by Corollary 9.21. For the converse, given a precompact open V with $\overline{V} \subset U$, we can choose $\phi \in C_c^\infty(U)$ with $\overline{\phi} = 1$ on a neighborhood of $\overline{V}$ by the C^∞ Urysohn lemma; then $\phi f \in H_s$ and $\phi f = f$ on V. (We have implicitly used Proposition 4.31 to obtain compact neighborhoods of $\text{supp}\,\phi$ and $\overline{V}$ in U.) ∎

We conclude this section with one of the classic applications of Sobolev spaces, a regularity theorem for certain partial differential operators.

If $L = \sum_0^m a_j(d/dx)^j$ is an ordinary differential operator with C^∞ coefficients such that a_m never vanishes, it is not hard to show that smooth data give smooth solutions. More precisely, if $Lu = f$ and f is C^k on an open interval I, then u is C^{k+m} on I. No such result holds for partial differential operators in general. For example, for any $f \in L^1_{\text{loc}}(\mathbb{R})$ the function $u(x,t) = f(x-t)$ satisfies the wave equation $(\partial_t^2 - \partial_x^2)u = 0$, but u has only as much smoothness as f. However, there is a large class of differential operators for which a strong regularity theorem holds. We restrict attention to the constant-coefficient case, although the results are valid in greater generality.

Let $P(D) = \sum_{|\alpha|\le m} c_\alpha D^\alpha$ (notation as in §8.7) be a constant-coefficient operator. We assume that m is the true order of $P(D)$, i.e., that $c_\alpha \ne 0$ for some α with $|\alpha| = m$. The **principal symbol** P_m is the sum of the top-order terms in its symbol:

$$P_m(\xi) = \sum_{|\alpha|=m} c_\alpha \xi^\alpha.$$

$P(D)$ is called **elliptic** if $P_m(\xi) \neq 0$ for all nonzero $\xi \in \mathbb{R}^n$. Thus, ellipticity means that, in a formal sense, $P(D)$ is genuinely mth order in all directions. (For example, the Laplacian Δ is elliptic on $\mathbb{R}^n$, whereas the heat and wave operators $\partial_t - \Delta$ and $\partial_t^2 - \Delta$ are not elliptic on $\mathbb{R}^{n+1}$.)

9.24 Lemma. *Suppose that $P(D)$ is of order m. Then $P(D)$ is elliptic iff there exist $C, R > 0$ such that $|P(\xi)| \geq C|\xi|^m$ when $|\xi| \geq R$.*

Proof. If $P(D)$ is elliptic, let C_1 be the minimum value of the principal symbol P_m on the unit sphere $|\xi| = 1$. Then $C_1 > 0$, and since P_m is homogeneous of degree m, we have $|P_m(\xi)| \geq C_1|\xi|^m$ for all ξ. On the other hand, $P - P_m$ is of order $m-1$, so there exists C_2 such that $|P(\xi) - P_m(\xi)| \leq C_2|\xi|^{m-1}$. Therefore,

$$|P(\xi)| \geq |P_m(\xi)| - |P(\xi) - P_m(\xi)| \geq \tfrac{1}{2}C_1|\xi|^m \text{ for } |\xi| \geq 2C_2C_1^{-1}.$$

Conversely, if $P(D)$ is not elliptic, say $P_m(\xi_0) = 0$, then $|P(\xi)| \leq C|\xi|^{m-1}$ for every scalar multiple ξ of ξ_0. ∎

9.25 Lemma. *If $P(D)$ is elliptic of order m, $u \in H_s$, and $P(D)u \in H_s$, then $u \in H_{s+m}$.*

Proof. The hypotheses say that $(1+|\xi|^2)^{s/2}\widehat{u} \in L^2$ and $(1+|\xi|^2)^{s/2}P\widehat{u} \in L^2$. By Lemma 9.24, for some $R \geq 1$ we have

$$\left(1+|\xi|^2\right)^{m/2} \leq 2^m|\xi|^m \leq C^{-1}2^m|P(\xi)| \text{ for } |\xi| \geq R,$$

and $(1+|\xi|^2)^{m/2} \leq (1+R^2)^{m/2}$ for $|\xi| \leq R$. It follows that

$$\left(1+|\xi|^2\right)^{(s+m)/2}|\widehat{u}| \leq C'\left(1+|\xi|^2\right)^{s/2}\left(|P\widehat{u}| + |\widehat{u}|\right) \in L^2,$$

that is, $u \in H_{s+m}$. ∎

9.26 The Elliptic Regularity Theorem. *Suppose that L is a constant-coefficient elliptic differential operator of order m, Ω is an open set in $\mathbb{R}^n$, and $u \in \mathcal{D}'(\Omega)$. If $Lu \in H_s^{\mathrm{loc}}(\Omega)$ for some $s \in \mathbb{R}$, then $u \in H_{s+m}^{\mathrm{loc}}(\Omega)$; and if $Lu \in C^\infty(\Omega)$, then $u \in C^\infty(\Omega)$.*

Proof. The second assertion follows from the first in view of Corollary 9.18, so by Proposition 9.23 we must show that if $Lu \in H_s^{\mathrm{loc}}(\Omega)$ and $\phi \in C_c^\infty(\Omega)$, then $\phi u \in H_{s+m}$. Let V be a precompact open set such that $\mathrm{supp}(\phi) \subset V \subset \overline{V} \subset \Omega$, and choose $\psi \in C_c^\infty(\Omega)$ such that $\psi = 1$ on $\overline{V}$. Then $\psi u \in \mathcal{E}'$, so it follows from Proposition 9.11 that $\psi u \in H_\sigma$ for some $\sigma \in \mathbb{R}$. By decreasing σ we may assume that $s + m - \sigma$ is a positive integer k. Set $\psi_0 = \psi$ and $\psi_k = \phi$, and choose recursively $\psi_1, \ldots, \psi_{k-1} \in C_c^\infty$ such that $\psi_j = 1$ on a neighborhood of $\mathrm{supp}(\phi)$ and $\mathrm{supp}(\psi_j)$ is contained in the set where $\psi_{j-1} = 1$. We shall prove by induction

that $\psi_j u \in H_{\sigma+j}$. When $j = k$, we obtain $\phi u = \psi_k u \in H_{\sigma+k} = H_m$, which will complete the proof.

The crucial observation is that for any $\zeta \in C_c^\infty$ the operator $[L, \zeta]$ defined by

$$[L, \zeta]f = L(\zeta f) - \zeta L f$$

is a differential operator of order $m-1$ whose coefficients are linear combinations of derivatives of ζ; in particular, these coefficients are C^∞ functions that vanish on any open set where ζ is constant. (This follows from the product rule for derivatives.) Thus, if $f \in H_t$, we have $\partial^\alpha f \in H_{t-(m-1)}$ for $|\alpha| \le m-1$ and hence $[L, \zeta]f \in H_{t-(m-1)}$ by Theorem 9.20.

For $j = 0$ we have $\psi_0 u \in H_\sigma$ by assumption. Suppose we have established that $\psi_j u \in H_{\sigma+j}$, where $0 \le j < k$. Then by the preceding remarks,

$$\begin{aligned} L(\psi_{j+1}u) = \psi_{j+1}Lu + [L, \psi_{j+1}]u &= \psi_{j+1}Lu + [L, \psi_{j+1}]\psi_j u \\ &\in H_s + H_{\sigma+j-(m-1)} = H_{\sigma+j+1-m}. \end{aligned}$$

Since $\psi_{j+1}u = \psi_{j+1}\psi_j u \in H_{\sigma+j}$, Lemma 9.25 (with $P(D) = L$) implies that $\psi_{j+1}u \in H_{\sigma+j+1}$, and we are done. ∎

Two classical special cases of this theorem are particularly noteworthy. First, every distribution solution of Laplace's equation $\Delta u = 0$ is a C^∞ function. (This fact is known as **Weyl's lemma**.) Second, if $L = \partial_1 + i\partial_2$ on $\mathbb{R}^2$, the equation $Lu = 0$ is the **Cauchy-Riemann equation**, whose solutions are the holomorphic (or analytic) functions of $z = x_1 + ix_2$. We thus recover the fact that holomorphic functions are C^∞.

Exercises

30. Let $f_s(\xi) = (1 + |\xi|^2)^{s/2}$. Then $|\partial^\alpha f_s(\xi)| \le C_\alpha(1 + |\xi|)^{s-|\alpha|}$.

31. If $k \in \mathbb{N}$, H_k is the space of all $f \in L^2$ that possess strong L^2 derivatives $\partial^\alpha f$, as defined in Exercise 8 in §8.2, for $|\alpha| \le k$; and these strong derivatives coincide with the distribution derivatives.

32. Suppose $r < s < t$. For any $\epsilon > 0$ there exists $C > 0$ such that $\|f\|_{(s)} \le \epsilon\|f\|_{(t)} + C\|f\|_{(r)}$ for all $f \in H_t$.

33. (**Converse of the Sobolev Theorem**) If $H_s \subset C_0^k$, then $s > k + \frac{1}{2}n$. (Use the closed graph theorem to show that the inclusion map $H_s \to C_0^k$ is continuous and hence that $\partial^\alpha \delta \in (H_s)^*$ for $|\alpha| \le k$.)

34. (**A Sharper Sobolev Theorem**) For $0 < \alpha < 1$, let

$$\Lambda_\alpha(\mathbb{R}^n) = \left\{ f \in BC(\mathbb{R}^n) : \sup_{x \ne y} \frac{|f(x) - f(y)|}{|x-y|^\alpha} < \infty \right\}.$$

a. If $s = \frac{1}{2}n + \alpha$ where $0 < \alpha < 1$, then $\|\tau_x\delta - \tau_y\delta\|_{(-s)} \le C_\alpha|x-y|^\alpha$. (We have $(\tau_x\delta)\widehat{\ }(\xi) = e^{-2\pi i\xi\cdot x}$. Write the integral defining $\|\tau_x\delta - \tau_y\delta\|_{(-s)}^2$ as the

sum of the integrals over the regions $|\xi| \le R$ and $|\xi| > R$, where $R = |x-y|^{-1}$, and use the mean value theorem to estimate $(\tau_x\delta - \tau_y\delta)\widehat{\ }$ on the first region.)

b. If $s = \frac{1}{2}n + \alpha$ where $0 < \alpha < 1$, then $H_s \subset \Lambda_\alpha(\mathbb{R}^n)$.

c. If $s = \frac{1}{2}n + k + \alpha$ where $k \in \mathbb{N}$ and $0 < \alpha < 1$, then

$$H_s \subset \{f \in C_0^k : \partial^\alpha f \in \Lambda_a(\mathbb{R}^n) \text{ for } |\alpha| \le k\}.$$

35. The Sobolev theorem says that if $s > \frac{1}{2}n$, it makes sense to evaluate functions in H_s at a point. For $0 \le s \le \frac{1}{2}n$, functions in H_s are only defined a.e., but if $s > \frac{1}{2}k$ with $k < n$, it makes sense to restrict functions in H_s to subspaces of codimension k. More precisely, let us write $\mathbb{R}^n = \mathbb{R}^{n-k} \times \mathbb{R}^k$, $x = (y, z)$, $\xi = (\eta, \zeta)$, and define $R : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^{n-k})$ by $Rf(y) = f(y, 0)$.

a. $(Rf)\widehat{\ }(\eta) = \int \widehat{f}(\eta, \zeta)\, d\zeta$. (See Exercise 20 in §8.3.)

b. If $s > \frac{1}{2}k$,

$$|(Rf)\widehat{\ }(\eta)|^2 \le C_s\big(1 + |\eta|^2\big)^{(k/2)-s} \int |\widehat{f}(\eta, \zeta)|^2 \big(1 + |\eta|^2 + |\zeta|^2\big)^s\, d\zeta.$$

c. R extends to a bounded map from $H_s(\mathbb{R}^n)$ to $H_{s-(k/2)}(\mathbb{R}^{n-k})$ provided $s > \frac{1}{2}k$.

36. Suppose that $0 \ne \phi \in C_c^\infty$ and $\{a_j\}$ is a sequence in $\mathbb{R}^n$ with $|a_j| \to \infty$, and let $\phi_j(x) = \phi(x - a_j)$. Then $\{\phi_j\}$ is bounded in H_s for every s but has no convergent subsequence in H_t for any t.

37. The heat operator $\partial_t - \Delta$ is not elliptic, but a weakened version of Theorem 9.26 holds for it. Here we are working on $\mathbb{R}^{n+1}$ with coordinates (x, t) and dual coordinates (ξ, τ), and $\partial_t - \Delta = P(D)$ where $P(\xi, \tau) = 2\pi i\tau + 4\pi^2|\xi|^2$.

a. There exist $C, R > 0$ such that $|\xi|\,|(\xi, \tau)|^{1/2} \le C|P(\xi, \tau)|$ for $|(\xi, \tau)| > R$. (Consider the regions $|\tau| \le |\xi|^2$ and $|\tau| \ge |\xi|^2$ separately.)

b. If $f \in H_s$ and $(\partial_t - \Delta)f \in H_s$, then $f \in H_{s+1}$ and $\partial_{x_i} f \in H_{s+(1/2)}$ for $1 \le i \le n$.

c. If $\zeta \in C_c^\infty(\mathbb{R}^{n+1})$, we have

$$[\partial_t - \Delta, \zeta]f = (\partial_t\zeta - \Delta\zeta)f - 2\sum(\partial_{x_i}\zeta)(\partial_{x_i} f).$$

d. If Ω is open in $\mathbb{R}^{n+1}$, $u \in \mathcal{D}'(\Omega)$, and $(\partial_t - \Delta)u \in H_s^{\mathrm{loc}}(\Omega)$, then $u \in H_{s+1}^{\mathrm{loc}}(\Omega)$. (Let ψ_j be as in the proof of Theorem 9.26. Show inductively that if $\psi_0 u \in H_\sigma$, then $\psi_j u \in H_{\sigma+(j/2)}$ and $\partial_{x_i}(\psi_j u) \in H_{\sigma+(j-1)/2}$ provided $\sigma + \frac{1}{2}j \le s$.)

38. Suppose $s_0 \le s_1$ and $t_0 \le t_1$, and for $0 \le \lambda \le 1$ let

$$s_\lambda = (1-\lambda)s_0 + \lambda s_1, \qquad t_\lambda = (1-\lambda)t_0 + \lambda t_1.$$

If T is a bounded linear map from H_{s_0} to H_{t_0} whose restriction to H_{s_1} is bounded from H_{s_1} to H_{t_1}, then the restriction of T to H_{s_λ} is bounded from H_{s_λ} to H_{t_λ} for

$0 \le \lambda \le 1$. (T is bounded from H_s to H_t iff $\Lambda_s T \Lambda_{-t}$ is bounded on L^2. Observe that Λ_z is well defined for all $z \in \mathbb{C}$ and Λ_z is unitary on every H_s if $\operatorname{Re} z = 0$. Let $s(z) = (1-z)s_0 + zs_1$, $t(z) = (1-z)t_0 + zt_1$, and for $0 \le \operatorname{Re} z \le 1$ and $\phi, \psi \in \mathcal{S}$ let $F(z) = \int [\Lambda_{t(z)} T \Lambda_{-s(z)} \phi]\psi$. Apply the three lines lemma as in the proof of the Riesz-Thorin theorem.)

39. Let Ω be an open set in $\mathbb{R}^n$, and let $G : \Omega \to \mathbb{R}^n$ be a C^∞ diffeomorphism. For any $\phi \in C_c^\infty(G(\Omega))$, the map $Tf = (\phi f) \circ G$ is bounded on H_s for all s; consequently, $f \circ G \in H_s^{\mathrm{loc}}(\Omega)$ whenever $f \in H_s^{\mathrm{loc}}(G(\Omega))$. Proceed as follows:

a. If $s = 0, 1, 2, \ldots$, use the chain rule and the fact that $f \in H_s$ iff $\partial^\alpha f \in L^2$ for $|\alpha| \le s$.

b. Use Exercise 38 to obtain the result for all $s > 0$.

c. For $s < 0$, use Proposition 9.16 and the fact that the transpose of T is another operator of the same type, namely, $T'f = (\psi f) \circ H$ where $H = G^{-1}$ and $\psi = (J\phi) \circ G$ with $J(x) = |\det D_x G|$.

40. State and prove analogues of the results in this section for the periodic Sobolev spaces

$$H_s(\mathbb{T}^n) = \left\{ f \in \mathcal{D}'(\mathbb{T}^n) : \sum (1 + |\kappa|^2)^s |\widehat{f}(\kappa)|^2 < \infty \right\}.$$

9.4 NOTES AND REFERENCES

The mathematical foundations for the theory of distributions were largely laid in the 1930s. On the one hand, several researchers in partial differential equations arrived at the notion of "weak derivatives" of functions; to wit, if $f, g \in L^1_{\mathrm{loc}}(U)$, g is the derivative $\partial^\alpha f$ in the weak sense if $\int g\phi = (-1)^{|\alpha|} \int f \partial^\alpha \phi$ for all ϕ in some suitable space of test functions. On the other, various attempts were made to extend the domain of the Fourier transform beyond $L^1 + L^2$. The idea of defining "generalized functions" as linear functionals on certain function spaces goes back to Sobolev [136], but it was Laurent Schwartz who systematically developed the theory of distributions and who introduced the spaces $\mathcal{S}$ and $\mathcal{S}'$ as natural domains for the Fourier transform. (See Dieudonné [33] for a more detailed historical account.)

Rudin [126] contains a good concise introduction to the theory of distributions that includes some functional-analytic points we have elided, such as the definition of the topology on $C_c^\infty(U)$ and the properties of convolution on $\mathcal{D}' \times \mathcal{E}'$. More extensive treatments can be found in Gelfand and Shilov [55], Schwartz [132], and Treves [150]. Hörmander [77] contains an excellent full-scale treatment of distribution theory with a view toward its applications to differential equations.

§9.2: See Folland [49] for a case study of the analytic techniques used in manipulating distributions and their Fourier transforms and applying them to differential equations.

§9.3: The spaces originally considered by Sobolev [137] are the spaces H_k^p of functions $f \in L^p$ whose distribution derivatives $\partial^\alpha f$ are in L^p for $|\alpha| \le k$. When

$1 < p < \infty$, it turns out that $H_k^p = \{f : \Lambda_s f \in L^p\}$ (although this is far from obvious when $p \neq 2$), and this characterization of H_k^p can be used to define H_s^p for all $s \in \mathbb{R}$. Sobolev's embedding theorem, in this setting, is that if $s < n/p$ then $H_s^p \subset L^q$ where $q^{-1} = p^{-1} - n^{-1}s$, and if $s > k + n/p$ then $H_s^p \subset C^k$. See Stein [140, §§V.2–3]. Further results on Sobolev space and their applications can be found in Adams [1] and Lieb and Loss [93].

In Rellich's theorem the hypothesis that K is compact can be replaced by the hypothesis that $m(K) < \infty$ when $s \geq 0$; see Lair [89].

A differential operator $L = \sum_{|\alpha| \leq m} a_\alpha D^\alpha$ with C^∞ coefficients is called **elliptic** on $\Omega \subset \mathbb{R}^n$ if $\sum_{|\alpha|=m} a_\alpha(x)\xi^\alpha \neq 0$ for all $x \in \Omega$ and all nonzero $\xi \in \mathbb{R}^n$. The elliptic regularity theorem remains true as stated for such operators; see Folland [48]. The L^p version of this theorem is also valid for $1 < p < \infty$, but not for $p = 1$ or $p = \infty$. It is not true, except in dimension 1, that if $Lu \in C^k(\Omega)$ then $u \in C^{k+m}(\Omega)$, but if $\partial^\alpha Lu$ is not just continuous but Hölder continuous of exponent λ $(0 < \lambda < 1)$ for $|\alpha| = k$, then $u \in C^{k+m}(\Omega)$ and $\partial^\beta u$ satisfies the same Hölder condition for $|\beta| = k + m$. See Taylor [147, Chapter XI].

10

Topics in Probability Theory

Probability theory, originally conceived to analyze games of chance, has developed into a broadly useful discipline with deep connections to other branches of mathematics and many applications to other subjects such as physics, statistics, and economics. The mathematical study of probability began some two and a half centuries ago, but until the advent of modern analytic tools the theory was limited to combinatorial theorems involving discrete sample spaces and a few other results of somewhat doubtful rigor. It is now recognized that the fundamental datum in probability theory is a measure space $(X, \mathcal{M}, \mu)$ such that $\mu(X) = 1$; such a measure μ is called a **probability measure**. X is to be considered as the set of all possible outcomes of some process, such as an experiment or a gambling game, and the measure of a set $E \in \mathcal{M}$ is interpreted as the probability that the outcome lies in E. Although measure spaces are the natural setting for the study of probability, it is hardly accurate to say that probability theory is a branch of measure theory, for its central ideas and many of its techniques are distinctively its own.

This brief chapter is intended not as a systematic introduction to probability theory but rather as an advertisement for the subject; it also serves to illustrate further some results of previous chapters.

10.1 BASIC CONCEPTS

Probability theory has its own vocabulary, which is partly a legacy of its development before the connection with measure theory was made explicit and partly a result of the

fact that the probabilistic point of view is different. We therefore begin by presenting a brief dictionary of probabilists' dialect.

Analysts' Term	*Probabilists' Term*
Measure space $(X, \mathcal{M}, \mu)$ $(\mu(X) = 1)$	Sample space $(\Omega, \mathcal{B}, P)$
(σ-)algebra	(σ-)field
Measurable set	Event
Measurable real-valued function f	Random variable X
Integral of f, $\int f\, d\mu$	Expectation or mean of X, $E(X)$
L^p [as adjective]	Having finite pth moment
Convergence in measure	Convergence in probability
Almost every(where), a.e.	Almost sure(ly), a.s.
Borel probability measure on $\mathbb{R}$	Distribution
Fourier transform of a measure	Characteristic function of a distribution
Characteristic function	Indicator function

Probabilists have an aversion to displaying the arguments of random variables. For example, $\{\omega : X(\omega) > a\}$ and $P(\{\omega : X(\omega) > a\})$ are commonly written as $\{X > a\}$ and $P(X > a)$.

Henceforth we shall, for the most part, adopt probabilistic language in this chapter, although we shall use the term "L^p random variable" in preference to the more cumbersome "random variable with finite pth moment." One more standard piece of terminology, which has no equivalent in classical analysis, is the following: If X is a random variable, its **variance** $\sigma^2(X)$ and **standard deviation** $\sigma(X)$ are defined by

$$\sigma^2(X) = \inf_{a \in \mathbb{R}} E\big[(X - a)^2\big], \qquad \sigma(X) = \sqrt{\sigma^2(X)}.$$

If $X \notin L^2$, then $\sigma^2(X) = \infty$. If $X \in L^2$, then $E[(X-a)^2] = E(X^2) - 2aE(X) + a^2$ is a quadratic function of a whose minimum occurs when $a = E(X)$; hence

$$\sigma^2(X) = E\big[(X - E(X))^2\big] = E(X^2) - E(X)^2 \qquad (X \in L^2).$$

$\sigma(X)$ is a measure of how widely X deviates from its mean $E(X)$.

At this point we must discuss a general measure-theoretic construction. Let $(\Omega, \mathcal{B}, P)$ be a probability space (or, for that matter, an arbitrary measure space), let $(\Omega', \mathcal{B}')$ be another measurable space, and let $\phi : \Omega \to \Omega'$ be a $(\mathcal{B}, \mathcal{B}')$-measurable map. Then the measure P induces an **image measure** P_ϕ on Ω' by

$$P_\phi(E) = P\big(\phi^{-1}(E)\big).$$

That this is indeed a measure follows from the fact that ϕ^{-1} commutes with unions and intersections.

10.1 Proposition. *With notation as above, if $f : \Omega' \to \mathbb{R}$ is a measurable function, then $\int_{\Omega'} f\, dP_\phi = \int_\Omega (f \circ \phi)\, dP$ whenever either side is defined.*

Proof. When $f = \chi_E$ with $E \in \mathcal{B}'$, this is just the definition of P_ϕ, since $\chi_E \circ \phi = \chi_{\phi^{-1}(E)}$. The general result follows by taking linear combinations and limits. ∎

If X is a random variable on Ω, then P_X is a probability measure on $\mathbb{R}$, called the **distribution** of X, and the function

$$F(t) = P_X\big((-\infty, t]\big) = P(X \le t)$$

(which determines P_X by Theorem 1.16) is called the **distribution function** of X. If $\{X_\alpha\}_{\alpha \in A}$ is a family of random variables such that $P_{X_\alpha} = P_{X_\beta}$ for all $\alpha, \beta \in A$, the X_α's are said to be **identically distributed**.

More generally, for any finite sequence $X_1, \ldots, X_n$ of random variables, we can consider $(X_1, \ldots, X_n)$ as a map from Ω to $\mathbb{R}^n$, and the measure $P_{(X_1,\ldots,X_n)}$ on $\mathbb{R}^n$ is called the **joint distribution** of $X_1, \ldots, X_n$. It is a general principle that all properties of random variables that are relevant to probability theory can be expressed in terms of their joint distributions. For example, by Proposition 10.1,

$$E(X) = \int t\, dP_X(t), \qquad \sigma^2(X) = \int \big(t - E(X)\big)^2\, dP_X(t),$$
$$E(X+Y) = \int (t+s)\, dP_{(X,Y)}(t,s).$$

In fact, given a Borel probability measure λ on $\mathbb{R}$, one can simply speak of the **mean** $\overline{\lambda}$ and **variance** σ^2 of λ,

$$\overline{\lambda} = \int t\, d\lambda(t), \qquad \sigma^2 = \int (t - \overline{\lambda})^2\, d\lambda(t),$$

which are the mean and variance of any random variable with distribution λ.

One of the most important concepts in probability theory, and the one that most clearly sets it apart from general measure theory, is that of (stochastic) independence. To motivate this idea, consider a probability space $(\Omega, \mathcal{F}, P)$ and an event E such that $P(E) > 0$. Then the set function $P_E(F) = P(E \cap F)/P(E)$ is a probability measure on Ω called the conditional probability on E; $P_E(F)$ represents the probability of the event F given that E occurs. If $P_E(F) = P(F)$, that is, if the probability of F is the same whether or not we restrict to E, then F is said to be independent of E. Thus, F is independent of E iff $P(E \cap F) = P(E)P(F)$; moreover, the latter condition is clearly symmetric in E and F and makes sense even if $P(E) = 0$.

With this in mind, we define a collection $\{E_\alpha\}_{\alpha \in A}$ of events in Ω to be **independent** if

$$P(E_{\alpha_1} \cap \cdots \cap E_{\alpha_n}) = \prod_1^n P(E_{\alpha_j}) \text{ for all } n \in \mathbb{N} \text{ and all distinct } \alpha_1 \ldots, \alpha_n \in A.$$

(For the events E_α to be independent it does *not* suffice for them to be pairwise independent; see Exercise 1.)

A collection $\{X_\alpha\}_{\alpha \in A}$ of random variables on Ω is called **independent** if the events $\{X_\alpha \in B_\alpha\} = X_\alpha^{-1}(B_\alpha)$ are independent for all Borel sets $B_\alpha \subset \mathbb{R}$. This condition can be neatly rephrased as follows. Observe that if $\alpha_1, \ldots, \alpha_n \in A$ and

we write $X_j = X_{\alpha_j}$, we have

$$\begin{aligned} P\big(X_1^{-1}(B_1) \cap \cdots \cap X_n^{-1}(B_n)\big) &= P\big((X_1, \ldots, X_n)^{-1}(B_1 \times \cdots \times B_n)\big) \\ &= P_{(X_1,\ldots,X_n)}(B_1 \times \cdots \times B_n), \end{aligned}$$

whereas

$$\prod_1^n P\big(X_j^{-1}(B_j)\big) = \prod_1^n P_{X_j}(B_j) = \Big(\prod_1^n P_{X_j}\Big)(B_1 \times \cdots \times B_n).$$

These quantities are equal for all Borel sets $B_j \subset \mathbb{R}$ iff

$$P_{(X_1,\ldots,X_n)} = \prod_1^n P_{X_j}.$$

That is, $\{X_\alpha\}_{\alpha \in A}$ *is an independent set of random variables iff the joint distribution of any finite set of* X_α*'s is the product of their individual distributions.*

The following proposition expresses the fact that functions of independent random variables are independent.

10.2 Proposition. *Let* $\{X_{nj} : 1 \le j \le J(n),\ 1 \le n \le N\}$ *be independent random variables, and let* $f_n : \mathbb{R}^{J(n)} \to \mathbb{R}$ *be Borel measurable for* $1 \le n \le N$. *Then the random variables* $Y_n = f_n(X_{n1}, \ldots, X_{nJ(n)})$, $1 \le n \le N$, *are independent.*

Proof. Let $\mathbf{X}_n = (X_{n1}, \ldots, X_{nJ(n)})$. If $B_1, \ldots, B_N$ are Borel subsets of $\mathbb{R}$, we have $Y_n^{-1}(B_n) = \mathbf{X}_n^{-1}(f_n^{-1}(B_n))$ and hence

$$\begin{aligned} (Y_1, \ldots, Y_N)^{-1}(B_1 \times \cdots \times B_N) &= \bigcap_1^N Y_n^{-1}(B_n) \\ &= (\mathbf{X}_1, \ldots, \mathbf{X}_N)^{-1}\big(f_1^{-1}(B_1) \times \cdots \times f_N^{-1}(B_N)\big). \end{aligned}$$

Therefore, by the independence of the X_{nj}'s and Fubini's theorem,

$$\begin{aligned} P_{(Y_1,\ldots,Y_n)}(B_1 \times \cdots \times B_n) &= P_{(\mathbf{X}_1,\ldots,\mathbf{X}_N)}\big(f_1^{-1}(B_1) \times \cdots \times f_N^{-1}(B_N)\big) \\ &= \Big(\prod_{n=1}^N \prod_{j=1}^{J(n)} P_{X_{n_j}}\Big)\big(f_1^{-1}(B_1) \times \cdots \times f_N^{-1}(B_N)\big) \\ &= \prod_{n=1}^N P_{\mathbf{X}_n}\big(f_n^{-1}(B_n)\big) \\ &= \prod_{n=1}^N P_{Y_n}(B_n). \end{aligned}$$

∎

We now present some fundamental properties of independent random variables. For the first one we need the notion of convolutions of measures on $\mathbb{R}$ developed in §8.6. An easy induction on (8.47) shows that if $\lambda_1, \ldots, \lambda_n \in M(\mathbb{R})$, then $\lambda_1 * \cdots * \lambda_n$ is given by

$$(10.3) \quad \lambda_1 * \cdots * \lambda_n(E) = \int \cdots \int \chi_E(t_1 + \cdots + t_n) d\lambda_1(t_1) \cdots d\lambda_n(t_n).$$

10.4 Proposition. *If $\{X_j\}_1^n$ are independent random variables, then*

$$P_{X_1+\cdots+X_n} = P_{X_1} * \cdots * P_{X_n}.$$

Proof. Let $A(t_1, \ldots, t_n) = \sum_1^n t_j$. Then $X_1 + \cdots + X_n = A(X_1, \ldots, X_n)$, so

$$P_{X_1+\cdots+X_n} = \left(P_{(X_1,\ldots,X_n)}\right)_A = \left(\prod_1^N P_{X_j}\right)_A,$$

and by (10.3), the last expression equals $P_1 * \cdots * P_n$. ∎

10.5 Proposition. *Suppose that $\{X_j\}_1^n$ are independent random variables. If $X_j \in L^1$ for all j, then $\prod_1^n X_j \in L^1$, and $E(\prod_1^n X_j) = \prod_1^n E(X_j)$.*

Proof. We have $\prod_1^n |X_j| = f(X_1, \ldots, X_n)$ where $f(t_1, \ldots, t_n) = \prod_1^n |t_j|$. Hence

$$E\Big(\prod_1^n |X_j|\Big) = \int f \, dP_{(X_1,\ldots,X_n)} = \int f \, d\Big(\prod_1^n P_{X_j}\Big)$$
$$= \prod_1^n \int |t_j| \, dP_{X_j}(t_j) = \prod_1^n E(|X_j|).$$

This proves the first assertion, and once this is known, the same argument (with the absolute values removed) proves the second one. ∎

10.6 Corollary. *If $\{X_j\}_1^n$ are independent and in L^2, then $\sigma^2(X_1 + \cdots + X_n) = \sum_1^n \sigma^2(X_j)$.*

Proof. Let $Y_j = X_j - E(X_j)$. Then $\{Y_j\}_1^n$ are independent and have mean zero, so

$$E(Y_j Y_k) = E(Y_j)E(Y_k) = 0 \qquad (j \neq k).$$

Therefore,

$$\sigma^2(X_1 + \cdots + X_n) = E\big((Y_1 + \cdots + Y_n)^2\big) = \sum_{j,k} E(Y_j Y_k)$$
$$= \sum_j E(Y_j^2) = \sum_j \sigma^2(X_j).$$

∎

These results show that independence is a very stringent property. For one thing, it is usually not the case that the product of two L^1 functions is in L^1. For another, suppose that X and Y are independent and $E(X) = 0$. Then for any Borel measurable function f on $\mathbb{R}$ such that $f \circ Y \in L^1$ we have

$$E\big(X \cdot (f \circ Y)\big) = E(X)E(f \circ Y) = 0.$$

In other words, X is orthogonal (in the L^2 sense) to *every* function of Y. This indicates that, for example, if one tries to construct a sequence of independent random variables on $[0, 1]$ with Lebesgue measure by using the familiar functions of calculus, one will probably not succeed. (Perhaps the simplest example is $X_n(x) =$ the nth digit in the decimal expansion of x; see Exercise 23.) Rather, the natural setting for independence is product spaces.

Indeed, suppose

$$\Omega = \Omega_1 \times \cdots \times \Omega_n, \quad \mathcal{B} = \mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_n, \quad P = P_1 \times \cdots \times P_n.$$

Then any random variables $X_1, \ldots, X_n$ on Ω such that X_j depends only on the jth coordinate are independent, for if $X_j = f_j \circ \pi_j$ where $\pi_j : \Omega \to \Omega_j$ is the coordinate map,

$$(X_1, \ldots, X_n)^{-1}(B_1 \times \cdots \times B_n) = \prod_1^n f_j^{-1}(B_j)$$

and hence

$$\begin{aligned} P_{(X_1,\ldots,X_n)}(B_1 \times \cdots \times B_n) &= P\Big(\prod_1^n f_j^{-1}(B_j)\Big) \\ &= \prod_1^n P_j\big(f_j^{-1}(B_j)\big) = \prod_1^n P_{X_j}(B_j). \end{aligned}$$

The same idea can be made to work for infinitely many factors; see §10.4.

As an application of these ideas, we present Bernstein's constructive proof of the Weierstrass approximation theorem.

10.7 Theorem. *Given $f \in C([0, 1])$, let*

$$B_n(x) = \sum_{k=0}^n f(k/n) \frac{n!}{k!(n-k)!} x^k (1-x)^{n-k}.$$

Then $B_n \to f$ uniformly on $[0, 1]$ as $n \to \infty$.

Proof. Given $x \in [0, 1]$, let $\lambda = x\delta_1 + (1-x)\delta_0$ where δ_t is the point mass at t. Let $\Omega = \mathbb{R}^n$, $P = \lambda \times \cdots \times \lambda$, and $X_j =$ the jth coordinate function on $\mathbb{R}^n$. Then $X_1, \ldots, X_n$ are independent and have the common distribution λ. It is easy to check (Exercise 7) that

$$\lambda * \cdots * \lambda = \sum_0^n \frac{n!}{k!(n-k)!} x^k (1-x)^{n-k}\, \delta_k,$$

and hence, in view of Proposition 10.4,

$$B_n(x) = E\left[f\Big(\frac{X_1 + \cdots + X_n}{n}\Big)\right].$$

Now, $\sum_0^n n![k!(n-k)!]^{-1}x^k(1-x)^{n-k} = 1$ by the binomial theorem, so

$$(10.8) \quad \big|f(x) - B_n(x)\big| \le \sum_0^n \big|f(x) - f(k/n)\big| \frac{n!}{k!(n-k)!} x^k(1-x)^{n-k}.$$

Given $\epsilon > 0$, by the uniform continuity of f on $[0,1]$ there exists $\delta > 0$, independent of x and y, such that $|f(x) - f(y)| \le \epsilon$ whenever $|x - y| \le \delta$. The sum of the terms in (10.8) such that $|x - (k/n)| \le \delta$ is at most ϵ, while the sum of the remaining terms is at most

$$2\|f\|_u P\left(\Big|\frac{X_1 + \cdots + X_n}{n} - x\Big| > \delta\right).$$

But

$$\sigma^2(X_j) = E(X_j^2) - E(X_j)^2 = x - x^2 \le 1,$$

so by Corollary 10.6,

$$E\left[\Big(\frac{X_1 + \cdots + X_n}{n} - x\Big)^2\right] = \sigma^2\Big(\frac{X_1 + \cdots + X_n}{n}\Big) \le \frac{n}{n^2} = \frac{1}{n},$$

and Chebyshev's inequality therefore gives

$$\big|f(x) - B_n(x)\big| \le \epsilon + \frac{2\|f\|_u}{n\delta^2},$$

which is less than 2ϵ provided n is sufficiently large. ∎

Exercises

1. Let Ω consist of four points, each with probability $\frac{1}{4}$. Find three events that are pairwise independent but not independent. Generalize.

2. Let $\{X_j\}$ be a sequence of independent identically distributed positive random variables such that $E(X_j) = a < \infty$ and $E(1/X_j) = b < \infty$, and let $S_n = \sum_1^n X_j$.

a. $E(X_j/S_n) = 1/n$ if $j \le n$, and $E(X_j/S_n) = aE(1/S_n)$ if $j > n$.

b. $E(S_m/S_n) = m/n$ if $m \le n$ and $E(S_m/S_n) = 1 + (m-n)aE(1/S_n)$ if $m > n$.

3. Suppose that $\{E_\alpha\}_{\alpha \in A}$ is a collection of events in Ω.

a. If $E_{\alpha_1}, \ldots, E_{\alpha_n}$ are independent, so are $E_{\alpha_1}, \ldots, E_{\alpha_{n-1}}, E_{\alpha_n}^c$.

b. If $\{E_\alpha\}$ is an independent set, so is $\{F_\alpha\}$, where each F_α is either E_α or E_α^c.

c. $\{E_\alpha\}$ is an independent set of events iff $\{\chi_{E_\alpha}\}$ is an independent set of random variables.

4. Let X, Y, Z be positive independent random variables with a common distribution λ, and let $F(t) = \lambda((0,t])$. The probability that the polynomial $Xt^2 + Yt + Z$ has real roots is $\int_0^\infty \int_0^\infty F(t^2/4s)\, d\lambda(t)\, d\lambda(s)$.

5. If X is a random variable with distribution $dP_X(t) = f(t)\, dt$ where $f(t) = f(-t)$, then the distribution of X^2 is $dP_{X^2}(t) = t^{-1/2} f(t^{1/2}) \chi_{(0,\infty)}(t)\, dt$.

6. For $a, u > 0$, let $d\gamma_{a,u}(t) = [\Gamma(a)]^{-1} u^a t^{a-1} e^{-ut} \chi_{(0,\infty)}(t)\, dt$, the **gamma distribution** with parameters a and u.

a. The mean and variance of $\gamma_{a,u}$ are a/u and a/u^2, respectively.

b. $\gamma_{a,u} * \gamma_{b,u} = \gamma_{a+b,u}$. (Use Exercise 60 in §2.6.)

c. If $X_1, \ldots, X_n$ are independent and all have the distribution $dP_X(t) = (2\pi)^{-1/2} e^{-t^2/2}\, dt$, then $X_1^2 + \cdots + X_n^2$ has the distribution $\gamma_{n/2,1/2}$. (Use (b) and Exercise 5. $\gamma_{n/2,1/2}$ is called the **chi-square distribution** with n degrees of freedom.)

7. Let δ_t denote the point mass at $t \in \mathbb{R}$. Given $0 < p < 1$, let $\beta_p = p\delta_1 + (1-p)\delta_0$, and let β_p^{*n} be the nth convolution power of β_p. Then

$$\beta_p^{*n} = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \delta_k,$$

and the mean and variance of β_p^{*n} are np and $np(1-p)$. β_p^{*n} is called the **binomial distribution** on $\{0, \ldots, n\}$ with parameter p.

8. Let δ_t denote the point mass at $t \in \mathbb{R}$. Given $a > 0$, let $\lambda_a = e^{-a} \sum_0^\infty (a^k/k!)\delta_k$, the **Poisson distribution** with parameter a.

a. The mean and variance of λ_a are both equal to a.

b. $\lambda_a * \lambda_b = \lambda_{a+b}$.

c. The binomial distribution $\beta_{a/n}^{*n}$ (Exercise 7) converges vaguely to λ_a as $n \to \infty$. (Use Proposition 7.19.)

9. Suppose that $\{X_n\}_1^\infty$ is a sequence of random variables. If $X_n \to X$ in probability, then $P_{X_n} \to P_X$ vaguely. (Use Proposition 7.19.)

10. (**The Moment Convergence Theorem**) Let $X_1, X_2, \ldots, X$ be random variables such that $P_{X_n} \to P_X$ vaguely and $\sup_n E(|X_n|^r) < \infty$, where $r > 0$. Then $E(|X_n|^s) \to E(|X|^s)$ for all $s \in (0, r)$, and if also $s \in \mathbb{N}$, then $E(X_n^s) \to E(X^s)$. (By Chebyshev's inequality, if $\epsilon > 0$, there exists $a > 0$ such that $P(|X_n| > a) < \epsilon$ for all n. Consider $\int \phi(t)|t|^s\, dP_{X_n}(t)$ and $\int [1-\phi(t)]|t|^s\, dP_{X_n}(t)$ where $\phi \in C_c(\mathbb{R})$ and $\phi(t) = 1$ for $|t| \le a$.)

10.2 THE LAW OF LARGE NUMBERS

If one plays a gambling game many times, one's average winnings or losses per game should be roughly the the expected winnings or losses in each individual game;

more generally, if one plays a sequence of possibly different games, one's average winnings or losses should be roughly the average of the expected winnings or losses in the individual games. In symbols: If $\{X_j\}_1^\infty$ is a sequence of independent random variables and $E(X_j) = \mu_j$, then the average $n^{-1}\sum_1^n X_j$ should be close to the constant $n^{-1}\sum_1^n \mu_j$ when n is large.

The law of large number is a precise formulation of this idea. It comes in several versions, depending on the hypotheses one wishes to make. The first version, with the weakest hypotheses and conclusions, has a very simple proof.

10.9 The Weak Law of Large Numbers. *Let $\{X_j\}_1^\infty$ be a sequence of independent L^2 random variables with means $\{\mu_j\}$ and variances $\{\sigma_j^2\}$. If $n^{-2}\sum_1^n \sigma_j^2 \to 0$ as $n \to \infty$, then $n^{-1}\sum_1^n (X_j - \mu_j) \to 0$ in probability as $n \to \infty$.*

Proof. $n^{-1}\sum_1^n (X_j - \mu_j)$ has mean 0 and variance $n^{-2}\sum_1^n \sigma_j^2$ (the latter by Corollary 10.6). Hence by Chebyshev's inequality, for any $\epsilon > 0$ we have

$$P\left(\left|n^{-1}\sum_1^n (X_j - \mu_j)\right| > \epsilon\right) \le (n\epsilon)^{-2}\sum_1^n \sigma_j^2 \to 0 \text{ as } n \to \infty.$$

∎

Under slightly stronger hypotheses, we can obtain the sharper conclusion that $n^{-1}\sum_1^n (X_j - \mu_j) \to 0$ almost surely. To establish this, we need the following two lemmas, which are of interest in their own right.

10.10 The Borel-Cantelli Lemma. *Let $\{A_n\}_1^\infty$ be a sequence of events.*

a. If $\sum_1^\infty P(A_n) < \infty$, then $P(\limsup A_n) = 0$.

b. If the A_n's are independent and $\sum_1^\infty P(A_n) = \infty$, then $P(\limsup A_n) = 1$.

Proof. We recall that $\limsup A_n = \bigcap_{k=1}^\infty \bigcup_{n=k}^\infty A_n$, so that

$$P(\limsup A_n) \le P\Big(\bigcup_{n=k}^\infty A_n\Big) \le \sum_{n=k}^\infty P(A_n),$$

and the latter sum tends to zero as $k \to \infty$ if $\sum P(A_n)$ converges. On the other hand, suppose that $\sum P(A_n)$ diverges and the A_n's are independent. We must show that

$$P\big((\limsup A_n)^c\big) = P\Big(\bigcup_{k=1}^\infty \bigcap_{n=k}^\infty A_n^c\Big) = 0,$$

and for this it is enough to show that $P(\bigcap_{n=k}^\infty A_n^c) = 0$ for all k. But the A_n^c's are independent (Exercise 3), so since $1 - t \le e^{-t}$,

$$P\Big(\bigcap_{n=k}^K A_n^c\Big) = \prod_k^K [1 - P(A_n)] \le \prod_k^K e^{-P(A_n)} = \exp\Big(-\sum_k^K P(A_n)\Big).$$

The last expression tends to zero as $K \to \infty$, which yields the desired result. ∎

10.11 Kolmogorov's Inequality. *Let $X_1, \ldots, X_n$ be independent random variables with mean 0 and variances $\sigma_1^2, \ldots, \sigma_n^2$, and let $S_k = X_1 + \cdots + X_k$. For any $\epsilon > 0$,*

$$P\Big(\max_{1 \le k \le n} |S_k| \ge \epsilon\Big) \le \epsilon^{-2} \sum_1^n \sigma_k^2.$$

Proof. Let A_k be the set where $|S_j| < \epsilon$ for $j < k$ and $|S_k| \ge \epsilon$. Then the A_k's are disjoint and their union is the set where $\max |S_k| \ge \epsilon$, so

$$P\big(\max |S_k| \ge \epsilon\big) = \sum_1^n P(A_k) \le \epsilon^{-2} \sum_1^n E(\chi_{A_k} S_k^2),$$

because $S_k^2 \ge \epsilon^2$ on A_k. On the other hand,

$$\begin{aligned} E(S_n^2) &\ge \sum_1^n E(\chi_{A_k} S_n^2) \\ &= \sum_1^n E\Big(\chi_{A_k}[S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2]\Big) \\ &\ge \sum_1^n E(\chi_{A_k} S_k^2) + 2\sum_1^n E\big(\chi_{A_k} S_k(S_n - S_k)\big). \end{aligned}$$

It will suffice to show that $E(\chi_{A_k} S_k(S_n - S_k)) = 0$ for all k, for then we have

$$P\big(\max |S_k| \ge \epsilon\big) \le \epsilon^{-2} E(S_n^2) = \epsilon^{-2} \sum_1^n \sigma_k^2$$

by Corollary 10.6, since the X_k's have mean zero. But χ_{A_k} is a measurable function of $S_1, \ldots, S_k$ and hence of $X_1, \ldots, X_k$, whereas $S_n - S_k$ is a measurable function of $X_{k+1}, \ldots, X_n$. Moreover, $E(S_k) = \sum_1^k E(X_j) = 0$ for all k. Therefore, by Propositions 10.2 and 10.5,

$$E\big(\chi_{A_k} S_k(S_n - S_k)\big) = E(\chi_{A_k} S_k) E(S_n - S_k) = E(\chi_{A_k} S_k) \cdot 0 = 0.$$

∎

10.12 Kolmogorov's Strong Law of Large Numbers. *If $\{X_n\}_1^\infty$ is a sequence of independent L^2 random variables with means $\{\mu_n\}$ and variances $\{\sigma_n^2\}$ such that $\sum_1^\infty n^{-2}\sigma_n^2 < \infty$, then $n^{-1}\sum_1^n (X_j - \mu_j) \to 0$ almost surely as $n \to \infty$.*

Proof. Let $S_n = \sum_1^n (X_j - \mu_j)$. Given $\epsilon > 0$, for $k \in \mathbb{N}$ let A_k be the set where $n^{-1}|S_n| \ge \epsilon$ for some n such that $2^{k-1} \le n < 2^k$. Then on A_k we have $|S_n| \ge \epsilon 2^{k-1}$ for some $n < 2^k$, so by Kolmogorov's inequality,

$$P(A_k) \le (\epsilon 2^{k-1})^{-2} \sum_1^{2^k} \sigma_n^2.$$

Therefore,

$$\sum_1^\infty P(A_k) \le \frac{4}{\epsilon^2}\sum_{k=1}^\infty\sum_{n=1}^{2^k} 2^{-2k}\sigma_n^2 = \frac{4}{\epsilon^2}\sum_{n=1}^\infty\Big(\sum_{k\ge\log_2 n} 2^{-2k}\Big)\sigma_n^2 \le \frac{8}{\epsilon^2}\sum_{n=1}^\infty \frac{\sigma_n^2}{n^2} < \infty,$$

so $P(\limsup A_k) = 0$ by the Borel-Cantelli lemma. But $\limsup A_k$ is precisely the set where $n^{-1}|S_n| \ge \epsilon$ for infinitely many n, so

$$P\big(\limsup n^{-1}|S_n| < \epsilon\big) = 1.$$

Letting $\epsilon \to 0$ through a countable sequence of values, we conclude that $n^{-1}S_n \to 0$ almost surely. ∎

The hypotheses of this theorem are a bit stronger than those of the weak law (Exercise 11). They are certainly satisfied when the X_n's are identically distributed L^2 random variables, since then σ_n^2 is independent of n. However, in the identically distributed case the assumption that $X_n \in L^2$ can be weakened.

10.13 Khinchine's Strong Law of Large Numbers. *If $\{X_n\}_1^\infty$ is a sequence of independent identically distributed L^1 random variables with mean μ, then $n^{-1}\sum_1^n X_j \to \mu$ almost surely as $n \to \infty$.*

Proof. Replacing X_n by $X_n - \mu$, we may assume that $\mu = 0$. Let λ be the common distribution of the X_j's; we are thus assuming that

$$\int |t|\,d\lambda(t) < \infty, \qquad \int t\,d\lambda(t) = 0.$$

Let $Y_j = X_j$ on the set where $|X_j| \le j$ and $Y_j = 0$ elsewhere. Then

$$\begin{aligned}\sum_1^\infty P(Y_j \ne X_j) = \sum_1^\infty P(|X_j| > j) &= \sum_1^\infty \lambda(\{t : |t| > j\}) \\ &= \sum_{j=1}^\infty\sum_{k=j}^\infty \lambda(\{t : k < |t| \le k+1\}).\end{aligned}$$

Since $\sum_{j=1}^\infty\sum_{k=j}^\infty = \sum_{k=1}^\infty\sum_{j=1}^k$, interchanging the order of summation yields

$$\sum_1^\infty P(X_j \ne Y_j) = \sum_{k=1}^\infty k\lambda(\{t : k < |t| \le k+1\}) \le \int |t|\,d\lambda(t) < \infty.$$

By the Borel-Cantelli lemma, then, with probability one we have $X_j = Y_j$ for j sufficiently large, and it therefore suffices to show that $n^{-1}\sum_1^n Y_j \to 0$ almost surely.

We have

$$\sigma^2(Y_n) \le E(Y_n^2) = \int_{|t|\le n} t^2\,d\lambda(t),$$

and hence

$$\sum_1^\infty n^{-2}\sigma^2(Y_n) \le \sum_{n=1}^\infty \sum_{j=1}^n n^{-2} \int_{j-1<|t|\le j} t^2\, d\lambda(t)$$
$$\le \sum_{n=1}^\infty \sum_{j=1}^n jn^{-2} \int_{j-1<|t|\le j} |t|\, d\lambda(t).$$

Reversing the order of summation again and using the fact that $\sum_{n=j}^\infty n^{-2} \le 2j^{-1}$ (by comparison to $\int_j^\infty x^{-2}\, dx$), we obtain

$$\sum_1^\infty n^{-2}\sigma^2(Y_n) \le 2\sum_{j=1}^\infty \int_{j-1<|t|\le j} |t|\, d\lambda(t) = 2\int_{-\infty}^\infty |t|\, d\lambda(t) < \infty.$$

By Theorem 10.12, therefore, if $\mu_j = E(Y_j)$ we have $n^{-1}\sum_1^n (Y_j - \mu_j) \to 0$ almost surely. However, by the dominated convergence theorem,

$$\mu_j = \int_{|t|\le j} t\, d\lambda(t) \to \int_{-\infty}^\infty t\, d\lambda(t) = 0,$$

and it follows easily (Exercise 12) that $n^{-1}\sum_1^n \mu_j \to 0$ also. Hence $n^{-1}\sum_1^n Y_j \to 0$ a.s., and the proof is complete. ∎

Thus far we have not shown how to construct sequences of random variables that satisfy the hypotheses of the theorems in this section. We shall do so in §10.4.

Exercises

11. If $\sum_1^\infty n^{-2}\sigma_n^2 < \infty$, then $\lim_{n\to\infty} n^{-2}\sum_1^n \sigma_j^2 = 0$. (Estimate $\sum_{j\le\epsilon n}\sigma_j^2$ and $\sum_{j>\epsilon n}\sigma_j^2$ separately.)

12. If $\{a_n\} \subset \mathbb{C}$ and $\lim a_n = a$, then $\lim n^{-1}\sum_1^n a_j = a$.

13. The weak law of large numbers remains valid if the hypothesis of independence is replaced by the (much weaker) hypothesis that $E[(X_j - \mu_j)(X_k - \mu_k)] = 0$ for $j \ne k$.

14. If $\{X_n\}$ is a sequence of independent random variables such that $E(X_n) = 0$ and $\sum_1^\infty \sigma^2(X_n) < \infty$, then $\sum_1^\infty X_n$ converges almost surely. (Apply Kolmogorov's inequality to show that the partial sums are Cauchy a.s.) Corollary: If the plus and minus signs in $\sum_1^\infty \pm n^{-1}$ are determined by successive tosses of a fair coin, the resulting series converges almost surely.

15. If $\{X_n\}$ is a sequence of independent identically distributed random variables that are not in L^1, then $\limsup_{n\to\infty} n^{-1}|\sum_1^n X_j| = \infty$ almost surely. (Let $A_n = \{|X_n| > n\}$. Using an idea from the proof of Theorem 10.13, show that $\sum_1^\infty P(A_n) = \infty$, and apply the Borel-Cantelli lemma.)

16. **(Shannon's Theorem)** Let $\{X_i\}$ be a sequence of independent random variables on the sample space Ω having the common distribution $\lambda = \sum_1^r p_j\delta_j$ where $0 <$

$p_j < 1$, $\sum_1^r p_j = 1$, and δ_j is the point mass at j. Define random variables $Y_1, Y_2, \ldots$ on Ω by

$$Y_n(\omega) = P(\{\omega' : X_i(\omega') = X_i(\omega) \text{ for } 1 \le i \le n\}).$$

a. $Y_n = \prod_1^n p_{X_i}$. (The notation is peculiar but correct: $X_i(\cdot) \in \{1, \ldots, r\}$ a.s., so p_{X_i} is well-defined a.s.)

b. $\lim_{n\to\infty} n^{-1} \log Y_n = \sum_1^r p_j \log p_j$ almost surely. (In information theory, the X_i's are considered as the output of a source of digital signals, and $-\sum_1^r p_j \log p_j$ is called the **entropy** of the signal.)

17. A collection or "population" of N objects (such as mice, grains of sand, etc.) may be considered as a smaple space in which each object has probability N^{-1}. Let X be a random variable on this space (a numerical characteristic of the objects such as mass, diameter, etc.) with mean μ and variance σ^2. In statistics one is interested in determining μ and σ^2 by taking a sequence of random samples from the population and measuring X for each sample, thus obtaining a sequence $\{X_j\}$ of numbers that are values of independent random variables with the same distribution as X. The nth **sample mean** is $M_n = n^{-1} \sum_1^n X_j$ and the nth **sample variance** is $S_n^2 = (n-1)^{-1} \sum_1^n (X_j - M_j)^2$. Show that $E(M_n) = \mu$, $E(S_n^2) = \sigma^2$, and $M_n \to \mu$ and $S_n^2 \to \sigma^2$ almost surely as $n \to \infty$. Can you see why one uses $(n-1)^{-1}$ instead of n^{-1} in the definition of S_n^2?

10.3 THE CENTRAL LIMIT THEOREM

Suppose $\mu \in \mathbb{R}$ and $\sigma > 0$. By Proposition 2.53 and some elementary calculus, the measure $\nu_\mu^{\sigma^2}$ on $\mathbb{R}$ defined by

$$d\nu_\mu^{\sigma^2}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{(t-\mu)^2/2\sigma}\, dt$$

is a probability measure that satisfies

$$\int t\, d\nu_\mu^{\sigma^2}(t) = \mu, \qquad \int (t-\mu)^2 d\nu_\mu^{\sigma^2}(t) = \sigma^2.$$

It is called the **normal** or **Gaussian distribution** with mean μ and variance σ^2. The special case ν_0^1 is called the **standard normal distribution**.

It is a matter of empirical observation that normal and approximately normal distributions are extremely common in applied probability and statistics. The theoretical explanation for this phenomenon is the central limit theorem, the idea of which is as follows. Suppose that $\{X_j\}$ is a sequence of independent identically distributed random variables with mean 0 and variance σ^2. Then $n^{-1} \sum_1^n X_j$ has mean 0 and variance $n^{-1}\sigma^2$, so there is a high probability that it is close to 0 when n is large; this is the content of the weak law of large numbers. On the other hand, $n^{-1/2} \sum_1^n X_j$ has mean 0 and variance σ^2 for all n, so one might ask if its distribution approaches some nontrivial limit as $n \to \infty$. The remarkable answer is that no matter what the

distribution of the X_j's is, this limit exists and equals the normal distribution with mean 0 and variance σ^2.

The central limit theorem is really a theorem in Fourier analysis. We shall state it as such and then translate it into probablity theory.

10.14 Theorem. *Let λ be a Borel probability measure on $\mathbb{R}$ such that*

$$\int t^2\, d\lambda(t) = 1, \qquad \int t\, d\lambda(t) = 0.$$

(The finiteness of the first integral implies the existence of the second.) For $n \in \mathbb{N}$ let $\lambda^{*n} = \lambda * \cdots * \lambda$ (n factors) and define the measure λ_n by $\lambda_n(E) = \lambda^{*n}(\sqrt{n}\, E)$, where $\sqrt{n}\, E = \{\sqrt{n}\, t : t \in E\}$. Then $\lambda_n \to \nu_0^1$ vaguely as $n \to \infty$.

Proof. The hypotheses on the measure λ imply that its Fourier transform $\widehat{\lambda}(\xi) = \int e^{-2\pi i \xi \cdot x}\, d\lambda(x)$ is of class C^2 and satisfies $\widehat{\lambda}(0) = 1$, $\widehat{\lambda}'(0) = 0$, and $\widehat{\lambda}''(0) = -4\pi^2$. (Differentiate the integral twice as in Theorem 8.22d.) Thus by Taylor's theorem,

$$\widehat{\lambda}(\xi) = 1 - 2\pi^2\xi^2 + o(\xi^2),$$

where $o(\alpha)$ denotes a quantity that satisfies $\alpha^{-1}o(\alpha) \to 0$ as $\alpha \to 0$. Moreover, $(\lambda^{*n})\widehat{\ } = (\widehat{\lambda})^n$, so by the obvious change of variable,

$$\widehat{\lambda}_n(\xi) = \left[\widehat{\lambda}(n^{-1/2}\xi)\right]^n = \left[1 - \frac{2\pi^2\xi^2}{n} + o\Big(\frac{\xi^2}{n}\Big)\right]^n.$$

Thus, since $\log(1+z) = z + o(z)$,

$$\log \widehat{\lambda}_n(\xi) = n \log\left[1 - \frac{2\pi^2\xi^2}{n} + o\Big(\frac{\xi^2}{n}\Big)\right] = -2\pi^2\xi^2 + n \cdot o\Big(\frac{\xi^2}{n}\Big),$$

which tends to $-2\pi^2\xi^2$ as $n \to \infty$. In other words, $\widehat{\lambda}_n(\xi) \to e^{-2\pi^2\xi^2}$ as $n \to \infty$ for all ξ, so the conclusion follows from Propositions 8.24 and 8.50. ∎

10.15 The Central Limit Theorem. *Let $\{X_j\}$ be a sequence of independent identically distributed L^2 random variables with mean μ and variance σ^2. As $n \to \infty$, the distribution of $(\sigma\sqrt{n})^{-1}\sum_1^n (X_j - \mu)$ converges vaguely to the standard normal distribution ν_0^1, and for all $a \in \mathbb{R}$,*

$$\lim_{n\to\infty} P\Big(\frac{1}{\sigma\sqrt{n}}\sum_1^n (X_n - \mu) \le a\Big) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^a e^{-t^2/2}\, dt.$$

Proof. Replacing X_j by $\sigma^{-1}(X_j - \mu)$, we may assume that $\mu = 0$ and $\sigma = 1$. If λ is the common distribution of the X_j's, then λ satisfies the hypotheses of Theorem 10.14, and in the notation used there, λ_n is the distribution of $n^{-1/2}\sum_1^n X_j$. The first assertion thus follows immediately, and the second one is equivalent to it by Proposition 7.19. ∎

As the reader may readily verify, the same argument yields the following more general result. Under the hypotheses of the central limit theorem, if $\{K_n\}$ is any sequence of finite subsets of $\mathbb{N}$ such that $k_n = \operatorname{card}(K_n) \to \infty$, then the distribution of $(\sigma\sqrt{k_n})^{-1}\sum_{j\in K_n}(X_j - \mu)$ converges vaguely to ν_0^1.

Exercises

18. A fair coin is tossed 10,000 times; let X be the number of times it comes up heads. Use the central limit theorem and a table of values (printed or electronic) of $\operatorname{erf}(x) = 2\pi^{-1/2}\int_0^x e^{-t^2}\,dt$ to estimate

a. the probability that $4950 \le X \le 5050$;

b. the number k such that $|X - 5000| \le k$ with probability 0.98.

19. If $\{X_j\}$ satisfies the hypotheses of the central limit theorem, the sequence $Y_n = (\sigma\sqrt{n})^{-1}\sum_1^n(X_j - \mu)$ does not converge in probability. (Use the remarks following the central limit theorem to show that $\{Y_{2^k}\}$ is not Cauchy in probability.)

20. If $\{X_j\}$ is a sequence of independent identically distributed random variables with mean 0 and variance 1, the distributions of

$$\sum_1^n X_j \Big/ \Big(\sum_1^n X_j^2\Big)^{1/2} \quad \text{and} \quad \sqrt{n}\sum_1^n X_j \Big/ \sum_1^n X_j^2$$

both converge vaguely to the standard normal distribution.

21. Let $\{X_n\}$ be a sequence of independent random variables, each having the Poisson distribution with mean 1 (Exercise 8). Let

$$S_n = \sum_1^n X_j, \qquad Y_n = \left(\frac{S_n - n}{\sqrt{n}}\right)^- = \max\left(\frac{n - S_n}{\sqrt{n}}, 0\right).$$

a. $E(Y_n) = e^{-n}\sum_{k=0}^n \frac{n-k}{\sqrt{n}}\frac{n^k}{k!} = \frac{n^{n+(1/2)}e^{-n}}{n!}$. (For the first equation, use Exercise 8b. As for the second, the sum telescopes.)

b. P_{Y_n} converges vaguely to $\frac{1}{2}\delta_0 + \chi_{(0,\infty)}\nu_0^1$. (Use Proposition 7.19.)

c. $\int t^2\,dP_{Y_n}(t) = E(Y_n^2) \le 1$ for all n.

d. $E(Y_n) = \int_0^\infty t\,dP_{Y_n}(t) \to \int t\,d\nu_0^1(t) = (2\pi)^{-1/2}$. (Use Exercise 10.) Combining this with (a), one obtains **Stirling's formula**:

$$\lim_{n\to\infty}\frac{n!}{n^n e^{-n}\sqrt{2\pi n}} = 1.$$

22. In this exercise we consider random variables with values in the circle $\mathbb{T}$, regarded as $\{z \in \mathbb{C} : |z| = 1\}$. The distribution of such a random variable is a measure on $\mathbb{T}$.

a. If $X_1, \ldots, X_n$ are independent, then $P_{X_1X_2\cdots X_n} = P_{X_1} * \cdots * P_{X_n}$.

b. If $\{X_j\}$ is a sequence of independent random variables with a common distribution λ, the distribution of $\prod_1^n X_j$ converges vaguely to the uniform

distribution on $\mathbb{T}$ (= arc length over 2π) unless λ is supported on a finite subgroup $Z_m = \{e^{2\pi ij/m} : 0 \le j < m\}$ of $\mathbb{T}$, in which case it converges to the uniform distribution $m^{-1}\sum_{z\in Z_m}\delta_z$ on Z_m. (Use Exercise 8.39.)

10.4 CONSTRUCTION OF SAMPLE SPACES

The preceding two sections have dealt with sequences of random variables whose joint distributions have certain properties. We now address the question of finding examples of such sequences, and more generally of constructing families $\{X_\alpha\}_{\alpha\in A}$ of random variables indexed by an arbitrary set A whose finite subfamilies have prescribed joint distributions.

If the index set A is finite, this is easy: given any Borel probability measure P on $\mathbb{R}^n$, P is by definition the joint distribution of the coordinate functions $X_1, \dots, X_n$ on the space $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, P)$. If A is infinite, however, the problem is more delicate. Suppose to begin with that $\{X_\alpha\}_{\alpha\in A}$ is a family of random variables on some sample space $(\Omega, \mathcal{B}, P)$, and for each ordered n-tuple $(\alpha_1, \dots, \alpha_n)$ of distinct elements of A ($n \in \mathbb{N}$) let $P_{(\alpha_1,\dots,\alpha_n)}$ be the joint distribution of $X_{\alpha_1}, \dots, X_{\alpha_n}$. Then the measures $P_{(\alpha_1,\dots,\alpha_n)}$ satisfy the following consistency conditions:

$$\text{If } \sigma \text{ is a permutation of } \{1,\dots,n\}, \text{ then} \quad dP_{(\alpha_{\sigma(1)},\cdots\alpha_{\sigma(n)})}(x_{\sigma(1)},\dots,x_{\sigma(n)}) = dP_{(\alpha_1,\dots,\alpha_n)}(x_1,\dots,x_n). \tag{10.16}$$

$$\text{If } k < n \text{ and } E \in \mathcal{B}_{\mathbb{R}^k}, \text{ then} \quad P_{(\alpha_1,\dots,\alpha_k)}(E) = P_{(\alpha_1,\dots,\alpha_n)}(E \times \mathbb{R}^{n-k}). \tag{10.17}$$

Conversely, given any family of measures $P_{(\alpha_1,\dots,\alpha_n)}$ satisfying (10.16) and (10.17), we shall show that there exist a sample space $(\Omega, \mathcal{B}, P)$ and random variables $\{X_\alpha\}$ on Ω such that $P_{(\alpha_1,\dots,\alpha_n)}$ is the joint distribution of $X_{\alpha_1}, \dots, X_{\alpha_n}$. To do this, it is convenient to make one minor technical modification: We replace $\mathbb{R}$ by its one-point compactification $\mathbb{R}^* = \mathbb{R} \cup \{\infty\}$. Any Borel measure on $\mathbb{R}^n$ can be regarded as a Borel measure on $(\mathbb{R}^*)^n$ that assigns measure zero to $(\mathbb{R}^*)^n \setminus \mathbb{R}^n$, and vice versa. In other words, we allow our random variables to assume the value ∞, although they will do so with probability zero. The point of this is that the space $(\mathbb{R}^*)^A$ is *compact* for any A, by Tychonoff's thoerem.

With this modification, the construction of the sample space $(\Omega, \mathcal{B}, P)$ in the case where the random variables X_α are independent, so that P should be the product of the P_α's, is contained in Theorem 7.28. The general case is achieved by a simple adaptation of the argument given there, which we review in detail for the convenience of the reader.

10.18 Theorem. *Let A be an arbitrary nonempty set, and suppose that for each ordered n-tuple of distinct elements of A ($n \in \mathbb{N}$) we are given a Borel probability measure $P_{(\alpha_1,\dots,\alpha_n)}$ on $\mathbb{R}^n$, or equivalently on $(\mathbb{R}^*)^n$, satisfying (10.16) and (10.17). Then there is a unique Radon probability measure P on the compact Hausdorff space*

$\Omega = (\mathbb{R}^*)^A$ such that $P_{(\alpha_1,\ldots,\alpha_n)}$ is the joint distribution of $X_{\alpha_1},\ldots,X_{\alpha_n}$, where $X_\alpha : \Omega \to \mathbb{R}^*$ is the αth coordinate function.

Proof. Let $C_F(\Omega)$ be the set of all $f \in C(\Omega)$ that depend only on finitely many coordinates. If $f \in C_F(\Omega)$, say $f(x) = F(x_{\alpha_1},\ldots x_{\alpha_n})$, let

$$I(f) = \int F\, dP_{(\alpha_1,\ldots,\alpha_n)}.$$

$I(f)$ is well defined because of (10.16) and (10.17): If we permute the variables or add some extra ones, the result is the same. Clearly $I(f) \geq 0$ if $f \geq 0$, and $|I(f)| \leq \|f\|_u$ with equality when f is constant.

Now, $C_F(\Omega)$ is clearly an algebra that separates points, contains constant functions, and is closed under complex conjugation, so by the Stone-Weierstrass theorem it is dense in $C(\Omega)$. Hence, the functional I extends uniquely to a positive linear functional on $C(\Omega)$ with norm 1, and the Riesz representation theorem therefore yields a unique Radon measure P on Ω such that $I(f) = \int f\, dP$ for all $f \in C(\Omega)$. Let X_α be the αth coordinate function on Ω and let $P'_{(\alpha_1,\ldots,\alpha_n)}$ be the joint distribution of $X_{\alpha_1},\ldots,X_{\alpha_n}$ on $(\mathbb{R}^*)^n$. If $F \in C((\mathbb{R}^*)^n)$ and $f = F \circ (X_{\alpha_1},\ldots,X_{\alpha_n})$ as above, then

$$\int F\, dP'_{(\alpha_1,\ldots,\alpha_n)} = \int f\, dP = I(f) = \int F\, dP_{(\alpha_1,\ldots,\alpha_n)}.$$

But $P'_{(\alpha_1,\ldots,\alpha_n)}$ and $P_{(\alpha_1,\ldots,\alpha_n)}$ are both Radon measures by Theorem 7.8, so they are equal by the uniqueness of the Riesz representation. ∎

The only property of $\mathbb{R}^*$ used in this proof is that it is a compact Hausdorff space in which every open set is σ-compact, so the theorem admits an obvious generalization. In particular, if for each α there is a compact set $K_\alpha \subset \mathbb{R}$ such that $P_{(\alpha_1,\ldots,\alpha_n)}$ is supported in $K_{\alpha_1} \times \cdots \times K_{\alpha_n}$ for all $\alpha_1,\ldots,\alpha_n$, we could take $\Omega = \prod_{\alpha\in A} K_\alpha$ and thus avoid introducing the point at infinity.

Of special interest is the independent case, in which $P_{(\alpha_1,\ldots,\alpha_n)} = P_{\alpha_1} \times \cdots \times P_{\alpha_n}$. We state it as a corollary:

10.19 Corollary. *Suppose $\{P_\alpha\}_{\alpha\in A}$ is a family of probability measures on $\mathbb{R}$. Then there exist a sample space $(\Omega, \mathcal{B}, P)$ and independent random variables $\{X_\alpha\}_{\alpha\in A}$ on Ω such that P_α is the distribution of X_α for every $\alpha \in A$. Specifically, we can take Ω to be $(\mathbb{R}^*)^A$ and X_α to be the αth coordinate function; if P_α is supported in the compact set $K_\alpha \subset \mathbb{R}$ for each α, we can take Ω to be $\prod_{\alpha\in A} K_\alpha$.*

Exercises

23. Given $b \in \mathbb{N} \setminus \{1\}$, let $B = \{0, 1, \ldots, b-1\}$, and let P_0 be the probability measure on B (or $\mathbb{R}$) that assigns measure b^{-1} to each point in B. Let P be the measure on $\Omega = B^{\mathbb{N}}$ given by Corollary 10.19, where $A = \mathbb{N}$ and $P_n = P_0$ for all $n \in \mathbb{N}$, and let $\{X_n\}_1^\infty$ be the coordinate functions on Ω.

a. If $A_1, \ldots, A_n \subset B$,

$$P\Big(\bigcap_1^n X_j^{-1}(A_j)\Big) = b^{-n} \prod_1^n \operatorname{card}(A_j),$$

and $P(\{\omega\}) = 0$ for all $\omega \in \Omega$.
b. Let $\Omega' = \{\omega \in \Omega : X_n(\omega) \neq 0 \text{ for infinitely many } n\}$. Then $\Omega \setminus \Omega'$ is countable and $P(\Omega') = 1$.
c. Define $F : \Omega \to [0,1]$ by $F(\omega) = \sum_1^\infty X_n(\omega) b^{-n}$ (so $F(\omega)$ is the number such that $\{X_n(\omega)\}$ is the sequence of digits in its base b decimal expansion). Then $F|\Omega'$ is a bijection from Ω' to $(0,1]$ that maps $\mathcal{B}_{\Omega'}$ bijectively onto $\mathcal{B}_{(0,1]}$. ($\mathcal{B}_{\Omega'}$ is generated by sets of the form $X_n^{-1}(A) \cap \Omega'$. The image under F of such a set is a finite union of intervals of the form $(jb^{-n}, kb^{-n}]$, and these sets generate $\mathcal{B}_{(0,1]}$.)
d. The image measure of P under F is Lebesgue measure.
e. (**Borel's Normal Number Theorem**) A number $x \in (0,1]$ is called **normal** in base b if the digits $0, 1, \ldots, b-1$ occur with equal frequency in its base b decimal expansion, that is, if

$$\lim_{n\to\infty} \frac{\operatorname{card}\{m \in \{1,\ldots,n\} : X_m(\omega) = j\}}{n} = \frac{1}{b} \text{ for } j = 0, 1, \ldots, b-1.$$

Almost every $x \in (0,1]$ (with respect to Lebesgue measure) is normal in base b for every b.

10.5 THE WIENER PROCESS

It is observed that small particles suspended in a fluid such as water or air undergo an irregular motion, known as Brownian motion, due to the collisions of the particles with the molecules of the fluid. A physical derivation of the statistical properties of Brownian motion was developed independently by Einstein and Smoluchowski, but the rigorous mathematical model for Brownian motion — in the limiting case where the motion is assumed to result from an infinite number of collisions with molecules of infinitesimal size — is due to Wiener. This model, called the Wiener process or Brownian motion process, has turned out to be of central importance in probability theory and its applications to physics and mathematical analysis.

One can consider Brownian motion in any number of space dimensions. We shall describe the theory in dimension one and indicate how to generalize it.

The position of a particle undergoing Brownian motion on the line at time $t > 0$ is considered to be a random variable X_t (on a sample space to be specified later) satisfying the following conditions. First, as a matter of normalization, we assume that the particle starts at the origin at time $t = 0$:

$$X_0 = 0 \text{ (almost surely).} \tag{10.20}$$

Second, since any given collision affects the particle by only an infinitesimal amount, it has no long-term effect, so the motion of the particle after time t should depend on its position X_t at that time but not on its previous history. Thus we assume:

(10.21) If $0 \le t_0 < t_1 < \cdots < t_n$, the random variables $X_{t_j} - X_{t_{j-1}}$ $(1 \le j \le n)$ are independent.

Third, since the physical processes underlying Brownian motion are homogeneous in time, we shall postulate that the distribution of $X_t - X_s$ depends only on $t - s$. If we divide the interval $[s, t]$ into n equal subintervals $[t_0, t_1], \ldots, [t_{n-1}, t_n]$ $(t_0 = s$, $t_n = t)$ and write $X_t - X_s = \sum_1^n (X_{t_j} - X_{t_{j-1}})$, it then follows from (10.21) that $X_t - X_s$ is a sum of n independent identically distributed random variables. Since n can be taken arbitrarily large, the central limit theorem suggests that the distribution of $X_t - X_s$ should be normal, and this conclusion is also supported by experimental evidence. Moreover, by Corollary 10.6, $\sigma^2(X_t - X_s) = n\sigma^2(X_{t_1} - X_{t_0})$, and it follows that $\sigma^2(X_t - X_s) = r\sigma^2(X_{t'} - X_{s'})$ whenever $t - s = r(t' - s')$ and r is rational; this strongly indicates that $\sigma^2(X_t - X_s)$ should be proportional to $t - s$. Finally, since the particle is as likely to move to the left as to the right, the mean of $X_t - X_s$ should be 0. Putting this all together, we are led to the third assumption:

(10.22) There is a constant $C > 0$ such that for $0 < s < t$, $X_t - X_s$ has the normal distribution $\nu_0^{C(t-s)}$ with mean 0 and variance $C(t - s)$.

The constant C, which expresses the rate of diffusion, is of course related to the physical parameters of the system. For simplicity, we shall henceforth take $C = 1$.

A family $\{X_t\}_{t \ge 0}$ of random variables satisfying (10.20)–(10.22), with $C = 1$, is called an **abstract Wiener process**. The generalization to n dimensions can now be described easily: an **n-dimensional abstract Wiener process** is a family of $\mathbb{R}^n$-valued random variables $\{\mathbf{X}_t\}_{t \ge 0}$, where $\mathbf{X}_t = (X_t^1, \ldots, X_t^n)$, such that (i) $\{X_t^j\}_{t \ge 0}$ is a one-dimensional abstract Wiener process for each j, and (ii) if Y_j is any function of the variables $\{X_t^j\}_{t > 0}$ for $j = 1, \ldots, n$, then $Y_1, \ldots, Y_n$ are independent. In other words, an n-dimensional abstract Wiener process is just a Cartesian product of n one-dimensional abstract Wiener processes. In particular, $\mathbf{X}_t - \mathbf{X}_s$ has the n-dimensional normal distribution $[\nu_0^{t-s}]^n$ for $t > s$:

$$d\left[\nu_0^{t-s}\right]^n(x_1, \ldots, x_n) = [2\pi(t - s)]^{-n/2} \exp\left(\frac{\sum_1^n x_j^2}{2(t - s)}\right) dx_1 \ldots dx_n,$$

which has the appropriate sort of spherical symmetry.

We return to the one-dimensional case. The conditions (10.20)–(10.22) completely determine the joint distributions of the X_t's as follows. If $t_1 < \cdots < t_n$, then $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent (since $X_0 = 0$ a.s.), so their joint distribution is the product measure

$$\nu_0^{t_1} \times \nu_0^{t_2 - t_1} \times \cdots \times \nu_0^{t_n - t_{n-1}}.$$

But

$$(X_{t_1}, \ldots, X_{t_n}) = T(X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}})$$

where

$$T(y_1, \ldots, y_n) = (y_1, y_1 + y_2, \ldots, y_1 + \cdots + y_n).$$

Since $\det T = 1$, Theorem 2.44 implies that the joint distribution $P_{(t_1,\ldots,t_n)}$ of $X_{t_1}, \ldots, X_{t_n}$ is given by

(10.23)

$$\begin{aligned} dP_{(t_1,\ldots,t_n)}(x_1, \ldots, x_n) &= d\nu_0^{t_n - t_{n-1}}(x_n - x_{n-1}) \cdots d\nu_0^{t_2 - t_1}(x_2 - x_1)\, d\nu_0^{t_1}(x_1) \\ &= \left[\prod_1^n 2\pi(t_j - t_{j-1})\right]^{-1/2} \exp\left(\sum_1^n \frac{(x_j - x_{j-1})^2}{2(t_j - t_{j-1})}\right) dx_1 \cdots dx_n, \end{aligned}$$

where $t_0 = x_0 = 0$. We thus know $P_{t_1,\ldots,t_n}$ when $t_1 < \cdots < t_n$, and we obtain it in the general case by permuting the variables according to (10.16). Also, it follows easily from (10.23) that (10.17) is satisfied. Therefore, by Theorem 10.18, abstract Wiener processes exist.

This situation leaves something to be desired, however. Physically, one expects the position of a particle to be a continuous function of time, so one would like the sample space for the Wiener process to be $C([0,\infty), \mathbb{R})$ (or some subset thereof) and the random variable X_t to be evaluation at t. Actually, Theorem 10.18 yields something along these lines: The sample space it provides is the space $(\mathbb{R}^*)^{[0,\infty)}$ of *all* functions from $[0,\infty)$ into the compactified line, and $X_t(\omega)$ is indeed $\omega(t)$. We can therefore achieve our goal by showing that the measure P of Theorem 10.18 is concentrated on $C([0,\infty), \mathbb{R})$, considered as a subset of $(\mathbb{R}^*)^{[0,\infty)}$. The resulting realization of the abstract Wiener process on $C([0,\infty), \mathbb{R})$ is what is usually called the **Wiener process**.

Henceforth we shall use the notation

$$\Omega = (\mathbb{R}^*)^{[0,\infty)}, \qquad \Omega_c = C([0,\infty), \mathbb{R}),$$

and P will denote the Radon measure on Ω whose finite-dimensional projections are given by (10.23).

To begin with, we need to make a few comments about the role of the point at infinity. The function $f(t,s) = |t - s|$ maps $\mathbb{R}^2$ to $[0, +\infty)$ (we write $+\infty$ to distinguish it from the point at infinity in $\mathbb{R}^*$), and we extend f to a map from $(\mathbb{R}^*)^2$ to $[0, +\infty]$ by declaring that $|t - \infty| = |\infty - t| = +\infty$ for $t \in \mathbb{R}$ and $|\infty - \infty| = 0$. When thus extended, f is of course discontinuous at ∞, but it is lower semicontinuous, as the reader may verify (Exercise 24). Thus for $a, t, s \in [0,\infty)$ the sets $\{\omega \in \Omega : |\omega(t) - \omega(s)| > a\}$ are open and the sets $\{\omega \in \Omega : |\omega(t) - \omega(s)| \le a\}$ are closed in Ω.

Next, we need to make some estimates in terms of the quantity

$$\rho(\epsilon, \delta) = \sup_{t \le \delta} \int_{|x| > \epsilon} d\nu_0^t(x) = \sup_{t \le \delta} \left[\frac{2}{\pi t}\right]^{1/2} \int_\epsilon^\infty e^{-x^2/2t}\, dx.$$

These estimates are contained in the following four lemmas, after which we come to the main theorem.

10.24 Lemma. *For each* $\epsilon > 0$, $\lim_{\delta \to 0} \delta^{-1}\rho(\epsilon, \delta) = 0$.

Proof. We have

$$\int_{\epsilon}^{\infty} e^{-x^2/2t}\, dx \le \int_{\epsilon}^{\infty} e^{-\epsilon x/2t}\, dt = \frac{2t}{\epsilon} e^{-\epsilon^2/2t},$$

whence $\rho(\epsilon, \delta) \le (2/\epsilon)(2\delta/\pi)^{1/2} e^{-\epsilon^2/2\delta}$. The exponential term tends to zero faster than any power of δ, so the result follows. ∎

10.25 Lemma. *Suppose* $\epsilon > 0$, $\delta > 0$, $0 \le t_1 < \cdots < t_k$, $t_k - t_1 \le \delta$, *and*

$$A = \{\omega : |\omega(t_j) - \omega(t_1)| > 2\epsilon \textit{ for some } j \in \{2, \ldots, k\}\}.$$

Then $P(A) \le 2\rho(\epsilon, \delta)$.

Proof. For $j = 1, \ldots, k$, let

$$\begin{aligned} B_j &= \{\omega : |\omega(t_k) - \omega(t_j)| > \epsilon\}, \\ D_j &= \{\omega : |\omega(t_j) - \omega(t_1)| > 2\epsilon \text{ and } |\omega(t_i) - \omega(t_1)| \le 2\epsilon \text{ for } i < j\}. \end{aligned}$$

If $\omega \in A$, then $\omega \in D_j$ for some $j \ge 2$; if ω is not also in B_j, then ω must be in B_1, for $|\omega(t_k) - \omega(t_1)| > \epsilon$ whenever $|\omega(t_j) - \omega(t_1)| > 2\epsilon$ and $|\omega(t_k) - \omega(t_j)| \le \epsilon$. In other words,

$$A \subset B_1 \cup \bigcup_2^k (B_j \cap D_j).$$

But $P(B_j) \le \rho(\epsilon, \delta)$ by (10.22), and B_j and D_j are independent events by (10.21); also, the D_j's are clearly disjoint. Therefore,

$$P(A) \le P(B_1) + \sum_2^k P(B_j)P(D_j) \le \rho(\epsilon, \delta)\Big[1 + \sum_2^k P(D_j)\Big] \le 2\rho(\epsilon, \delta).$$

∎

10.26 Lemma. *With the notation of Lemma 10.25, let*

$$E = \{\omega : |\omega(t_i) - \omega(t_j)| > 4\epsilon \textit{ for some } i, j \in \{1, \ldots, k\}\}.$$

Then $P(E) \le 2\rho(\epsilon, \delta)$.

Proof. If $|\omega(t_i) - \omega(t_j)| > 4\epsilon$, we have either $|\omega(t_i) - \omega(t_1)| > 2\epsilon$ or $|\omega(t_j) - \omega(t_1)| > 2\epsilon$. Thus $E \subset A$, so the result follows from Lemma 10.25. ∎

10.27 Lemma. *Suppose $\epsilon > 0$, $0 \le a < b$, and $b - a \le \delta$. Let*

$$V = \{\omega : |\omega(t) - \omega(s)| > 4\epsilon \text{ for some } t, s \in [a, b]\}.$$

Then $P(V) \le 2\rho(\epsilon, \delta)$.

Proof. If S is a finite subset of $[a, b]$, let

$$V(S) = \{\omega : |\omega(t) - \omega(s)| > 4\epsilon \text{ for some } t, s \in S\}.$$

Then the sets $V(S)$ are open in Ω by the remarks preceding Lemma 10.24, and their union is V. Also, by Lemma 10.26, $P(V(S)) \le 2\rho(\epsilon, \delta)$. Since the family $\{V(S) : S \text{ is a finite subset of } [a, b]\}$ is closed under finite unions, if K is any compact subset of V we have $K \subset V(S)$ for some S and hence $P(K) \le 2\rho(\epsilon, \delta)$. But then $P(V) \le 2\rho(\epsilon, \delta)$ by the inner regularity of P. ∎

10.28 Theorem. *Let $\Omega = (\mathbb{R}^*)^{[0,\infty)}$ and $\Omega_c = C([0, \infty), \mathbb{R})$, and let P be the Radon measure on Ω whose finite-dimensional projections are given by (10.23), according to Theorem 10.18. Then Ω_c is a Borel subset of Ω and $P(\Omega_c) = 1$.*

Proof. A real-valued function ω on $[0, \infty)$ is continuous iff it is uniformly continuous on $[0, n]$ for each n, and it is uniformly continuous on $[0, n]$ iff for every $j \in \mathbb{N}$ there exists $k \in \mathbb{N}$ such that $|\omega(t) - \omega(s)| \le j^{-1}$ (note the use of $\le$ rather than $<$) for all $s \in [0, n]$ and all $t \in [0, n] \cap (s - k^{-1}, s + k^{-1})$. Moreover, even if we only assume that ω is $\mathbb{R}^*$-valued, this last condition implies that it is real-valued unless it is identically ∞. Therefore, if ω_∞ denotes the function whose value is identically ∞, we have

$$\begin{aligned} &\Omega_c \cup \{\omega_\infty\} \\ (10.29) \quad &= \bigcap_{n=1}^{\infty} \bigcap_{j=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{s,t \in [0,n],\ |t-s|<1/k} \{\omega \in \Omega : |\omega(t) - \omega(s)| \le j^{-1}\}. \end{aligned}$$

By the remarks preceding Lemma 10.24, $\{\omega : |\omega(t) - \omega(s)| \le j^{-1}\}$ is closed for all s, t, and j. Hence $\Omega_c \cup \{\omega_\infty\}$ is an $F_{\sigma\delta}$ set, and Ω_c is therefore a Borel set.

Moreover, if for $\epsilon, \delta > 0$ and $n \in \mathbb{N}$ we set

$$U(n, \epsilon, \delta) = \{\omega \in \Omega : |\omega(t) - \omega(s)| > 8\epsilon \text{ for some } t, s \in [0, n] \text{ with } |t - s| < \delta\},$$

by (10.29) we have

$$\Omega \setminus \Omega_c = \left[\bigcup_{n=1}^{\infty} \bigcup_{j=1}^{\infty} \bigcap_{k=1}^{\infty} U(n, (8j)^{-1}, k^{-1})\right] \cup \{\omega_\infty\}.$$

Clearly $P(\{\omega_\infty\}) = 0$, so in order to show that $P(\Omega_c) = 1$, or equivalently that $P(\Omega \setminus \Omega_c) = 0$, it will suffice to show that $\lim_{k\to\infty} P(U(n, \epsilon, k^{-1})) = 0$ for all ϵ and n.

The interval $[0, n]$ is the union of the subintervals $[0, k^{-1}]$, $[k^{-1}, 2k^{-1}]$, ..., $[n - k^{-1}, n]$. If $\omega \in U(n, \epsilon, k^{-1})$, then $|\omega(t) - \omega(s)| > 8\epsilon$ for some t, s lying in the same subinterval or in adjacent subintervals, and hence, in the notation of Lemma 10.27, $\omega \in V$ where $\delta = k^{-1}$ and $[a, b]$ is one of the subintervals. (In the case of adjacent subintervals, use their common endpoint as an intermediate point.) As there are nk subintervals, Lemma 10.27 implies that $P(U(n, \epsilon, k^{-1})) \leq 2nk\rho(\epsilon, k^{-1})$. Lemma 10.24 then shows that $P(U(n, \epsilon, k^{-1})) \to 0$ as $k \to \infty$, which completes the proof. ∎

Exercises

24. The function $f : (\mathbb{R}^*)^2 \to [0, +\infty]$ defined by $f(t, s) = |t - s|$ for $t, s \in \mathbb{R}$, $f(\infty, t) = f(t, \infty) = +\infty$ for $t \in \mathbb{R}$, and $f(\infty, \infty) = 0$ is lower semicontinuous.

25. Let $\Omega = (\mathbb{R}^*)^{[0,\infty)}$, $\{X_t\}_{t \geq 0}$ the coordinate functions on Ω, and for any $A \subset [0, \infty)$, $\mathcal{M}_A =$ the σ-algebra generated by $\{X_t\}_{t \in A}$. (Thus $\mathcal{M}_{[0,\infty)}$ is the product σ-algebra on Ω corresponding to the Borel σ-algebras on the factors.)

a. Suppose $V \in \mathcal{M}_A$ and $\omega, \omega' \in \Omega$. If $\omega \in V$ and $\omega'(t) = \omega(t)$ for all $t \in A$, then $\omega' \in V$.

b. If $V \in \mathcal{M}_{[0,\infty)}$, then $V \in \mathcal{M}_A$ for some countable set A. (Use Exercise 5 in §1.2.)

c. The set $\Omega_c = C([0, \infty), \mathbb{R})$ is not in $\mathcal{M}_{[0,\infty)}$.

26. Let Ω_c and P be as in Theorem 10.30. If $\omega \in \Omega_c$, it can be shown that ω is almost surely not of bounded variation, so if f is a Borel measurable function on $[0, \infty)$, the integral $\int_0^\infty f(t)\, d\omega(t)$ apparently makes no sense. However:

a. If $f = \sum_1^n c_j \chi_{[a_j, b_j)}$ is a step function, define

$$I_f(\omega) = \int_0^\infty f(t)\, d\omega(t) = \sum_1^n c_j \left[\omega(b_j) - \omega(a_j)\right].$$

Then I_f is an L^2 random variable on Ω_c with mean 0 and variance $\int_0^\infty |f(x)|^2\, dx$. (Hint: The intervals $[a_j, b_j)$ may be assumed disjoint.)

b. The map $f \to I_f$ extends to an isometry from $L^2([0, \infty), m)$ to $L^2(\Omega_c, P)$.

c. If $f \in BV([0, \infty))$ is right continuous and $\operatorname{supp}(f)$ is compact, there is a sequence $\{f_n\}$ of step functions such that $f_n \to f$ in L^2 and $df_n \to df$ vaguely, where df denotes the Lebesgue-Stietjes measure defined by f. (By Exercise 41 in §8.6, there is a sequence $\{\mu_n\}$ of linear combinations of point masses such that $d\mu_n \to df$ vaguely. Consider $f_n(x) = \mu_n((0, x]) + f(0)$.)

d. If $f \in BV([0, \infty))$ is right continuous and $\operatorname{supp}(f)$ is compact, then $I_f(\omega) = -\int_0^\infty \omega(t)\, df(t)$ almost surely. (Check this directly when f is a step function and apply (b) and (c).)

10.6 NOTES AND REFERENCES

The development of probability theory as a rigorous mathematical discipline began in the early part of the 20th century, when the tools of measure theory and Lebesgue-Stieltjes integrals became available. In 1933 Kolmogorov [85] put the subject on a solid foundation by explicitly identifying sample spaces and random variables with measure spaces and measurable functions. Since then it has grown extensively.

More detailed accounts of probability theory on a level comparable to that of this book can be found in Billingsley [17], Chung [25], and Lamperti [90].

§10.3: An account of the long history of the central limit theorem can be found in Adams [2]. More general versions of this result exist in which the random variables are not assumed to be identically distributed; see the references given above. The form of Taylor's theorem used in the proof is explained in Folland [45].

The proof of Stirling's formula outlined in Exercise 21 is due to Wong [162]; see Blyth and Pathak [18] for some other probabilistic proofs of Stirling's formula.

There is one more major result about the asymptotic behavior of sums of independent identically distributed random variables that should be mentioned along with the law of large numbers and the central limit theorem:

> *The Law of the Iterated Logarithm:* Suppose that $\{X_n\}_1^\infty$ is a sequence of independent identically distributed L^3 random variables with mean μ and variance σ^2, and let $S_n = \sum_1^n X_j$. Then
>
> $$\limsup_{n\to\infty} \frac{S_n - n\mu}{\sigma\sqrt{2n\log\log n}} = 1 \text{ almost surely.}$$

The proof may be found in Chung [25], which gives a more general result; see also Lamperti [90] for the case in which the X_n's are assumed uniformly bounded.

§10.4: Theorem 10.18 is a variant due to Nelson [103] of the fundamental existence theorem of Kolmogorov [85]. In Kolmogorov's original construction, the sample space is $\mathbb{R}^A$ (which could be replaced by $(\mathbb{R}^*)^A$) and the σ-algebra on which the measure P is constructed is $\bigotimes_{\alpha\in A} \mathcal{B}_{\mathbb{R}}$. Theorem 10.18 is a decided improvement on Kolmogorov's theorem, both in the simplicity of its proof and in the fact that the Borel σ-algebra on $(\mathbb{R}^*)^A$ properly includes $\bigotimes_{\alpha\in A} \mathcal{B}_{\mathbb{R}}$ when A is uncountable. (The significance of the latter fact is evident from Exercise 25.)

§10.5: Wiener constructed his measure P on $C([0,\infty),\mathbb{R})$ in [159] and [160]; his approach is quite different from ours. Our proof of Theorem 10.30 follows Nelson [104]. See also Nelson [105] for some related material, including the derivation of the postulates (10.20)–(10.22) from physical principles.

A discussion of the many interesting properties of the Wiener process is beyond the scope of this book. We shall mention only one, as a complement to Theorem 10.30: The sample paths of the Wiener process are almost surely nowhere differentiable; in fact, with probability one, at each point they are Hölder continuous of every exponent $\alpha < \frac{1}{2}$ but not of exponent $\frac{1}{2}$. This fact may be startling at first, but it seems almost

inevitable when one reflects that $|\omega(t) - \omega(s)|/|t - s|^{1/2}$ has the standard normal distribution for all t, s.

Knight [84] is a good source for further information about the Wiener process.

11

More Measures and Integrals

In this chapter we discuss some additional examples of measures and integrals that are of importance in analysis and geometry: invariant measures on locally compact groups, geometric measures of lower-dimensional sets in $\mathbb{R}^n$, and integration of densities and differential forms on manifolds. Although we have grouped these topics together in one chapter, they are substantially independent of one another.

11.1 TOPOLOGICAL GROUPS AND HAAR MEASURE

A **topological group** is a group G endowed with a topology such that the group operations $(x, y) \mapsto xy$ and $x \mapsto x^{-1}$ are continuous from $G \times G$ and G to G. Examples include topological vector spaces (the group operation being addition), groups of invertible $n \times n$ real matrices (with the relative topology induced from $\mathbb{R}^{n\times n}$), and all groups equipped with the discrete topology. If G is a topological group, we denote the identity element of G by e, and for $A, B \subset G$ and $x \in G$ we define

$$xA = \{xy : y \in A\}, \qquad Ax = \{yx : y \in A\},$$
$$A^{-1} = \{x^{-1} : x \in A\}, \qquad AB = \{yz : y \in A,\ z \in B\}.$$

We say that $A \subset G$ is **symmetric** if $A = A^{-1}$.

Here are some of the basic properties of topological groups:

11.1 Proposition. *Let G be a topological group.*

a. The topology of G is translation invariant: If U is open and $x \in G$, then Ux and xU are open.

b. *For every neighborhood U of e there is a symmetric neighborhood V of e with $V \subset U$.*
c. *For every neighborhood U of e there is a neighborhood V of e with $VV \subset U$.*
d. *If H is a subgroup of G, so is $\overline{H}$.*
e. *Every open subgroup of G is also closed.*
f. *If K_1, K_2 are compact subsets of G, so is K_1K_2.*

Proof. (a) is equivalent to the continuity in each variable of the map $(x, y) \mapsto xy$, and (b) and (c) are equivalent to the continuity of $x \mapsto x^{-1}$ and $(x, y) \mapsto xy$ at the identity. (Details are left to the reader.) For (d), if $x, y \in \overline{H}$, there exist nets $\langle x_\alpha \rangle_{\alpha \in A}$, $\langle y_\beta \rangle_{\beta \in B}$ in H that converge to x and y. Then $x_\alpha^{-1} \to x^{-1}$ and $x_\alpha y_\beta \to xy$ (with the usual product ordering on $A \times B$), so x^{-1} and xy belong to $\overline{H}$. For (e), if H is an open subgroup, the cosets xH are open for all x, so that $G \setminus H = \bigcup_{x \notin H} xH$ is open and hence H is closed. Finally, (f) is true because K_1K_2 is the image of the compact set $K_1 \times K_2$ under the continuous map $(x, y) \mapsto xy$. ∎

If f is a continuous function on the topological group G and $y \in G$, we define the left and right translates of f through y by

$$L_y f(x) = f(y^{-1}x), \qquad R_y f(x) = f(xy).$$

(The point of using y^{-1} on the left and y on the right is to make $L_{yz} = L_y L_z$ and $R_{yz} = R_y R_z$.) f is called **left** (resp. **right**) **uniformly continuous** if for every $\epsilon > 0$ there is a neighborhood V of e such that $\|L_y f - f\|_u < \epsilon$ (resp. $\|R_y f - f\|_u < \epsilon$) for $y \in V$. (Some authors reverse the roles of L_y and R_y in this definition.)

11.2 Proposition. *If $f \in C_c(G)$, then f is left and right uniformly continuous.*

Proof. We shall consider right uniform continuity; the proof on the left is the same. Let $K = \operatorname{supp}(f)$ and suppose $\epsilon > 0$. For each $x \in K$ there is a neighborhood U_x of e such that $|f(xy) - f(x)| < \frac{1}{2}\epsilon$ for $y \in U_x$, and by Proposition 11.1(b,c) there is a symmetric neighborhood V_x of e with $V_x V_x \subset U_x$. Then $\{xV_x\}_{x \in K}$ covers K, so there exist $x_1, \ldots, x_n \in K$ such that $K \subset \bigcup_1^n x_j V_{x_j}$. Let $V = \bigcap_1^n V_{x_j}$; we claim that $|f(xy) - f(x)| < \epsilon$ if $y \in V$. On the one hand, if $x \in K$, then for some j we have $x_j^{-1}x \in V_{x_j}$ and hence $xy = x_j(x_j^{-1}x)y \in x_j U_{x_j}$; therefore,

$$|f(xy) - f(x)| \le |f(xy) - f(x_j)| + |f(x_j) - f(x)| < \epsilon.$$

On the other hand, if $x \notin K$, then $f(x) = 0$, and either $f(xy) = 0$ (if $xy \notin K$) or $x_j^{-1}xy \in V_{x_j}$ for some j (if $xy \in K$); in the latter case $x_j^{-1}x = x_j^{-1}xyy^{-1} \in U_{x_j}$, so that $|f(x_j)| < \frac{1}{2}\epsilon$ and hence $|f(xy)| < \epsilon$. ∎

One usually assumes that the topology of a topological group is Hausdorff. The following proposition shows that this is not much of a restriction.

11.3 Proposition. *Let G be a topological group.*

a. If G is T_1, then G is Hausdorff.

b. If G is not T_1, let H be the closure of $\{e\}$. Then H is a normal subgroup, and if G/H is given the quotient topology (i.e., a set in G/H is open iff its inverse image in G is open), G/H is a Hausdorff topological group.

Proof. (a) If G is T_1 and $x \neq y \in G$, by Proposition 11.1(b,c) there is a symmetric neighborhood V of e such that $xy^{-1} \notin VV$. Then Vx and Vy are disjoint neighborhoods of x and y, for it $z = v_1 x = v_2 y$ for some $v_1, v_2 \in V$, then $xy^{-1} = v_1^{-1} z z^{-1} v_2 \in V^{-1}V = VV$.

(b) H is a subgroup by Proposition 11.1d; it is clearly the smallest closed subgroup of G. It follows that H is normal, for if H' were a conjugate of H with $H' \neq H$, $H' \cap H$ would be a smaller closed subgroup. It is routine to verify that the group operations on G/H are continuous in the quotient topology, so that G/H is a topological group. If $\overline{e}$ is the identity element in G/H, then $\{\overline{e}\}$ is closed since its inverse image in G is H. But then every singleton set in G/H is closed by Proposition 11.1a, so G/H is T_1 and hence Hausdorff. ∎

In the context of Proposition 11.3b, it is easy to see that every Borel measurable function on G is constant on the cosets of H and hence is effectively a function on G/H. Thus for most purposes one may as well work with the Hausdorff group G/H. We shall be interested in the case where G is locally compact, and we henceforth use the term **locally compact group** to mean a topological group whose topology is locally compact and Hausdorff.

Suppose that G is a locally compact group. A Borel measure μ on G is called **left-invariant** (resp. **right-invariant**) if $\mu(xE) = \mu(E)$ (resp. $\mu(Ex) = \mu(E)$) for all $x \in G$ and $E \in \mathcal{B}_G$. Similarly, a linear functional I on $C_c(G)$ is called left- or right-invariant if $I(L_x f) = I(f)$ or $I(R_x f) = I(f)$ for all f. A **left** (resp. **right**) **Haar measure** on G is a nonzero left-invariant (resp. right-invariant) Radon measure μ on G. For example, Lebesgue measure is a (left and right) Haar measure on $\mathbb{R}^n$, and counting measure is a (left and right) Haar measure on any group with the discrete topology. (Other examples will be found in the exercises.) The following proposition summarizes some elementary properties of Haar measures; in it, and in the sequel, we employ the notation

$$C_c^+ = \{f \in C_c(G) : f \geq 0 \text{ and } \|f\|_u > 0\}.$$

11.4 Proposition. *Let G be a locally compact group.*

a. A Radon measure μ on G is a left Haar measure iff the measure $\widetilde{\mu}$ defined by $\widetilde{\mu}(E) = \mu(E^{-1})$ is a right Haar measure.

b. A nonzero Radon measure μ on G is a left Haar measure iff $\int f\, d\mu = \int L_y f\, d\mu$ for all $f \in C_c^+$ and $y \in G$.

c. If μ is a left Haar measure on G, then $\mu(U) > 0$ for every nonempty open $U \subset G$, and $\int f\, d\mu > 0$ for all $f \in C_c^+$.

d. If μ is a left Haar measure on G, then $\mu(G) < \infty$ iff G is compact.

Proof. (a) is obvious. The "only if" implication of (b) follows by approximating f by simple functions, and the converse is true by (7.3). As for (c), since $\mu \neq 0$, by the regularity of μ there is a compact $K \subset G$ with $\mu(K) > 0$. If U is open and nonempty, K can be covered by finitely many left translates of U, and it follows that $\mu(U) > 0$. If $f \in C_c^+$, let $U = \{x : f(x) > \frac{1}{2}\|f\|_u\}$. Then $\int f\,d\mu \geq \frac{1}{2}\|f\|_u \mu(U) > 0$.

Finally, we prove (d). If G is compact, then $\mu(G) < \infty$ since μ is Radon. If G is not compact and V is a compact neighborhood of e, then G cannot be covered by finitely many translates of V, so by induction we can find a sequence $\{x_n\}$ such that $x_n \notin \bigcup_1^{n-1} x_j V$ for all n. By Proposition 11.1(b,c) there is a symmetric neighborhood U of e such that $UU \subset V$. If $m > n$ and $x_n U \cap x_m U$ is nonempty, then $x_m \in x_n UU \subset x_n V$, a contradiction. Hence $\{x_n U\}_1^\infty$ is a disjoint sequence, and $\mu(x_n U) = \mu(U) > 0$ by (c), whence $\mu(G) \geq \mu(\bigcup_1^\infty x_m U) = \infty$. ∎

Our aim now is to prove the existence and uniqueness of Haar measures. In view of Proposition 11.4a, one can pass from left Haar measures to right Haar measures at will, so for the sake of definiteness we shall concentrate on left Haar measures. We begin with some motivation for the existence proof.

If $E \in \mathcal{B}_G$ and V is open and nonempty, let $(E : V)$ denote the smallest number of left translates of V that cover E, that is,

$$(E : V) = \inf\Big\{\#(A) : E \subset \bigcup_{x \in A} xV\Big\},$$

where $\#(A) = \operatorname{card}(A)$ if A is finite and $\#(A) = \infty$ otherwise. Thus $(E : V)$ is a rough measure of the relative sizes of E and V. If we fix a precompact open set E_0, the ratio $(E : V)/(E_0 : V)$ gives a rough estimate of the size of E when the size of E_0 is normalized to be 1. This estimate becomes the more accurate the smaller V is, and it is obviously left-invariant as a function of E. We might therefore hope to obtain a Haar measure as a limit of the "quasi-measures" $(E : V)/(E_0 : V)$ as V shrinks to $\{e\}$.

This idea can be made to work as it stands, but it is simpler to carry out if we think of integrals of functions instead of measures of sets. If $f, \phi \in C_c^+$, then $\{x : \phi(x) > \frac{1}{2}\|\phi\|_u\}$ is open and nonempty, so finitely many left translates of it cover $\operatorname{supp}(f)$, and it follows that

$$f \leq \frac{2\|f\|_u}{\|\phi\|_u} \sum_1^n L_{x_j}\phi \text{ for some } x_1, \ldots, x_n \in G.$$

It therefore makes sense to define the "Haar covering number" of f with respect to ϕ:

$$(f : \phi) = \inf\Big\{\sum_1^n c_j : f \leq \sum_1^n c_j L_{x_j}\phi \text{ for some } n \in \mathbb{N} \text{ and } x_1, \ldots, x_n \in G\Big\}.$$

Clearly $(f : \phi) > 0$; in fact, $(f : \phi) \geq \|f\|_u / \|\phi\|_u$.

11.5 Lemma. *Suppose that* $f, g, \phi \in C_c^+$.

a. $(f : \phi) = (L_x f : \phi)$ *for any* $x \in G$.

b. $(cf : \phi) = c(f : \phi)$ *for any* $c > 0$.

c. $(f + g : \phi) \le (f : \phi) + (g : \phi)$.

d. $(f : \phi) \le (f : g)(g : \phi)$.

Proof. We have $f \le \sum c_j L_{x_j}\phi$ iff $L_x f \le \sum c_j L_{xx_j}\phi$; this proves (a), and (b) is equally obvious. If $f \le \sum_1^m c_j L_{x_j}\phi$ and $g \le \sum_{m+1}^n c_j L_{x_j}\phi$, then $f + g \le \sum_1^n c_j L_{x_j}\phi$, so (c) follows by minimizing $\sum_1^m c_j$ and $\sum_{m+1}^n c_j$. Similarly, if $f \le \sum c_j L_{x_j} g$ and $g \le \sum d_k L_{y_k}\phi$, then $f \le \sum_{j,k} c_j d_k L_{x_j y_k}\phi$. Since $\sum_{j,k} c_j d_k = (\sum c_j)(\sum d_k)$, (d) follows. ∎

At this point we make a normalization by choosing $f_0 \in C_c^+$ once and for all and defining

$$I_\phi(f) = \frac{(f : \phi)}{(f_0 : \phi)} \text{ for } f, \phi \in C_c^+.$$

By Lemma 11.5(a–c), for each fixed ϕ the functional I_ϕ is left-invariant and bears some resemblance to a positive linear functional except that it is only subadditive. Moreover, by Lemma 11.5d it satisfies

$$(f_0 : f)^{-1} \le I_\phi(f) \le (f : f_0). \tag{11.6}$$

We now show that, in a certain sense, I_ϕ is approximately additive when $\operatorname{supp}(\phi)$ is small.

11.7 Lemma. *If* $f_1, f_2 \in C_c^+$ *and* $\epsilon > 0$, *there is a neighborhood* V *of* e *such that* $I_\phi(f_1) + I_\phi(f_2) \le I_\phi(f_1 + f_2) + \epsilon$ *whenever* $\operatorname{supp}(\phi) \subset V$.

Proof. Fix $g \in C_c^+$ such that $g = 1$ on $\operatorname{supp}(f_1 + f_2)$, and let δ be a positive number to be specified later. Set $h = f_1 + f_2 + \delta g$ and $h_i = f_i/h$ $(i = 1, 2)$, where it is understood that $h_i = 0$ outside $\operatorname{supp}(f_i)$. Then $h_i \in C_c^+$, so by Proposition 11.2 there is a neighborhood V of e such that $|h_i(x) - h_i(y)| < \delta$ if $i = 1, 2$ and $y^{-1}x \in V$. If $\phi \in C_c^+$, $\operatorname{supp}(\phi) \subset V$, and $h \le \sum_1^n c_j L_{x_j}\phi$, then $|h_i(x) - h_i(x_j)| < \delta$ whenever $x_j^{-1}x \in \operatorname{supp}(\phi)$, so

$$f_i(x) = h(x)h_i(x) \le \sum_j c_j \phi(x_j^{-1}x)h_i(x) \le \sum_j c_j \phi(x_j^{-1}x)[h_i(x_j) + \delta].$$

But then $(f_i : \phi) \le \sum c_j[h_i(x_j) + \delta]$, and since $h_1 + h_2 \le 1$,

$$(f_1 : \phi) + (f_2 : \phi) \le \sum_j c_j[1 + 2\delta].$$

Now, $\sum c_j$ can be made arbitrarily close to $(h : \phi)$, so by Lemma 11.5(b,c),

$$I_\phi(f_1) + I_\phi(f_2) \le (1 + 2\delta)I_\phi(h) \le (1 + 2\delta)\big[I_\phi(f_1 + f_2) + \delta I_\phi(g)\big].$$

In view of (11.6), therefore, it suffices to choose δ small enough so that

$$2\delta(f_1 + f_2 : f_0) + \delta(1 + 2\delta)(g : f_0) < \epsilon.$$

∎

11.8 Theorem. *Every locally compact group G possesses a left Haar measure.*

Proof. For each $f \in C_c^+$ let X_f be the interval $[(f_0 : f)^{-1}, (f : f_0)]$, and let $X = \prod_{f \in C_c^+} X_f$. Then X is a compact Hausdorff space by Tychonoff's theorem, and by (11.6), every I_ϕ is an element of X. For each compact neighborhood V of e, let $K(V)$ be the closure in X of $\{I_\phi : \operatorname{supp}(\phi) \subset V\}$. Clearly $\bigcap_1^n K(V_j) \supset K(\bigcap_1^n V_j)$, so by Proposition 4.21 there is an element I in the intersection of all the $K(V)$'s. Every neighborhood of I in X intersects $\{I_\phi : \operatorname{supp}(\phi) \subset V\}$ for all V; in other words, for any neighborhood V of e and any $f_1, \ldots, f_n \in C_c^+$ and $\epsilon > 0$ there exists $\phi \in C_c^+$ with $\operatorname{supp}(\phi) \subset V$ such that $|I(f_j) - I_\phi(f_j)| < \epsilon$ for $j = 1, \ldots, n$. Therefore, in view of Lemmas 11.5 and 11.7, I is left-invariant and satisfies $I(af + bg) = aI(f) + bI(g)$ for all $f, g \in C_c^+$ and $a, b > 0$. It follows easily, as in the proof of Lemma 7.15, that if we extend I to C_c by setting $I(f) = I(f^+) - I(f^-)$, then I is a left-invariant positive linear functional on $C_c(G)$. Moreover, $I(f) > 0$ for all $f \in C_c^+$ by (11.6). The proof is therefore completed by invoking the Riesz representation theorem. ∎

11.9 Theorem. *If μ and ν are left Haar measures on G, there exists $c > 0$ such that $\mu = c\nu$.*

Proof. We first present a simple proof that works when μ is both left- and right-invariant — in particular, when G is Abelian. Pick $h \in C_c^+$ such that $h \in C_c^+$ and $h(x) = h(x^{-1})$ (e.g., $h(x) = g(x) + g(x^{-1})$ where g is any element of C_c^+). Then for any $f \in C_c(G)$,

$$\begin{aligned}
\int h\,d\nu \int f\,d\mu &= \iint h(y)f(x)\,d\mu(x)\,d\nu(y) \\
&= \iint h(y)f(xy)\,d\mu(x)\,d\nu(y) = \iint h(y)f(xy)\,d\nu(y)\,d\mu(x) \\
&= \iint h(x^{-1}y)f(y)\,d\nu(y)\,d\mu(x) = \iint h(y^{-1}x)f(y)\,d\nu(y)\,d\mu(x) \\
&= \iint h(y^{-1}x)f(y)\,d\mu(x)\,d\nu(y) = \iint h(x)f(y)\,d\mu(x)\,d\nu(y) \\
&= \int h\,d\mu \int f\,d\nu,
\end{aligned}$$

so that $\mu = c\nu$ where $c = (\int h\,d\mu)/(\int h\,d\nu)$. ($\int h\,d\nu \neq 0$ by Proposition 11.4c, and Fubini's theorem is applicable since the functions in question are supported in sets which are compact and hence of finite measure. The same remarks apply below.)

Now, another proof for the general case. The assertion that $\mu = c\nu$ is equivalent to the assertion that the ratio $r_f = (\int f\,d\mu)/(\int f\,d\nu)$ is independent of $f \in C_c^+$. Suppose, then, that $f, g \in C_c^+$; we shall show that $r_f = r_g$.

Fix a symmetric compact neighborhood V_0 of e and set

$$A = [\operatorname{supp}(f)]V_0 \cup V_0[\operatorname{supp}(f)], \qquad B = [\operatorname{supp}(g)]V_0 \cup V_0[\operatorname{supp}(g)].$$

Then A and B are compact by Proposition 11.1f, and for $y \in V_0$ the functions $x \mapsto f(xy) - f(yx)$ and $x \mapsto g(xy) - g(yx)$ are supported in A and B, respectively. Next, given $\epsilon > 0$, by Proposition 11.2 there is a symmetric compact neighborhood $V \subset V_0$ of e such that $\sup_x |f(xy) - f(yx)| < \epsilon$ and $\sup_x |g(xy) - g(yx)| < \epsilon$ for $y \in V$. Pick $h \in C_c^+$ with $\operatorname{supp}(h) \subset V$ and $h(x) = h(x^{-1})$. Then

$$\begin{aligned}\int h\,d\nu \int f\,d\mu &= \iint h(y)f(x)\,d\mu(x)\,d\nu(y)\\ &= \iint h(y)f(yx)\,d\mu(x)\,d\nu(y),\end{aligned}$$

and since $h(x) = h(x^{-1})$,

$$\begin{aligned}&\int h\,d\mu \int f\,d\nu = \iint h(x)f(y)\,d\mu(x)\,d\nu(y)\\ &= \iint h(y^{-1}x)f(y)\,d\mu(x)\,d\nu(y) = \iint h(x^{-1}y)f(y)\,d\nu(y)\,d\mu(x)\\ &= \iint h(y)f(xy)\,d\nu(y)\,d\mu(x) = \iint h(y)f(xy)\,d\mu(x)\,d\nu(y).\end{aligned}$$

Thus,

$$\begin{aligned}\left|\int h\,d\nu \int f\,d\mu - \int h\,d\mu \int f\,d\nu\right| &= \left|\iint h(y)\big[f(xy) - f(yx)\big]\,d\mu(x)\,d\nu(y)\right|\\ &\le \epsilon\mu(A)\int h\,d\nu.\end{aligned}$$

By the same reasoning,

$$\left|\int h\,d\nu \int g\,d\mu - \int h\,d\mu \int g\,d\nu\right| \le \epsilon\mu(B)\int h\,d\nu.$$

Dividing these inequalities by $(\int h\,d\nu)(\int f\,d\nu)$ and $(\int h\,d\nu)(\int g\,d\nu)$, respectively, and adding them, we obtain

$$\left|\frac{\int f\,d\mu}{\int f\,d\nu} - \frac{\int g\,d\mu}{\int g\,d\nu}\right| \le \epsilon\left(\frac{\mu(A)}{\int f\,d\nu} + \frac{\mu(B)}{\int g\,d\nu}\right).$$

Since ϵ is arbitrary, we are done. ∎

We conclude this section by investigating the relationship between left and right Haar measures. If μ is a left Haar measure on G and $x \in G$, the measure $\mu_x(E) = \mu(Ex)$ is again a left Haar measure, because of the commutativity of left and right translations (i.e., the associative law). Hence, by Theorem 11.9 there is a positive number $\Delta(x)$ such that $\mu_x = \Delta(x)\mu$. The function $\Delta : G \to (0, \infty)$ thus defined

is independent of the choice of μ by Theorem 11.9 again; it is called the **modular function** of G.

11.10 Proposition. *Δ is a continuous homomorphism from G to the multiplicative group of positive real numbers. Moreover, if μ is a left Haar measure on G, for any $f \in L^1(\mu)$ and y in G we have*

$$\int (R_y f)\, d\mu = \Delta(y^{-1}) \int f\, d\mu. \tag{11.11}$$

Proof. For any $x, y \in G$ and $E \in \mathcal{B}_G$,

$$\Delta(xy)\mu(E) = \mu(Exy) = \Delta(y)\mu(Ex) = \Delta(y)\Delta(x)\mu(E),$$

so Δ is a homomorphism from G to $(0, \infty)$. Also, since $\chi_E(xy) = \chi_{Ey^{-1}}(x)$,

$$\int \chi_E(xy)\, d\mu(x) = \mu(Ey^{-1}) = \Delta(y^{-1})\mu(E) = \Delta(y^{-1}) \int \chi_E\, d\mu.$$

This proves (11.11) when $f = \chi_E$, and the general case follows by the usual linearity and approximation arguments. Finally, it is an easy consequence of Proposition 11.2 that the map $x \mapsto \int R_x f\, d\mu$ is continuous for any $f \in C_c(G)$ (Exercise 2), so the continuity of Δ follows from (11.11). ∎

Evidently, the left Haar measures on G are also right Haar measures precisely when Δ is identically 1, in which case G is called **unimodular**. Of course, every Abelian group is unimodular; remarkably enough, groups that are highly noncommutative are also unimodular. To be precise, let $[G, G]$ denote the smallest closed subgroup of G containing all elements of the form $[x, y] = xyx^{-1}y^{-1}$. $[G, G]$ is called the **commutator subgroup** of G; it is normal because $z[x, y]z^{-1} = [zxz^{-1}, zyz^{-1}]$, and it is trivial precisely when G is Abelian.

11.12 Proposition. *If $G/[G, G]$ is finite, then G is unimodular.*

Proof. Every continuous homomorphism (such as Δ) from G into an Abelian group must annihilate $[x, y]$ for all x, y and must therefore factor through $G/[G, G]$. If the latter group is finite, $\Delta(G)$ is a finite subgroup of $(0, \infty)$; but $(0, \infty)$ has no finite subgroups except $\{1\}$. ∎

11.13 Proposition. *If G is compact, then G is unimodular.*

Proof. For any $x \in G$, obviously $G = Gx$. Hence if μ is a left Haar measure, we have $\mu(G) = \mu(Gx) = \Delta(x)\mu(G)$, and since $0 < \mu(G) < \infty$ we conclude that $\Delta(x) = 1$. ∎

We observed above that if μ is a left Haar measure, $\widetilde{\mu}(E) = \mu(E^{-1})$ is a right Haar measure. We now show how to compute it in terms of μ and Δ.

11.14 Proposition. $d\widetilde{\mu}(x) = \Delta(x)^{-1}\, d\mu(x)$.

Proof. By (11.11), if $f \in C_c(G)$,

$$\int f(x)\Delta(x)^{-1}\,d\mu(x) = \Delta(y)\int f(xy)\Delta(xy)^{-1}\,d\mu(x)$$
$$= \int R_y f(x)\Delta(x)^{-1}\,d\mu(x).$$

Thus the functional $f \mapsto \int f\Delta^{-1}\,d\mu$ is right-invariant, so its associated Radon measure is a right Haar measure. However, this Radon measure is simply $\Delta^{-1}\,d\mu$ by Exercise 9 in §7.2; hence, by Theorem 11.9, $\Delta^{-1}\,d\mu = c\,d\widetilde{\mu}$ for some $c > 0$. If $c \neq 1$, we can pick a symmetric neighborhood U of e in G such that $|\Delta(x)^{-1} - 1| < \frac{1}{2}|c-1|$ on U. But $\widetilde{\mu}(U) = \mu(U)$, so

$$|c-1|\mu(U) = |c\widetilde{\mu}(U) - \mu(U)| = \left|\int_U (\Delta(x)^{-1} - 1)\,d\mu(x)\right| < \tfrac{1}{2}|c-1|\mu(U),$$

a contradiction. Hence $c = 1$ and $d\widetilde{\mu} = \Delta^{-1}\,d\mu$. ∎

11.15 Corollary. *Left and right Haar measures are mutually absolutely continuous.*

Exercises

1. If G is a topological group and $E \subset G$, then $\overline{E} = \bigcap\{EV : V$ is a neighborhood of $e\}$.

2. If μ is a Radon measure on the locally compact group G and $f \in C_c(G)$, the functions $x \to \int L_x f\,d\mu$ and $x \to \int R_x f\,d\mu$ are continuous.

3. Let G be a locally compact group that is homeomorphic to an open subset U of $\mathbb{R}^n$ in such a way that, if we identify G with U, left translation is an affine map — that is, $xy = A_x(y) + b_x$ where A_x is a linear transformation of $\mathbb{R}^n$ and $b_x \in \mathbb{R}^n$. Then $|\det A_x|^{-1}\,dx$ is a left Haar measure on G, where dx denotes Lebesgue measure on $\mathbb{R}^n$. (Similarly for right translations and right Haar measures.)

4. The following are special cases of Exercise 3.

a. If G is the multiplicative group of nonzero complex numbers $z = x + iy$, $(x^2 + y^2)^{-1}\,dx\,dy$ is a Haar measure.

b. If G is the group of invertible $n \times n$ real matrices, $|\det A|^{-n}\,dA$ is a left and right Haar measure, where dA = Lebesgue measure on $\mathbb{R}^{n\times n}$. (To see that the determinant of the map $X \mapsto AX$ is $|\det A|^n$, observe that if X is the matrix with columns $X^1, \ldots, X^n$, then AX is the matrix with columns $AX^1, \ldots, AX^n$.)

c. If G is the group of 3×3 matrices of the form

$$\begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} \qquad (x, y, z \in \mathbb{R}),$$

then $dx\,dy\,dz$ is a left and right Haar measure.

d. If G is the group of 2×2 matrices of the form

$$\begin{pmatrix} x & y \\ 0 & 1 \end{pmatrix} \qquad (x > 0,\ y \in \mathbb{R}),$$

then $x^{-2}dx\,dy$ is a left Haar measure and $x^{-1}\,dx\,dy$ is a right Haar measure.

5. Let G be as in Exercise 4d. Construct a Borel set in G with finite left Haar measure but infinite right Haar measure, and a left uniformly continuous function on G that is not right uniformly continuous.

6. Let $\{G_\alpha\}_{\alpha \in A}$ be a family of topological groups and $G = \prod_{\alpha \in A} G_\alpha$.

a. With the product topology and coordinatewise multiplication, G is a topological group.

b. If each G_α is compact and μ_α is the Haar measure on G_α such that $\mu_\alpha(G_\alpha) = 1$, then the Radon product of the μ_α's, as constructed in Theorem 7.28, is a Haar measure on G.

7. In Exercise 6, for each α let G_α be the multiplicative group $\{-1, 1\}$ with the discrete topology. Let μ be a Haar measure on G.

a. If $\pi_\alpha : G \to \{-1, 1\}$ is the αth coordinate function, then $\int \pi_\alpha \pi_\beta \, d\mu = 0$ for $\alpha \neq \beta$.

b. If A is uncountable, $L^2(\mu)$ is not separable even though $\mu(G) < \infty$.

8. Let $\mathbb{Q}$ have the relative topology induced from $\mathbb{R}$. Then $\mathbb{Q}$ is a topological group that is not locally compact, and there is no nonzero translation-invariant Borel measure on $\mathbb{Q}$ that is finite on compact sets.

9. Let G be a locally compact group with left Haar measure μ.

a. G has a subgroup H that is open, closed, and σ-compact. (Let H be the subgroup generated by a precompact open neighborhood of e.)

b. The restriction of μ to subsets of H is a left Haar measure on H.

c. μ is decomposable in the sense of Exercise 15 in §3.2.

d. If the topology of G is not discrete, then $\mu(\{x\}) = 0$ for all $x \in G$. In this case, μ is regular iff μ is semifinite iff μ is σ-finite iff G is σ-compact. (See Exercises 12 and 14 in §7.2.)

11.2 HAUSDORFF MEASURE

In geometric problems it is important to have a method for measuring the size of lower-dimensional sets in $\mathbb{R}^n$, such as curves and surfaces in $\mathbb{R}^3$. Differential-geometric techniques provide such a method that applies to smooth submanifolds of $\mathbb{R}^n$; see §11.4. However, there is also a measure-theoretic approach to the problem that applies to more general sets. Indeed, the basic ideas can be carried out just as easily in arbitrary metric spaces, so we begin by working in this generality.

Let (X, ρ) be a metric space. (See §0.6 for the relevant terminology.) An outer measure μ^* on X is called a **metric outer measure** if

$$\mu^*(A \cup B) = \mu^*(A) \cup \mu^*(B) \text{ whenever } \rho(A, B) > 0.$$

11.16 Proposition. *If μ^* is a metric outer measure on X, then every Borel susbset of X is μ^*-measurable.*

Proof. Since the closed sets generate the Borel σ-algebra, it suffices to show that every closed set $F \subset X$ is μ^*-measurable. Thus, given $A \subset X$ with $\mu^*(A) < \infty$, we wish to show that

$$\mu^*(A) \geq \mu^*(A \cap F) + \mu^*(A \setminus F).$$

Let $B_n = \{x \in A \setminus F : \rho(x, F) \geq n^{-1}\}$. Then B_n is an increasing sequence of sets whose union is $A \setminus F$ (since F is closed), and $\rho(B_n, F) \geq n^{-1}$. Therefore,

$$\mu^*(A) \geq \mu^*\big((A \cap F) \cup B_n\big) = \mu^*(A \cap F) + \mu^*(B_n),$$

so it will be enough to show that $\mu^*(A \setminus F) = \lim \mu^*(B_n)$. Let $C_n = B_{n+1} \setminus B_n$. If $x \in C_{n+1}$ and $\rho(x, y) < [n(n+1)]^{-1}$, then

$$\rho(y, F) \leq \rho(x, y) + \rho(x, F) < \frac{1}{n(n+1)} + \frac{1}{n+1} = \frac{1}{n},$$

so that $\rho(C_{n+1}, B_n) \geq [n(n+1)]^{-1}$. A simple induction therefore shows that

$$\begin{aligned}\mu^*(B_{2k+1}) &\geq \mu^*(C_{2k} \cup B_{2k-1}) = \mu^*(C_{2k}) + \mu^*(B_{2k-1}) \\ &\geq \mu^*(C_{2k}) + \mu^*(C_{2k-2} \cup B^{2k-3}) \cdots \geq \sum_1^k \mu^*(C_{2j}),\end{aligned}$$

and similarly $\mu^*(B_{2k}) \geq \sum_1^k \mu^*(C_{2j-1})$. Since $\mu^*(B_n) \leq \mu^*(A) < \infty$, it follows that the series $\sum_1^\infty \mu^*(C_{2j})$ and $\sum_1^\infty \mu^*(C_{2j-1})$ are convergent. But by subadditivity we have

$$\mu^*(A \setminus F) \leq \mu^*(B_n) + \sum_{n+1}^\infty \mu^*(C_j).$$

As $n \to \infty$, the last sum vanishes and we obtain

$$\mu^*(A \setminus F) \leq \liminf \mu^*(B_n) \leq \limsup \mu^*(B_n) \leq \mu^*(A \setminus F),$$

as desired. ∎

We are now ready to define Hausdorff measure. Suppose that (X, ρ) is a metric space, $p \geq 0$, and $\delta > 0$. For $A \subset X$, let

$$H_{p,\delta}(A) = \inf\Big\{\sum_1^\infty (\operatorname{diam} B_j)^p : A \subset \bigcup_1^\infty B_j \text{ and } \operatorname{diam} B_j \leq \delta\Big\},$$

with the convention that $\inf \varnothing = \infty$. As δ decreases the infimum is being taken over a smaller family of coverings of A, so $H_{p,\delta}(A)$ increases. The limit

$$H_p(A) = \lim_{\delta \to 0} H_{p,\delta}(A)$$

is called the **p-dimensional Hausdorff (outer) measure** of A.

Several comments on this definition are in order:

- The sets B_j in the definition of $H_{p,\delta}$ are arbitrary subsets of X. However, one obtains the same result if one requires the B_j's to be closed (because $\operatorname{diam} B_j = \operatorname{diam} \overline{B}_j$), or if one requires the B_j's to be open (because one can replace B_j by the open set $U_j = \{x : \rho(x, B_j) < \epsilon 2^{-j-1}\}$, whose diameter is at most $(\operatorname{diam} B_j) + \epsilon 2^{-j}$). Similarly, if $X = \mathbb{R}$, one can restrict the B_j's to be closed or open intervals.

- The intuition behind the definition of H_p is that if p is an integer and A is a "p-dimensional" subset of $\mathbb{R}^n$ such as a relatively open set in a p-dimensional linear subspace of $\mathbb{R}^n$, the amount of A that is contained in a region of diameter r should be roughly proportional to r^p.

- The restriction to coverings by sets of small diameter is necessary to provide an accurate measure of irregularly shaped sets; otherwise one could simply cover a set by itself, with the result that its measure would be at most the pth power of its diameter. Consder, for example, the curve $A_m = \{(x, \sin nx) : |x| \le 1\}$ in $\mathbb{R}^2$. Clearly $\operatorname{diam} A_m \le 2^{3/2}$ for all m, but the length of A_n tends to ∞ along with m. One needs to take $\delta \ll m^{-1}$ before $H_{1,\delta}(A)$ becomes an accurate estimate of the length of A_m.

We now derive the basic properties of H_p.

11.17 Proposition. *H_p is a metric outer measure.*

Proof. $H_{p,\delta}$ is an outer measure by Proposition 1.10, and it follows that H_p is an outer measure. If $\rho(A, B) > 0$ and $\{C_j\}$ is a covering of $A \cup B$ such that $(\operatorname{diam} C_j) \le \delta < \rho(A, B)$ for all j, then no C_j can intersect both A and B. Splitting $\sum (\operatorname{diam} C_j)^p$ into two parts according to whether $C_j \cap B = \varnothing$ or $C_j \cap A = \varnothing$ shows that $\sum (\operatorname{diam} C_j)^p \ge H_{p,\delta}(A) + H_{p,\delta}(B)$, and hence $H_{p,\delta}(A \cup B) \ge H_{p,\delta}(A) + H_{p,\delta}(B)$. As this inequality is valid whenever $\delta < \rho(A, B)$, the desired result follows by letting $\delta \to 0$. ∎

In view of Propositions 11.16 and 11.17, the restriction of H_p to the Borel sets is a measure, which we still denote by H_p and call p-dimensional Hausdorff measure.

11.18 Proposition. *H_p is invariant under isometries of X. Moreover, if Y is any set and $f, g : Y \to X$ satisfy $\rho(f(y), f(z)) \le C\rho(g(y), g(z))$ for all $y, z \in Y$, then $H_p(f(A)) \le C^p H_p(g(A))$ for all $A \subset Y$.*

Proof. The first assertion is evident from the definition of H_p. As for the second, given $\epsilon, \delta > 0$, cover $g(A)$ by sets B_j such that $\operatorname{diam} B_j \le C^{-1}\delta$ and

$\sum(\operatorname{diam} B_j)^p \le H_p(g(A)) + \epsilon$. Then the sets $B_j' = f(g^{-1}(B_j))$ cover $f(A)$, and $\operatorname{diam} B_j' \le C(\operatorname{diam} B_j) \le \delta$, so that

$$H_{p,\delta}(f(A)) \le \sum(\operatorname{diam} B_j')^p \le C^p H_p(g(A)) + C^p\epsilon.$$

The proof is completed by letting $\delta \to 0$ and $\epsilon \to 0$. ∎

11.19 Proposition. *If $H_p(A) < \infty$, then $H_q(A) = 0$ for all $q > p$. If $H_p(A) > 0$, then $H_q(A) = \infty$ for all $q < p$.*

Proof. It suffices to prove the first statement, as the second one is its contrapositive. If $H_p(A) < \infty$, for any $\delta > 0$ there exists $\{B_j\}_1^\infty$ with $A \subset \bigcup B_j$, $\operatorname{diam} B_j \le \delta$, and $\sum(\operatorname{diam} B_j)^p \le H_p(A) + 1$. But if $q > p$,

$$\sum(\operatorname{diam} B_j)^q \le \delta^{q-p}\sum(\operatorname{diam} B_j)^p \le \delta^{q-p}\big(H_p(A) + 1\big),$$

so $H_{q,\delta}(A) \le \delta^{q-p}(H_p(A) + 1)$. Letting $\delta \to 0$, we see that $H_q(A) = 0$. ∎

According to Proposition 11.19, for any $A \subset X$ the numbers

$$\inf\{p \ge 0 : H_p(A) = 0\} \quad \text{and} \quad \sup\{p \ge 0 : H_p(A) = \infty\}$$

are equal. Their common value is called the **Hausdorff dimension** of A.

From now on we restrict attention to the case $X = \mathbb{R}^n$. Our object is to show that for $p = 1, \dots, n$, H_p gives the geometrically correct notion of measure (up to a normalization constant) for p-dimensional submanifolds of $\mathbb{R}^n$. We begin with the case $p = n$.

11.20 Proposition. *There is a constant $\gamma_n > 0$ such that $\gamma_n H_n$ is Lebesgue measure on $\mathbb{R}^n$.*

Proof. H_n is a translation-invariant Borel measure on $\mathbb{R}^n$. If $Q \subset \mathbb{R}^n$ is a cube, it is easily verified that $0 < H_n(Q) < \infty$ (Exercise 10). It follows that $H_n \ne 0$ and that H_n is finite on compact sets, whence H_n is a Radon measure by Theorem 7.8. The desired result is therefore a consequence of Theorem 11.9. (The simple argument given there for the Abelian case can be read without going through the rest of §11.1: Simply read $x + y$ for xy and $-x$ for x^{-1}.) ∎

The normalization constant γ_n turns out to be the volume of a ball of diameter 1, which by Corollary 2.55 is $\pi^{n/2}/2^n\Gamma(\frac{1}{2}n + 1)$. We shall not give the proof here, as the value of γ_n is irrelevant for our purposes. (The hard part is proving the intuitively obvious fact that among all sets of diameter 1, the ball has the largest volume.) Many authors build γ_n into the definition of Hausdorff measure; that is, they define p-dimensional Hausdorff measure, for $p \in [0, \infty)$, to be $\gamma_p H_p$ where $\gamma_p = \pi^{p/2}/2^p\Gamma(\frac{1}{2}p + 1)$.

We now consider lower-dimensional sets in $\mathbb{R}^n$. If $1 \le k \le n$, a ***k*-dimensional C^1 submanifold** of $\mathbb{R}^n$ is a set $M \subset \mathbb{R}^n$ with the following property: For each

$x \in M$ there exist a neighborhood U of x in $\mathbb{R}^n$, an open set $V \subset \mathbb{R}^k$, and an injective map $f : V \to U$ of class C^1 such that $f(V) = M \cap U$ and the differential $D_x f$ — i.e., the linear map from $\mathbb{R}^k$ to $\mathbb{R}^n$ whose matrix is $[(\partial f_i/\partial x_j)(x)]$ — is injective for each $x \in V$. Such an f is called a **parametrization** of $M \cap U$. Every submanifold M can be covered by countably many U's for which $M \cap U$ has a parametrization, so for our purposes it will suffice to assume that $M = M \cap U$ has a global parametrization.

(There are other common definitions of "submanifold": M is a k-dimensional C^1 submanifold if it is locally the set of zeros of a C^1 map $g : \mathbb{R}^n \to \mathbb{R}^{n-k}$ such that $D_x g$ is surjective at each $x \in M$, or if it is locally the image of a ball in a k-dimensional linear subspace of $\mathbb{R}^n$ under a C^1 diffeomorphism of $\mathbb{R}^n$. The equivalence of these definitions with the one given above is a standard exercise in the use of the implicit function theorem.)

We begin our study of submanifolds with the linear case. If T is a linear map from $\mathbb{R}^k$ to $\mathbb{R}^n$ and $T^* : \mathbb{R}^n \to \mathbb{R}^k$ is its transpose, then T^*T is a positive semidefinite linear operator on $\mathbb{R}^n$. Its determinant is therefore nonnegative, and we may define

$$J(T) = \sqrt{\det(T^*T)}\,.$$

11.21 Proposition. *If $k \le n$, $A \subset \mathbb{R}^k$, and $T : \mathbb{R}^k \to \mathbb{R}^n$ is linear, then $H_k(T(A)) = J(T)H_k(A)$.*

Proof. If $k = n$, then $\det(T^*T) = (\det T)^2$, so $J(T) = |\det T|$ and the assertion reduces to Theorem 2.44 because of Proposition 11.20. If $k < n$, let R be a rotation of $\mathbb{R}^n$ that maps the range of T to the subspace $\mathbb{R}^k \times \{0\} = \{y \in \mathbb{R}^n : y_j = 0 \text{ for } j > k\}$, and let $S = RT$. Then $S^*S = T^*R^*RT = T^*T$, so that $J(S) = J(T)$, and $H_k(S(A)) = H_k(T(A))$ since H_k is rotation-invariant. But if we identify $\mathbb{R}^k \times \{0\}$ with $\mathbb{R}^k$, S becomes a map from $\mathbb{R}^k$ to itself, and the definition of S^*S is unchanged when this identification is made. We are therefore back in the equidimensional case, which was disposed of above. ∎

It is now easy to guess what the corresponding formula must be for a general smooth injection $f : \mathbb{R}^k \to \mathbb{R}^n$, since locally every smooth map is approximately linear. Our next lemma makes this idea precise.

11.22 Lemma. *Suppose M is a k-dimensional C^1 submanifold of $\mathbb{R}^n$ parametrized by $f : V \to \mathbb{R}^n$. For any $\alpha > 1$ there is a sequence $\{B_j\}$ of disjoint Borel subsets of V such that $V = \bigcup_1^\infty B_j$, and a sequence $\{T_j\}$ of linear maps from $\mathbb{R}^k$ to $\mathbb{R}^n$, such that*

$$\alpha^{-1}|T_j z| \le |(D_x f)z| \le \alpha|T_j z| \text{ for } x \in B_j,\ z \in \mathbb{R}^k \tag{11.23}$$

and

$$\alpha^{-1}|T_j x - T_j y| \le |f(x) - f(y)| \le \alpha|T_j x - T_j y| \text{ for } x, y \in B_j. \tag{11.24}$$

Proof. Let us fix $\epsilon > 0$ and $\beta > 1$ such that

$$\alpha^{-1} + \epsilon < \beta^{-1} < 1 < \beta < \alpha - \epsilon,$$

and let $\mathcal{T}$ be a countable dense subset of the set of linear maps from $\mathbb{R}^k$ to $\mathbb{R}^n$ (e.g., the set of matrices with rational entries). For $T \in \mathcal{T}$ and $m \in \mathbb{N}$ let $E(T,m)$ be the set of all $x \in V$ such that

$$\beta^{-1}|Tz| \le |(D_x f)z| \le \beta|Tz| \text{ for } z \in \mathbb{R}^k,$$
$$\alpha^{-1}|Tx - Ty| \le |f(x) - f(y)| \le \alpha|Tx - Ty| \text{ for all } y \in V \text{ with } |y - x| < m^{-1}.$$

The definition of $E(T,m)$ is unaffected if y and z are restricted to lie in countable dense subsets of V and $\mathbb{R}^k$; hence $E(T,m)$ is defined by countably many inequalities involving continuous functions, so it is a Borel set. It will therefore suffice to show that the sets $E(T,m)$ cover V. Indeed, each $E(T,m)$ is a countable union of sets $E_i(T,m)$ of diameter less than m^{-1}, so by disjointifying the countable collection $\{E_i(T,m) : T \in \mathcal{T},\ i,m \in \mathbb{N}\}$ we obtain the desired sets B_j and the associated maps T_j.

Suppose, then, that $x \in V$, and let $\delta_0 = \inf\{|(D_x f)z| : |z| = 1\}$. Since $D_x f$ is injective, δ_0 is positive. Choose $\delta > 0$ so that $\delta \le (\beta - 1)\delta_0$ and $\delta \le (1 - \beta^{-1})\delta_0$, and then pick $T \in \mathcal{T}$ such that $\|T - D_x f\| < \delta$. Then

$$|Tz| \le |(D_x f)z| + |Tz - (D_x f)z| \le |(D_x f)z| + \delta|z| \le \beta|(D_x f)z|,$$

and similarly $|Tz| \ge \beta^{-1}|(D_x f)z|$. This establishes the first inequality and also shows that T is injective, so $\eta = \inf\{|Tz| : |z| = 1\}$ is positive. Since f is differentiable at x, there exists $m \in \mathbb{N}$ such that

$$\big|f(y) - f(x) - (D_x f)(y - x)\big| \le \epsilon\eta|y - x| \le \epsilon|T(y - x)| \text{ for } |x - y| < m^{-1}.$$

But then

$$\begin{aligned}|f(y) - f(x)| &\le \big|f(y) - f(x) - (D_x f)(y - x)\big| + \big|(D_x f)(y - x)\big| \\ &\le \epsilon|Ty - Tx| + \beta|Ty - Tx| < \alpha|Ty - Tx|,\end{aligned}$$

and similarly $|f(y) - f(x)| > \alpha^{-1}|Ty - Tx|$. In short, $x \in E(T,m)$, so we are done. ∎

11.25 Theorem. *Let M be a k-dimensional C^1 submanifold of $\mathbb{R}^n$ parametrized by $f : V \to \mathbb{R}^n$. If A is a Borel subset of V, then $f(A)$ is a Borel subset of $\mathbb{R}^n$, and*

$$H_k(f(A)) = \int_A J(D_x f)\, dH_k(x). \tag{11.26}$$

Moreover, if ϕ is a Borel measurable function on M that is either nonnegative or in $L^1(M, H_k)$, then

$$\int_M \phi(y)\, dH_k(y) = \int_V \phi(f(x)) J(D_x f)\, dH_k(x). \tag{11.27}$$

Proof. Since V is an open subset of $\mathbb{R}^k$, it is σ-compact. It follows that if A is closed in V, then A is σ-compact, and hence $f(A)$ is σ-compact since f is continuous.

The collection of all $A \subset V$ such that $f(A)$ is Borel is therefore a σ-algebra that contains all closed sets, hence all Borel sets. We shall prove (11.26); this establishes (11.27) when $\phi = \chi_{f(A)}$, and the general result follows from the usual linearity and approximation arguments.

Given $\alpha > 1$, let $\{B_j\}$, $\{T_j\}$ be as in Lemma 11.22, and let $A_j = A \cap B_j$. It follows from (11.23) and Proposition 11.18 that

$$\alpha^{-k} H_k(T_j(E)) \le H_k\big((D_x f)(E)\big) \le \alpha^k H_k(T_j(E)) \qquad (x \in A_j,\ E \subset \mathbb{R}^k),$$

and hence by Proposition 11.21,

$$\alpha^{-k} J(T_j) \le J(D_x f) \le \alpha^k J(T_j) \qquad (x \in A_j).$$

But it also follows from (11.24) and Proposition 11.18 that

$$\alpha^{-k} H_k(T_j(A_j)) \le H_k(f(A_j)) \le \alpha^k H_k(T_j(A_j))$$

and from Proposition 11.21 that $H_k(T_j(A_j)) = J(T_j) H_k(A_j)$. Therefore,

$$\begin{aligned} \alpha^{-2k} H_k(f(A_j)) &\le \alpha^{-k} J(T_j) H_k(A_j) \le \int_{A_j} J(D_x f)\, dH_k(x) \\ &\le \alpha^k J(T_j) H_k(A_j) \le \alpha^{2k} H_k(f(A_j)). \end{aligned}$$

Summing over j, we obtain

$$\alpha^{-2k} H_k(f(A)) \le \int_A J(D_x f)\, dH_k(x) \le \alpha^{2k} H_k(f(A)),$$

so the proof is completed by letting $\alpha \to 1$. ∎

If both sides of the identities (11.26) and (11.27) are multiplied by the normalizing constant γ_k in Proposition 11.20, the integrals on the right become ordinary Lebesgue integrals, and we obtain the formula for measures and integrals on M given by Riemannian geometry (see §11.4). Moreover, if $k = n$ we have $J(D_x f) = |\det D_x f|$, so the result reduces to Theorem 2.47. See also Exercises 11 and 12.

There remains the question of whether p-dimensional Hausdorff measure is of any interest when $p \notin \mathbb{N}$. An affirmative answer will be provided in the next section.

Exercises

10. Show directly from the definition of H_n that if Q is a cube in $\mathbb{R}^n$, then $0 < H_n(Q) < \infty$. (Hint: There is a constant C such that if $E \subset \mathbb{R}^n$, the Lebesgue measure of E is at most $C(\operatorname{diam} E)^n$.)

11. If $f : (a, b) \to \mathbb{R}^n$ is a parametrization of a smooth curve (i.e., a 1-dimensional C^1 submanifold of $\mathbb{R}^n$), the Hausdorff 1-dimensional measure of the curve is $\int_a^b |f'(t)|\, dt$.

12. If $\phi : \mathbb{R}^k \to \mathbb{R}$ is a C^1 function, the graph of ϕ is a k-dimensional C^1 submanifold of $\mathbb{R}^{k+1}$ parametrized by $f(x) = (x, \phi(x))$. If $A \subset \mathbb{R}^k$, the k-dimensional volume of the portion of the graph lying above A is

$$\int_A \sqrt{1 + |\nabla\phi(x)|^2}\, dx.$$

(First do the linear case, $\phi(x) = a \cdot x$. Show that if $T : \mathbb{R}^k \to \mathbb{R}^{k+1}$ is given by $Tx = (x, a \cdot x)$, then $T^*T = I + S$ where $Sx = (a \cdot x)a$, and hence $\det(T^*T) = 1 + |a|^2$. Hint: The determinant of a matrix is the product of its eigenvalues; what are the eigenvalues of S?)

13. In any metric space, zero-dimensional Hausdorff measure is counting measure.

14. If A_m is a subset of a metric space X of Hausdorff dimension p_m for $m \in \mathbb{N}$, then $\bigcup_1^\infty A_m$ has Hausdorff dimension $\sup_m p_m$.

15. If $A \subset \mathbb{R}^n$ has Hausdorff dimension p, then $A \times A \subset \mathbb{R}^{2n}$ has Hausdorff dimension $2p$.

11.3 SELF-SIMILARITY AND HAUSDORFF DIMENSION

In this section we produce some geometrically interesting examples of sets of fractional Hausdorff dimension. The sets we consider are "self-similar," which means roughly that each small part of the set looks like a shrunken copy of the whole set. We begin by establishing the terminology necessary to discuss such sets. Our definitions will be more restrictive than is really necessary, since we aim only to give the flavor of the theory and to display some examples.

For $r > 0$, a **similitude with scaling factor** r is a map $S : \mathbb{R}^n \to \mathbb{R}^n$ of the form $S(x) = rO(x) + b$, where O is an orthogonal transformation (a rotation or the composition of a rotation and a reflection) and $b \in \mathbb{R}^n$. Suppose that $\mathbf{S} = (S_1, \ldots, S_m)$ is a finite family of similitudes with a common scaling factor $r < 1$. If $E \subset \mathbb{R}^n$, we define

$$\mathbf{S}^0(E) = E, \qquad \mathbf{S}(E) = \bigcup_1^m S_j(E), \qquad \mathbf{S}^k(E) = \mathbf{S}(\mathbf{S}^{k-1}(E)) \text{ for } k > 1.$$

E is called **invariant** under $\mathbf{S}$ if $\mathbf{S}(E) = E$. In this case, $\mathbf{S}^k(E) = E$ for all k, which means that for every $k \geq 1$, E is the union of m^k copies of itself that have been scaled down by a factor of r^k. If, in addition, these copies are disjoint or have negligibly small overlap, E can be said to be "self-similar."

Before proceeding with the theory, let us examine some of the standard examples of self-similar sets. They are all obtained by starting with a simple geometric figure, applying a family of similitudes repeatedly to it, and passing to the limit.

- Given $\beta \in (0, \frac{1}{2})$, let C_β be the Cantor set obtained from $[0, 1]$ by successively removing open middle $(1 - 2\beta)$ths of intervals, as discussed at the end of §1.5.

That is, $C_\beta = \bigcap_0^\infty \mathbf{S}^k(E)$ where

$$E = [0,1], \quad \mathbf{S} = (S_1, S_2), \quad S_1(x) = \beta x, \quad S_2(x) = \beta x + 1 - \beta.$$

See Figure 11.1a.

- The **Sierpiński gasket** Γ is the subset of $\mathbb{R}^2$ obtained from a solid triangle by dividing it into four equal subtriangles by bisecting the sides, deleting the middle subtriangle, and then iterating. Thus, if we take the initial triangle Δ to be the closed triangular region with vertices $(0,0)$, $(1,0)$, and $(\frac{1}{2},1)$, then $\Gamma = \bigcap_0^\infty \mathbf{S}^k(\Delta)$, where $\mathbf{S} = (S_1, S_2, S_3)$ with

$$S_j(x) = \tfrac{1}{2}x + b_j, \qquad b_1 = (0,0),\ b_2 = (\tfrac{1}{2},0),\ b_3 = (\tfrac{1}{4},\tfrac{1}{2}).$$

See Figure 11.1b.

- The **snowflake curve** Σ is the subset of $\mathbb{R}^2$ obtained from a line segment by replacing its middle third by the other two legs of the equilateral triangle based on that middle third (there are two such triangles; make a definite choice), and then iterating. That is, let L be the broken line joining $(0,0)$ to $(\frac{1}{3},0)$ to $(\frac{1}{2},\frac{1}{6}\sqrt{3})$ to $(\frac{2}{3},0)$ to $(1,0)$, and let

$$\begin{gathered}\mathbf{S} = (S_1,\dots,S_4), \qquad S_j(x) = \tfrac{1}{3}O_j(x) + b_j,\\ b_1 = (0,0), \quad b_2 = (\tfrac{1}{3},0), \quad b_3 = (\tfrac{1}{2},\tfrac{1}{6}\sqrt{3}), \quad b_4 = (\tfrac{2}{3},0),\\ O_1 = O_4 = I, \quad O_2 = R_{\pi/3}, \quad O_3 = R_{-\pi/3},\end{gathered}$$

where R_θ denotes the rotation through the angle θ. Then $\Sigma = \lim_{k\to\infty} \mathbf{S}^k(L)$; more precisely, $\Sigma = \bigcup_1^\infty \mathbf{S}^k(L) \setminus \bigcup_1^\infty \mathbf{S}^k(M)$, where M is the union of the open middle thirds of the line segments that constitute L. See Figure 11.1c. (The actual "snowflake" is made by joining three rotated and reflected copies of Σ in the same way in which one can join three copies of the initial figure L to make a six-pointed star.)

The Cantor sets, the Sierpiński gasket, and the snowflake curve are all clearly invariant under the families of similitudes used to generate them. The condition of negligible overlap of the rescaled copies is also satisfied, for in all cases $S_i(E) \cap S_j(E)$ is either empty or a single point.

We now return to the general theory. Suppose that $\mathbf{S} = (S_1,\dots,S_m)$ is a family of similitudes with scaling factor $r < 1$. We introduce some notation for the actions of the iterations of $\mathbf{S}$ on points, sets, and measures: For $x \in \mathbb{R}^n$, $E \subset \mathbb{R}^n$, $\mu \in M(\mathbb{R}^n)$, and $i_i,\dots,i_k \in \{1,\dots,m\}$, we set

$$\begin{gathered}x_{i_1\cdots i_k} = S_{i_1} \circ \cdots \circ S_{i_k}(x), \quad E_{i_1\cdots i_k} = S_{i_1} \circ \cdots \circ S_{i_k}(E),\\ \mu_{i_1\cdots i_k}(E) = \mu\big((S_{i_1} \circ \cdots \circ S_{i_k})^{-1}(E)\big).\end{gathered}$$

It is an important property of compact sets that are invariant under a family of similitudes that they carry measures with a corresponding invariance property.

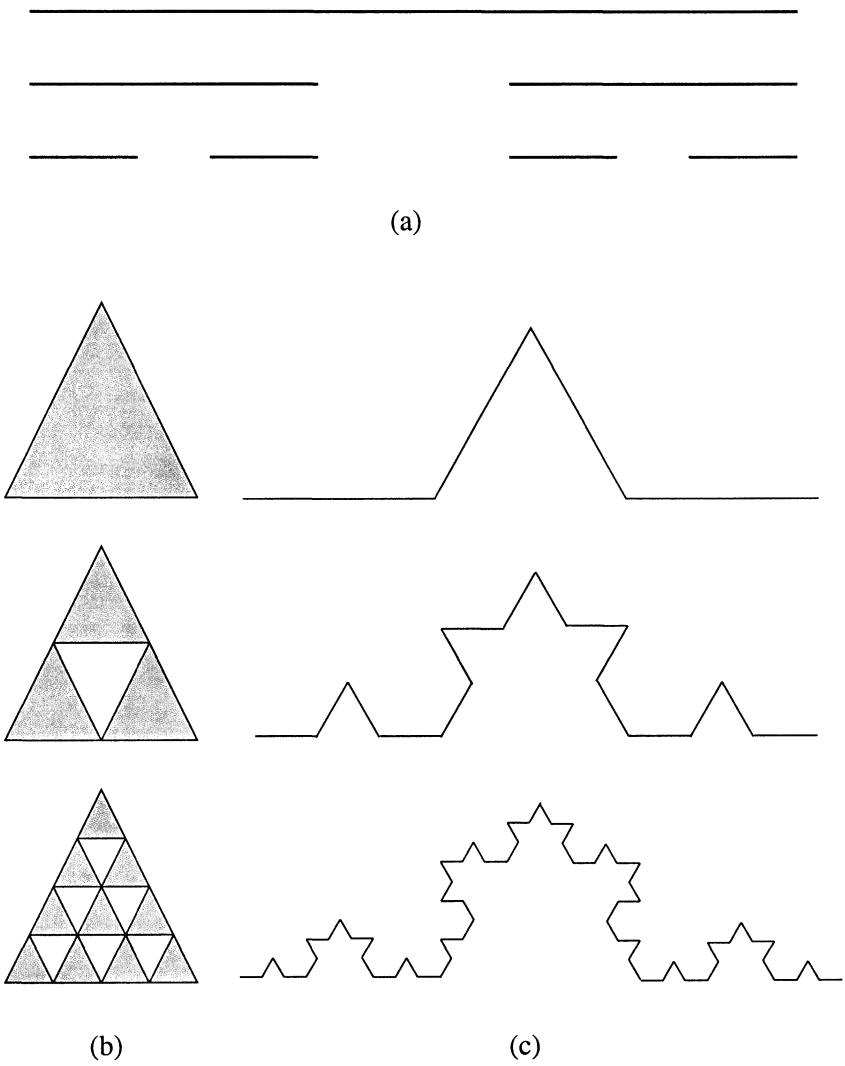

Fig. 11.1 The first three approximations to (a) the Cantor set $C_{3/8}$, (b) the Sierpiński gasket, and (c) the snowflake curve.

11.28 Theorem. *Suppose that* $\mathbf{S} = (S_1, \ldots, S_m)$ *is a family of similitudes with scaling factor* $r < 1$ *and that* X *is a nonmempty compact set that is invariant under* $\mathbf{S}$. *Then there is a Borel measure* μ *on* $\mathbb{R}^n$ *such that* $\mu(\mathbb{R}^n) = 1$, $\operatorname{supp}(\mu) = X$, *and for all* $k \in \mathbb{N}$,

$$\mu = \frac{1}{m^k} \sum_{i_1,\ldots,i_k=1}^{m} \mu_{i_1\cdots i_k}. \tag{11.29}$$

Proof. We shall construct μ as a measure on X, extending it to $\mathbb{R}^n$ by setting $\mu(X^c) = 0$. Pick $x \in X$, let δ_x be the point mass at x, and for $k \in \mathbb{N}$ define $\mu^k \in M(X)$ by

$$\mu^k = \frac{1}{m^k} \sum_{i_1,\ldots,i_k=1}^{m} [\delta_x]_{i_1\cdots i_k},$$

that is, for $f \in C_(X)$,

$$\int f\, d\mu^k = \frac{1}{m^k} \sum_{i_1,\ldots,i_k=1}^{m} f(x_{i_1\cdots i_k}).$$

Thus each μ^k is a probability measure on X. We claim that the sequence $\{\mu^k\}$ converges vaguely as $k \to \infty$. Indeed, given $f \in C(X)$ and $\epsilon > 0$, there exists $K > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $x, y \in X$ and $|x - y| < r^K(\operatorname{diam} X)$. Suppose $l > k \geq K$. Since $x_{i_1\cdots i_l} \in X_{i_1\cdots i_k}$ and $\operatorname{diam} X_{i_1\cdots i_k} = r^k(\operatorname{diam} X)$, we have

$$|f(x_{i_1\cdots i_k}) - f(x_{i_1\cdots i_k i_{k+1}\cdots i_l})| < \epsilon.$$

Summing over $i_{k+1}, \ldots, i_l$ gives

$$\left| f(x_{i_1\cdots i_k}) - \frac{1}{m^{l-k}} \sum_{i_{k+1},\ldots,i_l} f(x_{i_1\cdots i_k i_{k+1}\cdots i_l}) \right| < \epsilon,$$

and then summing over $i_1, \ldots, i_k$ yields $|\int f\, d\mu^k - \int f\, d\mu^l| < \epsilon$. Thus the sequence $\{\int f\, d\mu^k\}$ converges for every f, and the limit defines a positive linear functional on $C(X)$.

Let μ be the associated Radon measure, according to the Riesz representation theorem. Clearly $\mu(X) = \int 1\, d\mu = \lim \int 1\, d\mu^k = 1$. Also, we have $x_{i_1\cdots x_k} \in X_{i_i\cdots i_k}$, $\operatorname{diam} X_{i_1\cdots i_k} = r^k(\operatorname{diam} X)$, and $X = \bigcup X_{i_1\cdots i_k}$, so the points $x_{i_1\cdots i_k}$ ($k \in \mathbb{N}$) are dense in X; it follows that $\operatorname{supp}(\mu) = X$. Also, from the definition of μ^k we have

$$\mu^{k+l} = \frac{1}{m^k} \sum_{i_1,\ldots,i_k=1}^{m} [\mu^l]_{i_1\cdots i_k}$$

As $l \to \infty$, μ^{k+l} and μ^l both tend vaguely to μ, and composition with similitudes preserves vague convergence, so (11.29) follows. ∎

The existence of an invariant measure on the invariant set X requires no special hypotheses on the similitudes S_j, but in order to be able to compute the Hausdorff dimension of X, we need to impose an extra condition which (as we shall see in Theorem 11.33b) guarantees that the sets $S_j(X)$ have negligibly small overlap. Namely, we require $\mathbf{S}$ to possess a **separating set**: a nonempty bounded open set U such that

$$\mathbf{S}(U) \subset U, \qquad S_i(U) \cap S_j(U) = \varnothing \text{ if } i \neq j. \tag{11.30}$$

The existence of a separating set is more delicate than it might seem at first; the first condition in (11.30) will fail if U is too small, and the second one will fail if U is too big. However, all the examples considered above admit separating sets. As the reader may easily verify, for the Cantor sets C_β one can take $U = (0,1)$, for the Sierpiński gasket one can take U to be the interior of the initial triangular region Δ, and for the snowflake curve one can take U to be the interior of the triangular region with vertices $(0,0)$, $(1,0)$, and $(\frac{1}{2}, \frac{1}{6}\sqrt{3})$, i.e., the interior of the convex hull of the initial figure L.

11.31 Proposition. *Suppose that* $\mathbf{S}$ *is a family of similitudes with scaling factor* $r < 1$ *that admits a separating set* U. *Then there is a unique nonempty compact set that is invariant under* $\mathbf{S}$, *namely,* $\bigcap_0^\infty \mathbf{S}^k(\overline{U})$.

Proof. Since $\mathbf{S}(U) \subset U$ and the S_j's are continuous, we have

$$\overline{U} \supset \mathbf{S}(\overline{U}) \supset \mathbf{S}^2(\overline{U}) \supset \cdots.$$

It follows that $X = \bigcap_0^\infty \mathbf{S}^k(\overline{U})$ is a compact invariant set for $\mathbf{S}$, and it is nonempty by Proposition 4.21. If Y is another such set, let $d(Y,X) = \max_{y \in Y} \rho(y,X)$ be the maximum distance from points in Y to X. Since S_j decreases distances by a factor of r, we have $d(S_j(Y), S_j(X)) = rd(Y,X)$. But $Y = \bigcup_1^n S_j(Y)$, so $d(Y,X) \leq \max_j d(S_j(Y), X) \leq rd(Y,X)$. It follows that $d(Y,X) = 0$, which means that $Y \subset X$ since X and Y are compact. By the same reasoning, $X \subset Y$, so X is the unique compact invariant set. ∎

11.32 Lemma. *Let* c, C, δ *be positive numbers. Suppose that* $\{U_\alpha\}_{\alpha \in A}$ *is a collection of disjoint open sets in* $\mathbb{R}^n$ *such that each* U_α *contains a ball of radius* $c\delta$ *and is contained in a ball of radius* $C\delta$. *Then no ball of radius* δ *intersects more than* $(1+2C)^n c^{-n}$ *of the sets* $\overline{U}_\alpha$.

Proof. If B is a ball of radius δ and $B \cap \overline{U}_\alpha \neq \varnothing$, then $\overline{U}_\alpha$ is contained in the ball concentric with B with radius $(1+2C)\delta$. Hence, if N of the $\overline{U}_\alpha$'s intersect B, there are N disjoint balls of radius $c\delta$ contained in a ball of radius $(1+2C)\delta$. Adding up their Lebesgue measures, we see that $N(c\delta)^n \leq [(1+2C)\delta]^n$, so $N \leq (1+2C)^n c^{-n}$. ∎

11.33 Theorem. *Suppose that* $\mathbf{S} = (S_1, \ldots, S_m)$ *is a family of similitudes with scaling factor* $r < 1$ *that admits a separating set* U, *let* X *be the unique nonempty compact set that is invariant under* $\mathbf{S}$, *and let* $p = \log_{1/r} m$. *Then*

a. $0 < H_p(X) < \infty$; *in particular,* X *has Hausdorff dimension* p.

b. $H_p(S_i(X) \cap S_j(X)) = 0$ *for* $i \neq j$.

Proof. For any $k \in \mathbb{N}$ we have $X = \mathbf{S}^k(X) = \bigcup X_{i_1 \cdots i_k}$ and $\operatorname{diam} X_{i_1 \cdots i_k} = r^k(\operatorname{diam} X)$, so if $\delta_k = r^k(\operatorname{diam} X)$,

$$H_{p,\delta_k}(X) \leq \sum (\operatorname{diam} X_{i_1 \cdots i_k})^p = m^k r^{kp} (\operatorname{diam} X)^p = (\operatorname{diam} X)^p.$$

Since $\delta_k \to 0$ as $k \to \infty$, it follows that $H_p(X) \leq (\operatorname{diam} X)^p < \infty$.

Next we show that $H_p(X) > 0$. Choose positive numbers c and C so that the separating set U contains a ball of radius cr^{-1} and is contained in a ball of radius C, and let $N = (1 + 2C)^n c^{-n}$. We shall prove that $H_p(X) \geq N^{-1}$ by showing that if $\{E_j\}_1^\infty$ is any covering of X by sets of diameter ≤ 1, then $\sum (\operatorname{diam} E_j)^p \geq N^{-1}$. Since any set E with diameter δ is contained in a (closed) ball of radius δ, it is enough to show that if $X \subset \bigcup_1^\infty B_j$ where B_j is a ball of radius $\delta_j \leq 1$, then $\sum_1^\infty \delta_j^p \geq N^{-1}$.

Let μ be the measure on X given by Theorem 11.28. We claim that if B is any ball of radius $\delta \leq 1$, then $\mu(B) \leq N\delta^p$. The desired conclusion is an immediate consequence:

$$1 = \mu(X) \leq \sum \mu(B_j) \leq N \sum \delta_j^p.$$

To prove the claim, let k be the integer such that $r^k < \delta \leq r^{k-1}$. By (11.29),

$$\mu(B) = m^{-k} \sum_{i_1, \ldots, i_k = 1}^{m} \mu_{i_1 \cdots i_k}(B).$$

Since $X \subset \overline{U}$ by Proposition 11.31, we have $\operatorname{supp}(\mu_{i_1 \cdots i_k}) = X_{i_i \cdots i_k} \subset \overline{U}_{i_i \cdots i_k}$, so $\mu_{i_1 \cdots i_k}(B) = 0$ unless B intersects $\overline{U}_{i_1 \cdots i_k}$. On the other hand, iteration of (11.30) shows that the sets $U_{i_1 \cdots i_k}$ are all disjoint, and each of them contains a ball of radius $cr^{k-1} \geq c\delta$ and is contained in a ball of radius $Cr^k < C\delta$. By Lemma 11.32, B can intersect at most N of the $\overline{U}_{i_1 \cdots i_k}$'s. Therefore,

$$\mu(B) \leq N m^{-k} = N r^{kp} \leq N \delta^p,$$

as claimed.

Finally, since S_j decreases distances by a factor of r, we have $H_p(S_j(X)) = r^p H_p(X) = m^{-1} H_p(X)$ and hence $H_p(X) = \sum_1^m H_p(S_j(X))$. But since $X = \bigcup_1^m S_j(X)$, this can happen only if $H_p(S_i(X) \cap S_j(X)) = 0$ for $i \neq j$. ∎

11.34 Corollary. *The Cantor sets* C_β, *the Sierpiński gasket, and the snowflake curve have Hausdorff dimension* $\log_{1/\beta} 2$, $\log_2 3$, *and* $\log_3 4$, *respectively.*

Exercises

16. Modify the construction of the snowflake curve by using isosceles triangles rather than equilateral ones. That is, given $\frac{1}{4} < \beta < \frac{1}{2}$, let L be the broken line connecting $(0,0)$ to $(\beta,0)$ to $\frac{1}{2}(1,\sqrt{4\beta-1})$ to $(1-\beta,0)$ to $(1,0)$. Proceeding inductively, let L_k be the figure obtained by replacing each line segment in L_{k-1} by a copy of L, scaled down by a factor of β^k, and let Σ_β be the limiting set. (Thus $\Sigma_{1/3}$ is the ordinary snowflake curve.) Find the family of similitudes under which Σ_β is invariant, show that it possesses a separating set, and find the Hausdorff dimension of Σ_β.

17. Investigate analogues of the Sierpiński gasket constructed from squares, or higher-dimensional cubes, rather than triangles. (There are various possibilities here.)

18. Given $n \in \mathbb{N}$ and $p \in (0,n)$, construct Borel sets $E_1, E_2, E_3 \subset \mathbb{R}^n$ of Hausdorff dimension p, with the following properties.

a. $0 < H_p(E_1) < \infty$. (Exercise 15 or Exercise 17 could be used here.)

b. $H_p(E_2) = \infty$. (Use (a).)

c. $H_p(E_3) = 0$. (Use Exercise 14.)

19. The measure μ in Theorem 11.28 is unique.

11.4 INTEGRATION ON MANIFOLDS

This section is a brief essay on integration on manifolds for the benefit of those who are familiar with the language of differential geometry and wish to see how the geometric notions of integration fit into the measure-theoretic framework. The discussion and the notation will both be quite informal.

Let M be a C^∞ manifold of dimension m. (We work in the C^∞ category for convenience; C^1 would actually suffice.) Given a coordinate system $x = (x_1, \dots, x_m)$ on an open set $U \subset M$, one can consider Lebesgue measure $dx = dx_1 \cdots dx_m$ on U. This has no intrinsic geometric significance, for if $y = (y_1, \dots, y_m)$ is another coordinate system on U, by Theorem 2.47 we have $dy = |\det(\partial y/\partial x)|\, dx$ where $(\partial y/\partial x)$ denotes the matrix $(\partial y_i/\partial x_j)$. However, dy and dx are mutually absolutely continuous, and the Radon-Nikodym derivative $|\det(\partial y/\partial x)|$ is a C^∞ function. It therefore makes sense to define a **smooth measure** on M to be a Borel measure μ which, in any local coordinates x, has the form $d\mu = \phi^x\, dx$ where ϕ^x is a nonnegative C^∞ function. The representations of μ in different coordinate systems are then related by

$$\phi^x = |\det(\partial y/\partial x)|\phi^y. \tag{11.35}$$

(This conveniently sloppy notation is in the same spirit as the formula $du/dt = (du/dx)(dx/dt)$ for the chain rule. More precisely, if $y = F(x)$, then $\phi^x = |\det(\partial y/\partial x)|(\phi^y \circ F)$.)

Equation (11.35) may be interpreted in the language of geometry as follows. The functions $|\det(\partial y/\partial x)|$ are the transition functions for a line bundle on M; a section of this bundle, called a **density** on M, is an object that is represented in each local coordinate system x by a function ϕ^x, such that the functions for different coordinate systems are related by (11.35). In short, smooth measures can be identified with nonnegative densities on M. More generally, any density ϕ defines, at least locally, a smooth signed or complex measure μ on M, so the integral $\int_K \phi = \mu(K)$ is well defined for any compact $K \subset M$, as is $\int f\phi = \int f\, d\mu$ for any $f \in C_c(M)$.

Suppose now that M is equipped with a Riemannian metric. In any coordinate system x, the metric is represented by a positive definite matrix-valued function $g^x = (g^x_{ij})$. The matrix g^y for another coordinate system is related to g^x by

$$g^x_{ij} = \sum_{k,l} g^y_{kl} \frac{\partial y_i}{\partial x_k} \frac{\partial y_j}{\partial x_l}, \quad \text{or} \quad g^x = \left(\frac{\partial y}{\partial x}\right)^* g^y \left(\frac{\partial y}{\partial x}\right),$$

so that

$$\det g^x = \left[\det\left(\frac{\partial y}{\partial x}\right)\right]^2 \det g^y.$$

It follows that $\sqrt{\det g}$ is a positive density on M canonically associated to the metric g; it is called the **Riemannian volume density** on M. In particular, if M is a submanifold of $\mathbb{R}^n$ ($n \geq m$), it inherits a Riemannian structure from the ambient Euclidean structure. If M is parametrized by $f : V \to \mathbb{R}^n$ as in §11.2, the metric g is given in the coordinates induced by f by

$$g^x_{ij} = \sum_k \frac{\partial f_k}{\partial x_i} \frac{\partial f_k}{\partial x_j}, \quad \text{or} \quad g^x = \left(\frac{\partial f}{\partial x}\right)^* \left(\frac{\partial f}{\partial x}\right).$$

Theorem 11.25 therefore asserts that integration of the Riemannian volume density gives m-dimensional Hausdorff measure on M, up to the factor γ_m.

These ideas yield an easy construction of a left Haar measure on any Lie group, that is, any topological group that is a C^∞ manifold and whose group operations are C^∞. Namely, choose an inner product on the tangent space at the identity element and transport it to every other point by left translation. The result is a left-invariant Riemannian metric, and the associated volume density defines a left Haar measure.

The most popular things to integrate on manifolds are differential forms. For our purposes it will suffice to describe a differential m-form on an m-dimensional manifold M as a section of the line bundle whose transition functions are $\det(\partial y/\partial x)$. That is, a differential m-form ω is given in local coordinates x by a function ω^x, and the function ω^y for a different coordinate system is related to ω^x by $\omega^x = \det(\partial y/\partial x)\omega^y$. (The usual notation is $\omega = \omega^x\, dx_1 \wedge \cdots \wedge dx_m$.) Differential m-forms thus look just like densities if one restricts oneself to coordinate systems whose Jacobian matrices have positive determinant. If it is possible to do this consistently on all of M, M is called **orientable**. In this case, assuming that M is connected, the coordinate systems on subsets of M fall into two classes such that within each class one always has $\det(\partial y/\partial x) > 0$; a choice of one of these classes is an **orientation** of

M. (In $\mathbb{R}^3$, for example, one speaks of left-handed or right-handed coordinates.) If M is equipped with an orientation, therefore, differential m-forms may be identified with densities and as such may be integrated over compact subsets of M.

The notion of density can be generalized. If $0 \le \theta \le 1$, a **θ-density** on M is a section of the line bundle whose transition functions are $|\det(\partial y/\partial x)|^\theta$. (Thus, a 1-density is a density, and a 0-density is just a smooth function.) Suppose $\theta > 0$ and $p = \theta^{-1}$. If ϕ is a θ-density, $|\phi|^p$ is well defined as a nonnegative density and so can be integrated over M. The set of θ-densities ϕ such that $\|\phi\|_p = (\int |\phi|^p)^{1/p} < \infty$ is a normed linear space whose completion $L^p(M)$ is called the **intrinsic L^p space** of M. The duality results of Chapter 6 work in this setting: If ϕ_j is a θ_j-density for $j = 1, 2$, where $\theta_1 + \theta_2 = 1$, then $\phi_1\phi_2$ is a density and $|\int \phi_1\phi_2| \le \|\phi_1\|_{p_1}\|\phi_2\|_{p_2}$, where $p_j = \theta_j^{-1}$, and $L^{p_1}(M) \cong (L^{p_2}(M))^*$.

11.5 NOTES AND REFERENCES

§11.1: The existence and uniqueness of Haar measure were first proved by Haar [59] and von Neumann [155], respectively, for groups whose topology is second countable; the general case is due to Weil [158]. Our proof of existence and uniqueness follows Weil [158] and Loomis [94]. There is another proof, due to H. Cartan, which yields existence and uniqueness simultaneously and avoids the use of the axiom of choice (which we invoked via Tychonoff's theorem). This proof, as well as further references and historical remarks, can be found in Hewitt and Ross [75, §15].

Haar measure is the foundation for harmonic analysis on locally compact groups. The articles by Graham, Weiss, and Sally in Ash [7] provide a good introduction to this field; a more extensive treatment can be found in Folland [47].

§11.2: Proposition 11.16 is due to Carathéodory [22], and the theory of Hausdorff measure was developed in Hausdorff [69]. The computation of the constant γ_n can be found in Billingsley [17] or Falconer [39, §1.4]. There are other ways of defining lower-dimensional measures on $\mathbb{R}^n$, all of which agree on smooth submanifolds but sometimes differ on more irregular sets; see Federer [41].

The concept of Hausdorff measure can be generalized. If λ is any strictly increasing continuous function on $[0, \infty)$ such that $\lambda(0) = 0$, for a subset A of a metric space X one can define

$$H_{\lambda,\delta}(A) = \inf\Big\{\sum_1^\infty \lambda(\operatorname{diam} B_j) : A \subset \bigcup_1^\infty B_j,\ \operatorname{diam} B_j \le \delta\Big\}$$

and $H_\lambda(A) = \lim_{\delta\to 0} H_{\lambda,\delta}(A)$. Thus $H_o = H_\lambda$ where $\lambda(t) = t^p$. Rogers [119] contains a systematic treatment of these generalized Hausdorff measures.

§11.3: The computation of the Hausdorff dimension of the Cantor sets C_β goes back to Hausdorff [69]. The arguments presented here are due to Hutchinson [78]; they can easily be extended to families of similitudes with different scaling factors,

that is, $\mathbf{S} = (S_1, \ldots, S_m)$ where S_j has scaling factor $r_j < 1$. Such a family always has a unique nonempty compact invariant set X, and if it possesses a separating set, the Hausdorff dimension of X is the number p such that $\sum_1^m r_j^p = 1$. See Hutchinson [78] or Falconer [39, §8.3].

Self-similar sets are among the simplest examples of "fractals." Falconer [39] is a good reference for the geometric measure theory of fractals; see also Edgar [37], Falconer [40], and Mandelbrot [96] for other aspects of the theory of fractals.

Continuous curves of Hausdorff dimension > 1 can be constructed from non-differentiable functions. For example, if $f : [a, b] \to \mathbb{R}$ is Hölder continuous of exponent α, $0 < \alpha < 1$, the graph of f in $\mathbb{R}^2$ can have Hausdorff dimension as large as $2 - \alpha$, and the range of a sample path of the n-dimensional Wiener process ($n \geq 2$) almost surely has Hausdorff dimension 2. See Falconer [39, §§8.2,7].

§11.4: The theory of integration of differential forms can be found in a number of books such as Warner [157] and Loomis and Sternberg [95]; the latter book also has a discussion of densities.

Bibliography

1. R. A. Adams, *Sobolev Spaces,* Academic Press, New York (1975).

2. W. J. Adams, *The Life and Times of the Central Limit Theorem,* Kaedmon, New York (1974).

3. Weak convergence of linear functionals, *Bull. Amer. Math. Soc.* **44** (1938), 196.

4. Weak topologies of normed linear spaces, *Annals of Math.* **41** (1940), 252–267.

5. S. A. Alvarez, L^p arithmetic, *Amer. Math. Monthly* **99** (1992), 656–662.

6. C. Arzelà, Sulle funzioni di linee, *Mem. Accad. Sci. Ist. Bologna Cl. Sci. Fis. Mat.* (5) **5** (1895), 55–74.

7. J. M. Ash (ed.), *Studies in Harmonic Analysis,* Mathematical Association of America, Washington, D.C. (1976).

8. S. Banach, Sur le problème de la mesure, *Fund. Math.* **4** (1923), 7–33; also in Banach's *Oeuvres,* Vol. I, Polish Scientific Publishers, Warsaw (1967), 66–89.

9. S. Banach, *Théorie des Opérations Linéaires,* Monografje Matematyczne, Warsaw, 1932; also in Banach's *Oeuvres,* Vol. II, Polish Scientific Publishers, Warsaw (1967), 13–302.

10. S. Banach and H. Steinhaus, Sur le principe de la condensation de singularités, *Fund. Math.* **9** (1927), 50–61; also in Banach's *Oeuvres,* Vol. II, Polish Scientific Publishers, Warsaw (1967), 365–374.

11. S. Banach and A. Tarski, Sur la décomposition des ensembles de points en parties respectivement congruentes, *Fund. Math.* **6** (1924), 244–277; also in Banach's *Oeuvres,* Vol. I, Polish Scientific Publishers, Warsaw (1967), 118–148.

12. R. G. Bartle, An extension of Egorov's theorem, *Amer. Math. Monthly* **87** (1980), 628–633.

13. R. G. Bartle, Return to the Riemann integral, *Amer. Math. Monthly* **103** (1996), 625–632.

14. W. Beckner, Inequalities in Fourier analysis, *Annals of Math.* **102** (1975), 159–182.

15. C. Bennett and R. Sharpley, *Interpolation of Operators,* Academic Press, Boston (1988).

16. J. Bergh and J. Löfström, *Interpolation Spaces,* Springer-Verlag, Berlin (1976).

17. P. Billingsley, *Probability and Measure* (2nd ed.), Wiley, New York (1986).

18. C. B. Blyth and P. K. Pathak, A note on easy proofs of Stirling's theorem, *Amer. Math. Monthly* **93** (1986), 376–379.

19. N. Bourbaki, Sur les espaces de Banach, *C. R. Acad. Sci. Paris* **206** (1938), 1701–1704.

20. N. Bourbaki, *General Topology* (2 vols.), Hermann, Paris, and Addison-Wesley, Reading, Mass. (1966).

21. A. Brown, An elementary example of a continuous singular function, *Amer. Math. Monthly* **76** (1969), 295–297.

22. C. Carathéodory, *Vorlesungen über Reelle Funktionen,* Teubner, Leipzig (1918); 2nd ed. (1927), reprinted by Chelsea, New York (1948).

23. E. Čech, On bicompact spaces, *Annals of Math.* **38** (1937), 823–844.

24. P. R. Chernoff, A simple proof of Tychonoff's theorem via nets, *Amer. Math. Monthly* **99** (1992), 932–934.

25. K. L. Chung, *A Course in Probability Theory* (2nd ed.), Academic Press, New York (1974).

26. P. J. Cohen, *Set Theory and the Continuum Hypothesis,* Benjamin, New York (1966).

27. D. L. Cohn, *Measure Theory,* Birkhäuser, Boston (1980).

28. R. Coifman, M. Cwikel, R. Rochberg, Y. Sagher, and G. Weiss, Complex interpolation for families of Banach spaces, *Harmonic Analysis in Euclidean Spaces*

(*Proc. Symp. Pure Math.*, Vol. 35), part 2, American Mathematical Society, Providence, R.I. (1979), 269–282.

29. P. J. Daniell, A general form of integral, *Annals of Math.* **19** (1918), 279–294.

30. K. M. Davis and Y. C. Chang, *Lectures on Bochner-Riesz Means,* Cambridge University Press, Cambridge, U.K. (1987).

31. K. deLeeuw, The Fubini theorem and convolution formula for regular measures, *Math. Scand.* **11** (1962), 117–122.

32. J. DePree and C. Swartz, *Introduction to Real Analysis,* Wiley, New York (1988).

33. J. Dieudonné, *History of Functional Analysis,* North-Holland, Amsterdam (1981).

34. J. Dugundji, *Topology,* Allyn and Bacon, Boston (1966); reprinted by Wm. C. Brown, Dubuque, Iowa (1989).

35. N. Dunford and J. T. Schwartz, *Linear Operators* (3 vols.), Wiley-Interscience, New York (1958, 1963, and 1971).

36. H. Dym and H. P. McKean, *Fourier Series and Integrals,* Academic Press, New York (1972).

37. G. A. Edgar, *Measure, Topology, and Fractal Geometry,* Springer-Verlag, New York (1990).

38. R. Engelking, *General Topology* (rev. ed.), Heldermann Verlag, Berlin (1989).

39. K. J. Falconer, *The Geometry of Fractal Sets,* Cambridge University Press, Cambridge, U.K. (1985).

40. K. J. Falconer, *Fractal Geometry,* Wiley, New York (1990).

41. H. Federer, *Geometric Measure Theory,* Springer-Verlag, Berlin (1969).

42. C. Fefferman, Pointwise convergence of Fourier series, *Annals of Math.* **98** (1973), 551–571.

43. M. B. Feldman, A proof of Lusin's theorem, *Amer. Math. Monthly* **88** (1981), 191–192.

44. E. Fischer, Sur le convergence en moyenne, *C. R. Acad. Sci. Paris* **144** (1907), 1022–1024.

45. G. B. Folland, Remainder estimates in Taylor's theorem, *Amer. Math. Monthly* **97** (1990), 233–235.

46. G. B. Folland, *Fourier Analysis and Its Applications,* Wadsworth & Brooks/Cole, Pacific Grove, Cal. (1992).

47. G. B. Folland, *A Course in Abstract Harmonic Analysis,* CRC Press, Boca Raton, Fla. (1995).

48. G. B. Folland, *Introduction to Partial Differential Equations* (2nd ed.), Princeton University Press, Princeton, N.J. (1995).

49. G. B. Folland, Fundamental solutions for the wave operator, *Expos. Math.* **15** (1997), 25–52.

50. G. B. Folland and A. Sitaram, The uncertainty principle: a mathematical survey, *J. Fourier Anal. Appl.* **3** (1997), 207–238.

51. G. B. Folland and E. M. Stein, Estimates for the $\overline{\partial}_b$ complex and analysis on the Heisenberg group, *Commun. Pure Appl. Math.* **27** (1974), 429–522.

52. M. Fréchet, Sur quelques points du calcul fonctionnel, *Rend. Circ. Mat. Palermo* **22** (1906), 1–74.

53. M. Fréchet, Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait, *Bull. Soc. Math. France* **43** (1915), 248–265.

54. M. Fréchet, Des familles et fonctions additivies d'ensembles abstraits, *Fund. Math.* **5** (1924), 206–251.

55. I. M. Gelfand and G. E. Shilov, *Generalized Functions,* Academic Press, New York (1964).

56. K. Gödel, What is Cantor's continuum problem?, *Amer. Math. Monthly* **54** (1947), 515–525; revised and expanded version in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics,* Prentice-Hall, Englewood Cliffs, N.J. (1964), 258–273.

57. R. A. Gordon, *The Integrals of Lebesgue, Denjoy, Perron, and Henstock,* American Mathematical Society, Providence, R.I. (1994).

58. S. Grabiner, The Tietze extension theorem and the open mapping theorem, *Amer. Math. Monthly* **93** (1986), 190–191.

59. A. Haar, Der Massbegriff in der Theorie der kontinuerlichen Gruppen, *Annals of Math.* **34** (1933), 147–169; also in Haar's *Gesammelte Arbeiten,* Akadémiai Kiadó, Budapest (1959), 600–622.

60. H. Hahn, Über die Multiplikation total-additiver Mengefunktionen, *Annali Scuola Norm. Sup. Pisa* **2** (1933), 429–452.

61. H. Hahn and A. Rosenthal, *Set Functions,* University of New Mexico Press, Albuquerque, N.M. (1948).

62. P. R. Halmos, *Measure Theory,* Van Nostrand, Princeton, N.J. (1950); reprinted by Springer-Verlag, New York (1974).

63. P. R. Halmos, *Naive Set Theory,* Van Nostrand, Princeton, N.J. (1960); reprinted by Springer-Verlag, New York (1974).

64. G. H. Hardy and J. E. Littlewood, Some properties of fractional integrals I, *Math. Zeit.* **27** (1928), 565–606; also in Hardy's *Collected Papers,* Vol. III, Oxford University Press, Oxford (1969), 564–607.

65. G. H. Hardy and J. E. Littlewood, A maximal theorem with function-theoretic applications, *Acta Math.* **54** (1930), 81–116; also in Hardy's *Collected Papers,* Vol. II, Oxford University Press, Oxford (1967), 509–544.

66. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (2nd ed.), Cambridge University Press, Cambridge, U.K. (1952).

67. D. G. Hartig, The Riesz representation theorem revisited, *Amer. Math. Monthly* **90** (1983), 277–280.

68. F. Hausdorff, *Grundzüge der Mengenlehre,* Verlag von Veit, Leipzig (1914); reprinted by Chelsea, New York (1949).

69. F. Hausdorff, Dimension und äusseres Mass, *Math. Annalen* **79** (1919), 157–179.

70. T. Hawkins, *Lebesgue's Theory of Integration,* University of Wisconsin Press, Madison, Wisc. (1970).

71. J. Hennefeld, A nontopological proof of the uniform boundedness theorem, *Amer. Math. Monthly* **87** (1980), 217.

72. R. Henstock, *The General Theory of Integration,* Oxford University Press, Oxford, U.K. (1991).

73. E. Hewitt, On two problems of Urysohn, *Annals of Math* **47** (1946), 503–509.

74. E. Hewitt and R. E. Hewitt, The Gibbs-Wilbraham phenomenon: an episode in Fourier analysis, *Arch. Hist. Exact Sci.* **21** (1979), 129–160.

75. E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis,* Vol. I, Springer-Verlag, Berlin (1963).

76. E. Hewitt and K. Stromberg, *Real and Abstract Analysis,* Springer-Verlag, Berlin (1965).

77. L. Hörmander, *The Analysis of Linear Partial Differential Operators,* Vol. I, Springer-Verlag, Berlin (1983).

78. J. E. Hutchinson, Fractals and self-similarity, *Indiana U. Math. J.* **30** (1981), 713–747.

79. G. W. Johnson, An unsymmetric Fubini theorem, *Amer. Math. Monthly* **91** (1984), 131–133.

80. S. Kakutani, Concrete representation of abstract (M)-spaces, *Annals of Math.* **42** (1941), 994–1024.

81. S. Kakutani and J. C. Oxtoby, Construction of a non-separable invariant extension of the Lebesgue measure space, *Annals of Math.* **52** (1950), 580–590.

82. J. L. Kelley, The Tychonoff product theorem implies the axiom of choice, *Fund. Math.* **37** (1950), 75–76.

83. J. L. Kelley, *General Topology,* Van Nostrand, Princeton, N.J. (1955); reprinted by Springer-Verlag, New York (1975).

84. F. B. Knight, *Essentials of Brownian Motion and Diffusion,* American Mathematical Society, Providence, R.I. (1981).

85. A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung,* Springer-Verlag, Berlin (1933); translated as *Foundations of the Theory of Probability,* Chelsea, New York (1950).

86. H. König, *Measure and Integration,* Springer-Verlag, Berlin (1997).

87. T. W. Körner, *Fourier Analysis,* Cambridge University Press, Cambridge, U.K. (1988).

88. J. Kupka and K. Prikry, The measurability of uncountable unions, *Amer. Math. Monthly* **91** (1984), 85–97.

89. A. V. Lair, A Rellich compactness theorem for sets of finite volume, *Amer. Math. Monthly* **83** (1976), 350–351.

90. J. Lamperti, *Probability* (2nd ed.), Wiley, New York (1996).

91. H. Lebesgue, Intégrale, longueur, aire, *Annali Mat. Pura Appl.* (3) **7** (1902), 231–359; also in Lebesgue's *Oeuvres Scientifiques,* Vol. I, L'Enseignement Mathématique, Geneva (1972), 201–331.

92. H. Lebesgue, Sur l'intégration des fonctions discontinues, *Ann. Sci. Ecole Norm. Sup.* **27** (1910), 361–450; also in Lebesgue's *Oeuvres Scientifiques,* vol. II, L'Enseignement Mathématique, Geneva (1972), 185–274.

93. E. H. Lieb and M. Loss, *Analysis,* American Mathematical Society, Providence, R.I. (1997).

94. L. H. Loomis, *An Introduction to Abstract Harmonic Analysis,* Van Nostrand, Princeton, N.J. (1953).

95. L. H. Loomis and S. Sternberg, *Advanced Calculus,* Addison-Wesley, Reading, Mass. (1968); reprinted by Jones and Bartlett, Boston (1990).

96. B. Mandelbrot, *The Fractal Geometry of Nature,* Freeman, San Francisco (1983).

97. J. Marcinkiewicz, Sur l'interpolation d'opérateurs, *C. R. Acad. Sci. Paris* **208** (1939), 1272–1273.

98. A. Markov, On mean values and exterior densities [Russian], *Mat. Sbornik* **4(46)** (1938), 165–191.

99. R. M. McLeod, *The Generalized Riemann Integral,* Mathematical Association of America, Washington, D.C. (1980).

100. A. G. Miamee, The inclusion $L^p(\mu) \subset L^q(\nu)$, *Amer. Math. Monthly* **98** (1991), 342–345.

101. E. H. Moore and H. L. Smith, a general theory of limits, *Amer. J. Math.* **44** (1922), 102–121.

102. J. Nagata, *Modern General Topology* (2nd rev. ed.), North-Holland, Amsterdam (1985).

103. E. Nelson, Regular probability measures on function spaces, *Annals of Math.* **69** (1959), 630–643.

104. E. Nelson, Feynman integrals and the Schrödinger equation, *J. Math. Phys.* **5** (1964), 332–343.

105. E. Nelson, *Dynamical Theories of Brownian Motion,* Princeton University Press, Princeton, N.J. (1967).

106. D. J. Newman, Fourier uniqueness via complex variables, *Amer. Math. Monthly* **81** (1974), 379–380.

107. O. Nikodym, Sur une généralisation des intégrales de M. J. Radon, *Fund. Math.* **15** (1930), 131–179.

108. W. F. Pfeffer, *Integrals and Measures,* Marcel Dekker, New York (1977).

109. W. F. Pfeffer, *The Riemann Approach to Integration,* Cambridge University Press, London (1993).

110. M. Plancherel, Contribution à l'étude de la représentation d'une fonction arbitraire par des intégrales definies, *Rend. Circ. Mat. Palermo* **30** (1910), 289–335.

111. J. Radon, Theorie und Anwendungen der absolut additiv Mengenfunktionen, *S.-B. Math.-Natur. Kl. Kais. Akad. Wiss. Wien* **122**.IIa (1913), 1295–1438; also in Radon's *Collected Works,* vol. I, Birkhäuser, Basel (1987), 45–188.

112. M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis* (2nd ed.), Academic Press, New York (1980).

113. B. Riemann, Über die Hypothesen, welche der Geometrie zu Grunde liegen, in Riemann's *Gesammelte Mathematische Werke,* Teubner, Leipzig (1876), 254–269.

114. F. Riesz, Sur les systèmes orthogonaux de fonctions, *C. R. Acad. Sci. Paris* **144** (1907), 615–619; also in Riesz's *Oeuvres Complètes,* Vol. I, Akaémiai Kiadó, Budapest (1960), 378-381.

115. F. Riesz, Sur une espèce de géometrie analytique des systèmes de fonctions sommables, *C. R. Acad. Aci. Paris* **144** (1907), 1409–1411; also in Riesz's *Oeuvres Complètes,* Vol. I, Akaémiai Kiadó, Budapest (1960), 386–388.

116. F. Riesz, Sur les opérations fonctionnelles linéaires, *C. R. Acad Sci. Paris* **149** (1909), 974–977; also in Riesz's *Oeuvres Complètes,* Vol. I, Akaémiai Kiadó, Budapest (1960), 400–402.

117. F. Riesz, Untersuchungen über Systeme integrierbarer Funktionen, *Math. Annalen* **69** (1910), 449–497; also in Riesz's *Oeuvres Complètes,* Vol. I, Akaémiai Kiadó, Budapest (1960), 441–489.

118. M. Riesz, Sur les maxima des formes bilinéaires et sur les fonctionnelles linéaires, *Acta Math.* **49** (1926), 465–497; also in Riesz's *Collected Papers,* Springer-Verlag, Berlin (1988), 377–409.

119. C. A. Rogers, *Hausdorff Measures,* Cambridge University Press, Cambridge, U.K. (1970).

120. J. L. Romero, When is $L^p(\mu)$ contained in $L^q(\mu)$?, *Amer. Math. Monthly* **90** (1983), 203–206.

121. H. L. Royden, *Real Analysis* (3rd ed.), Macmillan, New York (1988).

122. L. A. Rubel, A complex-variables proof of Hölder's inequality, *Proc. Amer. Math. Soc.* **15** (1964), 999.

123. W. Rudin, Lebesgue's first theorem, in L. Nachbin (ed.), *Mathematical Analysis and Applications* (Advances in Math. Supplementary Studies, vol. 7B), Academic Press, New York (1981), 741–747.

124. W. Rudin, Well-distributed measurable sets, *Amer. Math. Monthly* **90** (1983), 41–42.

125. W. Rudin, *Real and Complex Analysis* (3rd ed.), McGraw-Hill, New York (1987).

126. W. Rudin, *Functional Analysis* (2nd ed.), McGraw-Hill, New York (1991).

127. S. Saeki, A proof of the existence of infinite product probability measures, *Amer. Math. Monthly* **103** (1992), 682–683.

128. S. Saks, *Theory of the Integral* (2nd ed.), Monografje Matematyczne, Warsaw (1937); reprinted by Hafner, New York (1938).

129. I. Schur, Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen, *J. Reine Angew. Math.* **140** (1911), 1–28; also

in Schur's *Gesammelte Abhandlungen,* Vol. I, Springer-Verlag, Berlin (1973), 464–491.

130. J. Schwartz, A note on the space L_p^*, *Proc. Amer. Math. Soc.* **2** (1951), 270–275.

131. J. Schwartz, The formula for change of variables in a multiple integral, *Amer. Math. Monthly* **61** (1954), 81–85.

132. L. Schwartz, *Théorie des Distributions* (2nd ed.), Hermann, Paris (1966).

133. 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), 514–520.

134. W. Sierpiński, Sur un problème concernant les ensembles mésurables superficiellement, *Fund. Math.* **1** (1920), 112–115; also in Sierpiński's *Oeuvres Choisis,* vol. II, Polish Scientific Publishers, Warsaw (1975), 328–330.

135. R. M. Smullyan and M. Fitting, *Set Theory and the Continuum Problem,* Oxford University Press, Oxford, U.K. (1996).

136. S. L. Sobolev, A new method for solving the Cauchy problem for normal linear hyperbolic equations [Russian], *Mat. Sbornik* **1(43)** (1936), 39–72.

137. S. L. Sobolev, On a theorem of functional analysis [Russian], *Mat. Sbornik* **4(46)** (1938), 471–496.

138. R. M. Solovay, A model of set theory in which every set of reals is Lebesgue measurable, *Annals of Math.* **92** (1970), 1–56.

139. S. M. Srivastava, *A Course in Borel Sets,* Springer-Verlag, New York (1998).

140. E. M. Stein, *Singular Integrals and Differentiability Properties of Functions,* Princeton University Press, Princeton, N.J. (1970).

141. E. M. Stein, *Harmonic Analysis,* Princeton University Press, Princeton, N.J. (1993).

142. E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces,* Princeton University Press, Princeton, N.J. (1971).

143. H. Steinhaus, Additive und stetige Funktionaloperationen, *Math. Zeit.* **5** (1919), 186–221; also pp. 252–288 in Steinhaus's *Selected Papers,* Polish Scientific Publishers, Warsaw (1985).

144. M. H. Stone, Applications of the theory of Boolean rings to general topology, *Trans. Amer. Math. Soc.* **41** (1937), 375–481.

145. M. H. Stone, The generalized Weierstrass approximation theorem, *Math. Mag.* **21** (1948), 167–184 and 237–254; reprinted in R. C. Buck (ed.), *Studies in Modern*

Analysis, Mathematical Association of America, Washington, D.C. (1962), 30–87.

146. K. Stromberg, The Banach-Tarski paradox, *Amer. Math. Monthly* **86** (1979), 151–161.

147. M. E. Taylor, *Pseudodifferential Operators,* Princeton University Press, Princeton, N.J. (1981).

148. H. J. ter Horst, Riemann-Stieltjes and Lebesgue-Stieltjes integrability, *Amer. Math. Monthly* **91** (1984), 551–559.

149. G. O. Thorin, An extension of a convexity theorem due to M. Riesz, *Kungl. Fysiografiska Saellskapet i Lund Forhaendlinger* **8** (1939), No. 14.

150. F. Treves, *Topological Vector Spaces, Distributions, and Kernels,* Academic Press, New York (1967).

151. A. Tychonoff, Über die topologische Erweiterung von Raümen, *Math. Annalen* **102** (1929), 544–561.

152. P. Urysohn, Über die Mächtigkeit der zusammenhängenden Mengen, *Math. Annalen* **94** (1925), 262–295.

153. P. Urysohn, Zum Metrisationsproblem, *Math. Annalen* **94** (1925), 309–315.

154. J. von Neumann, Mathematische Begründung der Quantenmechanik, *Göttinger Nachr.* (1927), 1–57; also in von Neumann's *Collected Works,* Vol. I, Pergamon Press, New York (1961), 151–207.

155. J. von Neumann, The uniqueness of Haar's measure, *Mat. Sbornik* **1(43)** (1936), 721–734; also in von Neumann's *Collected Works,* Vol. IV, Pergamon Press, New York (1962), 91–104.

156. L. E. Ward, A weak Tychonoff theorem and the axiom of choice, *Proc. Amer. Math. Soc.* **13** (1962), 757–758.

157. F. W. Warner, *Foundations of Differentiable Manifolds and Lie Groups,* Scott Foresman, Glenview, Ill. (1971); reprinted by Springer-Verlag, New York (1983).

158. A. Weil, *L'Intégration dans les Groupes Topologiques et ses Applications,* Hermann, Paris (1940).

159. N. Wiener, Differential-space, *J. Math. and Phys.* **2** (1923), 131–174; also in Wiener's *Collected Works,* Vol. I, MIT Press, Cambridge, Mass. (1976), 455–598.

160. N. Wiener, The average value of a functional, *Proc. London Math. Soc.* **22** (1924), 454–467; also in Wiener's *Collected Works,* Vol. I, MIT Press, Cambridge, Mass. (1976), 499–512.

161. N. Wiener, The ergodic theorem, *Duke Math. J.* **5** (1939), 1–18; also in Wiener's *Collected Works,* Vol. I, MIT Press, Cambridge, Mass. (1976), 672–689.

162. C. S. Wong, A note on the central limit theorem, *Amer. Math. Monthly* **84** (1977), 472.

163. K. Yosida, *Functional Analysis* (6th ed.), Springer-Verlag, New York (1980).

164. W. H. Young, On the multiplication of successions of Fourier constants, *Proc. Royal Soc.* (*A*) **87** (1912), 331–339.

165. A. C. Zaanen, Continuity of measurable functions, *Amer. Math. Monthly* **93** (1986), 128–130.

166. A. Zygmund, On a theorem of Marcinkiewicz concerning interpolation of operators, *J. Math. Pures Appl.* (9) **35** (1956), 223–248.

167. A. Zygmund, *Trigonometric Series* (2 vols., reprinted in 1 vol.), Cambridge University Press, Cambridge, U.K. (1968).

Index of Notation

For the basic notation used throughout the book for sets, mappings, numbers, and metric and topological spaces, see Chapter 0. Notation used only in the section in which it is introduced is, for the most part, not listed here.

Analysis on Euclidean space: $x \cdot y$ (dot product), 235. ∂^α, x^α, $\alpha!$, $|\alpha|$ (multi-index notation), 236. $\mathbb{T}^n$ (n-torus), 238.

Functions and operations on functions: $f^\pm$ (positive and negative parts), 46. sgn, 46. χ_E (characteristic function), 46. f_x, f^y (sections), 65. Γ, 58. $\operatorname{supp}(f)$ (support), 132, 284. λ_f (distribution function), 197. $\tau_y f$ (translation), 238. $f * g$ (convolution), 239, 285. ϕ_t (dilation), 242. $\widehat{f}$, $\mathcal{F}f$ (Fourier transform), 248, 249, 295. $\langle F, \phi \rangle$, 283. $\widetilde{f}$ (reflection), 283.

Integrals: The basic notation is developed in §2.2. $\int f(x)\,dx$ (Lebesgue integral), 57, 70. $\iint f\,d\mu\,d\nu$ (iterated integral), 67. $\int g\,dF$ (Stieltjes integral), 107.

Measures: μ_F (Lebesgue-Stieltjes measure), 35. m, m^n (Lebesgue measure), 37, 70. $\mu \times \nu$ (product), 64. σ (surface measure on sphere), 78. $\nu^\pm$ (positive and negative variations), 87. $|\nu|$ (total variation), 87, 93. $\mu \perp \nu$ (mutual singularity), 87. $\mu \ll \nu$ (absolute continuity), 88. $f\,d\mu$, 89. $d\mu/d\nu$ (Radon-Nikodym derivative), 91. $\operatorname{supp}(\mu)$ (support), 215. $\mu \mathbin{\widehat{\times}} \nu$ (Radon product), 227. $\mu * \nu$ (convolution), 270.

Norms and seminorms: $\|f\|_u$ (uniform norm), 121. $\|T\|$ (operator norm), 154. $\|f\|_p$ (L^p norm), 181. $\|f\|_\infty$ (L^∞ norm), 184. $[f]_p$ (weak L^p quasi-norm), 198. $\|\mu\|$ (measure norm), 222. $\|\phi\|_{(N,\alpha)}$ (Schwartz space norm), 237. $\|f\|_{(s)}$ (Sobolev norm), 302.

Probability theory: $E(X)$ (expectation), 314. $\sigma^2(X)$ (variance), 314. P_ϕ (image measure, distribution), 314. $\nu_\mu^{\sigma^2}$ (normal distribution), 325.

Sets: F_σ, $F_{\sigma\delta}$, G_δ, $G_{\delta\sigma}$, 22. E_x, E^y (sections), 65.

σ-algebras: $\mathcal{M}(\mathcal{E})$ (σ-algebra generated by $\mathcal{E}$), 22. $\mathcal{B}_X$ (Borel sets), 22. $\bigotimes_{\alpha\in A}\mathcal{M}_\alpha$, $\mathcal{M}\otimes\mathcal{N}$ (products), 22. $\mathcal{L}$, $\mathcal{L}^n$ (Lebesgue measurable sets), 37, 70. $\mathcal{B}_X^0$ (Baire sets), 215.

Spaces of functions, measures, etc.: L^+, 49. L^1, 54, 181. L^1_{loc}, 96. BV, 102. NBV, 103. $C(X,Y)$, 119. $B(X,\mathbb{R})$, 121. $BC(X,\mathbb{R})$, 121. $B(X)$, 121. $C(X)$, 121. $BC(X)$, 121. $C_c(X)$, 132. $C_0(X)$, 132. $L(\mathfrak{X},\mathfrak{Y})$, 154. $\mathfrak{X}^*$, 157. L^2, 172,181. l^2, 173, 181. L^p, 181, 184. l^p, 181. L^∞, 184. weak L^p, 198. $M(X)$, 222. C^k, 235. C^∞, 235. C_c^∞, 235. $\mathcal{S}$, 237. $\mathcal{D}'$, 282. $\mathcal{E}'$, 291. CS', 293. H_s, 301. H_s^{loc}, 306.

Index

A

Abel mean, 261
Abel summation, 261
Absolute continuity
  of a function, 105
  of a measure, 88
Absolute convergence, 152
Accumulation point, 114
Adjoint, 160, 177
A.e., 26
Alaoglu's theorem, 169
Alexandroff compactification, 132
Algebra
  Banach, 154
  of functions, 139
  of sets, 21
Almost every(where), 26
Almost sure(ly), 314
Almost uniform convergence, 62
Approximate identity, 245
Arcwise connected set, 124
Arzelà-Ascoli theorem, 137
A.s., 314
Axiom
  of choice, 6
  of countability, 116
  of separation, 116

B

Baire category theorem, 161
Baire classes, 83
Baire set, 215
Ball, 13
Banach algebra, 154
Banach space, 152
Banach-Tarski paradox, 20
Base for a topology, 115
Bessel's inequality, 175
Bijective mapping, 4
Binomial distribution, 320
Bochner integral, 179
Bochner-Riesz means, 279
Bolzano-Weierstrass property, 15
Borel isomorphism, 83
Borel measurable function, 44
Borel measure 33
Borel set, 22
Borel σ-algebra, 22
Borel's normal number theorem, 330
Borel-Cantelli lemma, 321
Boundary, 114
Bounded linear map, 153
Bounded set, 15
Bounded variation, 102

C

Cantor function, 39
Cantor set, 38
Cantor-Lebesgue function, 39
Carathéodory's theorem, 29
Cartesian product, 3–4
Cauchy net, 167
Cauchy sequence, 14
  in measure, 61
Cauchy-Riemann equation, 308
Central limit theorem, 326
Cesàro mean, 262
Cesàro summation, 262
Characteristic function, 46
  of a probability distribution, 314
Chebyshev's inequality, 193
Chi-square distribution, 320
Closed graph theorem, 163
Closed linear map, 163
Closed set, 13, 114
Closure, 13, 114
Closure operator, 119
Cluster point, 118, 126
Coarser topology, 114
Cofinal net, 127
Cofinite topology, 113
Commutator subgroup, 346
Compact set, 16, 128
Compact space, 128
  countably, 130
  locally, 131
  sequentially, 130
Compactification, 144
  one-point, 132
  Stone-Čech, 144
Compactly supported function, 132
Complement, 3
Complete measure, 26
Complete metric space, 14
Complete orthonormal set, 175
Complete topological vector space, 167
Completely regular algebra, 146
Completely regular space, 123
Completion
  of a measure, 27
  of a normed vector space, 159
  of a σ-algebra, 27
Complex measure, 93
Composition, 3
Condensation of singularities, 165
Conditional expectation, 93
Conjugate exponent, 183
Connected component, 119
Connected set, 118
Constant-coefficient operator, 273
Content, 72
Continuity, 14, 119
  absolute, 88, 105
  Hölder, 138
  Lipschitz, 108
  of linear maps on C_c^∞, 282
  of measures, 26
  uniform, 14, 238, 340
Continuous measure, 106
Continuum, 8
Continuum hypothesis, 17
Convergence
  absolute, 152
  almost uniform, 62
  in C_c^∞, 282
  in L^1, 54
  in measure, 61
  in probability, 314
  of a filter, 147
  of a net, 126
  of a sequence, 14, 116
  of a series, 152
Convex function, 109
Convolution
  of distributions, 285, 292
  of functions, 239
  of measures, 270
Coordinate, 4
Coordinate map, 4
Countable additivity, 24
Countable ordinals, 10
Countable set, 7
Countably compact space, 130
Counting measure, 25
Cover, 15
Cube, 71, 143

D

Daniell integral, 81
Decomposable measure, 92
Decreasing function, 12
Decreasing rearrangement, 199
DeMorgan's laws, 3
Dense set, 13, 114
Density, 362
  of a set, 100
  Riemannian volume, 362
Denumerable set, 7
Derivative
  L^p, 246
  of a distribution, 284
  Radon-Nikodym, 91
Diameter, 15
Diffeomorphism, 74
Difference of sets, 3
Differential operator, 273

Dirac measure, 25
Directed set, 125
Dirichlet kernel, 264
Dirichlet problem, 274
Disconnected set, 118
Discrete, 113
Discrete measure, 106
Discrete topology, 113
Disjoint sets, 2
Distribution function, 33, 197, 315
Distribution, 282
  homogeneous, 289
  joint, 315
  of a random variable, 315
  periodic, 297
  tempered, 293
Domain, 4
Dominated convergence theorem, 54
Dual space, 157

E

Egoroff's theorem, 62
Elementary family, 23
Elliptic differential operator, 307, 311
Elliptic regularity theorem, 307
Embedding, 120
Entropy, 325
Equicontinuity, 137
Equivalence class, 3
Equivalence relation, 3
Equivalent metric, 16
Equivalent norm, 152
Essential range, 187
Essential supremum, 184
Event, 314
Eventually, 126
Expectation, 314
  conditional, 93
Extended integrable function, 86
Extended real number system, 10

F

F_σ and $F_{\sigma\delta}$ sets, 22
Fatou's lemma, 52
Fejér kernel, 269
Field of sets, 21
Filter, 147
Finer topology, 114
Finite intersection property, 128
Finite measure, 25
Finite signed measure, 88
Finitely additive measure, 25
First category, 161
First countable space, 116
First uncountable ordinal, 10
Fourier integral, 253
Fourier inversion theorem, 251
Fourier series, 248
Fourier transform, 248–249
  of a measure, 272
  of a tempered distribution, 295
Fourier-Stieltjes transform, 272
Fractional integral, 77
Frequently, 126
Fréchet space, 167
Fubini's theorem, 67–68, 229
Fubini-Tonelli theorem, 67–68, 229
Function, 3
Fundamental theorem of calculus, 106

G

G_δ and $G_{\delta\sigma}$ sets, 22
Gamma distribution, 320
Gamma function, 58
Gauge, 82
Gauss kernel, 260
Gaussian distribution, 325
Generalized Cantor set, 39
Gibbs phenomenon, 268
Gram-Schmidt process, 175
Graph of a linear map, 162

H

H-interval, 33
Haar measure, 341
Hahn decomposition, 87
Hahn decomposition theorem, 86
Hahn-Banach theorem, 157–158
Hardy's inequalities, 196
Hardy-Littlewood maximal function, 96
Hausdorff dimension, 351
Hausdorff maximal principle, 5
Hausdorff measure, 350
Hausdorff space, 117
Hausdorff-Young inequality, 248
Heat equation, 275
Heine-Borel property, 15
Heisenberg's inequality, 255
Henstock-Kurzweil integral, 82
Hermite function, 256
Hermite operator, 256
Hermite polynomial, 257
Hilbert space, 172
Hilbert's inequality, 196
Homeomorphism, 119
Homogeneous distribution, 289
Hull of an ideal, 142
Hölder continuity, 138
Hölder's inequality, 182, 196

I

Ideal in an algebra, 142
Identically distributed random variables, 315
Iff, 1
Image, 3
Image measure, 314
Increasing function, 12
Independent events, 315
Independent random variables, 315
Indicator function, 46
Indiscrete topology, 113
Inequality
  Bessel's, 175
  Chebyshev's, 193
  Hardy's, 196
  Hausdorff-Young, 248
  Heisenberg's, 255
  Hilbert's, 196
  Hölder's, 182, 196
  Jensen's, 109
  Kolmogorov's, 322
  Minkowski's, 183, 194
  Schwarz, 172
  triangle, 151
  Wirtinger's, 254
  Young's, 240–241
Infimum, 9–10
Initial segment, 9
Injective mapping, 4
Inner measure, 32
Inner product, 171
Inner regular measure, 212
Integrable function, 53
  weakly, 179
  extended, 86
  locally, 95
Integral
  Bochner, 179
  Daniell, 81
  fractional, 77
  Henstock-Kurzweil, 82
  Lebesgue, 56
  Lebesgue-Stieltjes, 107
  of a complex function, 53
  of a nonnegative function, 50
  of a real function, 53
  of a simple function, 49
  Riemann, 57
Interior, 13, 114
Invariant set, 355
Inverse, 4
Inverse image, 3
Invertible linear map, 154
Isometry, 154
Isomorphism
  Borel, 83
  linear, 154
  order, 5
  unitary, 176

J

Jensen's inequality, 109
Joint distribution, 315
Jordan content, 72
Jordan decomposition
  of a function, 103
  of a signed measure, 87
Jordan decomposition theorem, 87

K

Kernel of a closed set, 142
Kolmogorov's inequality, 322
Krein extension theorem, 161

L

L^p derivative, 246
L^p norm, 181, 184
L^p space, 181, 184
  intrinsic, 363
  weak, 198
Laplacian, 273
Lattice of functions, 139
Law of large numbers
  strong, 322–323
  weak, 321
Law of the iterated logarithm, 336
LCH space, 131
Lebesgue decomposition, 91
Lebesgue differentiation theorem, 98
Lebesgue integral, 56
Lebesgue measurable function, 44
Lebesgue measurable set, 37, 70
Lebesgue measure, 37, 70
Lebesgue set, 97
Lebesgue-Radon-Nikodym theorem, 90, 93
Lebesgue-Stieltjes integral, 107
Lebesgue-Stieltjes measure, 35
Left continuous function, 12
Left-invariant measure, 341
Lemma
  Borel-Cantelli, 321
  Fatou's, 52
  monotone class, 66
  Riemann-Lebesgue, 249
  three lines, 200
  Urysohn's, 122, 131, 245
  Weyl's, 308
  Zorn's, 5
Limit inferior, 2, 11
Limit of a net, 126

Limit superior, 2, 11
Linear functional, 157
  positive, 211
Linear ordering, 5
Liouville's theorem, 300
Lipschitz continuity, 108
Localized Sobolev space, 306
Locally compact group, 341
Locally compact space, 131
Locally convex space, 165
Locally finite cover, 135
Locally integrable function, 95
Locally measurable set, 28
Locally null set, 192
Lower bound, 5
Lower semicontinuous function, 218
LSC function, 218
Lusin's theorem, 64, 217

M

Map, 3
Mapping, 3
Marcinkiewicz interpolation theorem, 202
Maximal element, 5
Maximal function, 96, 246
Maximal theorem, 96
Meager set, 161
Mean ergodic theorem, 178
Mean, 314–315
Measurable function, 44
Measurable mapping, 43
Measurable set, 25
  Lebesgue, 37, 70
  locally, 28
  with respect to an outer measure, 29
Measurable space, 25
Measure, 24
  Borel, 33
  complete, 26
  complex, 93
  continuous, 106
  counting, 25
  decomposable, 92
  dicrete, 106
  Dirac, 25
  finitely additive, 25
  Hausdorff, 350
  inner, 32
  inner regular, 212
  Lebesgue, 37, 70
  Lebesgue-Stieltjes, 35
  outer, 28
  outer regular, 212
  positive, 85
  Radon, 212
  regular, 99, 212
  semifinite, 25
  σ-finite, 25
  signed, 85
  singular, 87
  smooth, 361
Measure space, 25
Metric, 13
Metric outer measure, 349
Metric space, 13
Minimal element, 5
Minkowski's inequality, 183
  for integrals, 194
Modular function, 346
Moment convergence theorem, 320
Monotone class, 65
Monotone class lemma, 66
Monotone convergence theorem, 50
Monotone function, 12
Monotonicity of measures, 25
Multi-index, 236
Mutually singular measures, 87

N

Negative part of a function, 46
Negative set, 86
Negative variation
  of a function, 103
  of a signed measure, 87
Neighborhood, 114
Neighborhood base, 114
Net, 125
Norm, 152
  L^p, 181, 184
  operator, 154
  product, 153
  quotient, 153
  uniform, 121
Norm topology, 152
Normal distribution, 325
Normal number, 330
Normal space, 117
Normed linear space, 152
Normed vector space, 152
Nowhere dense set, 13, 114
Null set, 26, 86
  locally, 192

O

One-point compactification, 132
Open map, 162
Open mapping theorem, 162
Open set, 12–13, 114
Operator norm, 154
Order isomorphism, 5
Order topology, 118
Ordinal, 10

Orientable manifold, 362
Orientation, 362
Orthogonal projection, 177
Orthogonal set, 173
Orthonormal basis, 176
Orthonormal set, 175
Outer measure, 28
  metric, 349
Outer regular measure, 212

P

Paracompact space, 135
Parallelogram law, 173
Parametrization, 352
Parseval's identity, 175
Partial ordering, 4
Partition
  of an interval, 56
  of unity, 134
  tagged, 82
Periodic distribution, 297
Periodic function, 238
Plancherel theorem, 252
Point mass, 25
Pointwise bounded family, 137
Poisson distribution, 320
Poisson kernel, 260, 262
Poisson summation formula, 254
Polar coordinates, 78
Polar decomposition, 46
Positive definite function, 272
Positive linear functional, 211
Positive measure, 85
Positive part of a function, 46
Positive set, 86
Positive variation
  of a function, 103
  of a signed measure, 87
Pre-Hilbert space, 172
Precompact set, 128
Predecesor, 9
Premeasure, 30
Principal symbol, 306
Probability measure, 313
Product measure, 64
Product metric, 13
Product norm, 153
Product σ-algebra, 22
Product topology, 120
Projection, 4
  orthogonal, 177
Proper map, 135
Pythagorean theorem, 173

Q

Quotient norm, 153
Quotient space, 153
Quotient topology, 124

R

Radon measure, 212
  complex, 222
Radon product, 227
Radon-Nikodym derivative, 91
Radon-Nikodym theorem, 91
Random variable, 314
Range, 4
Real-analytic function, 263
Rectangle, 64
Refinement of a cover, 135
Reflexive space, 159
Regular measure, 99, 212
Regular space, 117
Relation, 3
Relative topology, 114
Rellich's theorem, 305
Residual set, 161
Reverse inclusion, 125
Riemann integrable function, 57
Riemann integral, 57
Riemann-Lebesgue lemma, 249
Riemannian volume density, 362
Riesz representation theorem, 212, 223
Riesz-Thorin interpolation theorem, 200
Right continuous function, 12
Right-invariant measure, 341
Ring of sets, 24

S

Sample mean, 325
Sample space, 314
Sample variance, 325
Sampling theorem, 255
Saturated measure, 28
Saturation of a measure, 28
Scalar product, 171
Schröder-Bernstein theorem, 7
Schwartz space, 236
Schwarz inequality, 172
Second category, 161
Second countable space, 116
Section of a set or function, 65
Semifinite measure, 25
Semifinite part, 27
Seminorm, 151
Separable space, 14, 116
Separating set, 359
Separation
  of points, 139
  of points and closed sets, 143
Sequence, 4
Sequentially compact space, 130

Shannon's theorem, 324
Shrink nicely, 98
Sides of a rectangle, 70
Sierpiński gasket, 356
σ-algebra, 21
  Borel, 22
  generated by a family of functions, 44
  generated by a family of sets, 22
  of countable or co-countable sets, 21
  product, 22
σ-compact space, 133
σ-field, 21
σ-finite measure, 25
σ-finite set, 25
σ-finite signed measure, 88
σ-ring, 24
Signed measure, 85
Similitude, 355
Simple function, 46
Singular measure, 87
Slowly increasing function, 294
Smooth measure, 361
Snowflake curve, 356
Sobolev embedding theorem, 303, 308
Sobolev space, 301
  localized, 306
Standard deviation, 314
Standard normal distribution, 325
Standard representation of a simple function, 46
Stirling's formula, 327
Stone-Čech compactification, 144
Stone-Weierstrass theorem, 139, 141
Strong law of large numbers, 322–323
Strong operator topology, 169
Strong type, 202
Stronger topology, 114
Subadditivity, 25
Subbase for a topology, 114
Sublinear functional, 157
Sublinear map, 202
Submanifold, 351
Subnet, 126
Subordination, 134
Subsequence, 4
Subspace of a vector space, 151
Support
  of a distribution, 284
  of a function, 132
  of a measure, 215
Supremum, 9–10
Surjective mapping, 4
Symbol, 273
  principal, 306
Symmetric difference, 3
Symmetric neighborhood, 339

T

$T_0, \ldots, T_4$ space, 116, 123
Tagged partition, 82
Tempered distribution, 293
Tempered function, 293
Theorem
  Alaoglu's, 169
  Arzelà-Ascoli, 137
  Baire category, 161
  Carathéodory's, 29
  central limit, 326
  closed graph, 163
  dominated convergence, 54
  Egoroff's, 62
  Fourier inversion, 251
  Fubini-Tonelli, 67–68, 229
  Hahn decomposition, 86
  Hahn-Banach, 157–158
  Jordan decomposition, 87
  Krein extension, 161
  Lebesgue differentiation, 98
  Lebesgue-Radon-Nikodym, 90, 93
  Liouville's, 300
  Lusin's, 64, 217
  Marcinkiewicz interpolation, 202
  maximal, 96
  moment convergence, 320
  monotone convergence, 50
  open mapping, 162
  Plancherel, 252
  Pythagorean, 173
  Radon-Nikodym, 91
  Rellich's, 305
  Riesz representation, 212, 223
  Riesz-Thorin interpolation, 200
  sampling, 255
  Schröder-Bernstein, 7
  Shannon's, 324
  Sobolev embedding, 303, 308
  Stone-Weierstrass, 139, 141
  Tietze extension, 122, 131
  Tychonoff's, 136
  Urysohn metrization, 145
  Vitali convergence, 187
  Vitali covering, 110
  Weierstrass approximation, 141, 318
Three lines lemma, 200
Tietze extension theorem, 122, 131
Tonelli's theorem, 67–68, 229
Topological group, 339
Topological space, 113
Topological vector space, 165
Topology, 113
  cofinite, 113
  generated by a family of sets, 114

indiscrete, 113
norm, 152
of uniform convergence, 133
of uniform convergence on compact sets, 133
product, 120
quotient, 124
relative, 114
strong operator, 169
trivial, 113
vague, 223
weak operator, 169
weak, 120, 168
weak*, 169
Zariski, 117
Torus, 238
Total ordering, 5
Total variation
of a complex measure, 93
of a function, 102
of a signed measure, 87
Totally bounded set, 15
Transfinite induction, 9
Transpose, 160
Triangle inequality, 151
Trivial topology, 113
Tychonoff space, 123
Tychonoff's theorem, 136

U

Uniform boundedness principle, 163
Uniform continuity, 238, 340
Uniform integrability, 92
Uniform norm, 121
Unimodular group, 346
Unitary map, 176
Upper bound, 5
Upper semicontinuous function, 218
Urysohn metrization theorem, 145
Urysohn's lemma, 122, 131, 245
USC function, 218

V

Vague topology, 223
Vanish at infinity, 132
Variance, 314–315
Vitali convergence theorem, 187
Vitali covering theorem, 110

W

Wave equation, 275
Weak convergence, 169
Weak L^p, 198
Weak law of large numbers, 321
Weak operator topology, 169
Weak topology, 120, 168
Weak type, 202
Weak* topology, 169
Weaker topology, 114
Weierstrass approximation theorem, 141, 318
Weierstrass kernel, 260
Well ordering principle, 5
Well ordering, 5
Weyl's lemma, 308
Wiener process, 332
abstract, 331
Wirtinger's inequality, 254

Y

Young's inequality, 240–241

Z

Zariski topology, 117
Zorn's lemma, 5

PURE AND APPLIED MATHEMATICS

A Wiley-Interscience Series of Texts, Monographs, and Tracts

Founded by RICHARD COURANT

Editor Emeritus: PETER HILTON and HARRY HOCHSTADT

Editors: MYRON B. ALLEN III, DAVID A. COX, PETER LAX, JOHN TOLAND

ADÁMEK, HERRLICH, and STRECKER—Abstract and Concrete Catetories
ADAMOWICZ and ZBIERSKI—Logic of Mathematics
AKIVIS and GOLDBERG—Conformal Differential Geometry and Its Generalizations
ALLEN and ISAACSON—Numerical Analysis for Applied Science
*ARTIN—Geometric Algebra
AZIZOV and IOKHVIDOV—Linear Operators in Spaces with an Indefinite Metric
BERMAN, NEUMANN, and STERN—Nonnegative Matrices in Dynamic Systems
BOYARINTSEV—Methods of Solving Singular Systems of Ordinary Differential Equations
BURK—Lebesgue Measure and Integration: An Introduction
*CARTER—Finite Groups of Lie Type
CASTILLO, COBO, JUBETE and PRUNEDA—Orthogonal Sets and Polar Methods in Linear Algebra: Applications to Matrix Calculations, Systems of Equations, Inequalities, and Linear Programming
CHATELIN—Eigenvalues of Matrices
CLARK—Mathematical Bioeconomics: The Optimal Management of Renewable Resources, Second Edition
COX—Primes of the Form $x^2 + ny^2$: Fermat, Class Field Theory, and Complex Multiplication
*CURTIS and REINER—Representation Theory of Finite Groups and Associative Algebras
*CURTIS and REINER—Methods of Representation Theory: With Applications to Finite Groups and Orders, Volume I
CURTIS and REINER—Methods of Representation Theory: With Applications to Finite Groups and Orders, Volume II
*DUNFORD and SCHWARTZ—Linear Operators
Part 1—General Theory
Part 2—Spectral Theory, Self Adjoint Operators in Hilbert Space
Part 3—Spectral Operators
FOLLAND—Real Analysis: Modern Techniques and Their Applications
FRÖLICHER and KRIEGL—Linear Spaces and Differentiation Theory
GARDINER—Teichmüller Theory and Quadratic Differentials
GREENE and KRANTZ—Function Theory of One Complex Variable
*GRIFFITHS and HARRIS—Principles of Algebraic Geometry
GRILLET—Algebra
GROVE—Groups and Characters
GUSTAFSSON, KREISS and OLIGER—Time Dependent Problems and Difference Methods
HANNA and ROWLAND—Fourier Series, Transforms, and Boundary Value Problems, Second Edition
*HENRICI—Applied and Computational Complex Analysis
Volume 1, Power Series—Integration—Conformal Mapping—Location of Zeros
Volume 2, Special Functions—Integral Transforms—Asymptotics—Continued Fractions

Volume 3, Discrete Fourier Analysis, Cauchy Integrals, Construction of Conformal Maps, Univalent Functions

*HILTON and WU—A Course in Modern Algebra
*HOCHSTADT—Integral Equations
JOST—Two-Dimensional Geometric Variational Procedures
*KOBAYASHI and NOMIZU—Foundations of Differential Geometry, Volume I
*KOBAYASHI and NOMIZU—Foundations of Differential Geometry, Volume II
LAX—Linear Algegra
LOGAN—An Introduction to Nonlinear Partial Differential Equations
McCONNELL and ROBSON—Noncommutative Noetherian Rings
NAYFEH—Perturbation Methods
NAYFEH and MOOK—Nonlinear Oscillations
PANDEY—The Hilbert Transform of Schwartz Distributions and Applications
PETKOV—Geometry of Reflecting Rays and Inverse Spectral Problems
*PRENTER—Splines and Variational Methods
RAO—Measure Theory and Integration
RASSIAS and SIMSA—Finite Sums Decompositions in Mathematical Analysis
RENELT—Elliptic Systems and Quasiconformal Mappings
RIVLIN—Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory, Second Edition
ROCKAFELLAR—Network Flows and Monotropic Optimization
ROITMAN—Introduction to Modern Set Theory
*RUDIN—Fourier Analysis on Groups
SENDOV—The Averaged Moduli of Smoothness: Applications in Numerical Methods and Approximations
SENDOV and POPOV—The Averaged Moduli of Smoothness
*SIEGEL—Topics in Complex Function Theory
Volume l—Elliptic Functions and Uniformization Theory
Volume 2—Automorphic Functions and Abelian Integrals
Volume 3—Abelian Functions and Modular Functions of Several Variables
SMITH and ROMANOWSKA—Post-Modern Algebra
STAKGOLD—Green's Functions and Boundary Value Problems, Second Editon
*STOKER—Differential Geometry
*STOKER—Nonlinear Vibrations in Mechanical and Electrical Systems
*STOKER—Water Waves: The Mathematical Theory with Applications
WESSELING—An Introduction to Multigrid Methods
WHITMAN—Linear and Nonlinear Waves
†ZAUDERER—Partial Differential Equations of Applied Mathematics, Second Edition

*Now available in a lower priced paperback edition in the Wiley Classics Library.
†Now available in paperback.

An in-depth look at real analysis and its applications—now expanded and revised

This new edition of the widely used analysis book continues to cover real analysis in greater detail and at a more advanced level than most books on the subject. Encompassing several subjects that underlie much of modern analysis, the book focuses on measure and integration theory, point set topology, and the basics of functional analysis. It illustrates the use of the general theories and introduces readers to other branches of analysis such as Fourier analysis, distribution theory, and probability theory.

This edition is bolstered in content as well as in scope—extending its usefulness to students outside of pure analysis as well as those interested in dynamical systems. The numerous exercises, extensive bibliography, and review chapter on sets and metric spaces make *Real Analysis: Modern Techniques and Their Applications, Second Edition* invaluable for students in graduate-level analysis courses. New features include:

- Revised material on the n-dimensional Lebesgue integral
- An improved proof of Tychonoff's theorem
- Expanded material on Fourier analysis
- A newly written chapter devoted to distributions and differential equations
- Updated material on Hausdorff dimension and fractal dimension

GERALD B. FOLLAND is Professor of Mathematics at the University of Washington in Seattle. He has written extensively on mathematical analysis, including Fourier analysis, harmonic analysis, and differential equations.

Cover Design: Adrienne Weiss

WILEY-INTERSCIENCE
John Wiley & Sons, Inc.
Scientific, Technical, and Medical Divisi
605 Third Avenue, New York, N.Y. 10158
New York • Chichester • Weinheim
Brisbane • Singapore • Toronto

0-471-31716-0
IBC S-04 1999
00068495
QM 1/1
430418-1
P5,522.00

ISBN 0-471-31716-0
90000
9 780471 317166